HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 2 (2023)

Convolutional Neural Network Approach to Classifying the CIFAR-10 Dataset

Chaoyi Jiang

Dulwich College Beijing

ABSTRACT

Convolutional neural network (CNN) is a powerful tool that can be used in many applications of machine learning. This paper demonstrates the effectiveness of using a CNN to classify images in the CIFAR-10 dataset. The model achieved an accuracy of 0.6276 and a loss of 1.116452 on the validation set. It was observed that the accuracy of predictions varied from class to class, and this paper discusses the potential causes for this variation, such as similar classes sharing common features. Further research in this field could lead to improvement in driving assistance technology and eventually automated driving.

Introduction

Image classification is defined as the task of analyzing groups of pixels within an image to classify the object being shown (*Contextual image classification* 2019). Image classification, along with object localization – which is the task of identifying and drawing bounding boxes around the positions of one or more objects in an image – forms the basis of object recognition (Brownlee, *A gentle introduction to object recognition with deep learning* 2021). Object recognition is the technology that is used in autonomous vehicles that allows them to "see" the real world. Image classification also has a variety of applications outside of autonomous vehicles, including facial recognition and character recognition.

Image classification problems are often solved with machine learning. Machine learning is a subset of artificial intelligence that allows computers to use known data as input to develop a computational model of that data. This model can then be used to make predictions or decisions, the accuracy of which will improve over time without being explicitly programmed to do so (Selig, *What is machine learning? A definition.* 2022) (Burns, *What is machine learning and why is it important?* 2021).

Machine learning is a powerful tool with a wide range of applications, including search engines that provide the user with search results which other users found helpful, spam detection that helps protect users' email inboxes from being flooded with unwanted emails, recommendation algorithms in streaming platforms that suggest new content which users may enjoy, and malware protection applications that learn to recognize the newest threats to a user's system.

The scope of this study is the application of machine learning on the image classification problem of classifying real-life objects. The goal of this article is to develop a model using a viable dataset which can predict the classes of new images not found in the original dataset.

Research Question

How can supervised machine learning be applied as a technique on a convolutional neural network to solve the image classification problem of recognizing and classifying images in the CIFAR-10 dataset?

HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 2 (2023)

Method

CIFAR-10 is a dataset with 60,000 32×32 color images split between 10 mutually exclusive classes, with 6,000 images per class. Each class contains 5,000 training images and 1,000 test images.

Classifying images in the CIFAR-10 dataset posed a unique challenge: while the images were small enough to make modeling a relatively rapid process, they did not provide enough detail to make any patterns clearly identifiable. This meant that drawing decision boundaries was a difficult task.

Before the dataset was used to develop a model, it was first preprocessed using one-hot encoding. When one-hot encoding was applied, the labels of every example were replaced with a 0-indexed list of 10 numbers. The numbers were all equal to 0 except for the number whose position corresponded to the value of the label, which instead had a value of 1. As the list of numbers was 0-indexed, the k^{th} element in the list was in position $(k - 1)$.

Then, feature normalization of the training and test set was performed by dividing the values of all pixels in the image by 255, which meant that the pixel values were changed from integer values ranging between 0 and 255 to float values ranging between 0 and 1. This process helped to improve the performance of the neural network model by making it easier to find the local or global minimum.

A convolutional neural network was used to solve this classification problem. The architecture of the neural network is described below in Table 1.

Table 1. Convolutional neural network architecture.

Input
Conv + ReLU filters: 64, kernel = 4×4, stride = 1
Max Pooling (kernel = 2×2)
Flatten
Dense (channel: 1024) + ReLU
Dense (channel: 10) + Softmax

The model was trained for 5 epochs. An epoch means training the neural network with all the training data for one cycle (Baeldung, *Epoch in neural networks* 2022). The neural network learns to optimize the calculations it performs based on the patterns it recognizes in the input data. By increasing the number of cycles performed on the training set through increasing the number of epochs, the model can use its optimizations from a previous cycle to make better calculations, which will yield better predictions (Baeldung, *The difference between epoch and iteration in Neural Networks* 2022).

Since an expected challenge when developing the model was the lack of clear patterns between examples of the same class, a high number of filters in the convolutional layer and a high number of channels in the dense layer with ReLU were used.

To evaluate the performance of the model, a confusion matrix was used, which is a table used to describe and visualize the performance of a classification model (Markham, *Simple guide to confusion matrix terminology* 2020). Metrics to be reported included accuracy, precision, recall, and F1-score.

Results

In the 5th epoch, the model produced a training set accuracy of 0.74712, a training set loss of 0.720038, a validation set accuracy of 0.6276, and a validation set loss of 1.116452. These metrics indicate the performance of the model over all the classes in the dataset. Table 2 shows the previously mentioned statistics over 5 epochs.

Table 2. Classification metrics of the model across all the classes over 5 epochs.

Epoch Number	Train Loss	Train Accuracy	Test Loss	Test Accuracy
1	1.436270	0.48784	1.245459	0.5627
2	1.104715	0.61280	1.061381	0.6270
3	0.944148	0.66936	1.100277	0.6182
4	0.833673	0.70740	1.027856	0.6395
5	0.720038	0.74712	1.116452	0.6276

Figures 1 and 2 show how the classification metrics change as the number of epochs increases by plotting training loss, testing loss, training accuracy, and testing accuracy against the number of epochs.

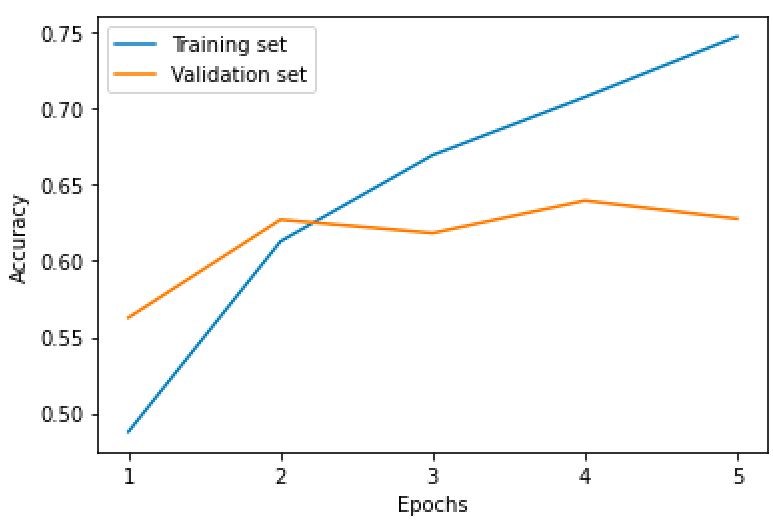

Figure 1. Graph of training accuracy and testing accuracy over 5 epochs.

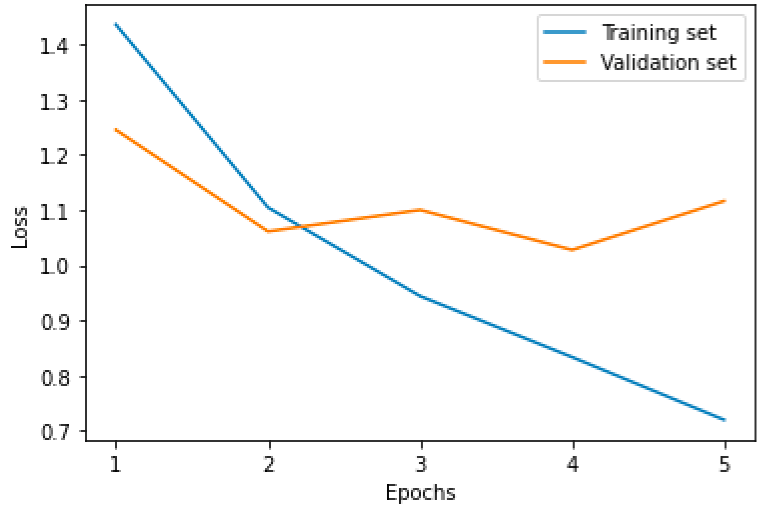

Figure 2. Graph of training loss and testing loss over 5 epochs.

Figure 3 shows the heat-mapped confusion matrix produced by the model. Along the vertical axis is the true label of an example, and along the horizontal axis is the predicted label of an example. Only when the predicted label is the same as the true label is a result on the confusion matrix a true positive. As the confusion matrix is heat-mapped, the darker the color of blue of a combination of labels, the more frequently the model predicted that combination.

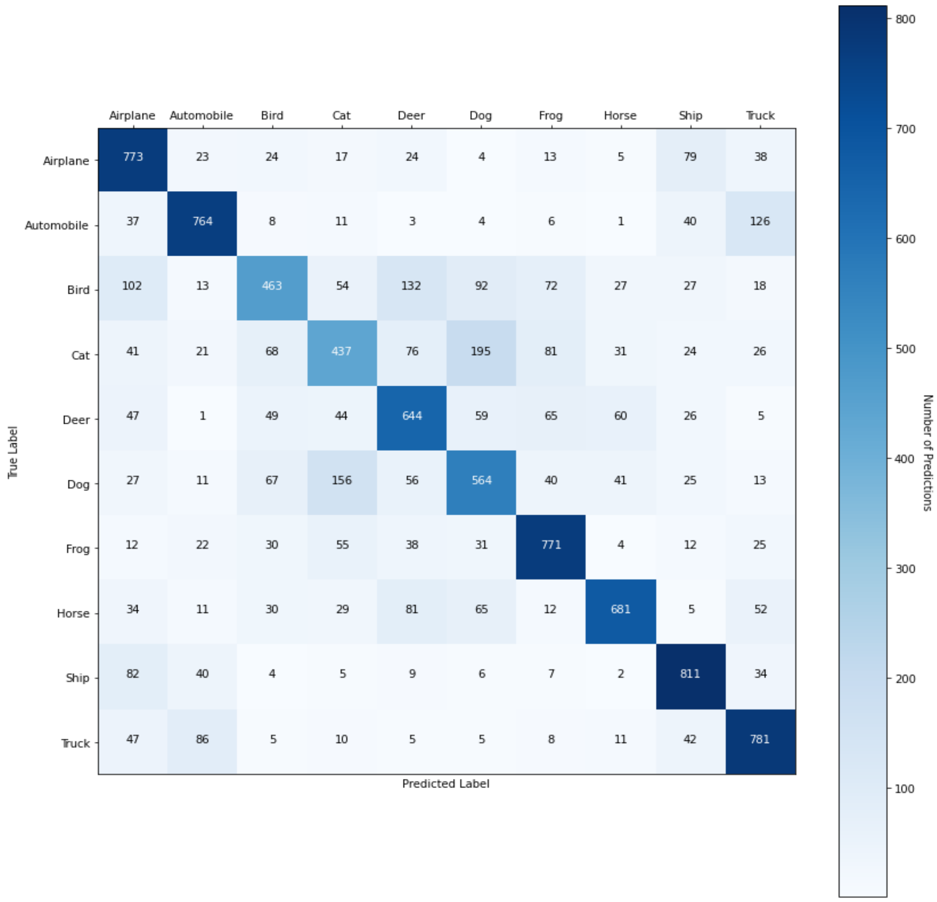

Figure 3. Heat-mapped confusion matrix of the model.

Figure 4 shows the classification report of the model. This report indicates the precision, recall, and F1 score of each class in the final epoch of training.

	precision	recall	f1-score	support
0	0.64	0.77	0.70	1000
1	0.77	0.76	0.77	1000
2	0.62	0.46	0.53	1000
3	0.53	0.44	0.48	1000
4	0.60	0.64	0.62	1000
5	0.55	0.56	0.56	1000
6	0.72	0.77	0.74	1000
7	0.79	0.68	0.73	1000
8	0.74	0.81	0.78	1000
9	0.70	0.78	0.74	1000
micro avg	0.67	0.67	0.67	10000
macro avg	0.67	0.67	0.66	10000
weighted avg	0.67	0.67	0.66	10000
samples avg	0.67	0.67	0.67	10000

Figure 4. Classification report produced by the model.

Discussion

There are some clear patterns that can be noticed in the predictions made by the model from the confusion matrix. Firstly, the model makes correct predictions more frequently than it does incorrect predictions. This is in line with the model's accuracy of 66.89% on the validation set on epoch 5. This indicates that the model is making informed decisions in assigning labels as predictions and can draw some decision boundaries. Secondly, the model tends to make more incorrect guesses between images of similar objects compared to images of completely different objects. For instance, when the true class is automobile, the model predicts truck (126 predictions) and ship (40 predictions) more frequently than it predicts cat (11 predictions) and dog (4 predictions). This is likely due to the similar features between similar objects (sharp, inorganic corners shared between all vehicle classes, for instance). It is expected that if the number of classes is reduced by combining similar classes (such as merging the automobile class and truck class to create a new road vehicle class), the performance of the model is to be improved. This is because similar features shared by similar classes can be more easily associated with the generalized class without having to consider the differences between the individual classes. For instance, wheels could be more easily associated with objects in the road vehicle class without having to consider factors such as the number of wheels or the number of sharp vertices, which may be different in the separate, unmerged classes. Thirdly, the model is better at predicting the classes of images depicting inorganic objects than predicting the classes of images depicting organic objects. For example, the distribution of predictions when the true label is bird (463 accurate predictions) is noticeably more spread out than the distribution when the true label is ship (811 accurate predictions). This may be caused by the similarities between the features expressed in classes showing organic objects (eyes, ears, nose, etc.).

If we examine the graphs of the classification metrics (Figure 1 and Figure 2), we also notice some patterns evident in the data. Training accuracy has a clear positive trend, and training loss has a clear negative trend. This is good, as models should aim to maximize accuracy and minimize loss on the training set. Past epoch 1, the difference between the value of training accuracy and test accuracy and the difference between the value of training loss and test loss increases over every epoch. This is likely caused by overfitting, where the model's decision boundaries are drawn too close to the features of the training set. This is not the best sign, as it indicates that perhaps the model has too much to analyze and is focusing too much on the random noise and slight differences between different images in the dataset. Test accuracy and loss have no clear trend, with test accuracy having a slight positive trend and test loss having a slight negative trend. This may also indicate overfitting.

While CNN is considered a powerful machine learning method, one of the limitations is that this study did not compare the prediction power of various methods. It is possible that other methods achieve similar results with lower computational power. To address this, future research should focus on the direct comparison between CNN and other methods. Furthermore, research should also be done to determine if it is a valid strategy to combine different approaches when creating a model to achieve better results. Another limitation is that this study did not focus on optimizing the hyperparameters of the model. Further studies should focus on the strategy of fine-tuning hyperparameters to allow the model to perform better. However, despite the limitations, the model was able to perform at a reasonable accuracy to achieve valid results after a short period of training time.

Conclusion

This paper demonstrated the utility of convolutional neural networks (CNNs) in identifying real-life objects in photos where there is not enough detail to clearly draw decision boundaries. It was found that a convolutional neural network can predict the class of similar-looking objects to a sufficiently high accuracy. The successful identification of low-resolution images can lead to improvement in driving assistance technology in conditions where obtaining high-resolution feed is not an option (i.e., during nighttime or trying to capture images of distant objects) and could eventually lead to fully automated driving.

Acknowledgments

I would like to thank my advisor Guillermo Goldsztein for the valuable insight provided to me on this topic.

References

Baeldung. (2022, June 20). *Epoch in neural networks*. Baeldung on Computer Science. Retrieved October 2022, from https://www.baeldung.com/cs/epoch-neural-networks

Baeldung. (2022, June 20). *The difference between epoch and iteration in Neural Networks*. Baeldung on Computer Science. Retrieved October 2022, from https://www.baeldung.com/cs/neural-networks-epoch-vs-iteration

Brownlee, J. (2021, January 26). *A gentle introduction to object recognition with deep learning*. Machine Learning Mastery. Retrieved September 2022, from https://machinelearningmastery.com/object-recognition-with-deep-learning/

Burns, E. (2021, March 30). *What is machine learning and why is it important?* SearchEnterpriseAI. Retrieved September 2022, from https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

Krizhevsky, A. (2009). *Learning Multiple Layers of Features from Tiny Images*. , , 32-33.

Markham, K. (2020, February 3). *Simple guide to confusion matrix terminology*. Data School. Retrieved September 2022, from https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Selig, J. (2022, July 4). *What is machine learning? A definition.* expert.ai. Retrieved September 2022, from https://www.expert.ai/blog/machine-learning-definition/

Team, K. (n.d.). *Keras Documentation: CIFAR10 small images classification dataset*. Keras. Retrieved September 2022, from https://keras.io/api/datasets/cifar10/

Wikimedia Foundation. (2019, July 30). *Contextual image classification*. Wikipedia. Retrieved September 2022, from https://en.wikipedia.org/wiki/Contextual_image_classification