

Image classification

Roman Micuda

August 2, 2024

Abstract

Image classification is a computer vision task used to classify input images into desired classes. The MNIST dataset [LeCnd] consists of handwritten digits. Our task is to develop a simple classifier.

1 Introduction

Image classification represents a crucial task within the field of computer vision, finding utility across various domains including self-driving cars, verification systems, and identification systems. We are currently seeking a classification system tailored specifically for the recognition of handwritten digits [Elg20].

2 Image classification

The MNIST dataset has been chosen for image recognition, as it is a widely used dataset for image classification encompassing digits ranging from 0 to 9.

3 Architecture

For the design of the convolutional neural network, we utilized the capabilities of chatbots Gemini [Dee24] and ChatGPT [Ope24].

Layer (type:depth-idx)	Output Shape	Param #
SimpleCNN	[1, 10]	–
Conv2d: 1-1	[1, 32, 28, 28]	320
MaxPool2d: 1-2	[1, 32, 14, 14]	–
Conv2d: 1-3	[1, 64, 14, 14]	18,496
MaxPool2d: 1-4	[1, 64, 7, 7]	–
Conv2d: 1-5	[1, 128, 7, 7]	73,856
Linear: 1-6	[1, 128]	802,944
Linear: 1-7	[1, 10]	1,290
Total params		896,906
Trainable params		896,906
Non-trainable params		0
Total mult-adds (M)		8.30
Input size (MB)		0.00
Forward/backward pass size (MB)		0.35
Params size (MB)		3.59
Estimated Total Size (MB)		3.94

Table 1: Model summary showing layer types, output shapes, and parameter counts.

4 Result

In Fig. 1 is the confusion matrix for the MNIST dataset.

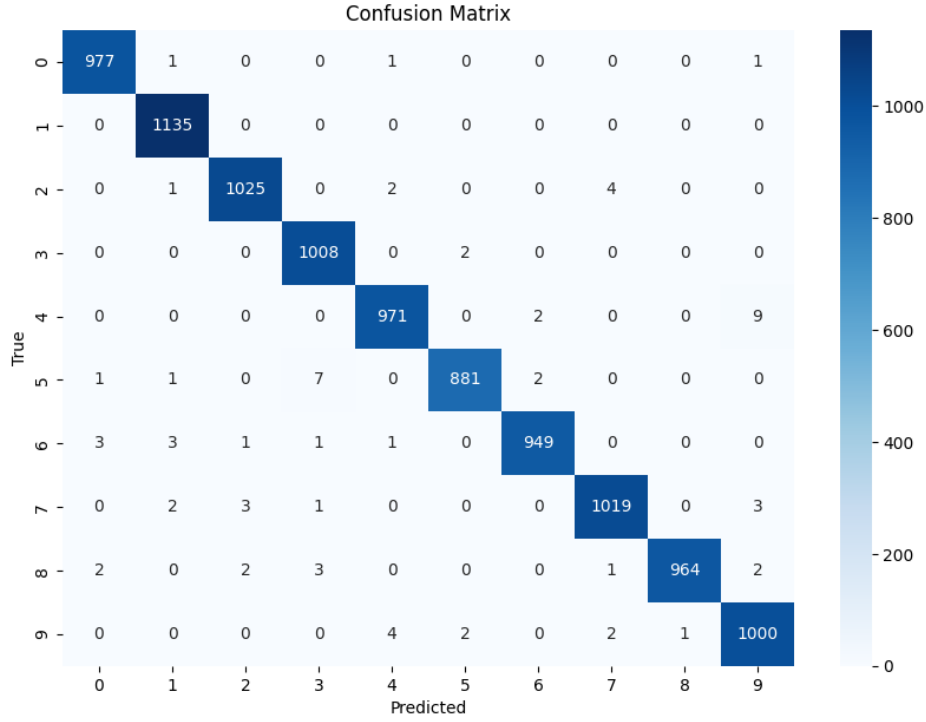


Figure 1: Confusion matrix.

In Fig. 2 is the plot of training and validation loss and accuracy.

In tab. 2 is the classification report.

	Precision	Recall	F1-Score	Support
0	0.99	1.00	1.00	980
1	0.99	1.00	1.00	1135
2	0.99	0.99	0.99	1032
3	0.99	1.00	0.99	1010
4	0.99	0.99	0.99	982
5	1.00	0.99	0.99	892
6	1.00	0.99	0.99	958
7	0.99	0.99	0.99	1028
8	1.00	0.99	0.99	974
9	0.99	0.99	0.99	1009
Accuracy			0.99	10000
Macro Avg	0.99	0.99	0.99	10000
Weighted Avg	0.99	0.99	0.99	10000

Table 2: Classification report with precision, recall, F1-score, and support.

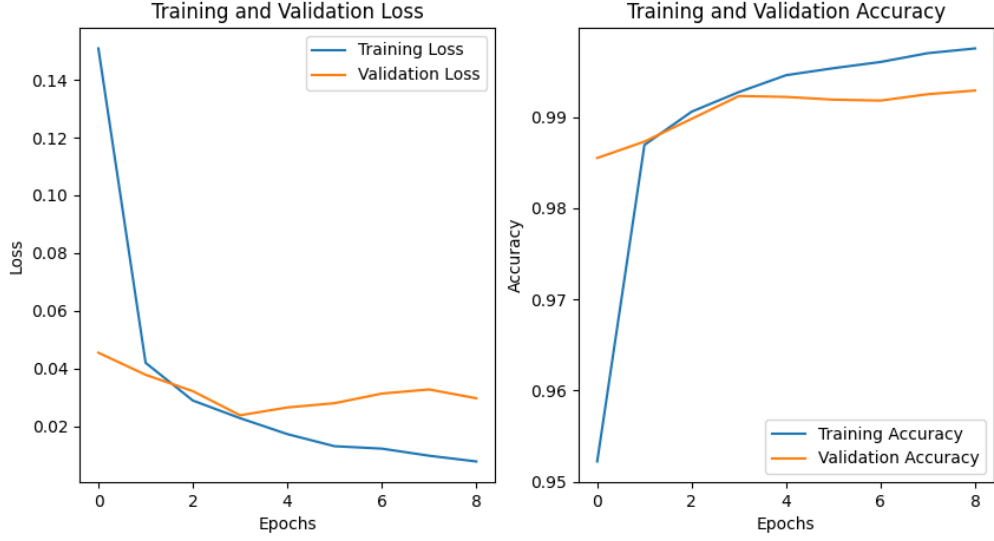
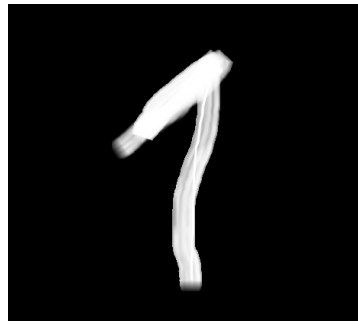


Figure 2: Training and validation loss and accuracy.

References

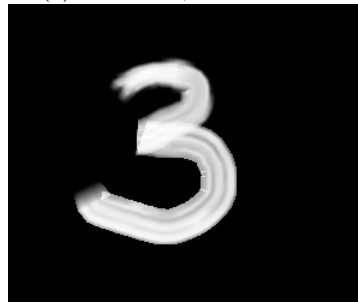
- [Dee24] Google DeepMind. Gemini. Accessed: 2024-08-02, 2024. AI conversational agent.
- [Elg20] Mohamed Elgendy. *Deep learning for vision systems*. Simon and Schuster, 2020.
- [LeCnd] Yann LeCun. The mnist database of handwritten digits, n.d. Accessed: 2024-08-02.
- [Ope24] OpenAI. Chatgpt. Accessed: 2024-08-02, 2024. Large language model.



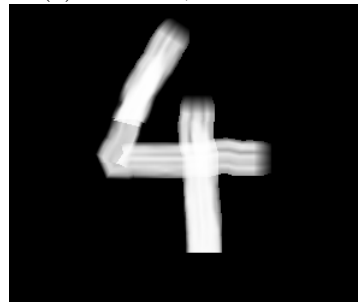
(a) Actual: 1, Predicted: 1



(b) Actual: 2, Predicted: 2



(c) Actual: 3, Predicted: 3



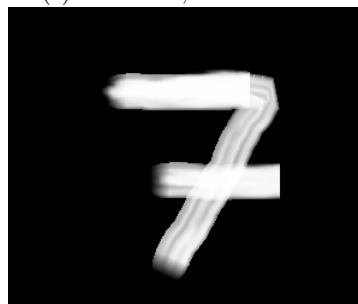
(d) Actual: 4, Predicted: 4



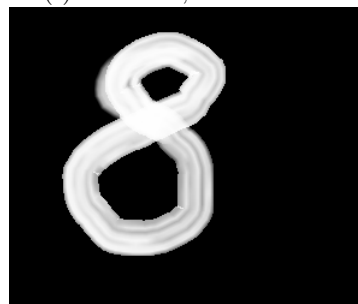
(e) Actual: 5, Predicted: 5



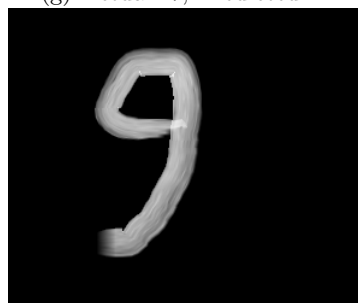
(f) Actual: 6, Predicted: 6



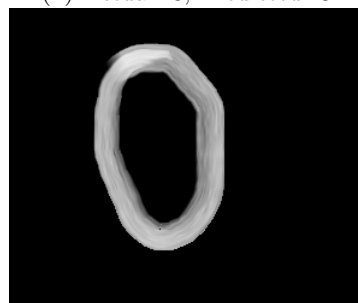
(g) Actual: 7, Predicted: 1



(h) Actual: 8, Predicted: 3



(i) Actual: 9, Predicted: 1



(j) Actual: 0, Predicted: 0

Figure 3: Test handwritten digits.