CP 322 – Optical Character Recognition ML Project Report

Dec 10, 2022

Obi Ihejirika - 190970850

Anshul Khatri - 193313680

Yash Rojiwadia - 203039360

Abstract

Optical character recognition is a technology that has been around since the 1970s and has seen very significant improvements since then. It has wide ranging applications that include digitizing physical media such as books and newspapers, automating data entry, enhancing accessibility and enhancing security. Recent advancements in machine learning have made this technology far more efficient.

Introduction

This report explores the design, training, and performance of a machine learning-based optical character recognition (OCR) program applied to a task of recognizing handwritten digits from other physical mediums. OCR is a technology with a wide range of applications, such as automating data entry, record-keeping, security analysis, and even converting printed media into Braille for blind people. This technology provides an efficient solution to time-intensive and error-prone tasks such as data entry, digitizing printed materials, and digitizing handwritten documents.

Description of Applied Problem

This project's objective is to be able to recognize handwritten characters through optical character recognition. The ability to convert real-world text and handwriting into digital text has wide-ranging applications. It solves the issue of having to manually convert written text into digital form, saving time and effort.

However, this is a challenging problem because text can appear in many different fonts, sizes, and colors, and can be distorted or obscured in various ways. By using machine learning algorithms and image pre-

processing techniques, OCR systems can improve their accuracy and performance in recognizing and classifying text in images.

Description of Available Data

In this project, we used the MNIST ("Modified National Institute of Standards and Technology" dataset), a widely-used dataset for training and testing machine learning algorithms designed to recognize handwritten digits. The dataset consists of a collection of images of handwritten digits, along with the corresponding labels that indicate the correct digit represented in each image.

Analysis techniques: Model selection and evaluation

The OCR model chosen for this application was an Artificial Neural Network (ANN) model from the Python-based Keras open-source library. Neural networks are especially suitable for this task due to their ability to handle huge amounts of data and scale to even larger datasets, which is a crucial factor in image recognition tasks with high-dimensional input data. To evaluate the model, we observed and displayed its prediction accuracy and plotted a confusion matrix to illustrate the model's performance for each of the 10-digit classes.

Technique & results

Stages in the Model Training and Testing Process

Libraries and Imports

First, we import the necessary dependencies and datasets for our program. Some of the important libraries used are Keras, Matplotlib, Numpy and Sklearn.

Data Preprocessing

We start the data preprocessing by loading the MNIST dataset into variables and normalizing the input values. To feed the data into the neural network, we must normalize the pixel values of the images, which is done by converting the data type of the tensors from uint8 to float32 and dividing the tensors by 255. We then convert the labels for each image into categorical representations by converting the data type of

the tensors from uint8 to float32 and dividing the tensors by 255: this uses the **to_categorical** function from Keras, which takes the integer labels and converts them to a binary matrix, where an element with value 1 indicating it that the image belongs to that class.

Model Initialization

The architecture of the neural network is comprised of two dense (fully connected) layers. The first layer has 10 neurons and utilizes the sigmoid activation function, while the second layer consists of 10 neurons and employs the “softmax” activation function. This produces a matrix of 10 probabilities for each image, indicating the probability that the image belongs to each of the 10-digit classes.

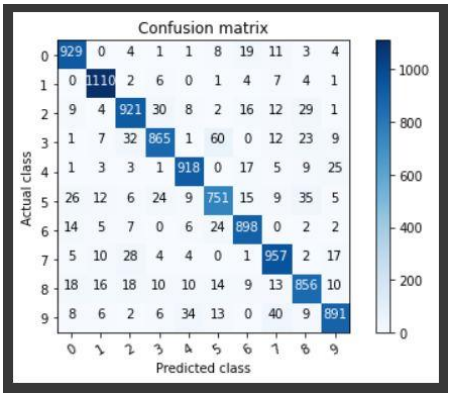
Model Training and Testing

The neural network is then trained using the stochastic gradient descent optimization algorithm, a widely used optimization technique for training neural networks. It was trained using 6679 epochs from the dataset with a batch size of 100. To evaluate the model's performance, the application then configured to accept user input and select images from the dataset for testing.

Visualization Techniques: *Visualization of result such as confusion matrix, Plotting of the loss function, accuracy score etc.*

To evaluate the model, we observed and displayed its prediction accuracy - the ratio of images that were correctly classified. The prediction accuracy while training the model for 6679 epochs was 94.26%.

In addition to evaluating the prediction accuracy we plotted a confusion matrix to illustrate the models for each of the 10-digit classes.



References

Agarwal, R. (2022, September 26). Deep learning based OCR for text in the wild. Nanonets AI & Machine Learning Blog. Retrieved December 10, 2022, from https://nanonets.com/blog/deep-learning-ocr/

Berchmans, D., & Kumar, S. S. (2014). Optical character recognition: An overview and an insight. 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). https://doi.org/10.1109/iccicct.2014.6993174

Berchmans, D., & Kumar, S. S. (2014). Optical character recognition: An overview and an insight. 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). https://doi.org/10.1109/iccicct.2014.6993174

Confusion matrix. scikit. (n.d.). Retrieved December 10, 2022, from https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Eikvil, L. (1993). Optical Character Recognition.

Mithe, R., Indalkar, S., & Divekar, N. (2013). Optical Character Recognition.