Analysis of Image Classification with Modified MNIST Dataset Comprised of Handwritten Digits

Anton Gladyr Adam Babs Saleh Bakhit

Kaggle Competition: Group 7

Abstract

Image classification became an area of intense scientific research in recent decades and currently, it is applied in numerous, crucial fields such as developing autonomous cars, security identification or medical diagnoses. Visual recognition is a process of classifying the image content into a set of predefined categories. As of today, some of the image classification problems have been solved and super-human accuracy has been achieved, notably, in certain cases, computers are able to perform even better than people do.

In this project, various methods have been applied to develop models and improve their accuracy. We demonstrate how image different techniques (e.g. image preprocessing, pixels normalization, adjusting hyperparameters) can change the recognition performance of implemented models.

1 Introduction

The main objective of this project is to categorize images from the modified MNIST dataset, which contains handwritten digits. More precisely, the model is supposed to label each image according to the highest numeric value found in the picture.

We experimented with applying various machine learning techniques. To perform the task, we used Support Vector Machines, Logistic Regression, Neural Networks and Convolutional Neural Networks (CNN). Ultimately, the implementation of the last-mentioned method resulted in the highest recognition performance, equal to 94.500%.

Additionally, to improve the performance of our models we changed the architecture of the implemented networks, scaled the image sizes, filtered out the background of each image, adjusted hyperparameters and applied different activation functions. The aforementioned solutions produced the expected results and improved the performance of the convolutional neural network we used to predict on the test dataset.

Convolutional Neural Networks

CNN is a popular deep learning algorithm commonly used for computer vision applications. Compared to multilayer perceptrons, CNNs are regularized and are able to learn useful hierarchical representations by integrating feature extraction and classification in the same model. CNNs are also less prone to overfitting due to their reduced number of parameters. This was brought about by shared weights.[1] CNNs are popular for their shift/space invariance.[2][3].

2 Related work

Numerous past papers describe the MNIST dataset handwritten digits recognition and many of them describe a very high accuracy that was achieved. The images from the aforementioned dataset seem to be easier to classify as images contain just one digit and the image size is a lot smaller. wheres the dataset used in this project was modified and the task is more complex.
Scientific papers indicate how crucial and useful image classification has become. Many of the image analysis problems are considered as solved, as some models have achieved a super-human performance, although many problems related to image analysis have not been solved yet.

AlexNet, a convolutional neural network design by Alex Krizhevsky, is one of the most influential papers in computer vision. It consists of five convolutional layers (some were followed by max-pooling layers) and three fully-connected layers. The main outcome of the paper was that the depth of the network was essential to its performance. This was only computationally feasible by employing GPUs during training the model.[4] Henceforth, CNNs became widely used in pattern and image-recognition problems as they have a number of advantages compared to other techniques.[5]

Samer Hijazi, Rishi Kumar, and Chris Rowen also discuss the advantage of convolutional neural networks over other methods in image classification tasks. They describe that using a standard neural network that would be equivalent to a CNN is more difficult and harder to train, because the number of parameters would be much higher and the training time would also increase proportionately. In a CNN, since the number of parameters is drastically reduced, training time is proportionately reduced. Also, assuming perfect training, we can design a standard neural network whose performance would be same as a CNN. But in practical training, a standard neural network equivalent to CNN would have more parameters, which would lead to more noise addition during the training process. Hence,

the performance of a standard neural network equivalent to a CNN will always be poorer.[5]

Tianming Yu, Jianhua Yang and Wei Lu discuss how useful background subtraction can be in image recognition. Background subtraction plays a fundamental role for anomaly detection in video surveillance, which is able to tell where moving objects are in the video scene. As an excellent classifier, a deep convolutional neural network is able to tell what those objects are. Therefore, we combined background subtraction and a convolutional neural network to perform anomaly detection for pumping-unit surveillance. In the proposed method, background subtraction was applied to first extract moving objects.[6]

3 Dataset and setup

The handwritten digit MNIST dataset is a well-known dataset, comprising of thousands of handwritten digits from 0 to 9. It has been used for many image processing and machine learning tasks. The dataset, this paper is concerned with is a modified version of the original MNIST, where images contain three handwritten digits and additional patterns were added to the background.

Figure 1. Training Dataset Distribution of Classes

The training data consists of 50,000 gray-scale images and the testing data consists of 10,000 images. Figure 1 shows the distribution of classes in the training dataset. We notice that the frequency increases linearly towards the higher digits, hence the dataset is imbalanced. Effectively, it biases the model towards the higher numerical values.

The paper presents the result of preprocessing the images prior to training, using binary thresholding. This acts as a filtering layer where we only keep the pixels of interest, i.e. the digits. This was achieved using the OpenCV library with a threshold value of 240. Figures 2 and 3 show sample images before and after thresholding.

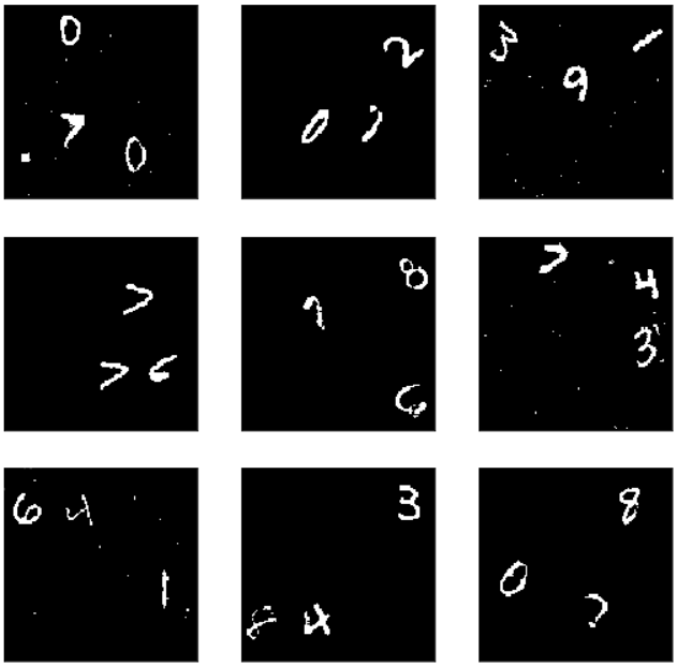

Figure 2. Images after filtering out the background

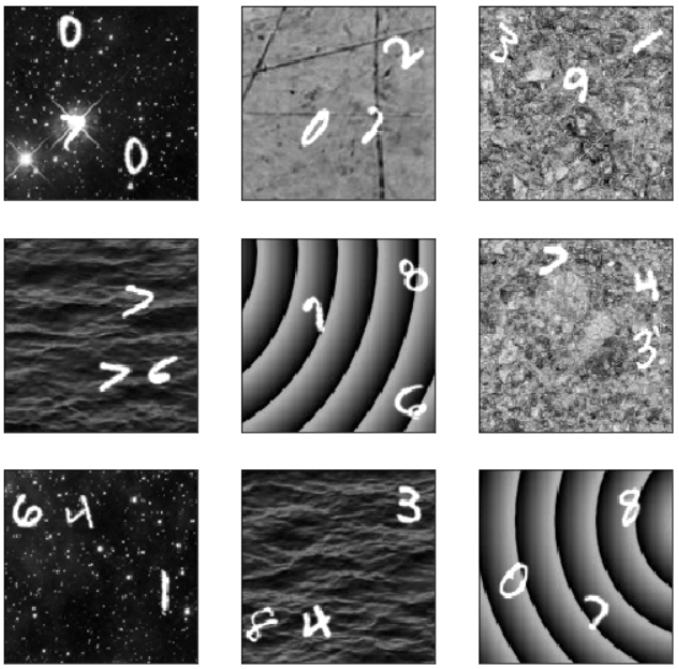

Figure 3. Images before filtering out the background

4 Proposed approach

To increase the recognition performance of our model we adjusted its hyperparameters and filtered out the background of images in the datasets.
The image filter applies the same threshold value for every pixel. In Eq. (1) shows the rule of thresholding, where if the pixel value is smaller than the threshold, it is set to 0, otherwise, it is set to a maximum value.

$$I'(x,y) = \begin{cases} \text{maxval}, & \text{if } I(x,y) > \text{THRESH} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Since all digits are the same color, and in one channel image the maximum value of a pixel is equal to 255, the threshold was set to 240. Results are presented in Figures 2 and 3.
As the related work section demonstrates, the best results on image classification tasks are achieved by using CNNs, thus it is reasonable to apply this type of models on the modified MNIST dataset. In the proposed implementation the architectural idea of CNN is based on VGG-16 Net, which has 13 convolutional layers and 3 fully connected layers. The model uses 3x3 kernels for convolution, rectified linear unit (ReLU) activation function, and 2x2 kernel for pooling. The architecture of the convolutional neural network implemented in this project was modified. We used fewer layers and adjusted hyperparameters. Hyperparameters that were used are presented in Table 1. The full architecture is seen in Figure 6 in the appendix section.

Hyperparemeter	Value
Learning rate	0.001
Number of epochs	100
Batch size	256
Activation function	ReLU
Number of hidden layers	12
Dropout	p=0.5
Optimizer	Adam
Criterion	Cross-Entropy loss

Table 1. Hyperparameters used in the Neural Network design

Furthermore, we experimented with different architectures of the implemented convolutional neural network. We tried various activation functions, changed the number of convolutional layers, added dropout, pooling layers, and applied batch normalization.
We also implemented different linear models, although, their accuracy was not even close to the performance of our Neural Network.

5 Results

Twelve-layered convolutional neural network provided the highest accuracy achieved in this project.
Highest accuracy achieved: 94.500%

Table 1 presents the recognition performance of all models used in the project. Notably, scores achieved are very low, it shows the advantage of using Neural Networks in the image recognition tasks.

Model	Accuracy
Logistic Regression	19.12%
Complement Naive Bayes	18.96%
Multinomial Naive Bayes	10.94%
Random Forest Tree	27.4%
Linear SVC Tree	17.3%
Ensembling of standard models	17.74%
Six-layered CNN	90.80%
Twelve-layered CNN	94.50%

Table 2. Accuracy of models used in the project

Figures 4 and 5 visualize the learning process of the model. As presented on the plots, the model stopped learning after approximately 10 iterations and for another 30 iterations, it did not improve its performance on the validation set as the curve oscillated between 93% and 95%.

Cross-entropy loss on the validation set, settled around 0.50 after 8 iterations. Both cross-entropy loss and accuracy are strongly correlated, hence the similarity of both curves is not surprising. It indicates that the model stopped learning after around 10 epochs. The validation set used in the presented figures comprised of 10,000 samples.

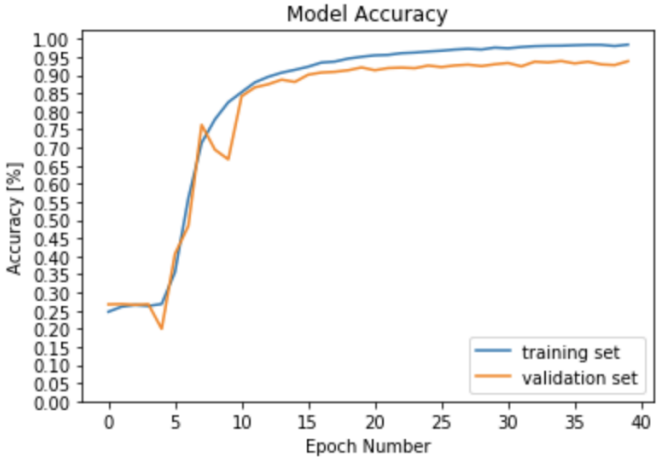

Figure 4. Model accuracy on validation and training sets, over 40 iterations

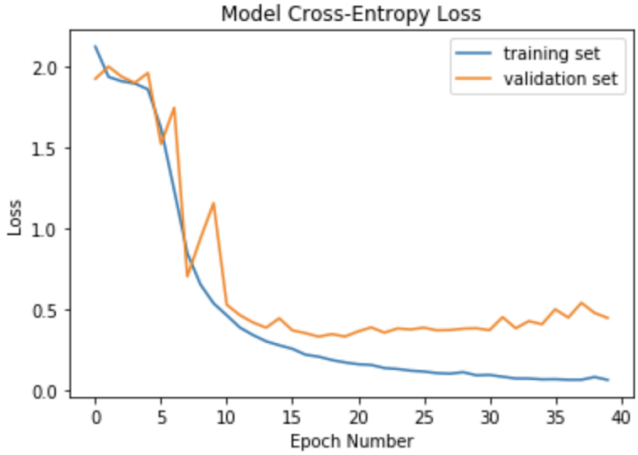

Figure 5. Model loss on validation and training sets, over 40 iterations

6 Discussion and Conclusion

During the process of setting up the architecture of our models, we discovered that convolutional neural networks performed significantly better than other deep learning methods. CNNs require fewer parameters, hence the training time is reduced and less memory is required. Further-

more, less noise is being added during the training process which contributes to improved classification accuracy.

According to our expectations, we observed how removing the background of the image improved the classifying accuracy of our model. It seemed reasonable to do so as it allowed the model to ignore the uninformative pixels, which in some cases could be misleading.

We observed how setting up a correct architecture and a proper adjustment of hyperparameters influences the net's performance. At the beginning we designed models that did not train on the training test and performed poorly on the validation set- accuracy oscillated between 20% and 25%. After adjusting the architecture to the image size and digits size, our model started extracting crucial features, which led to a significant improvement in the classifying performance.

In the design process, we experimented with models that achieved very high performance on the standard MNIST dataset, however, they performed poorly on the modified data and hardly learned. It proved that an accurate design is conducive to the final result.

Considering all our findings and experiments, we noticed how crucial is the correct setup of the architecture and adjusting its hyperparameters. We have also noticed that some changes in the data may affect the final outcome (e.g. values normalization and removing the background).

7 Statement of Contribution

During the first part of the project, the team decided to take a parallel approach. Each member proceeded to experiment with the task until one model was able to exceed the baseline. Then team then shifted focus to optimize the best model and prepare the write-up.

References

[1] Buda M, Maki A, Mazurowski MA. *A systematic study of the class imbalance problem in convolutional neural networks.*, 2017.

[2] Zhang, Wei. *Shift-invariant pattern recognition neural network and its optical architecture.* Proceedings of Annual Conference of the Japan Society of Applied Physics, 1988.

[3] Zhang, Wei. *Parallel distributed processing model with local space-invariant interconnections and its optical architecture*, 1990.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. *Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems*, 2012.

[5] Samer Hijazi, Rishi Kumar, and Chris Rowen, IP Group, Cadence. *Using Convolutional Neural Networks for Image Recognition*, 2005.

[6] Tianming Yu, Jianhua Yang and Wei Lu. *Combining Background Subtraction and Convolutional Neural Network for Anomaly Detection in Pumping-Unit Surveillance,* 2019.

Appendix

Figure 6. Visual Representation of the Twelve-Layered Convolutional Neural Network Architecture which provided the highest performance.