1st World Congress on Industrial Engineering and Operations Management
October 9-11, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/WC01.20240168
Published: October 09, 2024,

Implementation and Comparison of CNN Models on CIFAR-10 Dataset

Seonyul Shin, Shin Dong Ho
Student and Professor, My Paul School
12-11, Dowontongmi-gil, Cheongcheon-myeon, Goesan-gun
Chungcheongbuk-do, Republic of Korea
eavatar@hanmail.net

Jeongwon Kim
Department of Economics, College of Economics, Nihon University
3-2 Kanda-Misakicho, 1-chome, Chiyoda-ku, Tokyo, Japan
shinphys@naver.com

Abstract

Recent advancements in deep learning have yielded remarkable results in various fields such as image classification, object detection, and natural language processing. This study focuses on image classification and implements and compares two different Convolutional Neural Network (CNN) models based on the AlexNet and ResNet architectures using the CIFAR-10 dataset. The motivation for this research stems from the remarkable achievements of deep learning and the significance of image classification. The CIFAR-10 dataset, containing various categories of objects and animals, is utilized as a benchmark dataset to evaluate the generalization capability of the models. The first model, based on AlexNet, exhibits a gradual increase in training accuracy with increasing epochs. However, the test accuracy plateaus after approximately 20 epochs, indicating the challenge of achieving high accuracy with a relatively simple architecture. In contrast, the second model, based on ResNet, effectively addresses the gradient vanishing problem, enabling the training of deeper neural networks. Experimental results show stable increases in both training and test accuracy, with the test accuracy maintaining a high level. The ResNet-based model achieves an accuracy of 90.69% in the final test, demonstrating superior performance compared to AlexNet. This study particularly validates the effectiveness of deep neural network architectures, including ResNet. Future research will explore methods to further enhance performance using various datasets and models, addressing the observed potential limitations.

Keywords
Privacy Security, Privacy Paradox, Large Language Model, Natural Language Processing and Data Protection,

1. Introduction
In Currently, deep learning is achieving remarkable results in various application fields such as image classification, speech recognition, and natural language processing, and many companies are offering services in these forms. These achievements are due to the advancement of large-scale data and powerful model architectures. In particular, Convolutional Neural Networks (CNNs) have shown excellent performance in image classification tasks, with the CIFAR-10 dataset being a useful benchmark for evaluating and improving model performance. CIFAR-10 consists of 60,000 color images of size 32x32, divided into 10 different classes, with 6,000 images per class. This dataset, composed of classes rich in complex visual features and diversity, is suitable for evaluating a model's generalization capabilities.

In this paper, we discuss techniques for implementing a CNN model to perform image classification tasks using the CIFAR-10 dataset. Specifically, we aim to effectively perform the training process of the model by choosing Stochastic Gradient Descent (SGD) as the optimization algorithm. SGD is an optimization technique for effective learning, where at each training iteration, a subset of data samples is randomly selected for computation and weight updating. This allows the model to converge quickly while reducing computational costs. This paper presents experiments and results on how the CNN model implemented using the CIFAR-10 dataset is trained through SGD, demonstrating the effectiveness of the proposed technique. The objective is to provide an effective and practical solution to image classification tasks.

2. Body

Deep learning is a subfield of machine learning. As shown in Figure 1, artificial intelligence encompasses both machine learning and deep learning, and machine learning includes deep learning.

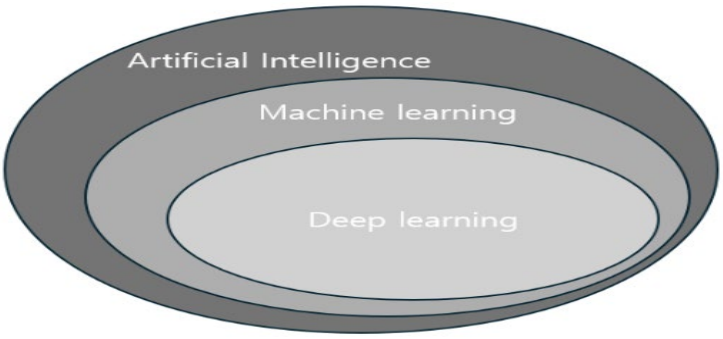

Figure 1. Scope of Deep Learning

Artificial intelligence generally refers to the imitation or replication of human behavior through machines or computer programs. The term artificial intelligence, or AI, is used in various subfields such as expert systems, pattern recognition systems, and robotics. AI-based systems utilize a variety of methods, including statistical algorithms, heuristic procedures, artificial neural networks (ANN), and other machine learning variants, to mimic or model human behavior and decision-making structures.As shown in Figure 1, deep learning is a subfield of machine learning that uses artificial neural networks to learn complex patterns and extract representations from data. Inspired by the workings of the human brain, these models consist of multiple layers of artificial neurons. These artificial neurons accept inputs, process them, and then generate outputs. Deep learning is used to recognize and predict patterns based on data.

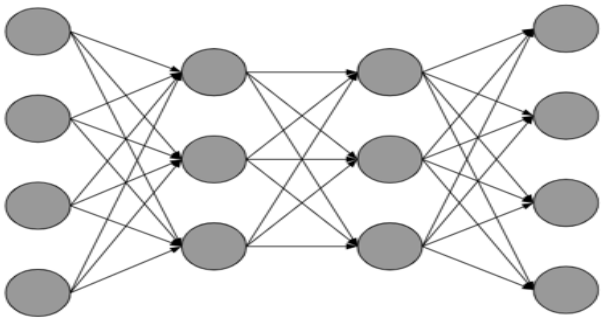

Figure 2. Artificial Neural Networks

A Convolutional Neural Network (CNN) is composed of several layers, primarily consisting of Convolutional Layers, Pooling Layers, and Fully Connected Layers. First, the Convolutional Layer extracts feature necessary for image classification. The Convolutional Layer contains many filters that extract features through these filters in Figure 3 shows the structure of a CNN.

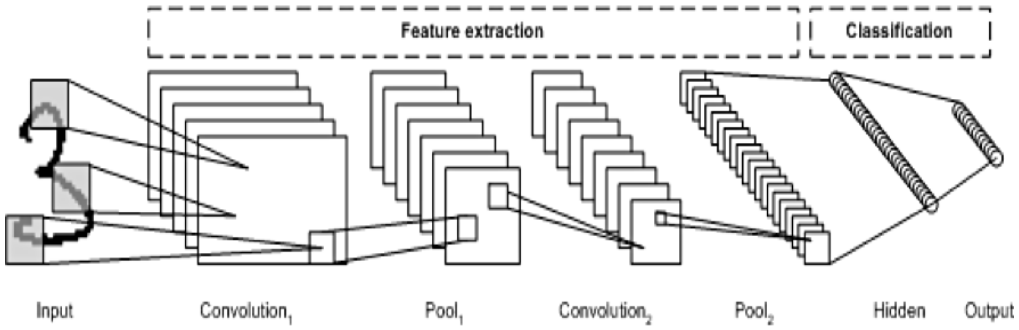

Figure 3. CNN Architecture

To extract features, it is necessary to determine which features to extract, with the contours of objects being a representative example. While humans can easily predict objects by looking at just the contours, computers cannot intuitively determine the answer. Therefore, they use convolution filters to identify what the object is.

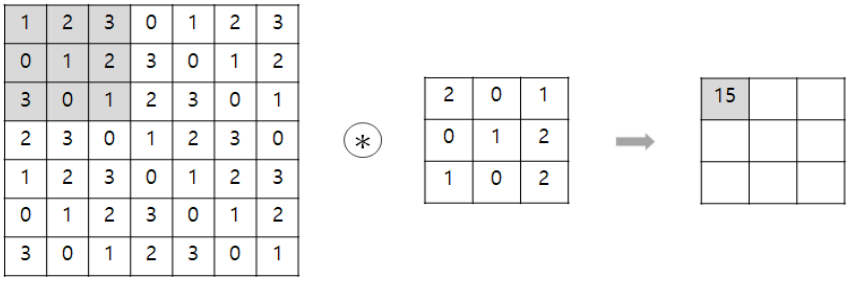

Figure 4. Convolution Operation

A convolution filter creates a feature map through the process of element-wise multiplication and summation between the filter and the input data. By moving the filter across the input data and performing element-wise multiplication with a sub-region of the input, then summing these values to produce a single value, the process generates a feature map for all positions of the input data. Figure 4 illustrates the convolution operation.

.

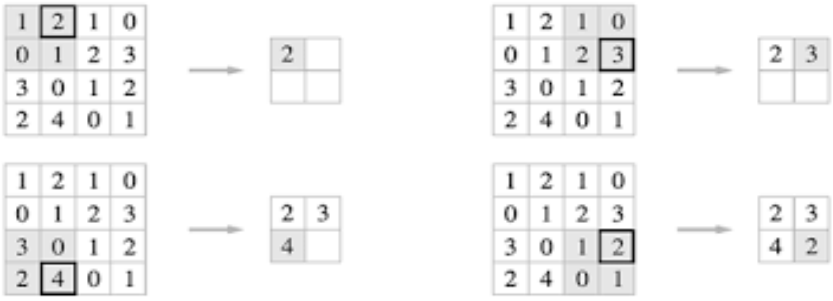

Figure 5. Max Pooling

In CNNs, edge filters are commonly used to detect contours. As the filter traverses each pixel, areas where the convolution filter aligns with edges result in high values, thus identifying contours. For example, if five filters are used, the first convolution layer in Figure 3 will produce five values, referred to as a feature map. Following the convolution layer, a pooling layer is necessary. The Pooling Layer reduces the spatial dimensions and computational load. It is mainly used in image processing, helping to retain important information while reducing the size of the feature map. Max pooling and average pooling are commonly used methods

Pooling operation, as shown in Figure 5, extracts values from specific regions of each feature map to create a smaller map than before. Figure 5 illustrates max pooling, where the largest value in the region is selected. In contrast, average pooling selects the average value. As pooling is performed, the image becomes increasingly abstract, making it easier to identify the shape of objects. When features are extracted in this manner, a feature map resembling the object is created, leading to classification. This process uses a fully connected (FC) layer. The FC layer is primarily used for classification tasks, predicting class probabilities, or performing classification based on the input feature vectors. Neurons in the FC layer learn the features of the input vector, computing a linear combination with weights and biases. A preprocessing step called flattening is performed to compare these values. Flattening, as shown in [Figure 6], converts each layer into a one-dimensional array, connecting all nodes into one. This array is then processed through a function called SoftMax to output the class with the highest probability .

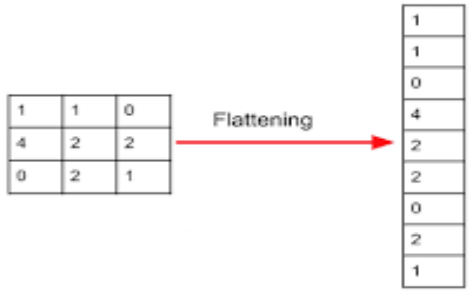

Figure 6. Flatten Operation

The dataset used in this study is the CIFAR-10 dataset. CIFAR-10 consists of 60,000 32x32 color images categorized into 10 unique classes. Each category comprises 6,000 images, encompassing various objects and animal categories. The 10 classes of CIFAR-10 include 'airplane', 'automobile', 'bird', 'cat', 'deer', 'dog', 'frog', 'horse', 'ship', and 'truck'. This diverse class composition covers a wide range of real-world objects, making it useful for evaluating the model's generalization ability.The models employed in the study are AlexNet and ResNet. AlexNet is one of the pioneering neural network architectures that led the advancement of deep learning. It won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet comprises 8 layers, consisting of 5 convolutional layers and 3 fully connected layers. AlexNet utilizes ReLU activation function, Dropout, and Local Response Normalization (LRN). In this research, the AlexNet model is defined, and classes are implemented by inheriting from PyTorch's nn.Module. In the forward method, input data passes through the model to generate output. Preprocessing is performed, and the CIFAR-10 dataset is loaded, divided into training and testing data. Data is batched and loaded, and stochastic gradient descent is set as the optimizer. Over a total of 100 epochs, a test accuracy of 85.54% was achieved. The graph of train accuracy and test accuracy for the model based on AlexNet is as follows.

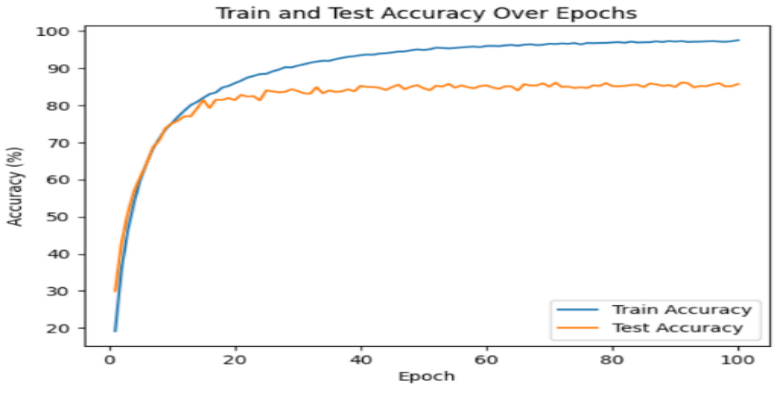

Figure 7. Accuracy of AlexNet-based Model

In Figure 7, we observe a gradual increase in Train Accuracy, while Test Accuracy remains stagnant from around the 20th epoch. The disparity between Train Accuracy and Test Accuracy suggests overfitting, and due to AlexNet's relatively low complexity, it appears challenging to achieve high accuracy. The table below depicts Train Accuracy and Test Accuracy based on the number of epochs for the model built on AlexNet. As mentioned earlier, since AlexNet is a relatively shallow model, a deeper model, ResNet, was implemented for comparison.

Table 1. Accuracy of Alex Net-based Model by Epoch

Epoch	Train Accuracy	Test Accuracy
10	72.84%	72.36%
20	86.01%	82.14%
30	90.53%	84.12%
40	93.12%	84.09%
50	95.05%	84.81%
60	95.82%	85.74%
70	96.24%	84.94%
80	96.76%	84.79%
90	96.94%	84.59%
100	97.34%	86.34%

ResNet (Residual Network) is a deep neural network architecture introduced by Microsoft Research in 2015. It was proposed to address the vanishing gradient problem that arises when increasing the depth of deep learning models. The main idea of ResNet is the introduction of residual connections. In typical CNN structures, multiple convolution layers are stacked between input and output. In this setup, the output of convolution layers contains transformed features from the input. This means that input values can skip intermediate layers altogether. In this study, ResNet8 was employed, which is a simplified form of ResNet consisting of 8 layers. The input layer extracts initial image features and stabilizes training through normalization. ReLU activation function introduces non-linearity, enabling the model to learn non-linear features. Multiple nested residual blocks are used in the Residual Block layer, and the Residual blocks are created through the self.make_layer method. This repetition allows the model to learn features hierarchically. Adaptive average pooling and Fully Connected layers perform adaptive average pooling, reducing spatial dimensions to 1x1. The layer then generates outputs corresponding to the number of classes, receiving 256 input features. Consequently, the model maps input images to probabilities for each class.

In the forward method, input data passes through each layer to the final output. After passing through Convolution-BatchNorm-ReLU layers, the data goes through Residual Block layers sequentially to generate the final prediction. Data preprocessing is performed similarly to when creating an AlexNet-based model. Over a total of 100 epochs, a Test Accuracy of 90.69% was achieved. The graph of Train Accuracy and Test Accuracy for the model based on ResNet is as follows.

Figure 8. Accuracy of ResNet-based Model

In Figure 8, a similar pattern to Figure 7 is observed, where both Train Accuracy and Test Accuracy show gradual improvement. However, they are slightly higher here. Similar to the AlexNet-based model, overfitting seems to occur, and from around the 10th epoch, a widening gap between Train Accuracy and Test Accuracy is noticeable.

Table 2. Accuracy of ResNet-based Model by Epoch

Epoch	Train Accuracy	Test Accuracy
10	87.89%	85.46%
20	93.27%	87.04%
30	95.91%	89.46%
40	97.24%	90.14%
50	97.89%	90.27%
60	98.19%	91.24%
70	98.35%	90.25%
80	98.26%	90.65%
90	98.33%	90.03%
100	98.75%	90.85%

Table 2 presents Train Accuracy and Test Accuracy based on the number of epochs for the model based on ResNet. Upon comparison with Table 1, it is evident that both Train Accuracy and Test Accuracy are consistently higher in the summary of the performance of the ResNet-based model in Table 2.

2.4 Research Findings

In this study, two different deep learning models based on AlexNet and ResNet were implemented and evaluated using the CIFAR-10 dataset. The CIFAR-10 dataset covers various categories of objects and animals, making it a valuable resource for evaluating the generalization ability of models.For the model based on AlexNet, the performance evaluation showed that while the training accuracy gradually increased with the number of epochs, the test accuracy plateaued after around 20 epochs. This indicated the onset of overfitting from around the 20th epoch. Due to its relatively low complexity, AlexNet exhibited a tendency to struggle in achieving high accuracy.

Similarly, the performance evaluation of the model based on ResNet showed a pattern similar to AlexNet, but with higher accuracy. ResNet introduced skip connections to address the vanishing gradient problem, allowing for the construction of deeper neural networks. The gap between training accuracy and test accuracy began to widen around the 10th epoch, indicating the occurrence of overfitting. Combining the results of experiments over 100 epochs, the AlexNet-based model achieved a final test accuracy of 85.54%, while the ResNet-based model achieved 90.69%

accuracy. This demonstrated that ResNet outperformed AlexNet, showcasing the quantitative performance improvement resulting from the advantages of deeper neural network architectures.

3. Conclusion

This study focused on image classification, one of the modern achievements of deep learning, utilizing the CIFAR-10 dataset. Two different base models, AlexNet and ResNet, were implemented and compared. The ResNet-based model exhibited superior performance and stability, highlighting the importance of model architecture in model selection and implementation in deep learning tasks. These results emphasize the significance of model selection in image classification tasks using the CIFAR-10 dataset, as well as the ability of deep neural networks to achieve high generalization performance. Future research could explore methods to improve performance using a wider range of datasets and models.

References

Krizhevsky A., Sutskever I., Hinton G. E., Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25. 2012.

Kim S. W. "Trends and Prospects of Robot Development as a New Growth Engine." Convergence Security Journal [KCI Indexed], 17(2), 153-158. 2017.

LG Electronics Newsroom Press Release. "LG Robotic Vacuum Cleaner Equipped with Intelligence Level of 6-7-Year-Old Child." Retrieved from: https://social.lge.co.kr/newsroom/lg_roboking_0717/, 2017.

TensorFlow Blog. "What is Deep Learning?" Retrieved from: https://tensorflow.blog/

Vlog.io. "Getting Started with Deep Learning Part 1." Retrieved from: https://vlog.io/@feelgi/

Kim T., Summary of CNN, Convolutional Neural Network. Retrieved from: http://taewan.kim/post/cnn/

He K., Zhang X., Ren S., and Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). 2016.

Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9). 2015.

Huang G., Liu, Z., Van Der Maaten L., and Weinberger K. Q., Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). 2017.

Taigman Y., Yang, M., Ranzato M. A., & Wolf L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708). 2014

Simonyan K., & Zisserman A., Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.

Ioffe S., and Szegedy C., Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). 2015

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., I. Attention is all you need. Advances in neural information processing systems, 30. 2017.

Kingma D. P., & Ba J. A., A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

Efficient Convolution Operation with im2col. Retrieved from: https://amber-chaeeunk.tistory.com/31

Pooling Layer in CNN. Retrieved from: http://computing.or.kr/14766/pooling-layer /

Bakrey M.. A Simplified Explanation of CNN Component Layers in Deep Learning. Retrieved from: https://mohamedbakrey094.medium.com/a-simplified-explanation-of-cnn-component-layers-in-deep-learning-518b5c45ec8c, 2021.

Biographies

Seonyul Shin is student in MY PAUL SCHOOL. He is interested in artificial intelligence, deep learning, cryptography, robots, autonomous vehicles, etc., and is conducting related research.

Jeongwon Kim is graduate in College of Economics, Nihon University. She is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, etc., and is conducting related research.

Shin Dong Ho is Professor and Teacher in MY PAUL SCHOOL. He obtained his Ph.D. in semiconductor physics in 2000. He is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, mechanical engineering, the Internet of Things, metaverse, virtual reality, and space science, and is

conducting related research.