Using Convolutional Neural Network for Image Classification and Segmentation

Ashara Emmanuel Chimaobi, Lu Ye
School Of Information and Electronic Engineering
Zhejiang University of Science and Technology, Hangzhou, CN.

Abstract
Image segmentation and Image Classification are two fundamental tasks in computer vision. In this thesis, a novel segmentation algorithm based on a deformable model and robust estimation is introduced to produce reliable segmentation results. The algorithm is extended to handle touching objects and partially occluded image segmentation. Although current conventional image classification methods have been widely applied to realistic problems, there are some issues with their implementation, including unsatisfactory results, poor classification accuracy, and a lack of adaptive capacity. This approach has been used to isolate the two processes of image feature extraction and classification into two stages. The deep learning model possesses a strong learning capability, which enables it to incorporate the feature extraction and classification processes, thereby improving the image classification accuracy. This thesis explores various machine learning methods to improve the model's performance. The primary objective is to discover the accuracy of the various networks on the datasets and to evaluate the consistency of each of these deep learning predictions. Nonetheless, there are limitations to this approach: first, it is difficult to perform accurate approximation in the advanced model. The second point is that the deep learning model comes with poor accuracy in its classifier. So, this paper introduces the idea of using different datasets and models of the deep learning network and comprehensively utilizes it to determine the best test accuracy for the images. In this paper, a deep neural network primarily based on Keras and TensorFlow is deployed using python. The two datasets are used to compare to determine which has the maximum accurate and fine time for processing. And a VGG-16 model method based on the optimized kernel function is proposed to replace the classifier in the deep learning model. The experimental results show that the proposed method not only has higher average accuracy than other mainstream methods but also can be good adapted to various image databases. Compared with other deep learning methods, it can better solve the problems of complex function approximation and poor classifier effectiveness, thus further improving image classification accuracy.
 Keywords: Image Classification, Deep Learning, TensorFlow
DOI: 10.7176/CEIS/13-1-03
Publication date: January 31st 2022

1.0 INTRODUCTION

Recently, image classification is growing and becoming a trend among technology developers especially with the growth of data in different parts of the industry such as e-commerce, automotive, healthcare, and gaming. The most obvious example of this technology is applied to Facebook. Facebook now can detect up to 98% accuracy to identify your face with only a few tagged images and classify it into your Facebook album. The technology itself almost beats the ability of humans in image classification or recognition. One of the dominant approaches for this technology is deep learning. Deep learning falls under the category of Artificial. Image segmentation is the basis of computer vision and other high-level image processing and is also the key step of image recognition and registration. It separates the target from the complex background according to the characteristics of gray scale, texture, or shape in the image [1]. Image segmentation methods can be divided into the following two categories according to the principle of model design and the different information they depend on: one is based on image information, morphology, topology, partial differential equation, etc. The other is the segmentation method based on deep learning, the image segmentation method based on region selection, the image segmentation method based on RNN and the segmentation method based on up sampling. The partial differential equation method of image segmentation is one of the more successful methods. According to the image characteristics, an energy functional is defined for the evolution curve. To minimize the energy function, the numerical solution of the equation is then used to solve the equation, and the numerical solution of the equation is the desired segmentation curve [2]. This kind of method can not only deal with the change of topology structure of evolution curve effectively but also can deal with the given image directly and does not need a lot of training data, repeated adjustment, and network learning. Intelligence is where it can act or think like a human. Normally, the system itself will be set with hundreds or maybe thousands of input data to make the 'training' session to be more efficient and fast. It starts by giving some sort of 'training' with all the input data. Machine learning is also a frequent system that has been applied to image classification. However, there are still parts that can be improved within machine learning. Therefore, image classification is going to be occupied with a deep learning system.

Machine Vision has its context when it comes to Image Classification. This technology can recognize people,

objects, places, actions, and writing in images. The combination of artificial intelligence software and machine vision technologies can achieve the outstanding result of image classification. The fundamental task of image classification is to make sure all the images are categorized according to their specific sectors or groups. Classification is easy for humans but it has proved to be a major problem for machines. It consists of unidentified patterns compared to detecting an object as it should be classified into the proper categories. The various applications such as vehicle navigation.

1.2 Purpose of the study

In this paper, we focused on image thresh holding which is mainly used in the pre-processing Classification and segmentation stages respectively, where our implementation is performing well enough in comparison to existing work), followed by secured transmission of the image data between multiple platforms and to the best of our knowledge this design belongs to a class of advanced implementation. Image Classification and segmentation help determine the relations between objects, as well as the context of objects in an image. Applications include face recognition, number plate identification, and satellite image analysis. Industries like retail and fashion use image segmentation, for example, in image-based searches. Image classification and segmentation are two fundamental problems in image analysis. Segmenting an image consists in dividing the image into homogeneous zones delimited by boundaries to separate the different entities visible in the image. The classification consists in labeling the various components visible in an image.

2.0 LITERATURE REVIEW
2.2 Theoretical Framework of Machine Learning

Machine learning techniques are nowadays routinely used in commercial systems for speech recognition, computer vision, and spam detection. To date, the primary theoretical advances in machine learning have been for passive supervised learning problems [1], where a target function (a classification rule) is estimated using labeled examples only. For example, in spam detection, and automatic classifier to label emails as "spam" or "not spam" would be trained using a sample of previous emails labeled by a human user. The goal here is then to get as high accuracy as possible using as little labeled data as possible. For most modern practical problems, however, there is often useful additional information available in the form of cheap and plentiful unlabeled data: e.g., unlabeled emails for the spam detection problem. As a consequence, there has recently been substantial practical interest in using this unlabeled data together with labeled data for learning, since any useful information that reduces the amount of labeled data needed can be a significant benefit. A variety of algorithms for doing this have been developed, and many successful experimental results have been reported. Some of these algorithms simply use raw unlabeled data in addition to labeled data, while others interact with the human labeler and adaptively identify specific informative unlabeled examples to be labeled. In parallel with this work, as the types of applications of machine learning have grown more and more diverse, the issue of how to represent data to the learning algorithm has become increasingly crucial. Typically, this representation is done using features: for example, for spam detection one might represent an email message by features indicating the presence or absence of various keywords in the message. However, in many cases, the problem of identifying high-quality features can in itself be quite difficult. This has led to the development of a powerful technique known as kernel methods. Kernel methods allow the user to specify a particular kind of pairwise function between data objects, known as a kernel function, which is used by the algorithm instead of explicit features. An example of a typical kernel for document classification would be the number of content words shared in common between two documents. Many well-understood and well-optimized algorithms such as SVMs can be used with kernels, allowing for their application to complex types of data. 2 Overall, incorporating unlabeled data in the learning process, adding interaction capabilities to the learning algorithm, and using kernels and similarity functions, are all areas that have been extensively explored in the machine learning community over the past few years. However, their theory has been lacking in several substantial ways. The idea of image retrieval was created by the Database Management community in a conference organized [2]. The early schemes consisted of annotating the images by text to consequently use database management systems. However, these kinds of approaches have problems in generating descriptive texts for large collections of images. Automatic generation is not yet feasible, thus a lot of labor is required to manually annotate the images, which is an expensive task. Moreover, manually image annotations (e.g. metadata) are affected by the subjectivity of human perception and different people might perceive images in different ways. Faced with these limitations, the computer vision (CV) community introduces [3]. Image Classification is based on visual features of images and puts textual representations in the second plan. After that and until today, a lot of new techniques were developed around the concept of Image Classification.
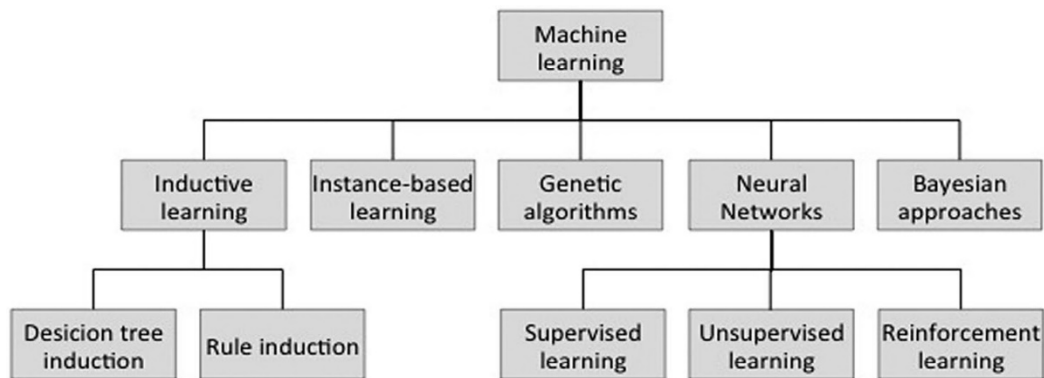
Figure 1. Classification of main Ml techniques according to [2].

Active learning is often used for problems where it is difficult (expensive and/or time-consuming) to obtain labeled training data. The advantage is to be able to achieve good performance needing less training data than other learners due to the sequentially identified useful examples by the active learner. Active learning is mostly applied within supervised ML scenarios but was also found to be of value within certain RL problems). Some researchers like [4] focus only on supervised classification techniques and group NN as a learning algorithm as part of supervised learning. However, NN algorithms can also be applied in unsupervised learning and RL [5]; [2]. This corresponds basically with [2] when the notion on top of the hierarchy is seen as 'Supervised ML' instead of the 'Machine learning' they originally stated.

2.2.1 Unsupervised Machine Learning

Unsupervised machine learning is another large area of research. The defining attribute is that within unsupervised learning, there is no feedback from an external teacher/ expert. The algorithm itself is supposed to identify clusters from existing data based on, e.g. conceptual cohesiveness of attributes.[6] introduced the rule that if instances are unlabeled (no known labels and corresponding correct outputs), it is most likely unsupervised learning. The goal is to discover unknown classes of items by clustering **[7]**whereas supervised learning is focused on classification (known labels). Unsupervised ML describes any ML process that tries to learn 'structure in the absence of either an identified output [e.g. supervised ML] or feedback [e.g. RL]. Three typical examples of unsupervised learning are clustering, association rules, and self-organizing maps' 8].

Especially in the Big Data context, unsupervised methods are becoming increasingly important. However, as in manufacturing applications, the main assumption is that experts can provide feedback on the classification of states to identify the learning set to train the algorithm [9]. Thus, the focus will be laid on supervised methods. However, some aspects of unsupervised learning may be beneficial in manufacturing applications after all. First, there is the possibility that in some cases there might be no expert feedback available or, in the future, desirable. Another aspect is to realize hybrid approaches, combing the 'best of both worlds' which gain importance due to the fast increase in unlabeled data especially in manufacturing [8]. And finally, unsupervised methods can be and are being used to, e.g. identify outliers in manufacturing data.

2.2.2 Reinforcement Learning

RL is defined by the provision of the training information by the environment. The information on how well the system performed in the respective turn is provided by a numerical reinforcement signal [10]. Another defining characteristic is that the learner has to uncover which actions generate the best results (numerical reinforcement signal) by trying instead of being told. This distinguishes RL from most of the other ML methods. However, RL is seen by some researchers as 'a special form of supervised learning' [12]. However, different from supervised learning problems, RL problems can be described by the absence of labeled examples of 'good' and 'bad' behavior. RL, based on the sequential environmental response, emulates the process of learning of humans. This 'reward signal,' which can be perceived in RL differentiates it from unsupervised ML.

A very specific challenge for RL is the tradeoff between exploration and exploitation. To achieve the goal, the agent has to 'exploit' the actions it learned to prefer and to identify those it has to 'explore' by actively trying new ways. In manufacturing, RL is not widely applied and just a few examples of successful applications exist as of today [19]. In the majority of manufacturing applications today, expert feedback is available. Therefore, even though RL is applicable in manufacturing applications, the focus in the following is on supervised techniques. The Reinforcement Learning setting introduces several challenges that we need to overcome, and potentially make trade between. That is, can it learn whether some actions are good/bad in previously unseen states?, We also need to consider delayed consequences of the agent's actions, that is, if it receives a high reward, was it because of an action it just took, or because of an action taken much earlier?

2.2.3 Supervised machine learning

This is partly due to the availability of (a) expert feedback (e.g. quality) and (b) the labeled instances. Supervised ML is applied in different domains of manufacturing, monitoring, and control being a very prominent one among

them [11] The general process of supervised ML contains several steps handling the data and setting up the training and test data-set by the teacher, hence supervised [13]. Based on a given problem, the required data are identified and (if needed) pre-processed. An important aspect is the definition of the training set, as it influences the later classification results to a large extent. Even so, it often appears as if the algorithm selection is always following the definition of the training data-set, the definition of the training data also has to take the requirements of the algorithm selection into account. Some algorithms allow for a so-called 'kernel selection' to adapt the algorithm to the specific nature of the problem. This highlights the adaptability of ML applications and the variety of problems that can be tackled. Similar requirements stand to some extent also true for the identification and pre-processing of the data as different algorithms have certain strengths and weaknesses concerning the handling of different data sets (e.g. format, dimensions, etc.

2.3 Artificial Intelligence

Artificial Intelligence (A I) is any technique that aims to enable computers to show similar human behavior, including, machine learning, natural language processing (NLP), speech synthesis, artificial vision, robotics, analysis sensor, optimization, and simulation. Machine learning (ML) is a subset of artificial intelligence techniques by which computer systems can do this learn from previous experiences (i.e., observe data) and improve one's behavior for a particular task. ML techniques include support for vector machines (SVM), decision trees, Bayesian learning, k-means grouping, learn to join rules, regression, neural networks, and more. Artificial neural networks are a subset of ML techniques freely inspired by biological neural networks.

DL techniques are based on data that will be analyzed to make this data known through relevant information. For the problem being analyzed. Today, the culture of big data encompasses all disciplines and areas of research (including IT, medicine, finance, etc.) For your potential in all these areas. Generation change and harvesting the data also led to changes in the data processing. The definition of big data is characterized by many Vs, such as Volume, Speed, and variations, as well as truthfulness, variability, visualization, value, etc.

At first glance, the general public associates big data processing with distributed platforms like Apache and Hadoop Spark to shine and it also speeds up background processing and reasoning. In the new era of big data, data analytics must change. The nature of large-scale data requires new approaches and new tools to adapt it to different data structures, different spatial and temporal scales. A wave of large amounts of information, especially with the Variety function, to process ML and data mining algorithms requires a new transformation of parallel and distributed processing solutions that are computed efficiently and effectively.

2.3.1 Deep Learning

Machine learning technologies have been powering many aspects of our daily life for years. Things as content filtering, web search, adding recommendations, and object identification in images, among others. Traditional machine learning was limited in the way it processes data, as many functionalities needed specific programming to perform certain tasks, not being able to receive raw data and transform it into a suitable representation without human intervention. This disadvantage is where Deep Learning (DL) shines. DL is a machine learning subset characterized by being able to process raw data and automatically learn the features needed to perform determined tasks. This ability is based on stacking several non-lineal transformation modules that convert the raw input data into a higher-level, more abstract representation. These layers vary depending on the function wanted to be performed. For example, in classification tasks, the high-level layers will amplify relevant aspects dismissing the less important variations. In images, the first layers start by detecting edges in particular locations, the second ones detect edges independently of their location, the third ones assemble these edges into bigger combinations, and so on.

The important point of the previous paragraph is that the weights of these layers of the feature are not designed by humans. Instead, they are learned from the data by a general-purpose learning technique.

Computer vision techniques are used in such systems to recognize people. In the most general context, computer vision systems create a model of the world from digital images [14]. The scope of the model depends on the task the computer vision system aims to solve and can be predefined by the designer of the system or partially be learned from the available data. Most often computer vision systems can be decomposed into the following three parts.

2.4 Machine Learning

When the majority hear "Machine Learning," nothing comes to mind but robots or robotics: depending on who you ask, a trustworthy butler or a lethal Terminator. Machine learning, though, is situated just an ultramodern delusion, it's right here. Algorithms for artificial intelligence and machine learning aren't new. The AI field goes back to the fifties. One of the earliest machine learning systems was created by IBM researcher Arthur Lee Samuels: a self-learning software for playing checkers. Currently, the word machine learning was invented by him. In a paper printed in the Research and Development Journal of IBM in 1959, he explained his approach to machine learning.

In general, machine learning algorithms can model complex class signatures, accept a variety of input predictor data, and make no assumptions about data distribution (i.e., are nonparametric). A variety of revisions have commonly instituted that compared to conventional parametric classifiers, these approaches tend to yield higher accuracy, particularly for multifaceted data with a high-dimensional feature space the likes of [14]

ML, on the other hand, avoids beginning with a data model and instead uses an algorithm to learn the link between the answer and its predictors. The ML method, on the other hand, assumes that the mechanism of data generation is complex and uncertain and seeks to learn the response by analyzing inputs and answers and identifying dominant patterns. The machine learning workflow is described in the Figure below.
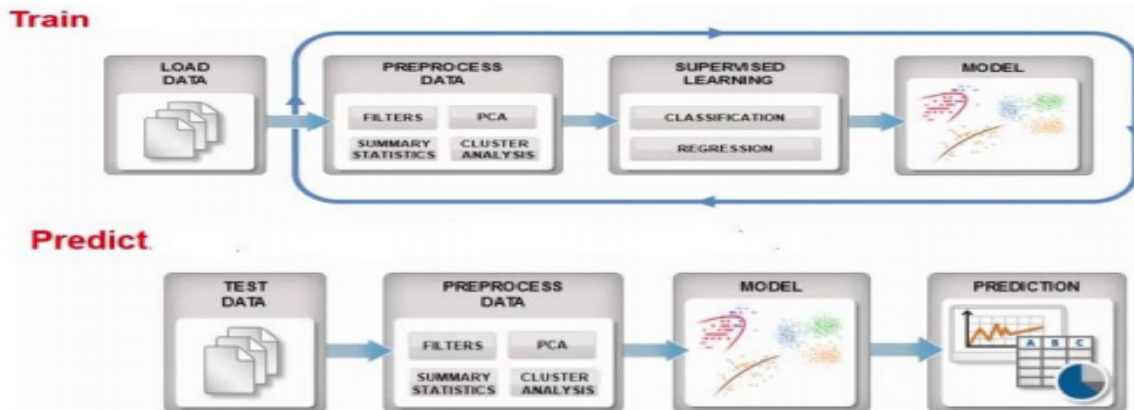

Figure 1. Machine learning workflow

Machine Learning focuses on what is expected, the ability of the model to predict well, and how to assess the success of the prediction. Machine learning structures are used in many facets of contemporary society: online search, content filtering on social networks, e-commerce website recommendations. ML is present today in consumer possessions such as cameras and smartphones. In computer vision, machine learning structures are used to transliterate speech into text, align news stories, articles, or products with the preferences of users, and pick suitable search results.

2.4.1 The Machine Learning Process

The application of DM in many areas of life has led to a cycle of the inter-industry standard data mining process (CRISP-DM 1999), The CRISP-DM cycle (Figure 1) consists of six phases:

1. Business insight is mainly based on the formulations provided for research and data description.

2. Understanding of the data is based on the submitted data and accompanying documentation.

3. Data preparation consists of data transformation, data analysis (EDA), and data engineering.

Features Each of them can be divided into smaller secondary passages; for example design, a feature is feature selection.

4. In the modeling phase, different ML algorithms with different parameter calibrations can be applied. Combining Variability of data and parameters can lead to extensive repetition of the train test model evaluation cycle, if the data to a large extent, the modeling phase imposes time-consuming and budget-intensive requirements Phase.

5. Evaluations can be conducted according to different criteria for a thorough examination of the AA model to select the best model for the implementation phase.

6. The implementation phase, also known as the production phase, involves the use of a trained ML model for its management functionality, as well as how to create a data pipeline in production. The group of the first five phases, called the development phase, can be repeated with different scenarios based on evaluation results. The implementation phase is crucial for the actual production below recurring requests; this includes evaluation, monitoring, model maintenance, diagnosis, and online retraining. The need to emphasize that ML algorithms learn from data. Therefore, in practice, the stages of understanding and preparation are expected and the data can take up a large part of the total DM project time.

2.5 Convolution Neural Network (CNN)

By creating multiple layers of convolution layers, CNN routinely extracts features of an image, ensuing in a hierarchy of characteristics. To learn about the image's local features, such as color and shape, the front convolution layer uses a smaller domain of perception while the back convolution layer uses a larger domain of perception, and from this data is then easily learned the consistent crop of image features (such as object size, location, and direction information). The neural network is a collection of artificial neurons interconnected with each other that exchange messages. [15] Connections have numerical weights which are tuned during the training process, so that when obtainable with an image or pattern to be identified, a suitably trained network responds correctly. The network consists of several features detecting "neuron" layers.

There are several neurons in each layer that respond to various input combinations from the previous layers. The trained model has layers stacked one on top of another with each one succeeding on the findings of the previous. CNN's usually use 5 to 25 different patterns matching layers as shown in the figure below.

2.7 Convolution Neural Network

CNN's are a type of Neural Network [15] designed to process data that come in the form of multiple arrays. This is useful for processing things such as natural language, audio, video, and images. Inspired by the animal visual cortex organization [16], they are developed around four concepts: local connections, shared weights, pooling, and the use of many layers.

Convolutional Neural Networks (CNNs, or ConvNets) [17] are neural network architectures specifically designed for handling data with some spatial topology (e.g. images, videos, sound spectrograms in speech processing, character sequences in text, or 3D voxel data). In each of these cases, an input example x is a multi-dimensional array (i.e. a tensor).
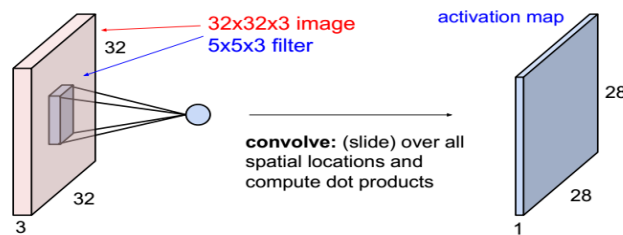

Figure 8: Illustration of convolving a 5↝5 filter

For E.g. a 256x256 color image is Figure 2.4: Illustration of convolving a 5↝5 filter (which we will eventually learn) over a 32↝32↝3 input array with stride 1 and with no input padding. The filters are always small spatially (5 vs. 32), but always span the full depth of the input array (3). There are 28↝28 unique positions for a 5↝5 filter in a 32↝32 input, so the convolution produces a 28↝28 activation map, where each element is the result of a dot product between the filter and the input.

As a deep neural network, CNN can capture complex features from image data. Several structures of CNN have become into being and they outperform traditional methods in the area of image recognition to a great extent [17]). They have been pre-trained using images from the ImageNet database which contains approximately 1.2 million images of 1000 ordinary objects [18] these network structures. Among all, VGGNet [19] and are the most frequently used structure. VGGNet contains many levels of networks with depth ranging from 11 to 19, and the commonly used ones are VGGNet-11, VGGNet-16 and VGGNet-19. VGGNet divides the network into five segments and each segment concatenates multiple 3×3 convolutional networks.

With the increase of network depth, network performance may become saturated and then degrade rapidly. To solve this problem, a deep residual learning framework was proposed. Shortcut connections providing identity mapping are introduced and their outputs are added to the outputs of the stacked layers [39].
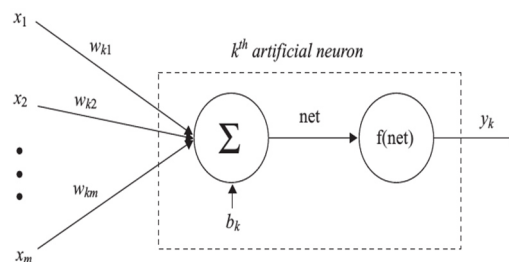

Figure 9. Simple Artificial Neuron.

As big data exploded and became widely available, deep learning systems were developed, which are neural networks with many layers. Deep learning techniques produce more and bigger data for a variety of use cases. It is this architecture and style of processing that we hope to incorporate in neural networks and, because of the emphasis on the importance of interneurons, this type of system is sometimes called connectionist. Neural networks are often referred to when discussing explanations of psychology-inspired models of human cognitive function. However, we use the term neural network generally to refer to any artificial neural network. "network" will be used to describe the neural network being discussed. This system may be a singular node in a small network or a collection of nodes in which each node is connected to every other node in the system. One type of network is shown in Figure 18 Each node is shown by a circle, but utilizes weight information on all connections. This model is only one of several to choose from and is typically used to place an input pattern into one of several

classes according to the resulting pattern of outputs. If the input contains the patterns of light and dark only in a handwritten alphabet image, the output. The layer contains 26 nodes—one for each letter of the alphabet—to determine which letter class the input character belongs to. The output of one node for each class will only be fired whenever a proper pattern of the corresponding class is received.

2.9 Concept of Deep Learning

You probably know deep learning is on the rise. It is like today's vernacular and is used a lot by young people on social media. But why is this observed? With a little digging, you can discover the actual cause. When searching for deep learning on Wikipedia, it will be apparent that the word is not new. The technology was developed for artificial neural networks by Rina Dechter in the '80s and applied to machine learning by Igor Aizenberg and colleagues in the '00s. Even though the idea of deep learning has its roots in the last decade, the term "deep learning" nor the approach itself were so prominent before 2012. Breakthroughs come after the research and hard work of three scientists that are now well known as the fathers of Deep Learning, their names are [21], and [22]. They are the founding fathers of Deep Learning. [20] These technological advances came about when several other factors, such as computer vision, speech recognition, and more, played a role. In March of this year, all together they were awarded the Turing Prize for groundbreaking innovations in the design of deep neural networks that have revolutionized artificial intelligence.

2.10 Image Classification and Segmentation
2.10.1 Image Classification

This section is devoted to describing the method for Image Classification that was used in the final experiment when testing how the pre-processing algorithms in this thesis can enhance the results when classifying objects. Because of certain time constraints for this project (about one or two weeks were devoted to the implementation of the classifier) we chose a relatively simple classifying algorithm, namely Template Matching by comparing pixel intensities. This has been tried before with successful results [23]. Many other choices could have been made, but we limited the thesis to cover only this one as the field of Classification is broad and many algorithms are too complex to be implemented in the time frame given. Recommended us to implement this algorithm because of this.

As with our other form of automated vision, the introduction of machine understanding, in particular, scene classification, recognition, and analyses remain crucial to this technology. In a vast and varied set of image and video databases, it is important to be able to sort and retrieve the images and videos in a way that is both efficient and effective. Only if the categories of images and/or their context are known to a user, the action of using e-cigarettes can be possible. This is why the ability to classify and distinguish scenes precisely is of the utmost importance. Among the advancement in scene recognition and classification algorithms is an ephemeral survey of the advances in scene recognition and classification. There are infinite ways in which to classify images, and each of them is different and complicated. In general, the classification of images into several classes involves two steps; the first is extraction and identification of features followed by classification of images based on the obtained features. Some of the ways to identify underwater objects are the closed and detailed monitoring of the artifacts and image analysis that provides the most probable solutions. Another one of the ways to achieve this is with machine learning. For supervised learning methods, the most prominent feature that the classifier is labeled on is "that he is a supervised learner.

Researchers and experts have made prodigious efforts in evolving progressive classification methods and techniques, which helps in classifying medical data more accurately [23]. Although much research and literature exist to support it, a comprehensive up-to-date review of classification methods and procedures is probably not yet available. However, helpful clarifications and references can be found in the literature. In a wide-ranging sagacity, image classification is the act of cataloging all pixels in an image data to obtain a given set of labels (such as classifying medium boreal forests as "green" or "brown" in color) [24]. [25] Discussed the process of classification as a complex process, requiring consideration of many factors. [26] found that data can be interpreted using classification to assign corresponding labels concerning homogeneous characteristics of groups. The foremost motive for classifying an image is, it can be easily and quickly distinguished from other similar images. Image classification can be referred to us as the function of extracting data classes from a raster image of a multi-band.

It is of interest what research methods have been tried already in similar tasks to those in this report. One interesting point is that different kinds of height sensing or similar techniques resulting in a three-dimensional image for analysis have often been applied in automated waste sorting and research [27]. It is also a frequent approach to the more general object recognition task, and video processing is sometimes chosen instead of using images [24, 25]. Numerous methods have been used for separating foreground from the background (segmentation in other words), for example, the background subtraction part color modeling has been tried), and also Bayesian Rules [26] which is a bit more novel than regular image processing methods. This is not to mention the approaches already written about in the section about background subtraction in this report. For example, histograms have

been used as an important tool [27]Also color-set back projection is a method that has been used, and it has worked.

Fourier transformations have been used for registering images in the sense that they become rotation invariant; in other words, similar problems to the image rotation problem [28] Interestingly enough, also histograms and wavelet transform [5] have been used to solve problems of similar character used gradient orientation histograms to compute image feature orientation for Classification). Thinning methods have been used as a part of rotation invariant algorithms [28]. It should also be mentioned that the problem of detecting skewness for printed characters has also been approached in several ways

2.10.2 Image Segmentation:

Image segmentation tasks can be categorized in the areas of semantic segmentation and instance segmentation. Whereas semantic segmentation CNN architectures process and segment the whole image as one instance, instance segmentation networks separate different image regions and process all detected objects individually. Semantic segmentation networks are offered in two methods: processing and classification of separate pixels or processing of the whole image at once. Pixel-wise implementations from [17] or [13] used sliding windows on the images to classify and segment the pixels in small batches. As shared features were not reused between the overlapping patches, these models resulted to be ineffective [10] presented a multi-path refinement network for the improvements of tasks like semantic segmentation. By implementation of residual connections and residual pooling, information like high-level semantic features over the down sampling process.

Today, the Mask R-CNN [10] as a state-of-the-art network architecture for instance segmentation outperforms other approaches made in this area. In segmentation, it is necessary to extract objects or fragments of an image based on certain conditions and classify them into certain regions. The result is the assignment of labels to all pixels in various regions defined by certain properties. Region growing is a type of image segmentation. In a region growing algorithm, the basic approach is to start from a set of "seed" points and grow regions by appending neighboring pixels that have predefined properties similar to the seed, such as intensity ranges. The result is a region of pixels that all share similar characteristics. This approach takes both intensity and relative location into account to result in better segmentation, compared to that of a simple threshold which only compares intensity information. The result is segmented pixels that have similar intensity values and have close spatial neighbors to identical elements. It is especially useful in determining the border between white and black elements as well as eliminating isolated bright spots that can severely hamper classification. Another type of segmentation is called image decomposition, which splits image regions into further subcategories and rebuilds the regions into ones that are more desirably segmented. This evenly spaced decomposition could prove a powerful tool in the barcode problem because of the size structure and layout rigidity that the elements of a barcode need to follow. Similar but non-uniform areas that are essential are only one of two possible regions. One approach called Quad tree decomposition breaks up an image into 4 regions called quad trees. These regions are tested against some criteria like intensity region uniformity. If regions fail the test, they are further subdivided by four and tested again. The process repeats until all quad trees or quad regions either pass the given test or is maximally divided for the given problem. This method, if applied correctly can break down an image into many small sub-regions, but the largest undivided blocks would occur over element centers. These blocks can be merged back into the most likely binary white and black region layout for the barcode.

2.11 Image Segmentation and Convolutional Neural Networks (CNNs)

Segmentation and Classification, achieving state-of-the-art results in these tasks. This is thanks to CNN's ability to learn a hierarchical representation of the raw input data. In the last years, one of the main fields of application of CNNs has been in medical image segmentation ([7], [1]). Apart from the 2D capabilities used to delineate organs, malformations [8], etc., CNNs have been impressed with their 3D abilities, helping also to process MIR scans [9]. Another interesting field where CNNs have been applied is in semantic segmentation thanks to these revolutionary results, CNNs have been the dominant approach to computer vision in recent times, obtaining near-human performance in some tasks. Some companies such as Google, Facebook, and Microsoft, among others, have been quickly adopting this technology, due to its reasonable computing performance and the hardware advancement done by companies like NVIDIA, Qualcomm, and Samsung that are even developing Systems on Chip (SoC) that dramatically accelerate the common operations used for DL and CNN networks.

A typical image appears on a complex background, usually with multiple objects to be identified. Thus, the classification system needs to be able to segment the image by recognizing individual patterns and identifying the physical limits of where the object exists in the image. In general, classification is only as good as both segmentation (to reduce data) and the features that can be identified in the segmented region. Segmentation is one of the most prolific problems in pattern recognition in that there is a wide variety and combination of pixel differences between similar objects in an image. Therefore, higher-level methods and approaches are required. It is believed the human brain is so successful in segmenting pieces of images is because it can focus on chunks of the image and quickly correlate it against a vast memory bank to see if it seems like a recognizable object (5).

3.0 Image Classification and Segmentation Techniques
3.2 Machine Learning

Moreover, regardless of whether humans are at the helm. *(Supervised, unsupervised, semi-supervised, and reinforcement learning).* Regardless of the fact they may be slow to adapt to the task at hand (online versus batch learning). Irrespective of whether they can work by just contrasting new information focuses with realized information focuses, or rather distinguish designs in the preparation information and construct a prescient model, much like researchers do. But I will concentrate more on supervised and unsupervised types of machine learning.

Supervised Learning: Supervised getting to know is fairly commonplace in class troubles due to the reality the intention is regularly to get the pc to examine a category machine that we've got created. Digit recognition, is a common example of classification studying. More generally, classification analyzing is appropriate for any hassle in which deducing a kind is beneficial and the type is easy to determine. In a few cases, it won't even be necessary to offer predetermined classifications to every example of trouble if the agent can work out the classifications for itself. Training statistics consists of both the information and therefore the preferred consequences. For a few examples, the best effects (objectives) are recognized and are given in input to the version for the duration of the mastering procedure. The construction of proper training, validation, and check set are essential. These techniques are generally fast and accurate. Have that permits you to generalize: deliver the best consequences whilst new data are given in input without understanding a concern the goal. The supervised machine learning systems are the one's algorithms that dream external help. The input dataset is split into teaching and take a peek at the dataset. The teach dataset has an output variable that wishes to be anticipated or classified. All algorithms search at some type of patterns from the education dataset and comply with them to the test dataset for prediction or kind. The work process of supervised machine learning algorithms is shown in figure 1 below. The most common supervised machine learning algorithm is Decision Tree, Naïve Bayes, and Support Vector Machine.
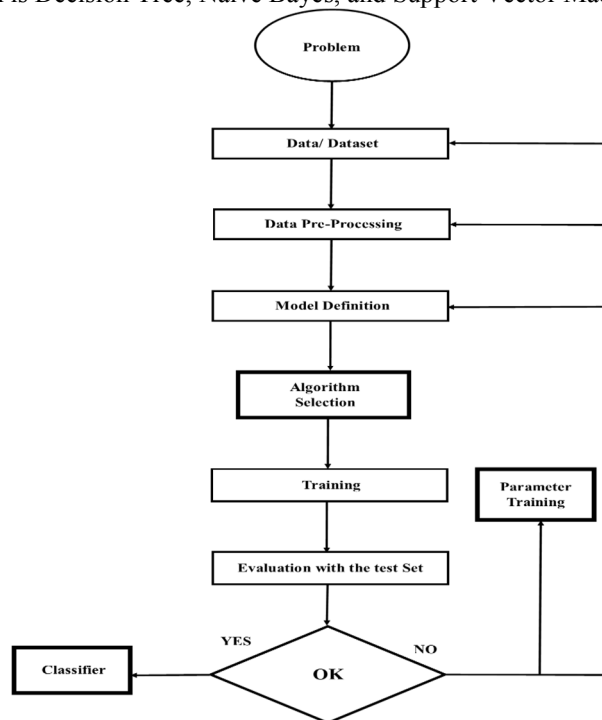

Figure 10. Work flow of Supervised Machine Learning.

Unsupervised Learning: Unsupervised studying appears tons harder: the purpose is to have the pc learn how to perform a little factor that we don't inform it a manner to do! There are certainly two tactics to unsupervised learning. The first method is to train the agent no longer through the way of giving express categorizations, but via the usage of a few forms of reward system to signify achievement. [18] Note that this kind of education will normally suit the selection hassle framework due to the fact the aim isn't always to provide a type but to make alternatives that maximize rewards. This approach well generalizes to the actual international, wherein outlets are probably rewarded for doing positive moves and punished for doing others. Unsupervised studying has produced many successes, which include international-champion satisfactory backgammon packages or even machines able to use cars! It is a robust technique while there is a clean manner to assign values to movements. In unsupervised studying, as you'll probably wager, the training data is unlabeled, the tool tries to investigate without a trainer. The second approach is called clustering, in this type of learning, the purpose isn't always to maximize a software program feature, but virtually to find out similarities inside the education statistics. The assumption is frequently that the clusters found will inform fairly because of an intuitive magnificence. For example, clustering human

beings based totally on demographics may result in a clump of the affluent one organization and the bad in every different. This is a data-driven approach that can work well when there is sufficient data to work with. Conferring to [10], unsupervised learning algorithms are built to extract the structure from data samples. The idea of an edifice is estimated by a cost capacity which is generally limited to construe ideal parameters portraying the concealed structure in the information. Unsupervised analyzing has produced many successes, inclusive of world-champion caliber backgammon applications and even machines able to use vehicles. It may be a powerful approach whilst there is a clean manner to assign values to actions. Clustering may be useful even as there may be enough information to form clusters   and specifically while extra statistics about contributors of a cluster may be used to offer similar results because of dependencies inside the information. Classification getting to know is powerful whilst the classifications are recognized to be accurate (as an instance, simultaneously as managing sicknesses, it's far generally instantly-in advance to decide the layout after the truth with the aid of a post-mortem), or whilst the classifications are arbitrary subjects that we would love the computer so that you can recognize for us.

3.3 Techniques of Image Classification and Segmentation
Image segmentation is a topic of great interest in computer vision and it has been utilized for a variety of tasks, such as object localization and recognition, boundary detection, autonomous driving, and medical imaging. Prior successful segmentation methods relied on some level of manual user initialization. This paper proposed effective architectures to fully automate the process of localizing and delineating the boundaries of objects in images. Many different Machine Learning algorithms or techniques are widely used in many areas of our life and they help us to solve some everyday problems. Algorithms can aid us not only to identify images, videos, and texts, but are also used to strengthen cybersecurity, improve medical solutions, customer service, and marketing.

3.5 Evaluation of Classification Performance
Applying machine learning algorithms for object recognition solves many problems even more effectively than human eyes. For example, convolutional neural networks are most commonly used in tasks for image analysis including Classification, classification, and recognition. The range of these tasks is gradually becoming wider. Due to this, the development of new network architectures, layers, and framework modifications remains of current interest. In our research, we often turn to methods using convolutional neural networks. In particular, we recently used a special type of neural network, fully convolutional neural networks for the receipt recognition project.

3.5.1 Feature Extraction and Selection
This section is devoted to featuring detection methods (or more concretely edge-, line-, and Points of Interest (POI) detection methods as we have limited the content to). To make it possible to see what an image contains, it might be necessary to make a residual image that is a simplified version of the original, for instance, black and white with white as foreground and black as a background, or in some other way clearly distinguish different objects and contours in an image. This is what the first section is about. Next follows line- and POI detection, which search for more specific features that stand out in some way, and mark them in the image. As will be seen, several of the methods presented here will more or less work as algorithms for deciding the orientation angle for a given image. This is intuitive since many of the algorithms which were eventually used for finding the rotation would be based on averaging some set of features (angle of the average line of the line detection, for instance), so this chapter covers algorithms for rotating the images consistently, as well as different processing algorithms.

3.5.1 Convolutional Neural Network and Convolution process
A convolutional neural network works based on basic neural networks which were described above. So what does the CNNs change? There are several variations on CNN's layers architecture: Convolutional Layer, Pooling Layer, and Fully-Connected Layer. A fully-Connected Layer is just acting as a neural network which we have already covered previously. CNN algorithm has two main processes: convolution and sampling, which will happen on convolutional layers and max-pooling layers. Every neuron takes inputs from a rectangular $n \times n$ section of the previous layer, the rectangular section is called the local receptive field.

$$x_{i,j}^{(l)} = \sigma(b + \sum_{r=0}^{n} \sum_{c=0}^{n} w_{r,c} x_{i+r,j+c}^{(l-1)})$$

Since every local receptive field takes the same weights or,c and biases b from the equation above, the parameters could be viewed as a trainable filter or kernel F, the convolution process could be considered as acting an image convolution and the convolutional layer is the convolution output of the previous layer. We sometimes call the trainable filter from the input layer to the hidden layer a feature map with shared weights and bias

3.6 Statistical Methods
3.6.1 Bayes Formula
The probabilistic approach to maximize correct classification is described by Bayes Theorem which states that the posterior probability is equal to the likelihood function times the prior probability scaled by the evidence factor (1). This is mathematically stated in equation 1. It is an ideal equation that can never be fully implemented because

the posterior probability cannot be known ahead of time. However, it is still useful to discuss the theory because the posterior probability can often be estimated to obtain good classification results.

$$P(Wj[x] \_ y \ p, \ p\{x\backslash wj)$$

is the likelihood that feature x will be observed, given th at the true state is wj and P(wj) is the prior probability that the next sample will be of state

Wj. P{wj\x)

is the probability that the true state is Wj given that feature x has been observed. It follows logically that in a two-state system, given x, whichever state probability is greater should be the decision rule. In terms of classification theory, the goal is to minimize incorrect decisions because they incur a loss to the system. The loss function can be stated as A(ai|u>j); the loss incurred for taking action a t given that the true state is Wj. The conditional loss R of making the decision a* given feature x can be described in equation 2.2 as the Bayes decision rule.

$$C \ R(aj|x) = X \ ] \ A(a * K \ )F \ K >) \ (2\text{-}2) \ 3\text{=}0$$

In a two-category or binary classification system, the likelihood ratio is expressed as, which focuses on the x-dependence of the probability densities. Using the likelihood ratio as well as experimentally attained prior probability densities, a decision boundary of equal probability can be created to minimize our classification error. Bayes Theorem is an ideal situation that cannot be described exactly because the likelihood and prior probabilities cannot be known, but they can be estimated with prior data. If the likelihood function

p(x\wj) was known, then the posterior P(wj\x)

could be calculated (The probability that if feature x is observed, the true state is Wj). This will be applied later, by using border element pixel probability distributions to estimate the posterior (P(wj\x)) distribution and then attempt to determine the most likely state of each element w based on features x in the present case.

3.7 Set up the Environment
3.7 1 Using Tensorflow
We have been using the latest versions of Tensorflow. It comes with some advantages, such as:
• We do not need to implement lower-level operations (such as convolutions). It allows us to focus on higher-level implementations, such as pruning, or factorization.
• Most of the operations are highly optimized for many platforms and devices. If we were to implement a model in C++, we'd have to spend considerable effort in optimizing it for efficient use of memory and processor. In such a case comparing various techniques and models would
take considerable time. And it comes with some disadvantages, such as:
• When we started our work, Tensorflow was in version 0.10. There have been 4 major releases that we had to modify our codebase for.
• Not all operations are properly implemented. For example, before version 1.2, Tensorflow implementation of separable convolutions was not very well optimized. They were as fast as convolution operations. Before that, we could only hope that they would optimize their implementation.
• It is difficult to implement new operations and modify the existing ones because the C++ internals and build procedures (as of Tensorflow 1.2) are not well documented.
• Tensorflow does not provide tools to implement low-bit variables. So it is not possible to implement some methods that make use of variable width decimals. This limitation makes some methods impossible to use or useless. For example, it is not possible to use methods that represent weights using variable width decimals. Also, storing low-bit weight indices in combination with a small global weight array to reduce the model size is useless. Since we cannot use low-bit integers to represent these indices, our model size does not shrink at all.
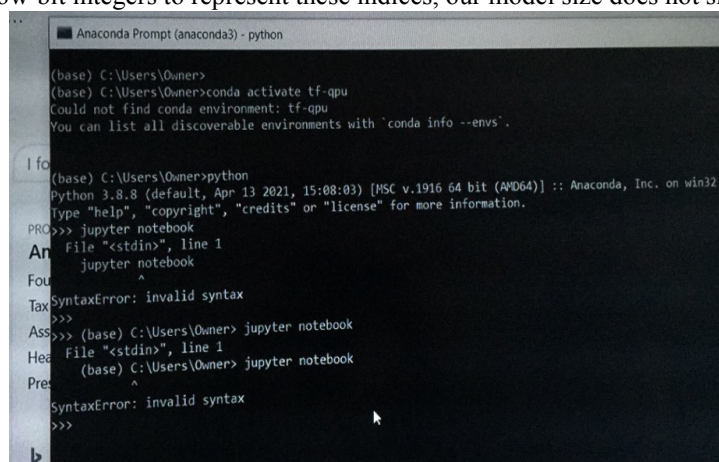

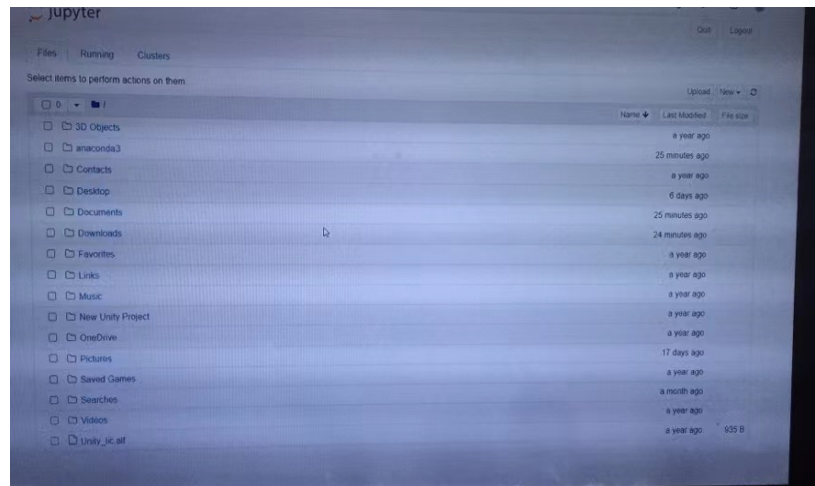Figure 11: Setting up the environment A
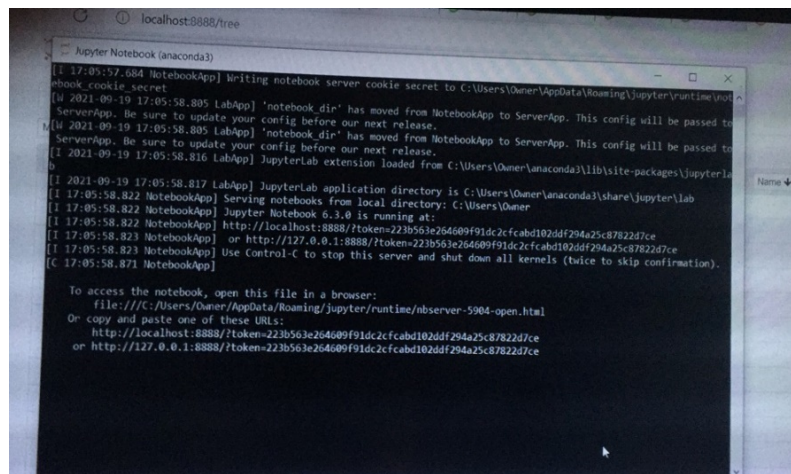
Figure 12: Setting up the environment B

Figure 11: Setting up the environment C

3.8 Model Training for Image Classification and Segmentations
3.8.1 Model Training for Image Classification

Although traditional methods have been widely used in practical problems, these difficulties persist. In this approach, the image features are extracted first, and then the classification is applied separately to each image. The deep learning model can learn features and classifications together, making it more powerful for image classification. This process has the following difficulties: From the outset, it is impossible to come close to replicating the deep model's complicated functions. Also, it comes with a shallow learning model which has low accuracy. The number of terabytes of global internet data will increase to 42 billion in 2020 according to the IDC. More than 70% of the information is transmitted via image or video; the rest is by writing, advertising, sound, or other media. Let's apply the information-gathering abilities of computers to image and video data analysis and see what they uncover. Currently, the computer vision field has gained momentum in the classification of images Based on this research, several image classification techniques have been suggested: Image classifications use: It's all about finding the least amount of error. There are traditional image classification methods such as color, texture, and distinctive local features: e.g., unique scale-invariant features. SIFT searches for the position, scale, scale, and rotation invariants at the extreme point Panoramic stitching and object recognition are very commonly used in its application. [9] proposed the use of deep learning. For the first time in any scientific journal, he proposed the idea of deep learning and uncovered the unknown secrets of the previously obscure feature learning. Considerable advances have been made concerning image classification in this research, leading many to believe that it may have greater potential than previously thought. The [11] and colleagues presented the AlexNet model at the 2012 International Conference on Visual Representation. Deep learning architectures such as these mainly focus on building a larger model, handling samples under overlapping activation, and the use of dropout, in my opinion.

3.8.2 Model Training for Image Classification and Segmentations

In this section, at first, since we have implemented a neural network without using any open source, we'll give a brief introduction to the neural network by explaining the feed forward and back propagation steps in mathematics.

Then we will describe how we use the basic neural network to do the classifying job on the CIFAR-10 dataset the choice of a model architecture is the commencement of the development of an ML network. The first step is to build an initial architecture containing suitable functions for the problem. Whereas for example loss function algorithms like MSE are suitable for linear regression problems, approaches like Log Loss are implemented for classification problems (as further explained. The feature interpretation for example of emotions in sentiment analysis is a sophisticated task, as this complex information is hardly comparable. In image detection problems, the data of the images are scanned and further processed by different filters. The evaluation of the most effective representation of features and the most important features being implemented also presents a task of optimization. In Neural Networks like CNNs, furthermore, the number of neurons and layers have to be chosen and adjusted. Neural network segmentation includes two important steps:

(1) Feature extraction: - The input data of the neural network is determined in this step. Some important features from images are extracted.

(2) Image segmentation:- The features that are extracted from the image are segmented in this step. Neural networks have fast computing and highly parallel computing ability making them suitable for real-time application. It improves segmentation results when the data deviated from a normal situation. It is high robustness that makes it immune to noise.

4.0 EXPERIMENTATION AND DISCUSSION
4.1 Implementation and Experimentation
To evaluate the two models, we conduct experiments on two standard datasets. As we use an unsupervised approach for image classification, we make use of the whole corpus of each dataset by aggregating training and test sets. In this, I used python as the programming language in creating the algorithm which was hosted on Jupiter notebook We evaluated the images by comparing with MNIST FASHION and CIFAR-10, algorithms in aspects of training accuracy, training loss and validation accuracy, validation loss on MNIST FASHION and CIFAR-10. The hardware I used is an NVIDIA GeForce 930MX with 20 GB memory and dual Intel Core i7, and 8GiB memory to compare the MNIST FASHION and CIFAR-10 model.

This compares all these features of the experiments on the training and validation of the two models. CNN's are one of the most commonly used Neural Networks for Image Classification and Recognition. This is the technique that we are going to use on the CIFAR-10 dataset to classify images in one of the 10 categories.

4.2 Classification and segmentation System for Training Sample
The following experiments were done to gain an understanding of what performance could be expected from the algorithms. The experiments are, as stated earlier, image segmentation, image rotation, and Image Classification in consecutive sections 4.2-4.4. The experiments were made on a Compaq 6720s laptop with an Intel Celeron Processor with a 1.73 GHz processor and 2 GB RAM (and with other processes in the background while running them). While the choice of the computer certainly may affect the performance, the idea was mainly to get a comparison of the different algorithms knowing that the final performance can probably be improved, while still getting results indicating the best algorithm choices. Even if the performance differs between different computers, it was believed that one could get results with the computer chosen here indicating whether the algorithms were probable to be effective in a real production setting.

4.3 Data Preprocessing
Post Processing Post-processing involves sorting through all the processed data and classification results to display the information to the user in the most efficiently desired manner. The methods used will vary significantly depending on the application and the user's specifications. A barcode reading application simply requires the output string message to be presented to the next stage of the system. Standardization is necessary to account for the variety of factors that can influence the appearance of the image. These include the different reflectivity of varying soil types and the changing angle and intensity of the sun from image to image. The standardization step operates under the assumption that the radiance values of the LROC NAC images follow a normal distribution. This step transforms the normal distribution into a standardized normal distribution with a mean of 0 and a standard deviation of 100 by using Map Algebra. These values are used instead of the standard normal, mean of 0, and standard deviation of 1 because the segmentation algorithm functions differently for different standard deviations.

4.5 Simulation and Training
The following experiments were done to gain an understanding of what performance could be expected from the algorithms. The experiments are, as stated earlier, image segmentation, and Image Classification in consecutive sections 4.2-4.4. The experiments were made on a laptop with an Intel Celeron Processor with a 1.73 GHz processor and 8 GB RAM (and with other processes in the background while running them). While the choice of the computer certainly may affect the performance, the idea was mainly to get a comparison of the different

algorithms knowing that the final performance can probably be improved, while still getting results indicating the best algorithm choices. Even if the performance differs between different computers, it was believed that one could get results with the computer chosen here indicating whether the algorithms were probable to be effective in a real production setting. Then if some algorithm would take 2 seconds to complete, for example, the 0.1-second limit could probably not be managed even by a stronger computer. So getting results for the sake of comparison and realistic evaluations was the main goal.

```
In [1]: from __future__ import print_function
        import keras
        from keras.datasets import cifar10
        from keras.preprocessing.image import ImageDataGenerator
        from keras.models import Sequential
        from keras.layers import Dense, Dropout, Activation
        from keras.layers import Conv2D, MaxPooling2D, GlobalMaxPooling2D,BatchNormalization

        batch_size = 32
        num_classes = 10
        #epochs = 1600
        data_augmentation = True

        # The data, shuffled and split between train and test sets:
        (x_train, y_train), (x_test, y_test) = cifar10.load_data()
        print('x_train shape:', x_train.shape)
        print(x_train.shape[0], 'train samples')
        print(x_test.shape[0], 'test samples')

        Using TensorFlow backend.

        x_train shape: (50000, 32, 32, 3)
        50000 train samples
        10000 test samples
```

4.51 Segmentation Dataset

This contains pet images, their classes, segmentation masks and head region-of-interest. You will only use the images and segmentation masks. This dataset is already included in TensorFlow Datasets and you can simply download it. The segmentation masks are included in versions 3 and above. The cell below will download the dataset and place the results in a dictionary named dataset. It will also collect information about the dataset and we'll assign it to a variable named.

Image Segmentation - Imports Libraries and Oxford-IIIT Pets dataset

```
In [ ]: try:
            # %tensorflow_version only exists in Colab.
            %tensorflow_version 2.x
        except Exception:
            pass

        import tensorflow as tf
        import tensorflow_datasets as tfds

        import matplotlib.pyplot as plt
        import numpy as np
```

```
In [ ]: # If you hit a problem with checksums, you can execute the following line first
        !python -m tensorflow_datasets.scripts.download_and_prepare --register_checksums --datasets=oxford_iiit_pet:3.1.0

        # download the dataset and get info
        dataset, info = tfds.load('oxford_iiit_pet:3.*.*', with_info=True)

        2021-05-29 03:11:08.533443: I tensorflow/stream_executor/platform/default/dso_loader.cc:53] Successfully opened dynamic library
        libcudart.so.11.0
        I0529 03:11:10.318339 140089609258880 download_and_prepare.py:200] Running download_and_prepare for dataset(s):
```

The model is now ready to make some predictions. You will use the test dataset you prepared earlier to feed input images that the model has not seen before. The utilities below will help in processing the test dataset and model predictions.

Prepare the Dataset

```
In [ ]: # Preprocessing Utilities

        def random_flip(input_image, input_mask):
            '''does a random flip of the image and mask'''
            if tf.random.uniform(()) > 0.5:
                input_image = tf.image.flip_left_right(input_image)
                input_mask = tf.image.flip_left_right(input_mask)

            return input_image, input_mask

        def normalize(input_image, input_mask):
            '''
            normalizes the input image pixel values to be from [0,1].
            subtracts 1 from the mask labels to have a range from [0,2]
            '''
            input_image = tf.cast(input_image, tf.float32) / 255.0
            input_mask -= 1
            return input_image, input_mask

        @tf.function
        def load_image_train(datapoint):
            '''resizes, normalizes, and flips the training data'''
            input_image = tf.image.resize(datapoint['image'], (128, 128), method='nearest')
            input_mask = tf.image.resize(datapoint['segmentation_mask'], (128, 128), method='nearest')
            input_image, input_mask = random_flip(input_image, input_mask)
            input_image, input_mask = normalize(input_image, input_mask)

            return input_image, input_mask

        def load_image_test(datapoint):
            '''resizes and normalizes the test data'''
            input_image = tf.image.resize(datapoint['image'], (128, 128), method='nearest')
            input_mask = tf.image.resize(datapoint['segmentation_mask'], (128, 128), method='nearest')
            input_image, input_mask = normalize(input_image, input_mask)

            return input_image, input_mask
```

Compile and Train the model

VGG networks have repeating blocks so to make the code neat, it's best to create a function to encapsulate this process. Each block has convolutional layers followed by a max-pooling layer which down samples the image.

You will use the following functions to create the TensorFlow datasets from the images in these folders. Notice that before creating the batches in the get_training_dataset () and get_validation_set(), the images are first preprocessed using the map_filename_to_image_and_mask() function you defined earlier.

The loss you will use is sparse_categorical_crossentropy. The reason is that the network is trying to assign each pixel a label, just like a multi-class prediction. In the true segmentation mask, each pixel has either a {0,1, 2}. The network here is outputting three channels. Essentially, each channel is trying to learn to predict a class and sparse_categorical_crossentropy is the recommended loss for such a scenario.

Compile and Train the model

```
In [ ]:  # configure the optimizer, loss and metrics for training
         model.compile(optimizer=tf.keras.optimizers.Adam(), loss='sparse_categorical_crossentropy',
                       metrics=['accuracy'])
```

```
In [ ]:  # configure the training parameters and train the model

         TRAIN_LENGTH = info.splits['train'].num_examples
         EPOCHS = 15
         VAL_SUBSPLITS = 5
         STEPS_PER_EPOCH = TRAIN_LENGTH // BATCH_SIZE
         VALIDATION_STEPS = info.splits['test'].num_examples//BATCH_SIZE//VAL_SUBSPLITS

         # this will take around 20 minutes to run
         model_history = model.fit(train_dataset, epochs=EPOCHS,
                                   steps_per_epoch=STEPS_PER_EPOCH,
                                   validation_steps=VALIDATION_STEPS,
                                   validation_data=test_dataset)
```

4.6 Results

In this research, we have investigated some methods to reduce the computational cost of convolutional neural networks. To do that, we experimented with some methods that could be used to define models with lower computational costs. We also experimented with some methods to reduce the computational complexity of a given model. To be able to experiment with pruning using larger models, we have implemented a tool to describe pruning routines. We have also implemented a tool that applies simple quantization, pruning, and factorization methods to trained models. Using these tools, we have observed that these methods reduce the computational cost of sufficiently large models. In our experiments, we have observed that the models using separable convolutions with non-linearity result with slightly better accuracy compared to models using convolution or kernel compositing convolution operations while requiring a significantly smaller number of operations. Using them, we have redefined residual blocks and designed a model that achieves similar results to ResNet-20 on the CIFAR-10 classification task. Our model is two times wider, however, it has fewer residual blocks, using two times fewer parameters and requiring 3 times fewer operations. However, more work needs to be done to achieve similar results using the ImageNet dataset. When developing models aimed at processing power-restricted environments, we think that designing and training small models based on the requirements is a more stable alternative to compressing large networks. We have seen that wider and shallower residual networks using separable residual blocks are one way of designing such models.

4.6.1 MNIST

In our experiments with the MNIST dataset, we have not seen a comparable difference between experiments with different operations. All of the experiments have resulted in $99 \pm 0.3\%$ top-1 accuracy, with no visible difference in terms of accuracy.
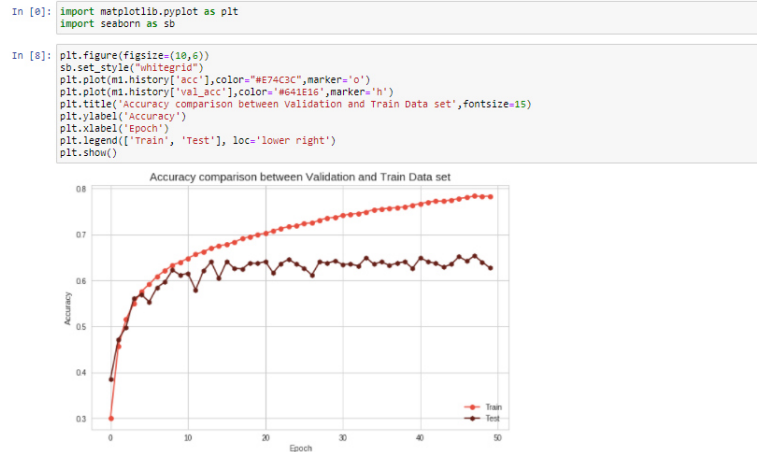
4.6.2 CIFAR-10

The results of our experiments in the CIFAR-10 dataset. As emphasized in the table, separable convolution operations with non-linearity performed slightly better than the rest of the operations. The models that use separable convolutions require 8 times fewer operations than the baseline. If you run this code, you would see that this classifier only achieves **38.6%** on CIFAR-10. That's more impressive than guessing at random (which would give 10% accuracy since there are 10 classes), but nowhere near-human performance or near state-of-the-art Convolutional Neural Networks that achieve about 95%, matching human accuracy.

4.7 Analysis and Discussion

Based on my contribution VGG-16 model is the best model for the classification of images, it seems as if the whole ImageNet classification industry is about to shut down. Unless a paradigm shift occurs, we will never be able to achieve higher accuracy on the ImageNet using the current deep learning approach. As a result, we're looking into some new areas, such as self-supervised or semi-supervised learning for large-scale visual recognition,

as a result of this. Meanwhile, engineers and entrepreneurs struggled to find a real-world application for this imperfect technology using existing methods. This section talks about the above results which describe all the results of both models. First, let's focus on the MNIST FASHION dataset. As shown in Table I, this dataset consumes the least time in the training process, when classifying the images, it takes less time to process but the accuracy is not good and the losses are also less. The test accuracy is 87.86%

```
In [0]: import matplotlib.pyplot as plt
        import seaborn as sb

In [8]: plt.figure(figsize=(10,6))
        sb.set_style("whitegrid")
        plt.plot(m1.history['acc'],color="#E74C3C",marker='o')
        plt.plot(m1.history['val_acc'],color='#641E16',marker='h')
        plt.title('Accuracy comparison between Validation and Train Data set',fontsize=15)
        plt.ylabel('Accuracy')
        plt.xlabel('Epoch')
        plt.legend(['Train', 'Test'], loc='lower right')
        plt.show()
```


5.0 Conclusion

To summarize, we first explained what CNNs are, beginning with the perceptron and extending to neural networks and finally convolutional networks. Next, an overview of person recognition was given. Afterward, we discussed the related work. Specifically, examples were brought for several CNN architectures and techniques such as dropout or batch normalization, and recurrent CNNs and fully convolutional networks were explained. Next, CNNs for object and pedestrian detection was introduced along with the RCNN pipeline, which is further extended with RPNs for fast and accurate Classification. The results and discussions are one of the utmost parts of the thesis which shows what happened. I have shown the comparative results and also discussed them in two different ways been qualitative and quantitative analysis. Prediction models estimated accuracy. However, the model makes errors by memorizing examples for the training set but has a poorer generalization performance on the test set. In computer vision or machine learning, Image classification has made great progress over the past decades and knowledge-based classification algorithms, incorporation of ancillary data into classification procedures, Exactness evaluation is a necessary part of a picture arrangement strategy. Accuracy assessment based on an error matrix is the utmost frequently employed approach. Uncertainty and error propagation in the image classification chain is important to factor influencing classification accuracy. Identifying the weakest links in the chain and then reducing the uncertainties is critical for the improvement of classification accuracy. The classification of images is a flagship instance of the Deep Learning technology's capability. This thesis evaluated the performances of the two datasets: CIFAR-10 and FASHION MNIST, which contain dissimilar types of objects and compares the test accuracy of different models on the Fashion-MNIST and CIFAR-10 dataset via comparing the results, we found out that CIFAR-10 is better at single-class datasets and perform better on real-world complex data and also has better test accuracy. And talked about the techniques involved in image classification, evaluation performance, model accuracy assessment, and selection of classification systems.

It has been shown that Foreground Segmentation can be done effectively in an environment where the background is partly static and partly dynamic. The method for this can be chosen as calculating an averaged background image based on a multitude of different background images and then making a pixel-wise threshold selection of pixels probable of being foreground because of the intensity difference, and finally marking dense areas of such pixels as foreground objects. Finally, it has been shown that using effective algorithms for segmentation and rotation can be helpful for the task of object recognition. It is possible to implement pre-processing algorithms that segment and rotate objects in the sense that they are likely to match templates previously stored in the classifier of the same object. This could then be a possibility of making object recognition easier.

REFERENCES

[1] Comber, A.; Fisher, P.; Brundson, C. &Khmag, A. 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sensing of Environment 127, 237–246. DOI: 10.1016/j. rse.2012.09.005

[2] Szegedy, C., Liu, W., Jia, Y., et al. (2015) Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, 7-12 June 2015, 1-9.https://doi.org/10.1109/CVPR.2015.7298594

[3]Roberts, Jacob (2016). "Thinking Machines: The Search for Artificial Intelligence". Distillations. Vol. 2 no. 2.

pp. 14–23.

[3] A. Krizhevsky, I. Sutskeevr, and G. E. Hinton, "ImageNet classification with deep Convolutional neural Networks," Advances in Neural Information Processing Systems, pp. 1106-1114, 2012.

[4] Kim, D., Kang, P., Cho, S., Lee, H., &Doh, S. (2012). Machine learning-based novelty detection for Faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications, 39*, 4075–4083. doi:http://dx.doi.org/10.1016/j.eswa.2011.09.088

[5] J. Günther, P.M. Pilarski, G. Helfrich, H. Shen, K. Diepold First steps towards an intelligent laser welding Architecture using deep neural networks and reinforcement learning ProcTechnol, 15 (2014) , pp. 474 – 483

[6] Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press.

[7] Pal, M., A. E. Maxwell, and T. A. Warner. 2013. "Kernel-Based Extreme Learning Machine for Image Classification." Remote Sensing Letters 4 (9): 853–862. doi:10.1080/ 2150704X.2013.805279

[8] Ghimire, B., J. Rogan, V. Rodríguez-Galiano, P. Panday, and N. Neeti. 2012. "An Evaluation of Bagging, Boosting, and Random Classification in Cape Cod, Massachusetts, USA.": 623–643. doi:10.2747/1548-1603.49.5.623

[9] Awad M., Khanna R. (2015) Machine Learning. In: Efficient Learning Machines. A press, Berkeley, CA. https://doi.org/10.1007/978-1-4302-5990-9_1

[10] Kheradpisheh S. R., Ganjtabesh M., Masquelier T. (2016a). Bio-inspired unsupervised learning of visual features leads to robust invariant object Recognition. Neurocomputing 205, 382–392. 10.1016/j.neucom.2016.04.029

[11] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556

[12] Russakovsky, O., et al. (2015) ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115, 211-252. https://doi.org/10.1007/s11263-015-0816-y.

[13] He et al. 2017. He, X., Y. Wada, N. Wanders, and J. Sheffield, 2017: Human water management intensifies hydrological drought in California. Geophys. Res. Lett., 44, no. 4, 1777-1785, doi:10.1002/2016GL071665.

[14] Yeboah, D., Sanoussi, M.S., & Agordzo, G. (2021). Image Classification Using TensorFlow GPU. 2021 International Conference on ICT for Smart Society (ICISS), 1-5.

[15] D. Cireşan, et al., "Multi-column deep neural networks for image classification," arXiv preprint arXiv:1202.2745, 2012.

[16] Phinn SR, Roelfsema CM, Mumby PJ (2012) Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. Int J Remote Sens 33:3768–3797

[17] Gong, Y.; Liu, Y.; Lin, Y.; Yang, J.; Duan, Z.; Li, G. Exploring spatiotemporal characteristics of intra-urban trips using metro smartcard records. In Proceedings of the 2012 20th International Conference on Geoinformatics (GEOINFORMATICS), Hong Kong, China, 15–17 June 2012; pp. 1–7.

[18] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," in Advances in neural information processing systems, 2015, pp. 919-927.

[19] Farabet et al., 2013) or Pinheiro and Collobert (Pinheiro and Collobert, 2014

[20] YE L, CHEN Cheng, Sugianto Sugianto, Agordzo, G.K., et al.Driver's distracted behavior recognition Research-based on image segmentation [J].2020,(03):-.[doi: 10.3969/j.issn.1671-8798.2020.03.008]

[21] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.

[22] Semantic networks, description logics, inheritance (including frames and scripts): * Russell &Norvig 2003, pp. 349–354, Representing events and time: Situation calculus, event calculus, fluent calculus (including solving the frame problem):

[23] Ye, L., Zhu, J., Zhou, W., Duan, T., Sugianto, S., Agordzo, G., Yeboah, D., & Kevin, M.T. (2020). DPNet: A Dual Path Network for Road Scene Semantic Segmentation. Trans. Edutainment, 16, 57-69.

[24] S. Park and B. Zhang. Large scale unstructured document classification using unlabeled data and syntactic Information. In PAKDD 2003, LNCS vol. 2637, pages 88–99. Springer, 2003. 2.2

[25] Girshick, R. (2015) Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[26] M Bai, R Urtasun. Proceedings of the IEEE Conference on Computer Vision and Pattern, 2017. 412: 2017: Upset: A unified panoptic segmentation network.

[27] Matthew D., Graham W. Taylor, and Rob Fergus. "Adaptive Convolutional Networks for mid and high-level feature learning." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[28] Y Zhou, Y Feng, T Liu, ZJ Shen. ... 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), 1968-1973, 2018. 8: 2018