Deep Residual Learning for Image Classification using Cross Validation

Kshitij Tripathi, Anil Kumar Gupta, Rajendra G Vyas

Abstract: Convolutional Neural Networks (CNN) are very common now especially in the image classification tasks as CNN’s have better classification accuracy than other techniques available in image classification. Another type of CNN called as Residual Neural Networks (RESNET) are gaining popularity because of better accuracy than normal CNN because of residual block available in it. In the present article the RESNET architecture is used for image classification on CIFAR-10 dataset using cross-validation approach that reflects a consistently better accuracy on the above dataset.

Keywords: Convolutional Neural Networks, CIFAR-10, Residual Neural Networks, Cross validation.

I. INTRODUCTION

CNNs [3, 4, 8] are based on deep learning algorithm [2, 3, 4] which consists of stack of repeated layers that are generally convolutional, relu [10], batch-normalization, max-pooling and dense [7]. In recent past artificial neural networks (ANN) [15] have proved themselves in solving wide variety of problems. CNNs are modified form of artificial neural network. In image classification, the image of respective size is given as input and its category or class is provided as for training the neural network. This process is applied multiple times till all samples with their respective classes are used for training. The network is trained multiple times and each complete exhaustive training of all samples is called an epoch.

Generally single epoch is not sufficient to obtain good classification accuracy so training is done multiple times/epochs to achieve high accuracy. Figure 1.1 and Figure 1.2 describes the architecture of RESNET [6] and CNN respectively. Other applications of CNN include object detection, image segmentation and object tracking etc.

II. RELATED WORK ON CIFAR 10

Krizhevsky [1, 9] works on CIFAR-10 [9] using deep CNN to achieve 89% accuracy. CIFAR 10 is created by Krizhevsky, Nair and Hinton [9]. The CIFAR-10 dataset contains 60,000 colour images of 32x32 dimensional and have 10 different classes. The 10 different classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks.

There are 6,000 images of each class. In 2016 RESNET CNN [6] came into existence. Several other variants of CNN like Alexnet [11] and VGGNET [5] have been proposed but RESNET achieve higher accuracy than pure CNN. Ciresan [11] discusses deep CNN to achieve good classification accuracy on CIFAR-10. Important aspect of RESNET is they have far fewer learning parameters and floating point operations. Another important characteristic is the introduction of skip-connections which results in better accuracy than pure CNN. Other variants of RESNET are also available and need to be experimented.

III. PROPOSED METHODOLOGY

The dataset used in the proposed RESNET is the CIFAR-10 [9] which is one of the benchmark dataset in image classification domain. In most of the literature available [12, 13, 14] for classification of CIFAR-10 dataset, the readily available training and test sets are used in which 50,000 samples are given in training set and 10,000 samples are given in test set that can be downloaded from website with the help of python code also. The total numbers of samples is 60,000.

From the various common approaches and literature available in the classification domain, the k-fold cross validation is the de-facto standard and preferred technique for determining the accuracy of classification. In the current article partial k-fold cross-validation is adapted. Most of the research work done on CIFAR-10 uses training and test set that is available from the respective source (website) i.e. 50,000 samples for training and 10,000 for test. In the proposed approach we convert 10,000 test samples into two parts of 5000 samples each i.e. total 60,000 samples into 12 folders where we opt 11th and 12th as a test set one by one.

We then apply normal k-fold approach such that in the first phase first 11 folds are used for training and 12th fold as test. In the second phase the 11th fold is used as test set and remaining folds (1 to 10 folds and 12th fold) are used for training. The RESNET is applied with cross validation approach and experiments are performed till 20 epochs due to resource and time constraints. The results are compared with normal approach in which 10,000 samples available in test set are used as single test set, with same epoch length (#epochs).

It is observed that cross-validation approach is consistently giving better results in each epoch than the normal approach. The results are shown in table 1-3. RESNET contain residual units [6] and mappings to improve the gradient. It is observed that residual networks with short-cut connections (Fig 1.1) gives better results in comparison to normal deep learning networks without short-cut connections. Figure 1.3 (a-c) describes the complete architecture of RESNET.

Revised Manuscript Received on March 30, 2020.

* Correspondence Author

Kshitij Tripathi*, Department of Computer Applications, The Maharaja Sayajirao University of Baroda, Vadodara, India. Email:Kshitij.tripathi-compapp@msubaroda.ac.in

Anil K Gupta, Department of Computer Science & Applications, Barkatullah University, Bhopal, India. Email:akgupta_bu@yahoo.co.in

Rajendra G Vyas, Department of Mathematics, The Maharaja Sayajirao University of Baroda, Vadodara, India. Email: vyas.rajendra.mathi@msubaroda.ac.in

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Retrieved Number: F413104962020098EIEESP
DOI: 10.35940/ijite.F4131.049620

1525
IV. RESULTS

As shown in table I when whole 10,000 images are used as test set the best training accuracy is 88% and best test accuracy is 84%. After cross-validation is applied and the 11th fold is taken as test set than it reflects training accuracy as 89% and test accuracy as 85%. Similarly when 12th fold is treated as test set than best training accuracy achieved is 89% and test accuracy as 86.5%. Further from table II and table III it is observed that cross-validation approach is consistently giving better accuracy than normal approach (table I).
| Layer (type)           | Output Shape   | Param # | Connected to                      |
|------------------------|---------------|---------|-----------------------------------|
| Input_1 (InputLayer)   | (None, 32, 32, 3) | 0       |                                   |
| conv2d_1 (Conv2D)      | (None, 32, 32, 32) | 896     | input_1[0][0]                    |
| batch_normalization_1  | (None, 32, 32, 32) | 128     | conv2d_1[0][0]                   |
| activation_1 (Activation) | (None, 32, 32, 32) | 0       | batch_normalization_1[0][0]      |
| conv2d_2 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_1[0][0]               |
| batch_normalization_2  | (None, 32, 32, 32) | 128     | conv2d_2[0][0]                   |
| activation_2 (Activation) | (None, 32, 32, 32) | 0       | batch_normalization_2[0][0]      |
| conv2d_3 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_2[0][0]               |
| batch_normalization_3  | (None, 32, 32, 32) | 128     | conv2d_3[0][0]                   |
| add_1 (Add)            | (None, 32, 32, 32) | 0       | activation_1[0][0]               |
|                        |               |         | batch_normalization_3[0][0]      |
| activation_3 (Activation) | (None, 32, 32, 32) | 0       | add_1[0][0]                      |
| conv2d_4 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_3[0][0]               |
| batch_normalization_4  | (None, 32, 32, 32) | 128     | conv2d_4[0][0]                   |
| activation_4 (Activation) | (None, 32, 32, 32) | 0       | batch_normalization_4[0][0]      |
| conv2d_5 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_4[0][0]               |
| batch_normalization_5  | (None, 32, 32, 32) | 128     | conv2d_5[0][0]                   |
| add_2 (Add)            | (None, 32, 32, 32) | 0       | activation_3[0][0]               |
|                        |               |         | batch_normalization_5[0][0]      |
| activation_5 (Activation) | (None, 32, 32, 32) | 0       | add_2[0][0]                      |
| conv2d_6 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_5[0][0]               |
| batch_normalization_6  | (None, 32, 32, 32) | 128     | conv2d_6[0][0]                   |
| activation_6 (Activation) | (None, 32, 32, 32) | 0       | batch_normalization_6[0][0]      |
| conv2d_7 (Conv2D)      | (None, 32, 32, 32) | 9248    | activation_6[0][0]               |
| batch_normalization_7  | (None, 32, 32, 32) | 128     | conv2d_7[0][0]                   |
| add_3 (Add)            | (None, 32, 32, 32) | 0       | activation_5[0][0]               |
|                        |               |         | batch_normalization_7[0][0]      |
| activation_7 (Activation) | (None, 32, 32, 32) | 0       | add_3[0][0]                      |
| conv2d_8 (Conv2D)      | (None, 16, 16, 64) | 18496   | activation_7[0][0]               |
| batch_normalization_8  | (None, 16, 16, 64) | 256     | conv2d_8[0][0]                   |
| activation_8 (Activation) | (None, 16, 16, 64) | 0       | batch_normalization_8[0][0]      |
| conv2d_9 (Conv2D)      | (None, 16, 16, 64) | 36928   | activation_8[0][0]               |
| conv2d_10 (Conv2D)     | (None, 16, 16, 64) | 18496   | activation_7[0][0]               |

Figure 1.3 (a) Proposed RESNET
Figure 1.3 (b) Proposed RESNET (Continuing part of Fig 3.2 (a))
Figure 1.3 (c) Proposed RESNET (Continuing part of Fig 3.2 (b))

Table I. Accuracy when whole 10000 instances used as validation set

| Epoch# | Training Accuracy | Validation accuracy |
|--------|-------------------|---------------------|
| 1      | 0.4366            | 0.5059              |
| 2      | 0.6081            | 0.5859              |
| 3      | 0.6850            | 0.6759              |
| 4      | 0.7291            | 0.7028              |
| 5      | 0.7571            | 0.6781              |
| 6      | 0.7807            | 0.7508              |
| 7      | 0.7967            | 0.7524              |
| 8      | 0.8121            | 0.7265              |
| 9      | 0.8211            | 0.7900              |
| 10     | 0.8399            | 0.8088              |
| 11     | 0.8413            | 0.7650              |
| 12     | 0.8493            | 0.8144              |
| 13     | 0.8530            | 0.7800              |
| 14     | 0.8622            | 0.7980              |
| 15     | 0.8676            | 0.7994              |
| 16     | 0.8738            | 0.8334              |
| 17     | 0.8779            | 0.8431              |
| 18     | 0.8777            | 0.8480              |
| 19     | 0.8846            | 0.8428              |
| 20     | 0.8862            | 0.8420              |

Table II. Accuracy when 11th fold is used as validation set

| Epoch# | Training Accuracy | Validation accuracy |
|--------|-------------------|---------------------|
| 1      | 0.4725            | 0.6038              |
| 2      | 0.6436            | 0.6726              |

Table III. Accuracy when 12th fold is used as validation set

| Epoch# | Training Accuracy | Validation accuracy |
|--------|-------------------|---------------------|
| 1      | 0.4703            | 0.5404              |
| 2      | 0.6132            | 0.6626              |
| 3      | 0.6926            | 0.6888              |
| 4      | 0.7403            | 0.6880              |
| 5      | 0.7650            | 0.7606              |
| 6      | 0.7896            | 0.7780              |
Deep Residual Learning for Image Classification using Cross Validation

|   |   |   |
|---|---|---|
| 1 | 0.8064 | 0.7610 |
| 2 | 0.8163 | 0.8024 |
| 3 | 0.8297 | 0.7992 |
| 4 | 0.8404 | 0.7150 |
| 5 | 0.8497 | 0.8076 |
| 6 | 0.8547 | 0.8134 |
| 7 | 0.8588 | 0.7970 |
| 8 | 0.8670 | 0.8242 |
| 9 | 0.8716 | 0.7772 |
| 10 | 0.8773 | 0.8474 |
| 11 | 0.8790 | 0.8074 |
| 12 | 0.8829 | 0.8322 |
| 13 | 0.8861 | 0.8656 |
| 14 | 0.8908 | 0.8334 |

V. Conclusion

It is revealed from the experiments and results obtained that k-fold cross validation approach is still the best choice as the results are unbiased and accuracy is also better. Residual Network has similar layers as in CNN networks that are convolution, pooling, activation and fully-connected layers stacked one over the other. The RESNET contains residual units where dense layer uses direct connections from convolutional layers to the output layer to improve the training by sharing information between the layers and improved gradient flow. The RESNET solve the problem of vanishing gradient by implementing identity connection between the layers as shown in figure 1.1.

REFERENCES

1. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.
2. I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, URL: http://www.deeplearningbook.org.
3. Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L.D. Jackel, Handwritten digit recognition with a backpropagation network, in: Advances in Neural Information Processing Systems (NIPS), 1990, pp. 396–404.
4. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
5. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations (ICLR), 2015.
6. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
7. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning (ICML), 2015, pp. 448–456.
8. Srihari Tripathi, Rajendra G. Vyas, Anil K. Gupta, “Deep Learning through Convolutional Neural Networks for Classification of Image: A Novel Approach Using Hyper Filter”, International Journal of Computer Sciences and Engineering, Vol.7, Issue.6, pp.164-168, 2019.
9. Krizhevsky, Alex, Vinod Nair, and Geoffrey Hinton. "The CIFAR-10 dataset." online: http://www.cs.toronto.edu/~kriz/cifar. html (2014).
10. V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: International Conference on Machine Learning (ICML), 2010, pp. 807–814. Relu
11. D. C. Ciresan, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. High-performance neural networks for visual object classification. Arxiv preprint arXiv:1102.0183, 2011.
12. M. Wistuba, A. Rawat, and T. Pedapati, “A survey on neural architecture search,” May 2019.
13. CIFAR-10 on Benchmarks.AI, https://benchmarks.ai/cifar-10, accessed: 2020-2-5.
14. N. Nayman, A. Noy, T. Rudnik, I. Friedman, R. Jin, and L. Zelnik-Manor, “XNAS: Neural architecture search with expert advice,” Jun. 2019.
15. Kshitij Tripathi, Rajendra G. Vyas, Anil K. Gupta, “The Classification of Data: A Novel Artificial Neural Network (ANN) Approach through Exhaustive Validation and Weight Initialization”, International Journal of Computer Sciences and Engineering, Vol.6, Issue.5, pp.241-254, 2018.

AUTHORS PROFILE

Mr. Kshitij Tripathi did his graduation in Mathematics and Masters in Computer Science from Barkatullah University Bhopal. Thereafter he completed M.C.A. He has obtained diploma in advance computing from C-DAC (A research & development organization belongs to Pune University). He has cleared Ugc accredited examination for the post of Assistant Professor. Currently, he is pursuing Ph.D. in Neurocomputing and working as a faculty in Department of Computer Applications, The Maharaja Sayajirao University of Baroda, Vadodara, India.

Dr. Anil Kumar Gupta is actively involved in Data Mining and Pattern Recognition research. His research work brings many new ideas in Classification and Clustering techniques. He has Ph.D. in Computer Science. He is currently serving as H.O.D of the Department of Computer Science and Applications, Barkatullah University, Bhopal. Currently he is guiding 7 Ph.D. research scholars. He has ten years of research and an over twenty years of teaching experience.

Mr. Rajendra G. Vyas did his Ph.D. in Mathematics, in the Field of Fourier Analysis from Department of Mathematics, The Maharaja Sayajirao University of Baroda, Vadodara, India in 1996. He is currently working as a Professor in Department of Mathematics, Faculty of Science, The Maharaja Sayajirao University of Baroda, Vadodara, India. He has published more than 40 research papers in national and international journals of repute. He is a member of the editorial board in many Mathematics journals, a reviewer for Mathematical Reviews, published by American Mathematical Society as well as a reviewer for Zentralblatt, published by European Mathematical Society.