**INTRODUCTION**

There is an ever-increasing amount of image data in the world, and the rate of growth itself is increasing. Info trends estimates that in 2016 still cameras and mobile devices captured more than 1.1 trillion images. According to the same estimate, in 2020 the figure will increase to 1.4 trillion. Many of these images are stored in cloud services or published on the Internet. In 2014, over 1.8 billion images were uploaded daily to the most popular platforms, such as Instagram and Facebook.

Going beyond consumer devices, there are cameras all over the world that capture images for automation purposes. Cars monitor the road, and traffic cameras monitor the same cars. Robots need to understand a visual scene in order to smartly build devices and sort waste. Imaging devices are used by engineers, doctors and space explorers alike.

To effectively manage all this data, we need to have some idea about its contents. Automated processing of image contents is useful for a wide variety of image-related tasks. For computer systems, this means crossing the so-called semantic gap between the pixel level information stored in the image files and the human understanding of the same images. Computer vision attempts to bridge this cap.

**Problem statement**

Digits contained in image files can be located and identified automatically. This is called object detection and is one of the basic problems of computer vision. As we will demonstrate, convolutional neural networks are currently the state-of-the-art solution for object detection. The main task of this thesis is to review and test convolutional object detection methods.

**Technology and Concepts**

**Machine Learning**

Learning algorithms are widely used in computer vision applications. Before considering image related tasks, we are going to have a brief look at basics of machine learning.

Machine learning has emerged as a useful tool for modelling problems that are otherwise di\_cult to formulate exactly. Classical computer programs are explicitly programmed by hand to perform a task. With machine learning, some portion of the human contribution is replaced by a learning algorithm. As availability of computational capacity and data has increased, machine learning has become more and more practical over the years, to the point of being almost ubiquitous.

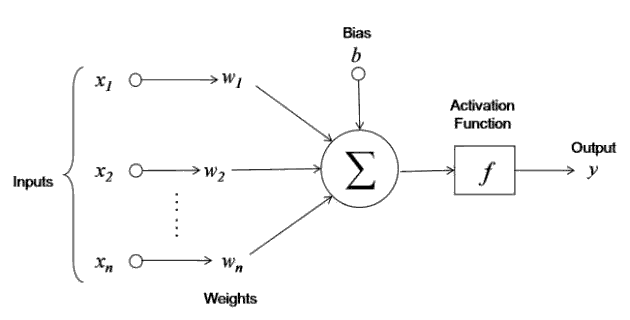
It can be used in two ways:

* *Supervised Learning*
* *Unsupervised Learning*

**Neural networks**

Neural networks are a popular type of machine learning model. A special case of a neural network called the convolutional neural network (CNN) is the primary focus of this thesis. Before discussing CNNs, we will discuss how regular neural networks work.

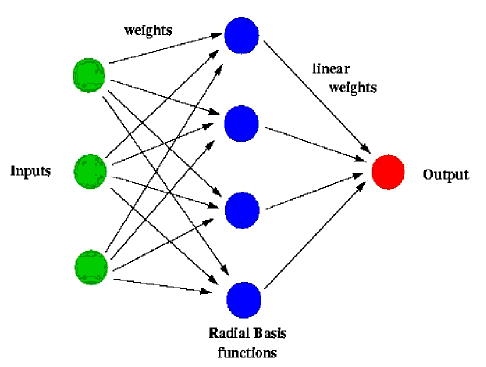
Neural networks were originally called arti\_cial neural networks, because they were developed to mimic the neural function of the human brain.



The neuron is trained by carefully selecting the weights to produce a desired output for each input.

**Multi-layer networks**

A neural network is a combination of arti\_cial neurons. The neurons are typically grouped into layers.

A multi-layer network typically includes three types of layers: an input layer, one or more hidden layers and an output layer. The input layer usually merely passes data along without modifying it. Most of the computation happens in the hidden layers. The output layer converts the hidden layer activations to an output, such as a classification. A multilayer feed-forward network with at least one hidden layer can function as a universal approximator.

**Computer vision**

Computer vision deals with the extraction of meaningful information from the contents of digital images or video. This is distinct from mere image processing, which involves manipulating visual information on the pixel level. Applications of computer vision include image classification, visual detection, 3D scene reconstruction from 2D images, image retrieval, augmented reality, machine vision and traffic automation.

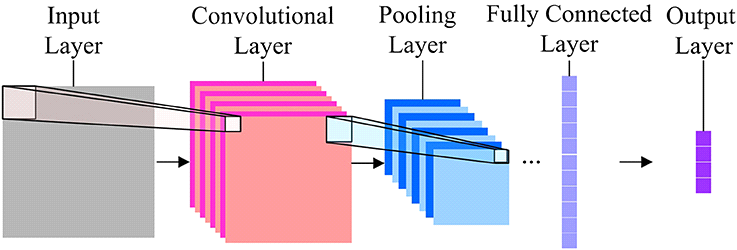
**Object detection**

Object detection is one of the classical problems of computer vision and is often described as a difficult task. In many respects, it is similar to other computer vision tasks, because it involves creating a solution that is invariant to deformation and changes in lighting and viewpoint. What makes object detection a distinct problem is that it involves both locating and classifying regions of an image [20]. The locating part is not needed in, for example,

whole image classification.

To detect an object, we need to have some idea where the object might be and how the image is segmented. This creates a type of chicken-and-egg problem, where, to recognize the shape (and class) of an object, we need to know its location, and to recognize the location of an object, we need to know its shape. [53] Some visually dissimilar features, such as the clothes and face of a human being, may be parts of the same object, but it is difficult to know this without recognizing the object first. On the other hand, some objects stand out only slightly from the background, requiring separation before recognition.

**Convolutional neural networks**

The basic idea of the CNN was inspired by a concept in biology called the receptive field. Receptive fields are a feature of the animal visual cortex. They act as detectors that are sensitive to certain types of stimulus, for example, edges. They are found across the visual field and overlap each other.

2.1 Description of the dataset

In this paper, we used the MNIST database consisting of offline handwritten digits ranging from 0-9. The database was constructed from Special Database 3 (SD-3) and Special Database 1 (SD-1) that contain binary images of handwritten digits. SD-3 was collected among Census Bureau employees, while SD-1 was collected among high-school students. For the results to be independent of both datasets, MNIST dataset was built by mixing NIST SD-1 and SD-3. The total number of digit image samples (70,000), the total number for training (60,000) and testing (10,000), and the subtotal number for each digit are shown in table 1. Each digit is a gray-level fixed-size image with a size of 28 x 28 (or 784 pixels) in total as the features.

275 *Proceedings of the International Conference on Industrial Engineering and Operations Management* Washington *DC, USA, September 27-29, 2018.*

© IEOM Society International

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1.** Dataset Digits | # Training | # Testing | Subtotal |
| 9 | 5949 | 1009 | 6958 |
| 8 | 5851 | 974 | 6825 |
| 7 | 6265 | 1028 | 7293 |
| 6 | 5918 | 958 | 6876 |
| 5 | 5421 | 892 | 6313 |
| 4 | 5842 | 982 | 6824 |
| 3 | 6131 | 1010 | 7141 |
| 2 | 5958 | 1032 | 6990 |
| 1 | 6742 | 1135 | 7877 |
| 0 | 5923 | 980 | 6903 |
| Total | 60,000 | 10,000 | 70,000 |

inal classification.

**Drawbacks**

In 2015 paper for Fast R-CNN [20], Girshick lists three main problems of R-CNN:

* Training consists of multiple stages, as described above.
* Training is expensive. For both SVM and region proposal training, features are extracted from each region proposal and stored on disk. This requires days of computation and hundreds of gigabytes of storage space.
* The most important, object detection is slow, requiring almost a minute for each image, even on a GPU.

**Fast R-CNN**

Fast R-CNN published in 2015 by Girshick provides a more practical method for object recognition. The main idea is to perform the forward pass of the CNN for the entire image, instead of performing it separately for each **RoI**.

*Proceedings of the International Conference on Industrial Engineering and Operations Management* Washington *DC, USA, September 27-29, 2018 .*

© IEOM Society International

removing the noise as edge is an important aspect of an image. Figure 3 shows the processed image of the median filter. An unwanted horizontal line connected to number “Zero” is removed after Median Filter.

**Figure 3.** Raw Image (Left) and image after Median Filter (Right)

2.3.3 Image Sharpening

The next step in preprocessing is the image sharpening technique which uses a blurred, or "unsharp", negative image to create a mask of the original image. The unshaped mask is then combined with the positive (original) image, creating an image that is sharper than the original. Sharpening uses a filter that amplifies the high-frequency components of a signal. It’s a necessary step taken after Median Filter as Median Filter not only remove noise, but also weaken the entire image in general. Sharpening can restore or enhance some of the useful information weakened by Median Filter. Figure 4 shows the image sharpening after processing Median Filter.

**Figure 4.** Median Filter and sharpened image

2.3.4 Image Attribute Reduction

Attribute reduction techniques is done to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. That is, mining on the reduced data set should be more efficient yet produce the same (or almost the same) analytical results. Each original image has a total of 784 attributes. It would be beneficial to reduce the total attributes to a relatively small amount so that it’s more data efficient and easier to be processed. A direct way to reduce the attribute is by dividing the image into each block and finding the mean of that block as one attribute. Each block can be treated as one attribute.

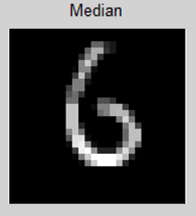
The number of blocks determines the number of attributes. This is done for 2 x 2, 4 x 4, 7 x 7, and 14 x14 attributes (blocks) of the image. Accuracy is improved as the no of blocked are increased (proved later in this report).

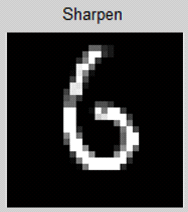
**Experimental setup**

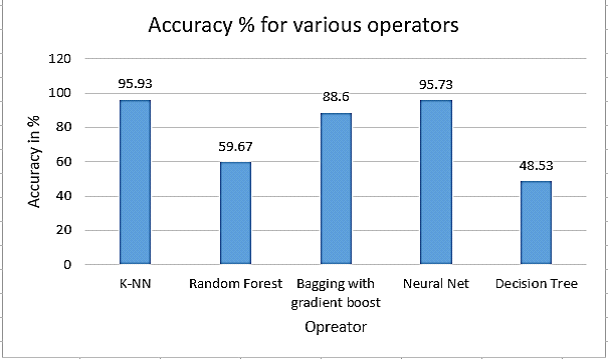
We selected Fast R-CNN as the core method for object detection experiments. Fast

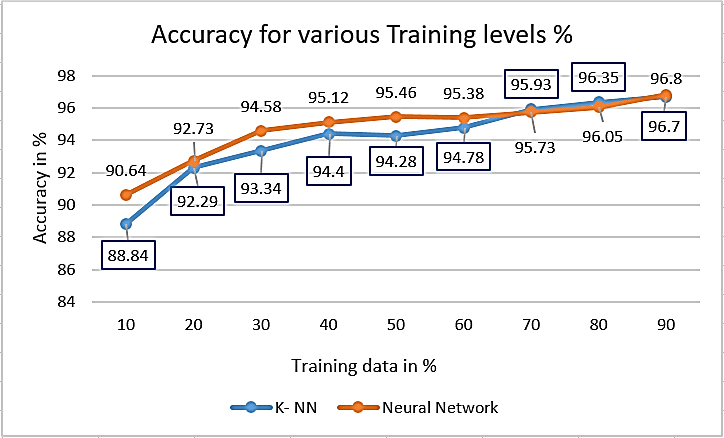
R-CNN is already well-established and has implementations and pretrained networks available for several di\_erent platforms. Even though the advanced methods, such as Faster R-CNN, provide a minor increase in accuracy, their main contribution is improved speed. Since the evaluation of execution-time was mostly left outside the scope of this thesis, these methods would have provided little additional value to the experiments.

Because of the implementation environment (which is described in more detail in the following chapter), we chose to skip the initial training and use a pretrained ResNet50 Faster R-CNN model.









3.3.1 Difficulties

Some of the difficulties faced during the preprocessing were discussed in this section. The figure 12 above on the left shows the handwritten image which is not even recognized by us. The images in the middle and the right were handwritten digit one and zero, when we preprocess these images in Bewilder one is recognized as 2 (2 (objects) - 0 (holes) = 2) and zero is recognized as one (1 (object) - 0 (hole) = 1).

282 *Proceedings of the International Conference on Industrial Engineering and Operations Management* Washington *DC, USA, September 27-29, 2018 .*

© IEOM Society International

**Figure 12.** Handwritten Images (From Left) - Unknown No. 8, No. 1, and Zero

In these cases, bewilder is not an effective tool to preprocess the handwritten image. Because of the errors in the raw dataset the accuracy of the classifier is affected.

**Figure 13.** Problem in Noise Removal stage

Another problem which we faced in this project is during the preprocessing stage. Figure 13 shows the Handwritten Image One and Three. In there, the median filter blurs the image. Sharpening helps restoring some details back after median filter, but the image is still blurred in an extent as compared to the raw image which exacerbates the quality of good image.

Another observed problem is the use of Binary technique. Binary technique transforms all image pixels to either 0 or 1. It simplifies the data and offers clean looking of the data if it is controlled well. However, from Figure 13 we can see that the good image (one and three) are in fact damaged by doing Binary process. That’s because the binary threshold was not set correctly, and image loss was resulted in. As it’s difficult to se

**Conclusions**

This paper has practiced different machine learning technique and different models for data training attempting to discover a representation of isolated handwritten digits that allow their effective recognition and to achieve the highest accuracy of predicting handwritten numeral. Thus, this study settled on classifying a given handwritten digit image as the required digit using five different algorithms and consequently testing its accuracy. This study built handwritten recognizers evaluated their performances on MNIST (Mixed National Institute of Standards and Technology) dataset and then improved the training speed and the recognition performance. In addition to develop a system for word based handwriting recognition system and test the handwriting of a given word and detect the writer by selecting which is being recognized for most of the user for a given training sample. This study discusses in detail all advances in the area of handwritten character recognition. The most accurate solution provided in this area directly or indirectly depends upon the quality as well as the nature of the material to be read. Various techniques have been described in this paper for character recognition in handwriting recognition system. The result of this study shows that accuracy is improved as the no of blocked are increased. Apart from that, a 4x4, 7x7 and 14x14 attribute reduction is performed separately to compare and find the optimal number of attributes that best represent the image. The initial hypothesis was answered with concrete data support. Comparing all the classification models we tested K-Nearest Neighbor is the preferred choice in terms of its high accuracy and computational efficiency. However, there is no single classifier that works best on all given problems. A result shows that probabilistic methods suit better for handwriting recognition. By varying the training and testing ratios (from 10% to 90%) we found that the larger training data size improves accuracy, but smaller testing dataset may also favor better accuracy. Preprocessing such as Attribute reduction (784 reduced to 196) reduce runtime and increase accuracy (from 93.2% to 95.9%). The proposed algorithm tries to address both the factors and well in terms of accuracy and time complexity. The general accuracy of tested K-NN was found to be 96.7% while 96.8% were achieved by Neural Network respectively. The overall highest accuracy 96.8% is achieved in the recognition process by Neural Network with the sacrifice of significantly extended runtime. This work is carried out as an initial attempt, and the aim of the paper is to facilitate for recognition of handwritten numeral without using any standard classification techniques. Image processing techniques Median filter, binary, Bweuler, and sharpening improve image quality, but we should be

**Bibliography**

* [https://www.kaggle.com](https://www.kaggle.com/)
* [https://www.analyticsvidhya.com](https://www.analyticsvidhya.com/)
* [https://www.medium.com](https://www.medium.com/)