

ResNet50 Based Classification on Mnist and Cifar10

Boshan Chen¹

¹Princeton University, Princeton NJ 08544, USA

chenboshan1@gmail.com

Abstract. Convolutional neural network (CNN) is gradually trending in the past few years due to its learning capacity and great generalization power. CNNs use a different version of regularization, which makes it on lower extreme. In this task, we studied the performance of Residual neural networks (ResNet50) on Mnist and Cifar10 datasets. We built the model mainly based on ResNet50 with cross-entropy loss. Then we trained Mnist and Cifar10 datasets on it separately and evaluated its prediction accuracy. The prediction accuracy of this model on the test set of Mnist and Cifar10 is 35.87% and 95.35% respectively.

Keywords: CNN, deep learning, ResNet50, Cifar10, Mnist.

1. Introduction

What is AI? A branch of computer science that combines the intelligence of algorithms and machine display ... There are many answers to this question. The common answer is that artificial intelligence (AI) is the effort we put on computers or machines to make them behave like human beings. More specifically, we try to make computers solve tasks intelligently and more efficiently and accurately by learning from past experiences. The use of AI was once limited by computing power and algorithms. In Intel's article, the intelligent behavior represented by AI is behind millions of calculations per second. This requires computing power that was not available at that time. Intel used an example of speech recognition to illustrate this point. According to Richard Mark Soley, 35 years ago, people needed to spend millions of dollars on computers, but computers could only do one thing well [1]. Although AI can complete the work quickly and intelligently, due to the limitation of computing power, it is mostly based on the theory of several decades ago. Nowadays, with the progress of GPUs and algorithms, the development speed of AI is getting faster and faster. AI has been applied in many fields, such as deep learning, computer vision and natural language processing. In addition, we can apply AI to other fields and this interdisciplinary combination produces new products. Making good use of AI can not only improve efficiency, but also improve performance. For example, the application of a recommendation system has many advantages, such as increasing sales or delivering relevant content [3]. In the report, the author also discussed the fields from transportation to entertainment

Machine learning focuses on computers' self-improvement of tasks through past experience and rules governing the system, thus imitating human intelligence. Machine learning has many applications, such as data analysis to big data, or natural language processing to computer vision. Broadly speaking, two types of machine learning models are classification and regression models. The three types of machine learning algorithms are supervised learning, unsupervised learning and reinforcement learning [2]. Among these methods, deep learning is outstanding for its convenience and performance. Deep learning

includes popular models such as CNNs, LSTMs, RNNs [3]. The reason why deep learning is better than other traditional machine learning methods comes from the progress of deep learning. Different from traditional machine learning methods, traditional machine learning methods require human beings to prepare data and features before running the model, while deep learning seems to be more automatic [4]. Therefore, it is reasonable to popularize deep learning. The use of deep learning is mainly related to the collection and analysis of big data. Two main characteristics of big data are, respectively, large-volume and constitution of different kinds of objects. Considering a large amount of data and many objects within it, deep learning can help us filter out useless features and extract knowledge such as trends or possible outcomes [5]. Specifically, we can use deep learning in industry to do natural language processing (NLP), which needs deep learning models to translate human voice into text, or computer vision (CV), which applies models to recognize image patterns.

Due to the development of online social networks, natural language processing has not only acquired data that can be used for analysis, but also gained the interest of industry and academia. At its most basic, NLP can be useful in solving many difficult tasks such as geolocation identification, public opinion mining, or trend analysis. The use of deep learning also stimulates the development of NLP, with techniques such as deep neural networks (DNN), convolutional neural networks (CNN), long short-term memory (LSTM), etc. Caption generation is one of the applications of NLP. In this paper, the author applied a technique to associate some parts of a given image with root words, and these parts were extracted via DNN based ImageNet dataset. Another interesting application is question answering. According to the authors, they first built a UIMA with Farasa toolkit which involves a Tree-Kernel based ranker. Next, they applied LSTM networks to select which question fragments would be used in the ranker [6].

In addition to voice, image is also an essential part of our life. Similarly, the progress of cameras and storage technology makes people more able to collect and store image information than ever before. However, we can not extract features manually. So this is the source of deep learning. For example, deep learning can be used for autonomous driving. With the help of deep learning, autonomous driving technology can now be tested on real public roads instead of traditional laboratory conditions. To evaluate the level of autonomous driving, there are currently five safety in automotive software (SAE) levels. From level 1 to level 5, level 5 refers to full automatic driving. That is, no human driver is needed [6].

Among learning algorithms, Convolutional Neural Networks (CNN) are better than most algorithms in understanding image content. Meanwhile, its performance in image segmentation, classification, detection and retrieval related tasks is also outstanding. CNNs 'most remarkable achievement is using fewer parameters than in Artificial Neural Network (ANN). The achievement has attracted researchers and developers to develop bigger models so that CNN can solve complex tasks. Companies such as Google and Microsoft are working to explore new architectures of CNN [7]. Another attractive aspect of CNN is its ability to recognize and use spatial or temporal correlation in data. When the input propagates toward deep layers, CNN is able to obtain abstract features. For example, in image classification, first layer detects the edges, and then the second layer detects simpler shapes, and then detects more advanced features in deeper layers. In addition, several studies have even been published in such fields as lesion detection, classification and segmentation [8].

2. Related work

The architecture of ResNet50 contains a convolutional layer, a max pooling, nine convolutional layers (1 * 1,64 kernel, 3 * 3,64 kernel, 1*1,256 kernel, all repeat 3 times), twelve layers (1 * 1,128 kernel, 3 * 3,128 kernel, 1 * 1,152 kernel, all repeat 4 times), eighteen layers (1 * 1,256 kernel, 3 * 3,256 kernel, 1 * 1,1024 kernel, all repeat 6 times), nine layers (1 * 1,512 kernel, 3 * 3,512 kernel, 1 * 1,2048 kernel, all repeat 3 times), and one layer with pooling and softmax function. On this project, we used cross-entropy loss. If the predicted probability moves away from the actual value, the cross-entropy loss also increases.

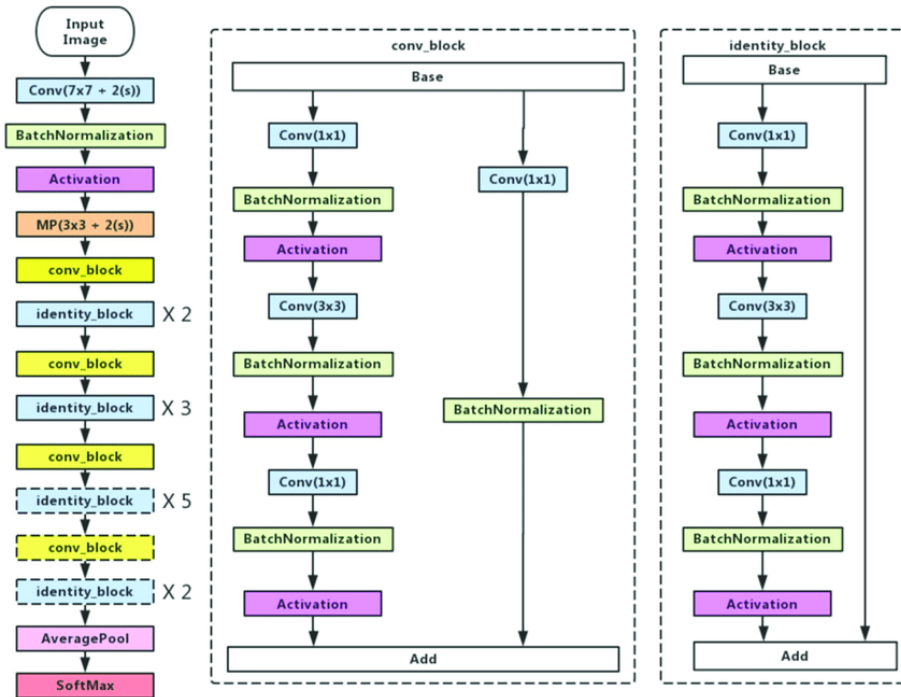


Figure 1. (Left) ResNet50 architecture. Blocks with dotted line represents modules that might be removed in our experiments. (Middle) Convolution block which changes the dimension of the input. (Right) Identity block which will not change the input. Dimension of the input, Figure from [9].

The MNIST database was mainly established from NIST's Special Database 3 and Special Database 1 which include binary images of handwritten digits. SD-3 was originally designated as the training set of NIST and SD-1 was designated as the test set. But SD-1 is not as clean and easy to recognize as SD-3. While SD-1 was collected among high-school students, SD-3 was collected among Census Bureau employees. This is the reason why SD-3 has a better performance. Result that is independent of the choice of training set and test set among the complete set of samples is the requirement before drawing sensible conclusions from learning experiments. Thus, building a new database through mixing NIST's datasets is necessary.



Figure 2. A subset of the MNIST database of handwritten digits, figure from [10].

The CIFAR-10 dataset is a standard benchmark dataset which is mainly used for image classification. It has 10 mutually exclusive classes and it is with 50,000 training examples and 10,000 test examples.

All examples are in RGB, with each image having a resolution of 32x32. with 10 mutually exclusive classes. The dataset consists of images in RGB channel.

3. Experiment and discussion

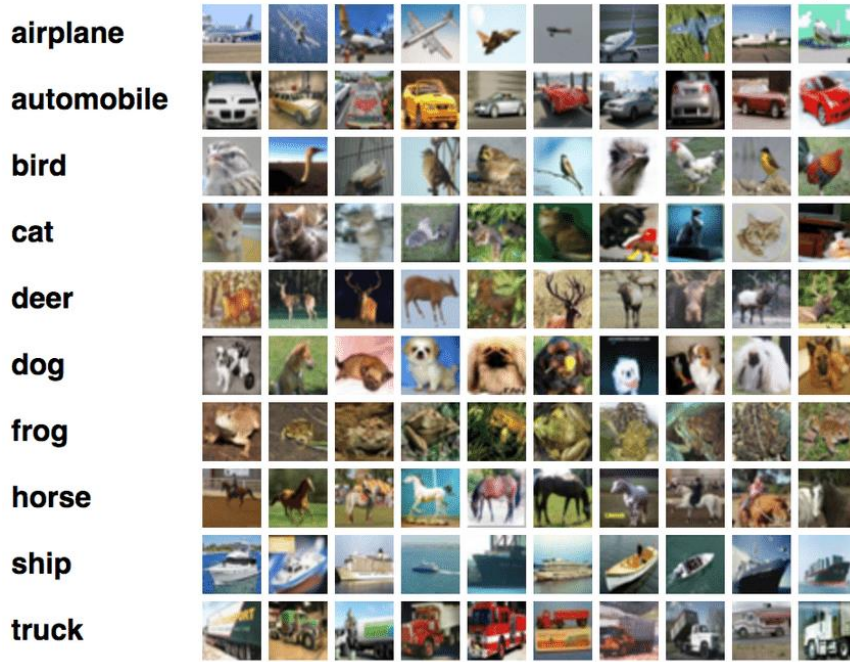


Figure 3. The CIFAR-10 dataset is a benchmark image classification dataset, figure from [11].

On this project, we used the Mnist and Cifar10 datasets from Keras. The dataset Mnist contains 60000 28x28 grayscale images of the handwritten 10 digits as training set, and a test set of 10000 images. Cifar10 consists of 6000 32x32 images for 10 classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck), 50000 of which are in training set and 10000 in test set.

To train Mnist on ResNet50, we need to expand new axis. For the new axis, we need 3 channels instead of 1.

Since the shape of Cifar10 (32x32x3) already satisfied our requirements, we don't need to change it. And the pixel values of Mnist and Cifar10 range from 0 to 255. To reduce the complexity of training, we did uniform normalization by dividing all values by 255, which converts all values ranging from 0 to 1.

The required minimum input size of ResNet50 is 32. For Mnist dataset, We need to resize the input shape and divide the labels into 10 classes (digits 0-9).

For implementation, the model consists of ResNet50, a max pooling layer and a classification layer as output layer. ResNet50 uses a pretrained weight, imagenet, to help the model converge in fewer epochs.

The loss function used is cross-entropy, which calculates the loss by the following formula.

$$Loss = - \sum_{i=1}^{output\ size} y_i \cdot \log \hat{y}_i \quad (1)$$

Where y_i is the target value, and $\hat{y}_i$ represents the i -th scalar value in the model output. Parameters used for ResNet50.

For model training, we use Adam optimizer. We train 2 epochs with batch size of 16 and verbose of 2.

For evaluation, we check the metric classification accuracy. We can see the model performance from Table 1.

Table 1. Performance of ResNet50 on Mnist and Cifar10.

Dataset	Classification Accuracy
Mnist	35.87%
Cifar10	93.53%

4. Conclusion

Cifar10 dataset is relatively simpler than Mnist dataset, and we could see that ResNet50 is doing well on Cifar10 but having hard time in generalization of Mnist, which means that a different model potential should be tried in the future for Mnist to further improve the model performance. In addition, data preprocessing, including unified normalization and channel reorganization, is an important aspect of training CNNs. In the future, we should try a more comprehensive model on these two data sets to obtain more comparisons results.

References

- [1] Intel. The Rise in Computing Power: Why Ubiquitous Artificial Intelligence Is Now A Reality. Forbes (2018).
- [2] Ray, S. Commonly used Machine Learning Algorithms. *AnalyticsVidhya* (2017).
- [3] Biswal, A. Deep Learning Algorithms You Should Know. *Simplilearn* (2021).
- [4] Stephenson, J. Why Deep Learning Is So Hot Right Now. *Logikk* (2019).
- [5] Zhang, Q., Yang, L.T., Chen, Z., Li, P. A survey on deep learning for big data. *ScienceDirect* (2018).
- [6] Al-Ayyoub, M., Nuseir, A., Alsmearat, K., Jararweh, Y., Gupta, B., Deep learning for Arabic NLP: A survey. *ScienceDirect* (2018).
- [7] Khan, A., Sohail, A., Zahoor, U., Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Springer* (2020).
- [8] Albawi, S., Mohammed, T.A., Al-Zawi, S. Understanding of a convolutional neural network. *IEEE* (2017).
- [9] Ji, Qingge & Huang, Jie & He, Wenjie & Sun, Yankui. Optimized Deep Convolutional Neural Networks for Identification of Macular Diseases from Optical Coherence Tomography Images. *Algorithms*. 12. 51. (2019).
- [10] Appiah, Kofi & Hunter, Andrew & Meng, Hongying & Yue, Shigang & Hobden, Mervyn & Priestley, Nigel & Hobden, Peter & Pettit, Cy. A binary Self-Organizing Map and its FPGA implementation. Neural Networks, IEEE - INNS - ENNS International Joint Conference on. 164-171. (2009).
- [11] Agarap, Abien Fred. Training Deep Neural Networks for Image Classification in a Homogenous Distributed System. (2019).