Deep Learning on Data Visualization

Topic: Hand Written Digits/ Text Detection

Abstract:
Nowadays Deep learning has became a really important and popular tool for solving complex problems, performing data visualizations and helping researchers get deep insights on large and complex datasets. In recent times, Deep learning has been used for a lot of practical applications like Driverless cars, MRI analysis, Image retrieval, gaming, etc.Handwritten digits and text detection is computer vision based task that includes numerous practical applications, such as optical character recognition and document analysis(for legal and research purposes).

In this 5 page essay, we will be discussing the use of deep learning techniques for handwritten digits and text detection. This would include pre-processing techniques and datasets that can be used for training and testing of our models. We will also be exploring various Deep learning models, such as CNNs, Recurrent Neural Networks (RNNs), and GANs, and their application to handwritten digits and text detection.
With the help of Deep learning techniques we are continuously improving the accuracy and efficiency of OCR and document analysis applications.
In this paper we will also discuss about the fundamentals of deep learning and it's state-of-art.

Introduction:

a)What is Deep learning?

Deep learning has brought about a change by providing a powerful tool for learning patterns for a large datasets. Deep learning is a type of machine learning(ML) that involves the use of artificial NN to analyze data. Deep learning falls under Machine learning that uses NN to model and solve complex problems. The way a deep learning algorithm works is by breaking down the data into smaller parts or sections and then processing it through a series of network to create a prediction or classification. Deep learning uses ANN(artificial neural networks), these networks are capable of processing and transmitting data.

ANN consists of neurons that process and transmit data to other neurons in the network. These connections or neurons are weighted allowing the network to learn based on the

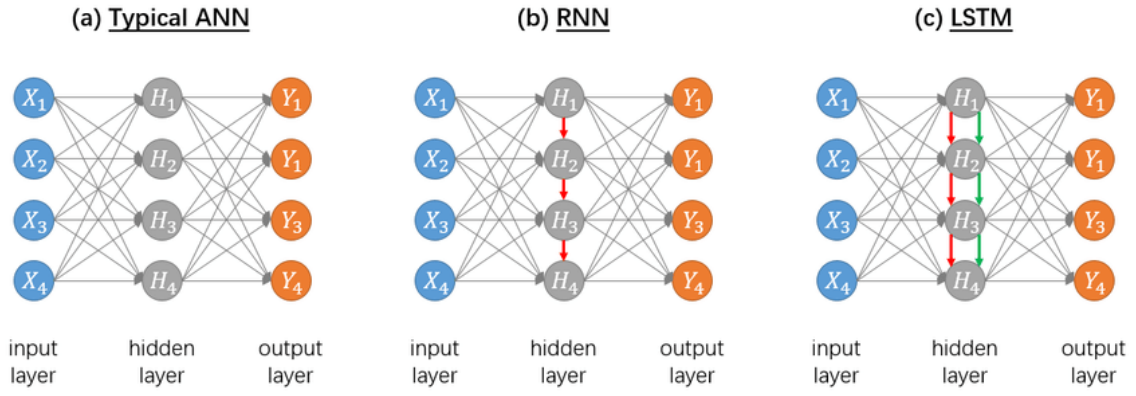

figure 1.1

input data from the neuron. There are several types of ANN each with its own architecture and formula. The most common ANN's that we have discussed in our lectures are FFNN(Feedforward neural networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural networks) and LSTM(Long short-term memory network) *(as shown in the figure 1.1)*

b)State of the art (How good is it?)

Deep learning has massively evolved over the recent years in the field of AI.Deep learning's state-of-the-art has also evolved greatly in turn leading to remarkable developments in a variety fields like Speech recognition, NLP, and Computer vision (OpenCV). These models have the ability to accomplish complicated tasks like speech and image recognition while also achieving high accuracy. Thanks to the feature extraction method in deep learning. All this was a bit challenging to do back in the days.

In NLP, RNN's have achieved state-of-the-art performance in tasks such as language translation, language modeling, and sentimental analysis. We have also seen the popular deep learning architectures such as LeNet, AlexNet,ResNet and GAN in our lectures.
The state-of-the-art methods for handwritten digit recognition involve convolutional neural networks (CNNs), which have achieved very high accuracy on benchmark datasets such as MNIST, USPS, and SVHN. Some popular architectures include LeNet-5, AlexNet, VGG, ResNet, and Inception. In this essay we will be looking at MNIST dataset for Handwritten digits and text detection. *(Figure 1.2)*

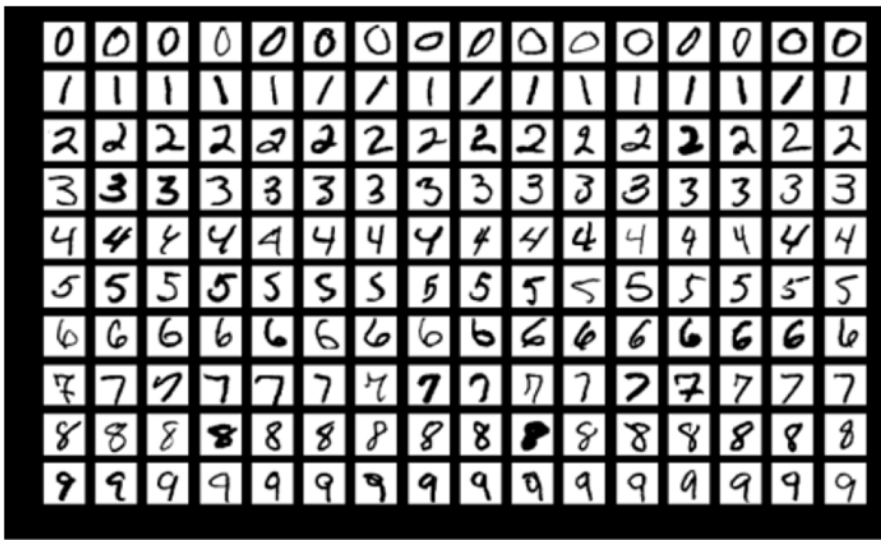

Figure 1.2- MNIST dataset sample.

State-of-the-art methods typically use deep learning techniques such as CNNs and recurrent neural networks (RNNs) if dealing with text detection model cases. One of the most popular methods for this problem is the Faster R-CNN framework and the YOLO (You Only Look Once) methods which are known for their speed and accuracy.
The state-of-the-art in deep learning has produced outstanding results in a number of deep learning fields, making it a potent tool for resolving challenging AI issues.
(Figure 1.3)

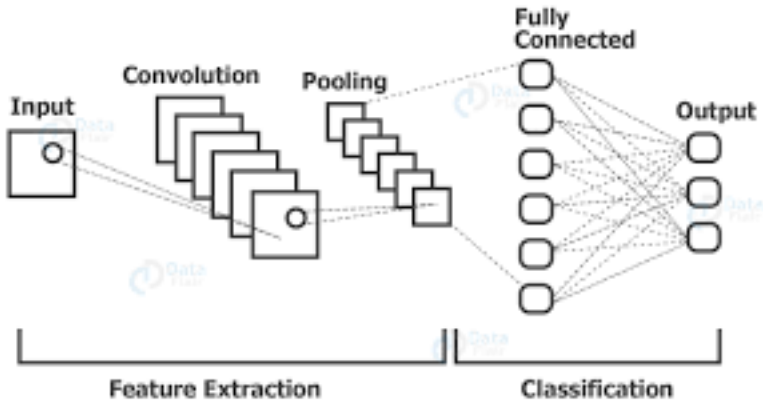

Figure 1.3: Handwritten character recognition using deep learning

c)Achievements of Deep Learning in Compute visualisations

Deep learning has enabled computers to understand and interpret visual information. In this way a great change has been brought like never before. We will be discussing some of the key achievements of deep learning in computer vision, including edge detection, segmentation, and image classification.
Edge detection is a fundamental tool in CV that involves identifying boundaries between objects in an image. There are various Edge detection methods, such as the Canny edge detector, Sobel Edge detector, Prewitt Edge detector, etc. Deep learning-based segmentation models like the Fully Convolutional Network (FCN) have greatly achieved significant improvements in accuracy as compared to other methods in the old times. Image classification is the task of assigning a label to an image from a predefined set of classes.Deep learning-based object detection models like the the region-based convolutional neural network (R-CNN) has also achieved higher accuracy compared to traditional object detection methods.

Handwritten digit and text recognition are the top most important applications of deep learning. Thanks to the advanced deep learning techniques like the CNN and RNN that now in the recent years, we see a significant progress in this technique. Common dataset for handwritten digit recognition is the MNIST dataset (link in the references) which contains 60,000 images for training and 10,000 for testing of handwritten digits. Several deep learning architectures have been tried and tested for the same task but the most successful of them all was the CNN architecture.

The general formula for a convolutional layer in a CNN is as follows:

$$h_i = f\left(\sum_{j=1}^{n} w_j x_{i+s_j} + b\right)$$

h_i :ioutput activation of the i-th neuron in the layer
f: activation function, such as ReLU or sigmoid
n: number of input neurons to the i-th neuron in the layer
w_j: weight associated with the j-th input neuron
x_{l+s_j}: input activation of the j-th neuron, which is offset by the stride s_j
b: bias term

Handwritten text recognition is a more challenging task than digit recognition because it involves recognizing variable-length sequences of characters. RNNs, and more specifically, the long short-term memory (LSTM) and gated recurrent unit (GRU) architectures, have been used successfully for this task. The reason why these models are well suited for handwritten text recognition is because the models mentioned above can process various sequences of the input data and also produce sequences of output data.

The formula for an RNN can be expressed as:

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = g(W_{hy}h_t + b_y)$$

IAM dataset is one of the most well-known datasets for handwritten text recognition that contains more than 7,000 pages of handwritten text. CNN-RNN architecture has been the most successful approaches amongst several deep learning approaches.

Overall, deep learning has made significant contributions to the field of handwritten digit and text recognition. We can see that the performance of these models continues to improve on a daily basis. New techniques and architectures are to be seen.
Edge detection, segmentation, image classification, object detection, and generative models are just a few of the areas where deep learning has made significant contributions. These achievements have opened up new possibilities for applications in healthcare, robotics, and autonomous vehicles fields.

Topic: How does it work?
Deep learning model for handwritten digit and text recognition typically involves a combination of CNN and RNN architecture. The first step involves pre-processing of the input image of the handwritten digit or text by converting it to a suitable format i.e. grayscale and normalized to a fixed ratio. The pre-processed image is then fed into a CNN, which is designed to automatically learn feature representations from the image. The CNN consists of multiple layers of convolutions, pooling, and nonlinear activations, which progressively extract features from the input image. The output of the CNN is a high-level feature representation of the input image.This feature representation is then fed into an RNN model, which is designed to capture the sequential dependencies within the different parts of the input image. The RNN consists of a series of recurrent layers, which maintain a hidden state that is updated at each time step. The hidden state captures the information from the previous time step and combines it with the current input to produce a new hidden state. The output of the RNN is a sequence of hidden states that represent the temporal dynamics of the input image. Finally, the output of the RNN is fed into a fully connected layer, which maps the hidden states to the output classes, such as the digits 0-9 or different characters in a text. The fully connected layer is trained with an activation function to produce a probability distribution over the possible output classes. The entire model is trained end-to-end using back propagation and gradient descent to minimize the loss function between the predicted and true output labels.*(As shown in Figure 1.4)*

Once the following model is trained, it its then used to recognize handwritten digits/text from new input images by applying the same pre-processing steps and feeding the input into the trained model.

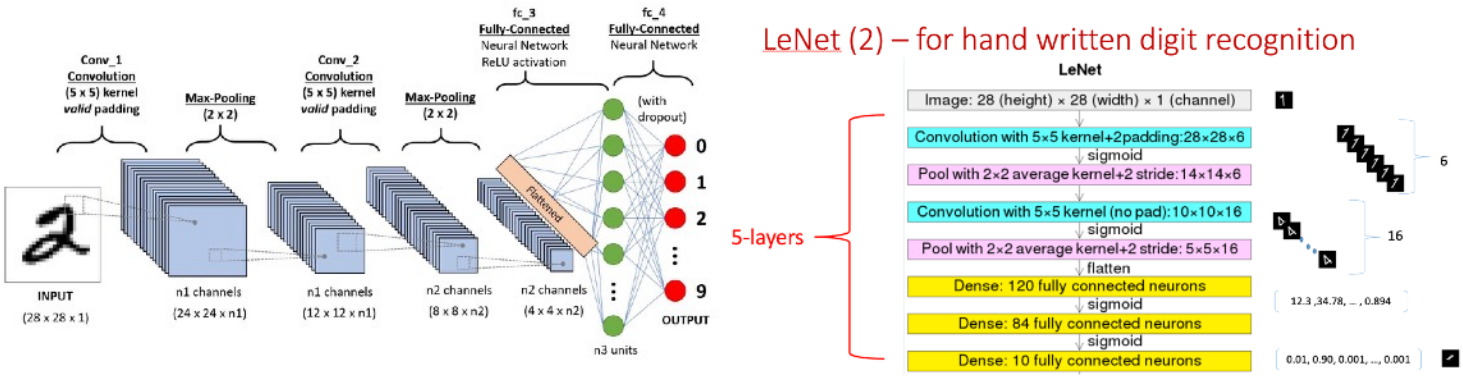

Figure 1.4 Working of the Handwritten digit detection.

Conclusion:

Deep learning has become an essential tool for solving complex problems and obtaining insights from large and complex datasets.We can see its successful applications in various fields like OpenCV, NLP, and speech recognition. It has achieved spectacular results in tasks such as image classification, object detection, handwriting recognition and various other ML tasks.

In this essay we explored various deep learning models, including ANNs, CNNs, RNNs, and GANs, along with their application to handwritten digits and text detection. The state of the art in deep learning has advanced largely.

However, deep learning models require large amounts of data and computing power to train.They also come with some limitations such as interpretability that can be challenging, but I'm sure it will be addressed in the coming future.

References

1) A Comparative Study on Handwriting Digit Recognition Using Neural Networks: https://ieeexplore.ieee.org/document/8109042
2) https://ijcrt.org/papers/IJCRT2106267.pdf
3) MNIST handwritten digit recognition dataset: http://yann.lecun.com/exdb/mnist/
4) USPS handwritten digit recognition dataset: https://www.kaggle.com/bistaumanga/usps-dataset.
5) SVHN dataset for digit recognition: http://ufldl.stanford.edu/housenumbers/
6) https://www.irjmets.com/uploadedfiles/paper//issue_3_march_2022/19432/final/fin_irjmets1646225071.pdf
7) https://paperswithcode.com/sota
8) https://medium.com/swlh/deep-learning-the-state-of-the-art-88153e69a602

Datasets for Handwritten Text Detection:

1. Faster R-CNN: https://arxiv.org/abs/1506.01497
2. YOLO: https://arxiv.org/abs/1506.02640
3. SSD: https://arxiv.org/abs/1512.02325
4. DETR: https://arxiv.org/abs/2005.12872
5. MNIST Dataset link: Mnist Dataset