IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

Image Recognition using Convolutional Neural Network

M M Girish[1], Jai Shankar G[2], Chandan B J[3]

Student, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[1,2,3]

Abstract: Deep Learning has emerged as a new area in machine learning and is applied to a number of signal and image applications. The main purpose of the work presented in this paper, is to apply the concept of a Deep Learning algorithm namely, Convolutional Neural Networks (CNN) in image classification. In this paper we built a simple Convolutional neural network on image classification. This simple Convolutional neural network completed the image classification. Our experiments are based on benchmarking datasets minist and cifar-10. On the basis of 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.

Keywords: Convolutional Neural Network; Deep Learning; Image Classification, Learning Rate

I. INTRODUCTION

Image Classification is an important topic in the field of artificial neural network and it has drawn significant amount of interest over the last ten years or so. This project aims at classifying the input image based on the visual content. This process includes several tasks such as image pre-processing, key feature extraction, image segmentation, information matching etc. Image classification has a huge scope in the field of medical, traffic identification, security, face recognition etc. Convolutional Neural Network (CNN) were inspired by the biological processes, which bring the connectivity between the neurons in the mammalian brain. The connectivity pattern in the CNN resembles the organization of the animal visual cortex. Individual cortical neurons respond to the stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field. CNN's use relatively little pre-processing compared to other image classification algorithms.

The visual system structure model based on the cat's visual cortex was proposed by Hubel and Wiesel in the 1950's. This model showed that the visual cortexes contain the neurons that individually respond to small regions of the visual field. They proposed two kinds of cells in the visual primary cortex called simple cell and the complex cell, and also proposed a cascading model of these two types of cells for use in pattern recognition tasks.

Inspired by the work of Hubel and Wiesel another model called the "neocognitron" was introduced by Kunihiko Fukushima. This was a hierarchical, multilayered artificial neural network. The neocognitron consists of multiple types of cells, the most important of which are called S-cells and C-cells. 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 networks 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. In this paper we proposed a simple Convolutional neural network on image classification.

II. LAYERS OF CNN

CNN's usually include four layers, namely Convolutional layer, RelU layer, pooling layer and fully connected layer. Convolutional Layer- The Conv layer is the core building block of a Convolutional Neural Network. The primary

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

purpose of Conv layer is to extract features from the input image.The above Figure 4.3 explains the convolution layer that uses feature detector to produce the feature map. The Conv Layer parameters consist of a set of learnable filters (kernels or feature detector). Filters are used for recognizing patterns throughout the entire input image. Convolution works by sliding the filter over the input image and along the way we take the dot product between the filter and chunks of the input image.

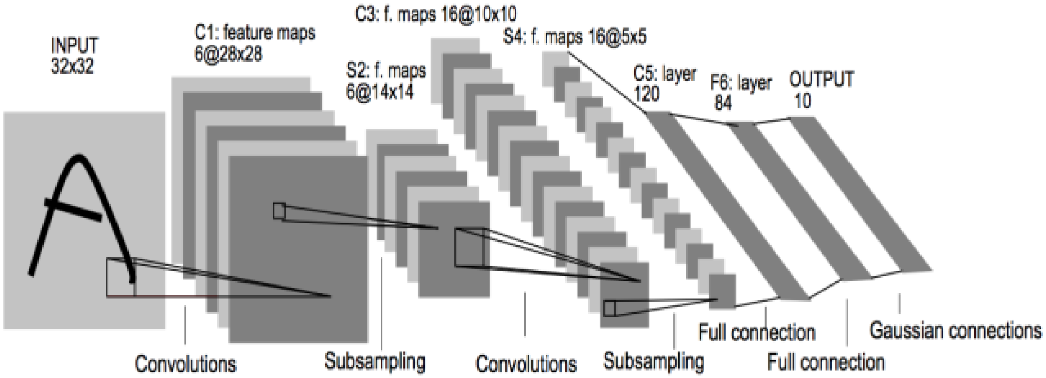

Figure 1. The architecture of CNN (LeNet-5)

RelU Layer- ReLU stands for Rectified Linear Unit and is a non-linear operation. ReLU is an element wise operation (applied per pixel) and replaces all negative pixel values in the feature map by zero.

$$\text{Output} = \text{Max}(\text{zero}, \text{Input})$$

The purpose of ReLU is to introduce non-linearity in our ConvNet, since most of the real-world data we would want our ConvNet to learn would be non-linear. Other non-linear functions such as tanh or sigmoid can also be used instead of ReLU, but ReLU has been found to perform better in most cases. Pooling Layer- Pooling layer reduce the size of feature maps by using some functions to summarize sub-regions, such as taking the average or the maximum value. Pooling works by sliding a window across the input and feeding the content of the window to a pooling function. The purpose of pooling is to reduce the number of parameters in our network (hence called down-sampling) and to make learned features more robust by making it more invariant to scale and orientation changes. Fully connected layer- The Fully Connected layer is configured exactly the way its name implies: it is fully connected with the output of the previous layer. A fully connected layer takes all neurons in the previous layer (be it fully connected, pooling, or convolutional) and connects it to every single neuron it has. Adding a fully-connected layer is also a cheap way of learning non-linear combinations of these features. Most of the features learned from convolutional and pooling layers may be good, but combinations of those features might be even better.

III. SYSTEM ARCHITECTURE

The input image to which we applied multiple different feature detector or also called filters to create these feature maps and this comprises our convolutional layer. On top of that convolutional layer we applied relu or rectified linear unit to remove any linearity or increase non-linearity in our images. Then we applied a pooling layer to our convolutional layer. So from every single feature map we create a pooled feature map. The main purpose of pooling layer is to make sure that we have a special invariant in our images. Pooling pick up the features plus pooling significantly reduces the size of our images. Also pooling helps with avoiding any kind of overfitting of our data but at the same time pooling preserves the main features and the pooling we have used is Max Pooling. Then we flattened all of the pooled images into one long vector or column of all these values and we input that into an artificial neural network and this comprises fully connected artificial neural network where all of these features are processed through a network and the we have final fully connected layer which performs the voting towards the classes and all of these are trained through a forward propagation and back propagation process.

3.1 Activity life cycle of CNN

Activity life cycle of CNN include few states through which it’s going to transit before displaying the output. First one being uploading the image. If the image is compatible then, it’s going to get next stage which is image pre-processing. A Filter is made to slide over the pre-processed image to get the feature map. Feature map is the second stage in this life cycle. Feature map is made to undergo pooling to remove the unwanted pixels in the image. This is nothing but

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

segmentation of the image. Recognition is done by back propagating through previous layers. Fully connected layer helps in classifying the image into different categories.

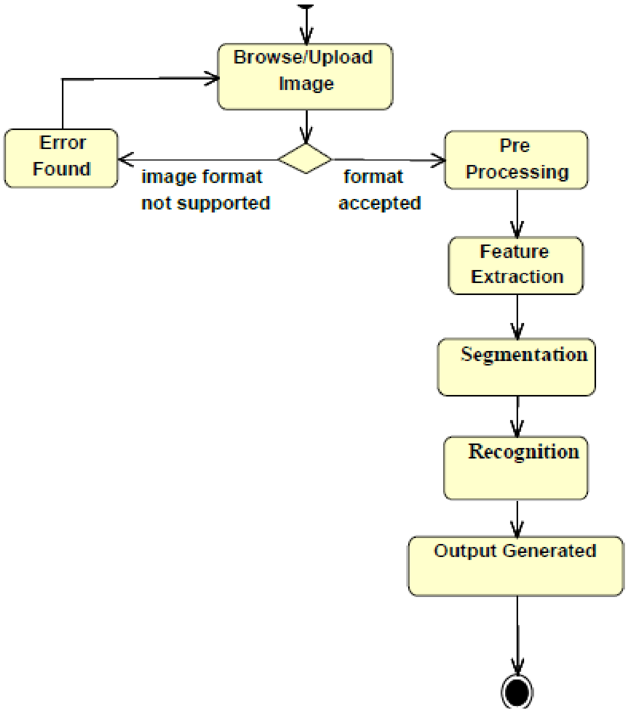

Fig. 2 Activity life cycle

IV. EXPERIMENTAL PROCEDURE AND RESULT

4.1 Preparing Database

The input is given as image itself. The images are converted to gray scale as data information is important for the network and not the color information. Also, the images are resized to 32x32. Since the images from the data sets are larger, pyramid reduction is done to make them of 32x32 in size. The image pyramid is a data structure designed to support efficient scaled convolution through reduced image representation. It

consists of a sequence of copies of an original image in which both sample density and resolution are decreased in regular steps.

4.2 Network Training and testing

The purpose of training algorithm as in [7] is to train a network such that the error is minimized between the network output and the desired output. The error function is as defined by the equation below and is same for weights as well as bias terms.

$$E(w) = \frac{1}{K \times N} \sum_{k=1}^{K} \sum_{n=1}^{N_L} (y_n^k - d_n^k)^2$$

The error gradient is computed through error sensitivities, which are defined as the partial derivatives of the error function with respect to the weighted sum input to a neuron. Once the error gradient ∇E(t) is derived, numerous optimization algorithms for minimizing the energy function can be applied to train the network. Here RPROP (resilient back propagation) is used. It is an efficient learning scheme that performs a direct adaptation of the weight step based on local gradient information.

4.2 Learning Rate Set and Algorithm of Solving the Optimal Parameters

The parameters of the convolutional neural network need to be solved by the gradient descent algorithm. In the iterative process, the basic learning rate needs to be adjusted, and the adjustment strategy has different choices. In the following, we give the influence of different solving strategies on the experimental results and give the corresponding analysis.

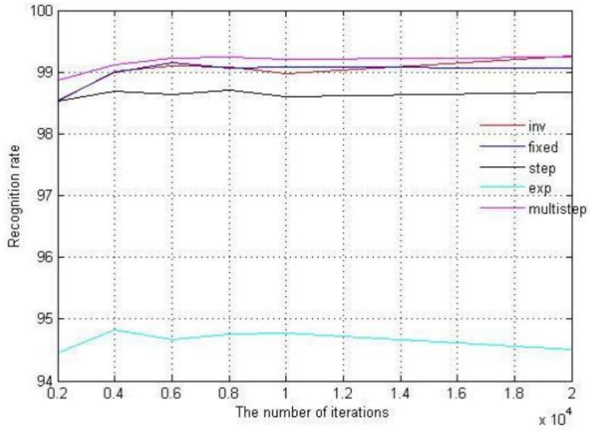

Fig. 3 different learning set in training data

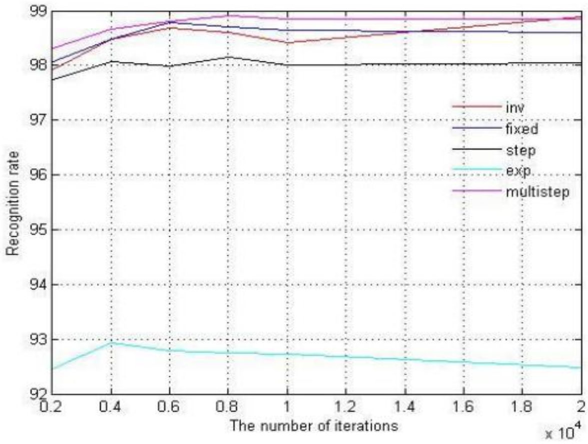

Fig 4. Different learning rate set in testing data

As can be seen from the figure above, with the increase of the number of iterations, the recognition rate of each. algorithm has improved and multistep is the best in those algorithms. The basic learning rate of multistep will be adjusted with the value of step value, so that we can learn more excellent parameters. The fixed method keeps the basic learning rate constant and has universal applicability. This setting has some empirical value, the general set to a smaller value because it will help the network learning and convergence.

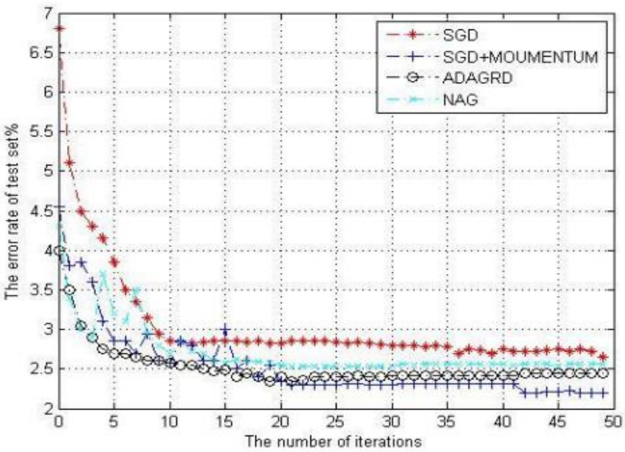

Fig. 5 different solving strategies in testing data

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

From the above experimental results, it can be seen that the SGD can reduce the error with increasing the number of iterations, but the descending speed and effect are general. SGD with momentum and NAG are better than SGD, but the gradient descends slowly in the early stage, and then gradually becomes stable with the increase of the iteration number, and SGD with momentum is better than NAG. The error rate of test set is decline relatively stable in ADAGRAD, the curve is smooth and the effect is better.

4.2 The Result of MNIST- The MNIST dataset consisit of 60000 28 u 28 grayscale images of handwritten digits 0 to 9. The classification results of our simple convolutional neural network are as follows:

TABLE I. THE RESULT OF MNIST

Method	Error rate
Regularization of Neural Networks using DropConnect[20]	0.21%
Multi-column Deep Neural Networks for Image Classification[21]	0.23%
APAC: Augmented Pattern Classification with Neural Networks[22]	0.23%
Batch-normalized Max-out Network in Network[23]	0.24%
Generalizing Pooling Functions in Convolutionalal Neural Networks: Mixed, Gated, and Tree[24]	0.29%
Fractional Max-Pooling[25]	0.32%
Max-out network (k=2) [26]	0.45%
Network In Network [27]	0.45%
Deeply Supervised Network [28]	0.39%
RCNN-96 [29]	0.31%
Our simple Convolutional neural network	0.66%

Compared with the existing methods, although our recognition rate is not the highest, but our network structure is simple, and parameters take up memory is small. We also verify that the shallow network also has a relatively good recognition effect.

V. CONCLUSION

In this paper we proposed a simple Convolutional neural network on image classification. This simple convolutional neural network imposes less computational cost. On the basis of 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. We also verify that the shallow network also has a relatively good recognition effect.

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 with Deep 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.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

[9]. Scott rozelle, Xue Lei Xu Yangming, etc. The depth of the study review [J]. Computer application research, 2012, 29 (8) : 2806-2810.
[10]. Lili guo, shifei ding. Deep learning research progress [J]. 2015.
[11]. FanYaQin Wang Binghao, wang wei, etc. Depth study domestic research review [J]. China distance education, (6) : 2015-27 to 33.
[12]. Liu Jinfeng. A concise and effective to accelerate the Convolutional of the neural network method [J]. Science, technology and engineering, 2014 (33) : 240-244.
[13]. Markoff J. How many computers to identify a cat? 16,000[J]. New York Times, 2012.
[14]. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(8): 1798-1828.
[15]. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in ICML, 2010, pp. 807–814.
[16]. T. Wang, D. Wu, A. Coates, and A. Ng, "End-to-end text recognition with Convolutional neural networks," in International Conference on Pattern Recognition (ICPR), 2012, pp. 3304–3308.
[17]. Y. Boureau, J. Ponce, and Y. Le Cun, "A theoretical analysis of feature pooling in visual recognition," in ICML, 2010, pp. 111–118.
[18]. Y. Tang, "Deep learning using linear support vector machines," arXiv prprint arXiv:1306.0239, 2013.
[19]. S. Wang and C. Manning, "Fast dropout training," in ICML, 2013.
[20]. Wan, L., et al. Regularization of neural networks using dropconnect. in Proceedings of the 30th International Conference on Machine Learning (ICML-13). 2013.
[21]. Ciregan,D.,U.Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. 2012. IEEE.
[22]. Sato, I., H. Nishimura, and K. Yokoi, APAC: Augmented PAttern Classification with Neural Networks. ar Xiv preprint ar Xiv:1505.03229, 2015.
[23]. Jia-Ren Chang,Y.-S.C.,Batch-normalized Maxout Network in Network.2015.ar Xiv:1511.02583v1.
[24]. Lee, C.-Y., P.W. Gallagher, and Z. Tu. Generalizing pooling functions in Convolutionalal neural networks: Mixed, gated, and tree. in International Conference on Artificial Intelligence and Statistics. 2016.
[25]. Graham, B., Fractional Max-Pooling. 2014. Ar Xiv e-prints, December 2014a.
[26]. Ian J. Goodfellow, D.W.-F., Mehdi Mirza,Aaron Courville,Yoshua Bengio, Maxout networks. 2013. 28: p. 1319-1327.
[27]. Lin, M., Q. Chen, and S. Yan, Network In Network. Co RR, 2013. abs/1312.4400.
[28]. Lee, C.-Y., Xie, Saining, Gallagher, Patrick W., Zhang, Zhengyou, and Tu, Zhuowen.,
[29]. Deeply supervised nets. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May9-12, 2015.
[30]. Hu, M la X., Recurrent Convolutionalal neural network for object recognition. 2015.