RESEARCH ARTICLE OPEN ACCESS

Image Classification and Regression Using Machine Learning Algorithm DNN

[1]Medak Srinivas , [2]Nallabelli Manoj,
[1] Assistant Professor, Department of ECE, JNTUHCEJ, JAGTIAL,Telangana, India
[2]Assistant Professor, Department of ECE, JNTUHCEJ, JAGTIAL, Telangana, India

ABSTRACT
In recent years, deep learning has been used in image classification, object tracking, pose estimation, text detection and recognition, visual saliency detection, action recognition and scene labeling. Image Classification is widely used in various fields such as Plant leaf disease classification, facial expression classification. To make bulky images handy, image classification is done using the concept of a deep neural network. Among different type of models, Convolutionalneural networks has been demonstrated high performance on image classification. In this paper we builded a simple Convolutional neural network on image classification. This simple Convolutional neural network completed the image classification.The paper contributes a methodology for a more accurate classification of images instead of image feature extraction or image segmentation. The proposed work established a promising accuracy of 99.89%.

Keywords: Deep Neural networks, Image classification, convolutional neural networks

---

Date of Submission: 01-09-2021 Date of Acceptance: 14-09-2021

I. INTRODUCTION

Image classification plays an important role in computer vision, it has a very important significance in our study, work and life. Image classification is process including image preprocessing, image segmentation, key feature extraction and matching identification. With the latest figures image classification techniques, we not only get the picture information faster than before, we apply it to scientific experiments, traffic identification, security, medical equipment, face recognition and other fields. During the rise of deep learning, feature extraction and classifier has been integrated to a learning framework which overcomes the traditional method of feature selection difficulties. The idea of deep learning is to discover multiple levels of representation, with the hope that high-level features represent more abstract semantics of the data. One key ingredient of deep learning in image classification is the use of Convolutional architectures.

Convolutional neural network design inspiration comes from the mammalian visual system structure[1]. Visualstructure model based on the cat visual cortex was proposedby Hubel and Wiesel in 1962. The concept of receptive field has been proposed for the first time. The first hierarchical structure Neocognition used to process images was proposed by Fukushima in 1980. TheNeocognition adopted the local connection between neurons, can make the network translation invariance. Convolutional neural network is first introduced by LeCun in [1] and improved in [2]. They developed a multi-layer artificial neural network called LeNet-5 which can classify handwriting number. Like other neural network, LeNet-5 has multiple layers and can be trained with the backpropagation algorithm [3]. However, due to the lack of large training data and computing power at that time. LeNet-5 cannot perform well on more complex problems, such as large-scale image and video classification. Since 2006, many methods have been developed to overcome the difficulties encountered in training deep neural networks. Krizhevsky propose a classic CNN architecture Alexnet [4] and show significant improvement upon previous methods on the image classification task. With the success of Alexnet [4], several works are proposed to improve its performance. ZFNet [5],VGGNet [6] and GoogleNet [7] are proposed.

In recent years, the optimization of Convolutional neural network are mainly concentrated in the following aspects: the design of Convolutional layer and pooling layer, the activation function, loss function, regularization and Convolutional neural network can be applied to practical problems.

II. PREDICTIVE MODELING

A. Classification Predictive Modeling

Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation.For example, an email of text can be classified as belonging to one of two classes: "spam" and "*not spam*".

- A classification problem requires that examples be classified into one of two or more classes.
- A classification can have real-valued or discrete input variables.
- A problem with two classes is often called a two-class or binary classification problem.
- A problem with more than two classes is often called a multi-class classification problem.
- A problem where an example is assigned multiple classes is called a multi-label classification problem.

It is common for classification models to predict a continuous value as the probability of a given example belonging to each output class. The probabilities can be interpreted as the likelihood or confidence of a given example belonging to each class. A predicted probability can be converted into a class value by selecting the class label that has the highest probability.

B. Regression Predictive Modeling

Regression predictive modeling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.

For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of 100,000to200,000.

- A regression problem requires the prediction of a quantity.
- A regression can have real valued or discrete input variables.
- A problem with multiple input variables is often called a multivariate regression problem.
- A regression problem where input variables are ordered by time is called a time series forecasting problem.
- Because a regression predictive model predicts a quantity, the skill of the model must be reported as an error in those predictions.

There are many ways to estimate the skill of a regression predictive model, but perhaps the most common is to calculate the root mean squared error, abbreviated by the acronym RMSE.

III. CNN FOR IMAGE CLASSIFICATION

A. Convolution Layer

The convolution layer is made up of a set of independent filters. Each filter slides over the image and creates feature maps that get different aspects of an image.CNN uses convolutions to joined fetch features from the local domain of given input. Most CNNs comprise a consortium of convolutional, pooling, and affine layers. CNN's offer fantastic performance on visual identification jobs, where they achieve the state of the art.

B. *Pooling Layer*

Pooling was at first developed to assist to make CNN layers put up distortions, as in the scale-invariant feature transform (SIFT) descriptor with a 4×4 sum pooling grid. This layer allows features to move relative to each other resulting in the deep meeting of features even in the light of small distortions. There are some other profits of pooling, as a reduction of the spatial dimension of the feature map degrading the number of parameters. This simplifies the overall complexity of the model. Though sum and max-pooling are a bit outdated as mostly, nowadays, stridden convolution is used. The motive of stridden convolution is to jump some domains during the convolution operation consequently resulting in capable convolution operation with the reduced spatial dimension of the output.

C. Fully Connected (FC) Layer

The fully connected layer in the CNN symbolizes the feature vector for input. This feature vector or tensor or layer takes information that is important to the input. When the network trains, this feature vector then uses for classification, regression, or make an input into another network like Recurrent Neural Network for translating into another type of output, etc. It is also being used as an encoded vector. During training, this feature vector is used to calculate the loss and helps the network to make it trained. The convolution layers before the fully connected layer keep information related to local features in the input image i.e. edges, blobs, shapes, etc. Every Convolutional layer keeps many filters that symbolize one of the local features. The fully connected layer keeps composite and aggregated information from all the important convolution layers.

D. Number of Parameters

The convolution layer consists of 2 types of parameters: biases and weights. The summation of weight and biases is an overall wide variety of parameters in the convolution layer. The variability of parameters is affected by kernel size. For a convolution layer with the clear out the length of 3 × 3, enter of length 25 with 3 channels, the quantity of weights in this layer is 3 × 3 × 25 × 3. The dimension of the input facts is the number of biases. Thus, the wide variety of parameters is 3 × 25 × 3 + 25 on this convolution layer. There can be no parameters since the hyperparameters include stride, pool size, and zero paddings for the pooling layer. In a DNN (Deep Neural Network) shape, the last pair of layers is often the FC (Fully Connected) layer. The numeral of biases is the variation of neurons in the presentation layer. Instead of a pair of layered (DNN) Deep Neural Network, the wide variety of parameters is the summation of parameters in every layer.

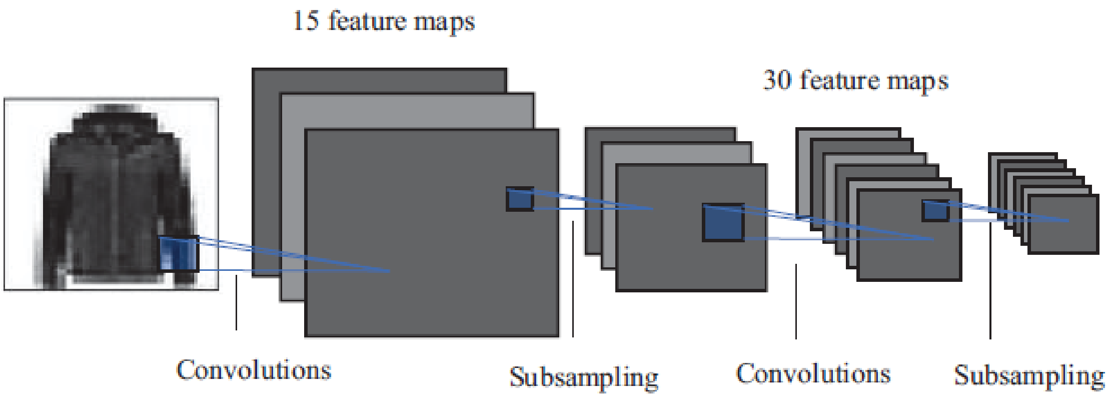

Fig1. Example of CNN structure

E. Visual Geometry Group (VGG16)

The Visual Geometry Group at Oxford University acquires the 16 layers VGG network for the state of the art results in the competition named ILSVRC-2014. The greater depth of its network is the fundamental characteristic of its structural design.

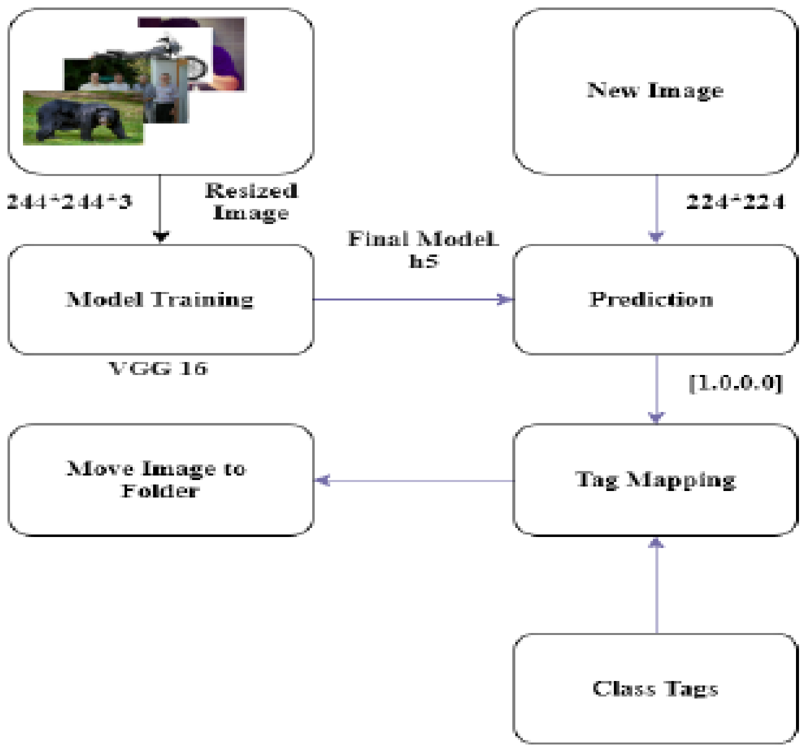

Figure2. Process flow diagram

In this the RGB images go through the five blocks of convolution layers and in every block, there are 3x3 numbers of channels/filters. The CNN layer is padded and stride fixed to 1 so that spatial resolution is conserved after convolution (which means that for 33 filters there is the padding of 1 pixel). The blocks are divided by the max- pooling layer. The stride is 2 in max-pooling which is executed over 22 windows. All the five blocks of the convolution layer are followed by FC (fully connected) layer. The output of the final layer gives class probabilities in soft max-layers.

F. Implementation of VGG 16 Model

The VGG 16 models consist of 16 layers which give a good output in the challenge of ImageNetimage classification. The model consists of 2 divisions: The feature extraction consists of VGG blocks and the classifier consists of FC and output layer. In this paper, a different classifier layer is added and utilize the feature extraction in design architecture. We didn't alter the weights of the convolutional layers, it is constant while training, and just trained new FC layers which will learn to infer the features extracted features aimed at binary classification from the design architecture or model. In the VGG-16 model, from the output side, the FC layer removed and added the new FC layers in output. The prerequisite of the design architecture is the specified input's shape and size which is (224, 224, and 3) in our case for the design model which referred that the updated design architecture lasts at the final max-pooling layer, then after the new classifier layer and a Flatten layer is added. The model ran in 10 epochs and attained an accuracy of 98.72%. This VGG16 model, trained on a particular ImageNet challenge dataset which is constituted to the input images with dimensions 224 x 224 pixels. The images are loaded from the image classification data set with this target size.

IV. RESULTS AND DISCUSSION

An initial interesting point is that the common design principles of the VGG models since it performed best in the competition called ILSVRC 2014[10] and it is very simple and easy to comprehend and implement this modular construction of the architecture. The architecture includes the piling convolutional layers through minor filters of 3×3 along with the max-pooling layer. These layers composed and form a block, and the number of filters in every block augmented beside the network's depth in these blocks for example 32, 64, 128, and 256 for the first four blocks of the design model. To confirm the width and height of the output feature maps counterparts with the inputs, padding is utilized on convolutional layers. Commonly for finest practices, every layer uses the weight initialization and activation function called "ReLU". For example, a VGG-based design of three blocks has a single pooling layer and convolutional layer. We can see the accuracy of each model concerning its activation function and loss function in table I.

Table1. Accuracy using different models

Model used	Accuracy (%)	Activation Function
Baseline CNN	55.075	Sigmoid
VGG3	74.56	Sigmoid
VGG3+Data Aug	61.40	Sigmoid
VGG16 proposed	98.97	Softmax

V. CONCLUSION

In this paper we proposed a simple Convolutional neuralnetwork on image classification. This simple convolutionalneural network imposes less computational cost. On the basisof the convolutional neural network, we also analyzed different methods of learning rate set and different optimization algorithm of solving the optimal parameters of the influence on image classification. In this study, the datasets are made from different sources and then all the images are categorically classified using VGG16. The images are classified into the two categories, firstly living and non-living and then further classified into subcategories such as living classified as nature, animal, human, and the human class categorized into group and selfie images. The non-living classified into the vehicle. Also using VGG16 we get 99.89% accuracy with minimal loss.

REFERENCES

[1]. LecunY, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[2]. Cun Y L, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]// Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers Inc. 1990:465.

[3]. Hecht-Nielsen R. Theory of the backpropagation neural network[M]// Neural

networks for perception (Vol. 2). Harcourt Brace & Co. 1992:593-605 vol.1.

[4]. Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification withDeep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.

[5]. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, 2014.

[6]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in ICLR, 2015.

[7]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with Convolutionals," Co RR, vol. abs/1409.4842, 2014.

[8]. Hinton G E. Deep belief networks[J]. Scholarpedia, 2009, 4(5): 5947.

[9]. Scott rozelle, Xue Lei XuYangming, etc. The depth of the study review [J]. Computer application research, 2012, 29 (8) : 2806-2810.

[10]. liliguo, shifei ding. Deep learning research progress [J]. 2015.

[11]. Bengio Y, Courville A, Vincent P. Representation learning: A reviewand new perspectives[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(8): 1798-1828.

[12]. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in ICML, 2010, pp. 807–814.

[13]. T. Wang, D. Wu, A. Coates, and A. Ng, "End-to-end text recognition with Convolutionalal neural networks," in International Conference on Pattern Recognition (ICPR), 2012, pp. 3304–3308.

[14]. Y. Boureau, J. Ponce, and Y. Le Cun, "A theoretical analysis of feature pooling in visual recognition," in ICML, 2010, pp. 111–118.

[15]. Y. Tang, "Deep learning using linear support vector machines," arXivprprint arXiv:1306.0239, 2013.

Medak Srinivas, et. al. "Image Classification and Regression Using Machine Learning Algorithm DNN." *International Journal of Engineering Research and Applications (IJERA)*, vol.11 (9), 2021, pp 42-45.