

Development in Deep Convolutional Neural Networks by using Machine Learning Framework

Yousuf Tarbez

M.Tech Student, Department of Computer Science and Engineering, Jamia Hamdard, India

Corresponding Author: yousuftarbez.100@hotmail.com

Received: 12-02-2023

Revised: 27-02-2023

Accepted: 20-03-2023

ABSTRACT

In recent years, the machine learning technology has drawn more interest in a variety of vision tasks such as image classification, image detection, and picture identification. Recent improvements in machine learning methods, in particular, stimulate the use of convolutional neural networks for image classification. CNNs are recognised as a potent class of models for image identification issues, sometimes even outperforming humans. The study described in this paper's major objective is to provide an overview of the rise and development of machine learning, deep learning, CNN, and the use of machine learning for image categorization. The CNN and conventional methods are contrasted at the conclusion.

Keywords: deep learning, neural networks, convolutional neural networks

I. INTRODUCTION

The ability of a computer to display "intelligence" is the simplest definition of artificial intelligence. John Mc Carthy, considered the inventor of artificial intelligence, defined it as "the science and engineering of creating intelligent machines, primarily intelligent computer programmes." Making machines seem to have human intelligence is the large and crucial field of artificial intelligence in computer science. Artificial intelligence is a subject of research, not a system. AI research's beating heart and soul are knowledge engineering. If machines had access to a wealth of information about the world, they could be programmed to behave and respond like people.

1.1 Computer Learning

An aspect of artificial intelligence is machine learning. Data science has a subfield called "Machine Learning" that allows computers to "learn" without having to be explicitly trained by humans. The Machine Learning model creates patterns by dissecting the historical data known as "training data" and uses these patterns to learn and predict the future. Such forecasts provided by ML models are becoming more accurate every day. Since the past decade, machine learning has helped us with self-driving cars, speech recognition, effective web search, and a significantly enhanced knowledge of the human social structure. These days, machine learning is so pervasive that many people utilise it without even realising it. It is also seen by many researchers as the most effective path towards AI that can compete with humans. The success of a machine depends on two factors: how much abstraction data speculation occurs, and how well the computer can use what it has learned to forecast the future. The goal of machine learning, which is closely related to the goal of AI, is to fully comprehend the concept of learning, including human learning and other types of learning, about the computational aspects of learning behaviours, and to instill the ability to learn in computer systems. Machine learning, which has applications in many fields of science, engineering, and society, is at the heart of artificial intelligence's success.

1.2 Techniques for Machine Learning

Nowadays, machine learning is employed extensively across all industries, even though some of these applications might not always be obvious. Machine learning's primary methods include:

- 1. Classification:** Classification relies on training data, observations with established categories, and predicts the category to which a new observation belongs. For instance, classifying the price of a house into the categories of very expensive, costly, affordable, cheap, or very cheap.
- 2. Regression:** It uses a continuous data set to forecast a value. For instance, estimating the cost of a property depending on its location, time of purchase, size, etc.
- 3. Clustering:** A set of data is divided into subsets (i.e., clusters) using the clustering approach so that the observations belonging to the same cluster are related in some way. Using Netflix as an example, viewers can be grouped into distinct groups based on their viewing preferences.

4. Recommendation Systems: It use machine learning algorithms to assist users in discovering new goods and services based on information about the user, the good or service, or both. For instance, YouTube may propose a certain movie based on user viewing habits, while Amazon may recommend things based on sales volume.

Identifying observations that do not fit an expected pattern or other dataset items is known as anomaly detection. For instance, a credit card transaction outlier (or abnormality) may indicate possible banking fraud.

Dimensionality Reduction: The procedure of reducing the number of randomly generated variables under investigation to get a set of variables that are statistically significant.

The three main categories of machine learning algorithms are as follows.

Predictive Models are used in supervised learning to make predictions about the future based on the historical data that is now accessible. In this technique, each training example has a pair that consists of a supervisory signal—a desired output value—and an input object, which is commonly a vector. To learn the mapping function between the input and output objects, many approaches are utilised. In supervised learning, there are two categories: classification and regression. Regression problems arise when the output variable is a real value, such as "rupees," "weight," or "temperature," as opposed to classification problems, which arise when the output variable is a category, such as "white" or "black" and "dog" or "cat."

Nevertheless, this technique also utilised a number of additional strategies, including support vector regression (SVR), Gaussian process regression (GPR), neural networks, naive bayes, and support vector machines, among others.

Image categorization, identity theft detection, weather forecasting, and other common supervised learning applications

Descriptive Models and Unsupervised Learning: The unsupervised learning challenge entails inferring a function to describe the unlabeled data, wherein classification or categorization is not included in the observation. Unsupervised learning differs from supervised and reinforcement learning in that the output of the applicable algorithm is inaccurate when the examples of learners provided are unlabeled. Clustering and association are additional ways to categorise unsupervised learning. Inherent groups of the data, such as student heights in a class or school, are what clustering is all about. However, association rules are about to create intriguing connections between datasets and variables. Applications of unsupervised learning in the real world include NASA remote sensing, micro UAVs, nano camera manufacturing technology, and others.

Reinforcement Learning (RL): This machine learning technique is highly distinct from supervised and unsupervised learning, making it one of the most well-known studies in the field of artificial intelligence. On the other hand, reinforcement learning learns everything from past experience where it recognises stuff that it has learned in the past without any cumbersome coding. Supervised learning provides feedback instantly whether the work has been done is right or wrong similarly the case with unsupervised learning. In order to maximise rewards, reinforcement learning, a component of human behavioural psychology, uses an agent to act in accordance with the environment.

1.3 In-Depth Learning

Artificial Neural Networks, a machine learning algorithm, are combined with deep learning. It supports the modelling of arbitrary functions by using the idea of the human brain. An enormous quantity of data is needed for ANN, yet this technique is quite versatile when modelling many outputs at once. Since deep learning has been around for a while, it is no longer a novel idea. But with all the focus now, deep learning is becoming more popular. Artificial intelligence's Deep Learning field generates game-changing outcomes. Deep learning is the term for neural networks with numerous hidden layers. It aims to replicate how the human brain functions. Similar to how little is known about the precise workings of the human brain, little is also known about the precise workings of deep learning. The input and output can be seen and are known, but the interior workings are still a mystery, making it similar to a black box. It's interesting to learn that data scientists think understanding how deep networks operate would bring us closer to understanding how the human brain functions. "Deep Learning is a new field of Machine Learning research, brought in with the aim of achieving the original goal of Machine Learning," says the author. "Deep Learning is a form of machine learning that achieves huge power, adaptability, and knowledge to constitute the world as an order of settled ideas characterised in connection to more straightforward ideas, and more conceptual representations registered in terms of less abstract ones. Two fundamental elements of deep learning are: 1) models with numerous layers or stages of processing information in a nonlinear way. Deep learning is about training multi-layered artificial neural networks to perform well out of data such as images, sound, and text. 2) techniques for learning to represent features at ever higher, more abstract levels, either supervised or unsupervised.

1.4 Overview of the CNN Building

CNNs are feed forward networks because information only flows from their inputs to their outputs in one direction. Both CNNs and artificial neural networks (ANN) are based on biological principles. They are made up of neurons with biases and weights that may be taught. Each neuron receives a certain amount of input, which is then multiplied by the appropriate weight and passed via the activation function [16]. From the unprocessed image pixels at one end to the class scores at the other, the complete system continues to represent a single differentiable scoring function. All

the techniques we used to learn NN still work because the loss function is still present on the final layer. Their construction is driven by the brain's visual cortex, which is composed of alternating layers of simple and complex cells (Hubel & Wiesel, 1959, 1962).

Although there are many different types of CNN designs, they typically consist of convolutional and pooling (or subsampling) layers that are put together into modules. These modules are followed by one or more fully linked layers, just like in a typical feedforward neural network. To create a deep model, modules are frequently stacked on top of one another. Figure shows a typical CNN design for classifying toy images [16]. The network receives an image as input, and then goes through a few convolution and pooling processes. Representations from these jobs are then sent into at least one fully connected layer. The class label is ultimately output by the last fully connected layer. Several design improvements have been suggested in recent years with the aim of enhancing picture classification accuracy or Deep convolutional decreasing processing costs[16], despite the fact that this is the most frequent basic architecture identified in the literature.

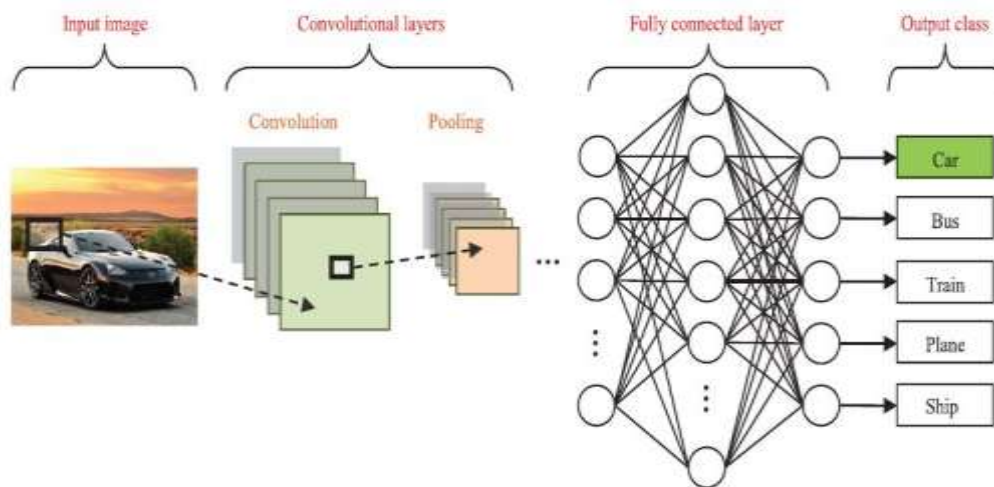


Figure 1: The pipeline for CNN image classification [16].

II. CONNECTED WORK

2.1 Learning from Machines: A Discipline

In this study [5], the authors provide some insights into how computers and machine learning can be used to create autonomous machines that can function without human supervision. Additionally, they discussed some of the uses for machine learning, such as robot control, computer vision, bio-surveillance, speech recognition, and computer vision. The function of machine learning in computer science has also been covered. They discussed the different difficulties in machine learning research as they came to a conclusion.

2.2 Scalable and Simple Deep Learning

Before Deep Learning is widely adopted in multimedia and other applications, it must overcome two challenges[1]. One is usability, which requires that different training and modelling procedures be easily completed by non-specialists, especially when the model is large and sophisticated. A user should have no trouble choosing the optimal model because different multimedia apps may benefit from different models. For example, deep auto encoders are suitable for multimodal data analysis, while deep convolutional neural networks and recurrent neural networks are used for language modelling and picture classification, respectively. One more crucial point: while these models could be overly complex and expensive, it shouldn't be necessary for a user to run them from scratch. GoogleNet, for instance, has 22 layers and 10 distinct types.

The second is scalability, which means that a deep learning system must be able to accommodate a significant increase in the demand for processing power for training large models on enormous datasets. Since larger models and larger datasets are used to increase accuracy, the memory requirement for training the model may actually exceed the processing capacity of a single CPU or GPU. For instance, with one GPU and 1.2 million training photos, training a deep convolutional neural network with 60 million parameters takes 10 days. We have designed SINGA, a distributed deep learning platform, to overcome these issues. SINGA is built on the common layer abstraction of deep learning models and features an intuitive programming architecture. We show that SINGA outperforms multiple cutting-edge deep learning systems and that it runs on both GPUs and CPUs. The stage is useable and adaptable, as demonstrated by our involvement in developing and readying deep learning models for authentic intelligent sight and sound applications in SINGA [1].

2.3 Survey of Machine Learning Techniques for Data Mining

Data mining is a common learning acquisition approach for knowledge dissemination. One of the data mining techniques is classification, which divides the data into preset categories and groups. For a data instance, it is used to call group enrollments. Data mining is more widely applicable in a number of industries, including those in the medical, advertising, telecommunications, and stock markets, among others. In this paper, various classification techniques are discussed, including Decision Tree, Bayesian Network, Nearest Neighbour, and Support Vector Machine (SVM) [8]. Generally speaking, the operational characteristics of decision trees and support vector machines differ; while one is exceedingly exact, the other isn't, and vice versa. However, the operating profile of rule classifiers and decision trees is similar. A combination of various algorithms will be used to characterise the data set. This document provides a concise overview of the numerous classification strategies used in various data mining disciplines. Various arrangement techniques are presented in this study. One classification method is always more useful than another in any given field.

2.4 Big Challenges with Big Data

They have outlined several ongoing challenges data scientists in the biological sciences face in this study [11], as well as the current methods used to address these challenges. Because of the pattern complexity and lack of measurability of the underlying algorithms, data analysis is a logical bottleneck in many biomedical applications due to data complexity, scale, heterogeneity, and timeliness [11]. Different boosted machine learning and data mining techniques are being developed to handle these difficulties. The complexity of the potential patterns may increase exponentially in relation to the complexity of the data, just as it did with the size of the pattern space. Machine learning and data mining calculations typically use a greedy approach to deal with scan for local ideal in the arrangement space or utilise a branch-and-bound approach to seek optimal solutions, and are frequently carried out as iterative or recursive methods to avoid a thorough hunt through the example space. These algorithms frequently use correlations between probable patterns to increase the number of calculations in memory and employ advanced hardware (such as GPU and FPGA) to increase speed. These lead to strong hardware, operating, and data dependencies, as well as occasionally ad hoc solutions that can't be applied to a wider range [11].

2.5 Description of CNN

This study shows that as ANN technology advances, there is a rising need for machine learning. CNN is one of the most amazing ANN architectures. In the beginning, CNN was employed to solve straightforward issues in image-driven pattern recognition, starting with simple architectures[12]. In contrast to other ANN types, CNN primarily concentrates on utilising the knowledge of a particular input.

The several layers employed in the CNN model, their description, and utilisation to create a framework for image analysis are all covered in this study. The many misconceptions concerning the complexity of the CNN models are clarified in this paper.

III. CONVOLUTIONAL NEURAL NETWORKS

The study discusses the rise in popularity of CNN in comparison to other techniques; CNN are mostly used for pattern and picture recognition, sometimes even outperforming humans in certain tasks. Using traffic sign recognition as an example, various challenges have been discussed in this paper to address the issue. Various algorithms and implementations that are developed by cadence can trade off computational load and energy for a slight decrease in sign identification rates have also been introduced. The difficulties of employing CNNs in embedded systems are discussed in this study, along with the essential properties of the Cadence Tensilica Vision P5 DSP for image and computer vision. For the GTSRB, hierarchical CNNs have been designed to recognise traffic signs [13]. For a CDR decline of less than 2%, the complexity of the algorithm used to compute performance vs complexity is lowered by a factor of 86. To run the DSP at 600MHZ, more than 850 traffic sign recognitions are required. CNNs can be operated using the tensilica vision P5 DSP from Cadence, which provides a comparatively ideal set of characteristics.

IV. FRAMEWORK FOR MACHINE LEARNING IN IMAGE CLASSIFICATION

In the context of picture classification and recognition applications, this study discusses extraction techniques and classification. The study discusses several machine learning framework for image categorization approaches and algorithms. On Caltech 101 image categories, several classifier techniques reveal the performance of training models[14]. We employ the SURF technique in contrast to global colour feature extraction for feature extraction. SURF local feature extractor approach and svm (cubic svm) training classifier perform best on average, according to experiments. The primary objective of this study is to provide the best machine learning framework techniques for stopping sign image recognition.

V. DEEP CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE CLASSIFICATION

A thorough analysis of CNNs for image classification problems is presented in this study. It divides their development into three categories: early development, participation in the deep learning renaissance, and recent rapid growth[16]. It specifically considers and analyses the majority of significant advancements in their architectures, supervision components, regularisation procedures, optimisation methods, and computation since 2012. Despite their triumphs in other areas, DCNNs have made tremendous progress in picture classification tasks, dominating various challenges and competitions including the job, and setting the state of the art on a number of difficult classification benchmarks. In fact, their performance has outperformed human performance on a number of single label image classification standards. But as DCNNs have become more popular recently, researchers have started to closely examine their classification performance, resilience, and computational properties, which has led to the discovery of a number of problems and trends to solve them. As a result, this review also summarises these unresolved problems and the patterns they are related with and, more importantly, offers various suggestions and research axes for further investigation[16].

VI. WEB PAGE CLASSIFICATION USING MACHINE LEARNING

This study compares different supervised learning methods to identify predetermined classes among web content. Several methods are used to execute an equivalence of behaviour on the classification problem for web pages, including ANN, random forest (RF), and AdaBoost. After testing, RF classifier accuracy was 87.26%, neural network accuracy was 84.82%, and AdaBoost accuracy was 81.7%[18]. The analysis of the findings showed that the RF algorithm outperformed the ANN and AdaBoost classifiers in terms of classification accuracy. When classifying the pages, FI has a higher value than the rest. AdaBoost is outperformed by the ANN in terms of performance. While NN architecture requires enormous data to generalise better, RF performs better on small data, maintaining a higher number of documents from several attributes. In contrast, RF allows even a large number of documents with numerous attributes to provide accurate results.

6.1 Comparison of Various Weed Recognition Methods from Photos based on Form Characteristics

In this study, k-nearest neighbours, decision tree learning, and support vector machine classifiers for weed recognition from photos are examined together with other classification algorithms. Data mining techniques are used to select the optimal subset of form features for categorization. Bispectral image processing was utilised to determine weed and crop densities while taking form features into account [22]. The classification process is a crucial part of finding weeds; several classification algorithms are explored and their effects on the outcome are taken into account. Barley and oil seed rape manual densities were used for the comparison. All compared classifiers can successfully classify a straightforward class schema, one for each species and noise. Subclasses of the species were created to account for single picture leaf segmentation and over laps of plants due to over- or undersegmentation. The effectiveness of the classifiers varies, and it may be determined by manually comparing the outcomes with weed infestation and utilising cross-validation approaches for classification accuracy. The subclasses of HORVS cannot be distinguished by KNN and decision tree models due to their insufficient model complexity[22].

6.2 Comparing Common Machine Learning Methods for Classifying Images

Several machine learning techniques, including the one that combines creating ensembles of highly randomised trees from the original photos, are examined in this research for image categorization. The other techniques are decision trees, bagging boosting, random forests, and support vector machines. These techniques are tested on four different classification problems with strictly defined test protocols. All of these techniques are applied directly to the pixel values without feature extraction in order to make the method generic. The generic sub-window technique accuracy is superior on three of the four problems and comparable to the paper's state-of-the-art techniques, but somewhat below the best known results[19]. According to our tests, it can be concluded that generic methods, in particular our sub-window methodology, can get quite close to specialised methods while yet maintaining their generality and conceptual simplicity.

6.3 Utilising a Convolutional Neural Network for Image Classification

CNN has been shown to be a potent class of models addressing issues with image recognition. This essay attempted to forecast the likelihood that a photo will receive many Instagram likes. On a fresh dataset of Instagram photos with the hashtag "me," a pre-trained Alexnet Imagenet CNN model was adjusted using caffe to forecast the likeability of pictures. Even though this work is challenging due to data's intrinsic noise[20], a cross-validation of 60% and the test accuracy of 57% were obtained in this research using several techniques. In this study, we taught the model to recognise specific aspects of photographs that garner more likes. Although there are other models like VGG that are more accurate by over 10%, the paper utilised the caffe CNN as the default base model.

6.4 Convolutional Neural Networks with Multi-Stage Features

In most cases, characteristics from the topmost layer of the CNN are used for classification; nevertheless, these features might not be strong enough to accurately anticipate a picture. CNN is now commonly employed for image

classification assignments. Similar to other times, the top layer sometimes conveys stronger traits than the lower layer[21]. As a result, it appears that using features from one layer solely to categorise objects does not fully utilise the potential discriminant power of learnt CNN. Due of this inherent trait, specific CNN models must combine features from several layers. To extract features from several layers, the CNN models that have already learned from the training images are employed again. CIFAR-10, NORB, and SVHN image classification benchmark data sets are used to assess the proposed fusion approach. This study demonstrates how the suggested strategy enhances the stated performances of the existing models by 0.38, 3.21, and 0.13 percent, respectively.

VII. CNN CLASSIFICATION ALGORITHMS COMPARISON WITH OTHERS

In actuality, the dataset and overall complexity of the problem influence the decision of the classifier to choose. Although setting up deep learning is more laborious than utilising alternative classifiers like random forests and SVMs, it excels at solving difficult problems like image classification, speech recognition, and natural language processing. Even so, it helps us to reduce our concern about feature engineering. Generic algorithms are a type of optimisation methods that are incapable of learning anything, whereas CNN are neural networks that require an optimisation method to train the network. CNN are a particular kind of ANN that, depending on the model, can either be deep or not. Contrarily, there is no such mechanism accessible in svm to enhance model complexity; instead, the model complexity is increased by adding more layers to the CNN network. When there is a lot of data available, CNN is employed. Less data is used using SVM[25]. Comparatively speaking to the other image classification techniques, CNN utilises less pre-processing. In order to handle computer vision problems, the CNN has an inherent advantage over traditional approaches: it can learn hierarchical features, or what characteristics are relevant and how to compute them. In some ways, CNN is superior to RNN because it has filters that function as feature detectors that resemble the human visual system; in other words, the convnet is well-suited for image domain. It is typically fairly quick to classify or predict something after training CNN.

VIII. SUMMARY OF WORK

One of the primary economic pillars for coastal nations, the fishing industry is one of the world's greatest food sectors. The strategy is to create a model using a convolutional neural network to automatically identify and categorise various fish species. The organisations are utilising cameras to record the fishing operations in order to protect this fishery in the future. The video will then be divided up into pictures, which will be used as an input to the model.

IX. CONCLUSION

On the larger dataset, convolutional neural networks outperform the other cutting-edge techniques. It not only automatically extracts the features, which lessens the effort of human feature extraction, but also quickly classifies the photos. Other approaches for various image classification issues have been investigated, including SVM, bagging, boosting, random forests, and classic decision trees. On a short collection of data, random forest can categorise more accurately[18]. CNN has an edge because it can adaptively learn the best features from photos. Experiments found that CNN significantly increased accuracy when compared to other picture categorization algorithms[16].

REFERENCES

1. Yann LeCun, Yoshua Bengio, & Geoffrey Hinton. (2015). Deep learning. *Nature*, 521(1), 436-444.
2. Geoffrey E. Hinton, Simon Osindero, & Yee-Whye Teh. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(1), 1527-1554.
3. Li Deng, & Dong Yu. (2013). Deep learning methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4), 197-387.
4. Tom M. Mitchell. (2006). The discipline of machine learning. *CMU-ML*, 06(1), 108.
5. Rich Caruana, & Alexandru Niculescu-Mizil. (2006). An empirical comparison of supervised learning algorithms. *International Conference on Machine Learning*, 23(1), pp. 1-8.
6. Alex Krizhevsky, & Georey E. Hinton. (2011). Using very deep autoencoders for content-based image retrieval. *University of Toronto - Department of Computer Science*, 2(18), 34-56.
7. Waseem Rawat & Zenghui Wang. (2017). *Deep convolutional neural networks for image classification; A comprehensive review*.