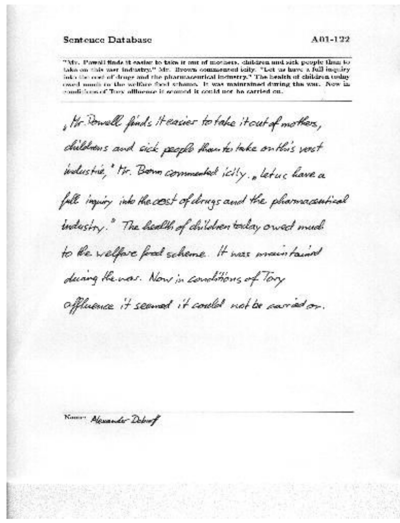
Deep Learning Assignment Grader Proposal

Joeny Bui
October 22, 2017



Sentence Database A01-122

"Mr. Powell finds it easier to take it out of mothers, children and sick people than to take on this vast industry." Mr. Brown commented icily. "Let us have a full inquiry into the cost of drugs and the pharmaceutical industry." The health of children today owed much to the welfare food scheme. It was maintained during the war. Now in conditions of Tory affluence it seemed it could not be carried on.

,,Mr. Powell finds it easier to take it out of mothers, childrens and sick people than to take on this vast industrie," Mr. Brown commented icily. ,,Let us have a full inquiry into the cost of drugs and the pharmaceutical industry." The health of children today owed much to the welfare food scheme. It was maintained during the war. Now in conditions of Tory affluence it seemed it could not be carried on.

Name: Alexander Dobrof

Proposal

Domain Background

Coming from a family of teachers I was exposed to the challenges from the perspectives of an educator in the American education system. One thing that is true today as it was a decade ago is that the individual workload for teachers is very high, from lesson planning, parent-teacher conference, teaching, and last but not least - grading. I always considered grading assignments as one of the most tedious and burdensome parts of teaching, where it takes away valuable time from teachers to spend with their students. There are a variety of solutions that are available for teachers, but most require the teachers to structure their assignments to the tools "format" or are cost prohibitive. I would like to propose a grading tool that conforms to each teacher's process but still provides the modern conveniences of tracking student performance of an assessment platform.

Most teachers still create their assignments on paper. Paper is still used widely in the education space because it's easy to create, it's versatile, and it's economical. Because of its versatility, it creates a challenging problem on creating a uniform grading system. One of the reason is that students have unique writing styles and few constraints are leverage on the students. To accomplish this, I proposed leveraging image recognition and deep learning to scan, analyze, and grade the assignments. For good reason, deep Learning has been the darling of machine learning in the past five years. Deep learning has been used across industry such as beating humans at Go, self-driving cars, diagnosing skin cancer, etc. Convolutional Neural Networks (CNN), a deep learning model, is a great choice to explore this problem due to automatic feature extraction and has been deployed across various image recognition problems.

Problem Statement

One of the most burdensome tasks in the education arena is the requirement to manually grade homework and test assignments. Even with massive investments made across the education space on integrated testing and grading platforms - a lot of assignments is still done on paper. By providing a tool that could take any paper assignment, extract the student's answer, record and grade it, the teacher will be able to track the student's capability down to each individual problem. The potential of tracking this information can be extremely valuable to teachers - from data analytics to help them improve lesson planning to target their student weaknesses. To approach this problem, I will be using computer vision and CNN model to analyze student answers. The CNN model will be used to extract words from scan images on their assignments.

Datasets and Inputs

The dataset I plan on using is the IAM Handwritten Dataset 3.0, the flagship database used for handwritten recognition assembled by the research group at University of Bern [1]. The source text was based on of Lancaster-Oslo/Bergen (LOB) corpus, which contains over a million English text. The database assembled handwritten text from over 600 different writers with over 10,000 unique and diverse words. The database contains forms of handwritten text with a scanned at a resolution of 300dpi and PNG images with 256 gray levels. Because words have varying lengths and characters have different sizes, the images come in at varying sizes. Therefore, it is required to preprocess the dataset to a uniform dataset size and a cleaner data.

Table 1: IAM Handwritten Dataset Metric

Labeled	Amount
Writers	657
Scanned Pages	1539
Isolated and Labeled Sentences	5687
Isolated and Labeled Text Lines	13,353
Isolated and Labeled Words	115,320

Solution Statement

To achieve the process, we will leverage deep convolutional neural network (CNN) to train and extract the features from IAM Handwritten Dataset. CNN has been a proven method used in many computer vision domains, including image classification, object detection, and facial recognition. The model will be using Keras on top of TensorFlow to build the different convolutional, max-pooling, dropout, and global pooling layer. The end results are to create a model that effectively predict the words across the assignment images that is pass to the model.

Benchmark Model

The model will be benchmarked against the results from Balci, Saadati, and Shiferaw [1]. They used three different CNN architecture to train the same dataset: VGG-19, RESNET-18, and RESNET-34. The goal is to have the model produce better accuracy than RESNET-34. Another metric that will be observed is the speed comparison for the different model set.

Table 1. Word Level and Character Level Classification Results [1]

Architecture	Training Accuracy	Validation Accuracy	Test Accuracy
VGG-19	28%	22%	20%
RESNET-18	31%	23%	22%
RESNET-34	35%	27%	25%

If the model doesn't reach a sufficiently high accuracy, I want to compare it against Google Vision API. Google Vision is the state of the art image recognition model that has been trained across many millions of different images from Google's' ML team. Suffice to say, I do not expect my model to outperform Google Vision, but it will show what a complete, production assignment grader can achieve with more data (either through a larger dataset or using reinforcement learning and feeding the images back into the dataset).

Evaluation Metrics

The evaluation metric for the word detection model will be how accurate the model is at predicting the correct word against the test images. Once the model is sufficiently trained, we will then run the model against actual homework assignment assembled from teachers.

Project Design

Model

The first step is to download the data collection from IAM Handwritten Database, which can be downloaded from their website. The IAM Handwritten Database originally is in a "form" images, but fortunately, the team also provided a breakdown of the files into sentences, lines, and words.

Due to the fact that the images come in different sizes, preprocessing is required to effectively train the model. To keep the images at the highest fidelity possible, instead of modifying the images by skewing or distorting it - I will instead pad all the images to the largest width and height dimensions. In addition, some words will require being rotated due to the fact that the forms did not have guided lines - some writer's started veering their writing upwards or downwards.

The neural network will be initial build with the same format that is used in the deep learning module from the dog breeding module. The model will use 5 convolutional layers and 4 fully connected layers (along with pooling and dropout within). This will be the first initial start, but the plan is to keep iterate amongst the different layers and even using transfer layers to improve the relative feature extractions.

The next step is to observe how the model will work amongst real-life assignments. I will collect samples from teachers in different grades and scan them in as images. The goal is to isolate the actual handwriting of the students and ignore the other parts of the images. To do this, I will be implementing background subtraction methods using OpenCV to isolate the actual answers. Afterward, I will regionalize the problem by cropping each problem to their specific borders and run the cropped image against the CNN model. The model will return the prediction probability for the image and it will then be validated against the answer key provided by the teacher. Finally, the data can be tabulated and graded.

References

1. U. Marti and H. Bunke. The IAM-database: An English Sentence Database for Off-line Handwriting Recognition. Int. Journal on Document Analysis and Recognition, Volume 5, pages 39 - 46, 2002.
2. Balci B., Saadati D., and Shiferaw D., “Handwritten Text Recognition using Deep Learning”
3. S. Johansson, G.N. Leech and H. Goodluck. Manual of Information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital Computers. Department of English, University of Oslo, Norway, 1978.
4. Machine Learning is Fun! Part 3: Deep Learning and Convolutional Neural Networks