Babylonian Journal of Machine Learning
Vol.2023, **pp**. 19–25
DOI: https://doi.org/10.58496/BJML/2023/004; ISSN: xxxxx
https://mesopotamian.press/journals/index.php/BJML

Review Article

Review of deep learning: Convolutional Neural Network Algorithm.

Abdulazeez Khlaif Shathir Alsajri[1,*], Abdullayev Vugar Hacimahmud[2],

[1] *University Arts, Sciences and Technology, Computer Science Department, Beirut, Lebanon.*
[2] *Azerbaijan State Oil and Industry University, Baku, Azerbaijan*

ARTICLE INFO

Article History
Received 02 Dec 2022
Accepted 06 Feb 2023
Published 14 Apr 2023

Keywords
Machine Learning
Deep Learning
CNN
ANN

ABSTRACT

Delves into the advanced field of image processing based on the use of neural networks to automatically and efficiently improve the quality and detail of images. The thesis explains that convolutional neural networks are one of the types of deep neural networks, and they are specially designed to gain knowledge from big data and extract complex features and patterns found in images. The different layers of the network are discussed in detail, as they handle images incrementally and extract various attributes in each layer. The thesis also highlights the ability of CNN to detect, learn, and improve important details found in images through convolutional, filtering, and data aggregation processes. The proposed CNN model for image enhancement was developed and tested on both medical and normal images. The images were enhanced using the proposed model and compared with other models. Different quality metrics were used to evaluate the results. The results showed that the proposed model can significantly improve the quality of images. The thesis also explores the potential applications of CNN in various fields such as medicine, photography, and space imaging. The use of CNN in these fields can lead to improved diagnosis and treatment in medicine, better image quality in photography, and more accurate and detailed images in space imaging.

1. INTRODUCTION

Machine learning (ML) has gained significant popularity in research and has been integrated into several applications such as text mining, spam detection, video recommendation, image classification, and multimedia concept retrieval . Deep learning (DL) is a widely used machine learning (ML) approach in several applications . DL can also be referred to as representation learning (RL). The ongoing emergence of new research in the domains of deep and distributed learning can be attributed to the combination of the unpredictable expansion in data acquisition capabilities and the remarkable advancements in hardware technologies, such as High Performance Computing (HPC)[1].

DL, or Deep Learning, is a more advanced version of the traditional neural network and significantly surpasses its earlier versions in terms of performance. Furthermore, DL utilizes both transformations and graph technologies concurrently to construct multi-layer learning models. Recent advancements in deep learning (DL) have demonstrated exceptional performance in various domains such as audio and speech processing, visual data processing, and natural language processing (NLP) [1]. The efficacy of a machine learning algorithm typically relies heavily on the integrity of the input data representation. Research has demonstrated that an appropriate data representation leads to enhanced performance in comparison to an inadequate data representation. Feature engineering has been a prominent study focus in machine learning for several years, contributing to countless research projects. The objective of this strategy is to generate characteristics from unprocessed data. Furthermore, it is highly specialized and often demands significant human labor. Various characteristics were introduced and examined in the computer vision environment, including histogram of oriented gradients (HOG) , scale-invariant feature transform (SIFT) , and bag of words (BoW) . Once a unique feature is presented and demonstrates strong performance, it creates a new research direction that is pursued for several decades. [2]

Deep learning (DL) is now one of the most popular research trends in the field of machine learning (ML) due to its significant success. This paper provides a comprehensive review of Deep Learning (DL), covering different aspects including key principles, architectures, problems, applications, computational tools, and the evolution matrix. The Convolutional Neural Network (CNN) is widely recognized as one of the most popular and often utilized Deep Learning (DL) networks . The popularity of deep learning has significantly increased in recent times, largely due to the influence of CNN. The primary benefit of CNN, in contrast to its predecessors, is its ability to autonomously identify relevant elements without

*Corresponding author. Email: aka104@live.aul.edu.lb

human supervision, making it the most widely utilized. Consequently, we have thoroughly examined CNN by outlining its key elements. In addition, we have included comprehensive explanations of the prevalent CNN architectures, commencing with the AlexNet network and concluding with the High-Resolution network (HR.Net) [3].

Several DL review papers have been published in recent years. Nevertheless, each of these studies has solely concentrated on a single aspect, emphasizing a certain application or topic. For instance, one study focused on examining CNN architectures], while another explored the use of DL for plant disease classification . Similarly, other studies delved into DL for object recognition and DL applications in medical picture analysis . While these studies touch upon important subjects, they fail to offer a comprehensive comprehension of deep learning themes, including principles, intricate research gaps, computational tools, and applications of deep learning. Prior to delving into the applications, it is imperative to grasp the fundamental components of Deep Learning, encompassing its principles, problems, and potential applications. In order to accomplish this, one must invest a significant amount of time and thoroughly study numerous research papers on deep learning, including topics such as research deficiencies and practical applications. Hence, we suggest conducting an in-depth analysis of deep learning in order to establish a more appropriate foundation for comprehensively grasping the subject, all within the confines of a single review paper. The purpose of our review was to encompass the fundamental aspects of deep learning, including its open challenges, applications, and computational tools from a comprehensive perspective. Moreover, our evaluation can serve as the initial stage for exploring other deep learning subjects[4].

2. LITERATURE REVIEW

Deep learning, which is a subset of machine learning, draws inspiration from the information processing patterns observed in the human brain. Deep learning operates without the need for manually created rules; instead, it leverages a vast amount of data to establish a mapping between the provided input and specific labels. Deep learning (DL) is constructed utilizing many tiers of algorithms, namely artificial neural networks (ANNs), where each tier offers a distinct analysis of the input data[5].

To accomplish the classification assignment using traditional machine learning approaches, one must follow a series of sequential phases, namely pre-processing, feature extraction, intelligent feature selection, learning, and classification. Moreover, the selection of features significantly influences the performance of machine learning approaches. Engaging in biased feature selection can result in inaccurate differentiation between classes. In contrast, Deep Learning possesses the capability to automate the acquisition of feature sets for multiple tasks, which is not possible with traditional Machine Learning approaches. Deep learning allows for simultaneous achievement of learning and classification in a single iteration. Deep learning has gained immense popularity as a machine learning method in recent years, mostly due to the significant expansion and advancement of the big data field. It is currently undergoing continual research to enhance its performance in many machine learning tasks. It has greatly facilitated advancements in several learning domains, including picture super-resolution, object detection, and image recognition. In recent times, deep learning performance has surpassed human performance in tasks such as image classification[6].

3. CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN or ConvNet) is an architecture for deep learning that learns directly from data. CNNs are particularly useful for finding patterns in images to recognize objects, classes, and categories. They can also be quite effective for classifying audio, time-series, and signal data Convolutional neural networks (CNNs) are increasingly being used to identify objects in images. They use principles from linear algebra, specifically matrix multiplication, to identify patterns within an image. That said, they can be computationally demanding, requiring graphical processing units (GPUs) to train models[7].

3.1 Convolutional neural networks function effectively.

A Convolutional Neural Network (CNN) is capable of including numerous layers, with each layer being responsible for learning and detecting distinct features within an input image. Each image undergoes a filtering process or is subjected to a kernel, resulting in an output that improves and becomes more intricate with each layer. With each subsequent layer, the filters become more intricate in order to detect distinct characteristics that accurately describe the input item. The result of convolving each image -- the image partially recognized after each layer -- is used as the input for the subsequent layer.

Convolution involves passing the input image through a series of filters. Every filter acquires the ability to recognize distinct characteristics, resulting in the repetition of operations over numerous layers, ranging from dozens to hundreds or even[7].

4. CONVOLUTIONAL NEURAL NETWORKS LAYERS

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

- Convolutional layer
- Pooling layer
- Fully-connected (FC) layer

4.1 Convolutional Layer

The convolutional layer is the fundamental component of a convolutional neural network (CNN), responsible for the majority of computational operations. The process necessitates several elements, including input data, a filter, and a feature map. Assuming the input will be a color image, composed of a three-dimensional matrix of pixels. The term used to describe this process is convolution.

The feature detector is a 2-D array of weights that corresponds to a certain portion of the image. The filter size is commonly a 3x3 matrix, which directly controls the extent of the receptive field. The dot product is computed by multiplying the input pixels with the filter and then stored in an output array. The ultimate result obtained from the sequence of dot products is referred to as a feature map, activation map, or convolved feature.

Following every convolution operation, a Convolutional Neural Network (CNN) performs a Rectified Linear Unit (ReLU) transformation to the feature map, therefore imparting nonlinearity to the model. The CNN's structure can exhibit a hierarchical nature, as the subsequent layers have access to the pixels located inside the receptive fields of preceding levels. For instance, let's suppose we are attempting to ascertain whether an image includes a bicycle. The bicycle can be conceptualized as a composite of several components, including a frame, handlebars, wheels, pedals, and so on. Each constituent component constitutes a sub-level configuration within the neural network, whereas the amalgamation of these components signifies more complex patterns at higher levels.

4.2 Pooling Layer

The pooling layer is employed to decrease the spatial dimensionality of the input image following convolution. The pooling layer does not have any parameters, however it does have two hyperparameters: the filter size (F) and the stride (S). By processing each individual slice of the representation separately, the amount of computation and weights needed is reduced. Additionally, it aids in diminishing the spatial dimensions of the representations, hence enhancing their legibility and interpretability.

Multiple methods are available for this objective, but the most widely used one is max pooling, which identifies and publishes the highest output value within the surrounding area. Max pooling operates by computing the average value of the rectangle region surrounding the central pixel.

4.3 Fully-connected (FC) layer

The primary distinction from a typical multilayer perceptron lies in the input layer, where an activation volume is used as the input instead of a vector. Fully connected layers of a convolutional network consist of a two or three layer Multi-Layer Perceptron (MLP) that is designed to map the activation volume, $m(l-1)1 \times m(l+1)2 \times m(-1)3$, obtained from the combination of previous layers.

The objective of a fully connected architecture is to build a probabilistic estimation for each class by combining the activation maps produced by the concatenation of convolutional, non-linear, rectification, and pooling layers. The functioning of individual completely connected layers is identical to the layers of the multilayer perceptron, except for the input layer[8].

5. ADVANTAGES AND DISADVANTAGES OF CNNs

5.1 Advantages of Convolutional Neural Networks

Here are the most significant advantages of (CNNs):

- The task of identifying important features does not require human supervision in CNNs.
- CNNs are very accurate at filtering and recognizing images.
- A key feature of CNNs is weight sharing.
- Convolutional neural networks also minimize computation in comparison with a regular neural network.
- CNNs can learn the locations of the images [9].

5.2 Disadvantages of Convolutional Neural Networks

Some of the disadvantages of using CNNs are that it requires a lot of training data to be effective and that they fail to encode the position and orientation of objects.

- They fail to encode the position and orientation of objects. They have a hard time classifying images with different positions.
- A lot of training data is needed for the CNN to be effective.
- CNNs tend to be much slower because of operations like maxpool.
- In case the convolutional neural network is made up of multiple layers, the training process could take a particularly long time if the computer does not have a good GPU.
- Convolutional neural networks will recognize the image as clusters of pixels which are arranged in distinct patterns. They don't understand them as components present in the image [9].

6. TYPES OF CONVOLUTIONAL NEURAL NETWORKS

- **LeNet:** LeNet is an early and extensively utilized Convolutional Neural Network (CNN) architecture. The development of this technology took place in 1998, spearheaded by Yann LeCun, Corinna Cortes, and Christopher Burges. Its purpose was to address the challenges associated with recognizing handwritten digits. LeNet is composed of several convolutional and pooling layers, which are then followed by a fully-connected layer. The model consists of five convolutional layers, which are then followed by two fully linked layers. The LeNet was devised almost two decades ago, although its structure remains pertinent in the present era and continues to be employed. The LeNet Convolutional Neural Network (CNN) is a straightforward yet robust model that has been employed for diverse applications, including the recognition of handwritten digits, traffic signs, and faces.
- **AlexNet :** the inaugural large-scale convolutional neural network (CNN), achieved victory in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet consists of 5 convolutional layers, which are accompanied by a combination of max-pooling layers. Additionally, it includes 3 fully connected layers and 2 dropout layers. The architecture contains over 60 million parameters in total.
- **ZF Net:** short for Zero-Phase Filtering Network, is a convolutional neural network (CNN) architecture that employs a blend of fully-connected layers and CNNs. ZF Net, created by Matthew Zeiler and Rob Fergus, emerged as the champion at ILSVRC 2013. The network possesses a comparatively smaller number of parameters in comparison to AlexNet, nevertheless it surpasses AlexNet's performance on the IlsVRC 2012 classification test. It attained exceptional precision using only 1000 photos per category.
- **GoogLeNet :** is the convolutional neural network (CNN) architecture developed by Google that was successful in winning the ILSVRC 2014 classification task. It has demonstrated a significantly lower error rate compared to former champions AlexNet and ZF-Net. Practical uses encompass the recognition of digits in Street View House Number (SVHN) images and the detection of objects on the side of the road.

- **VGGNet :** is a convolutional neural network (CNN) that consists of 16 layers. It has an impressive number of parameters, reaching up to 95 million. This network has been trained using an extensive dataset of over one billion photos, each belonging to one of the 1000 classes. The model can process input images with dimensions of 224 by 224 pixels and extract 4096 convolutional features. Some real-world applications/examples of VGGNet Convolutional Neural Network (CNN) include its use in the ILSVRC 2014 classification problem, where it competed against and was ultimately outperformed by the GoogleNet CNN design.
- **ResNet :** is a convolutional neural network (CNN) architecture created by Kaiming He et al. It was specifically designed to achieve high performance in the ILSVRC 2015 classification problem, where it achieved a top-five error rate of only 15.43%. The network consists of 152 layers and encompasses more than one million parameters, making it notably deep even for Convolutional Neural Networks (CNNs). An instance of ResNet being utilized in practical situations is Microsoft's machine comprehension system. This system employs CNNs to produce responses for a vast number of questions across several categories, exceeding 100,000 in total.
- **MobileNets :** are convolutional neural networks (CNNs) that are designed to be deployed on mobile devices. They are specifically optimized for tasks such as picture classification and object detection, while maintaining low latency. Typically, these CNN designs are compact, enabling efficient execution on embedded devices such as smartphones and drones. The architecture demonstrates flexibility as it has been successfully tested on Convolutional Neural Networks (CNNs) consisting of 100-300 layers. Notably, it outperforms other architectures such as VGGNet.
- **GoogLeNet_DeepDream :** is a convolutional neural network (CNN) architecture created by Alexander Mordvintsev, Christopher Olah, and others. The Inception network is employed to generate images by utilizing CNN features. The architecture is frequently employed in conjunction with the ImageNet dataset to produce hallucinatory pictures or abstract works of art [10].

7. DIFFERENCE BETWEEN (ANN) AND (CNN).

An Artificial Neural Network (ANN) is a computer software that simulates the functioning of a brain's neural network. Upon making an error, the system reevaluates its thinking and adjusts its approach, mimicking the behavior of a human. An Artificial Neural Network (ANN) primarily relies on weights and an activation function to carry out its main operations. However, it also has the capability to employ other strategies in order to make intricate decisions. The "layers" in an artificial neural network (ANN) consist of rows of data points that are processed by neurons within the same neural network. The weights are updated for each loop through each neuron in the artificial neural network (ANN), based on the accuracy determined by a "cost function"[11].

In comparison, CNN does not contain any neurons or weights. CNN employs a multi-layered approach to process images and utilizes filtration to examine the input images. The three levels are the math layer, corrected linear unit layer, and completely connected layer. The layers serve the goal of comprehending patterns that the network may perceive, processing the output of data, and generating an output in the form of an n-dimensional vector. The n-dimensional output is utilized to observe distinct characteristics and establish connections with the provided image input. Subsequently, it can provide the user with the classification output. Although they differ, both strategies utilize error metrics to enhance learning and generate epochs to evaluate the efficacy of developed models[12].

8. REAL-WORLD APPLICATIONS OF (CNN)

Here are some real-world applications of CNN:

- Facial recognition: A highly accurate neural network has been created to detect facial features, including eyes, nose, and mouth, while minimizing distortions caused by factors like angles or shadows. Convolutional Neural Networks (CNNs) can be employed to mitigate facial deformation. According to multiple research articles, this algorithm has demonstrated enhanced precision compared to previous algorithms, successfully detecting faces 97% of the time.
- Facial emotion recognition: Convolutional Neural Networks (CNNs) have been employed to accurately differentiate various facial expressions, including anger, sadness, and happiness. Convolutional Neural Networks (CNNs) may be effectively adjusted to achieve optimal performance under different lighting conditions and facial angles seen in photos.

- Object detection: Convolutional Neural Networks (CNN) have been employed to accurately determine the location and recognition of objects depicted in photographs. Additionally, CNNs may generate several perspectives of these objects, which can be utilized in applications like drones or autonomous vehicles. Object detection is accomplished by the utilization of methods such as semantic or instance segmentation. Convolutional neural network (CNN) models have been developed with the ability to accurately identify a diverse array of objects, ranging from common objects like food, animals, and humans, to less common objects such as dollar bills and firearms.
- Self-driving or autonomous cars: CNN has been employed in the realm of automated vehicles to facilitate obstacle detection and street sign interpretation. CNNs have been employed in tandem with reinforcement learning, which emphasizes the utilization of positive and negative input from the environment. Auto translation refers to the use of machines, such as CNN, to accurately translate between two languages, such as English and French. Convolutional Neural Networks (CNNs) can autonomously perform language translation tasks, such as converting text from Chinese to English, without requiring human intervention.
- Sentence completion using Convolutional Neural Networks (CNNs) has been employed to forecast the subsequent word in a sentence based on contextual information. CNN models have the ability to analyze several sentences and acquire knowledge about the usual word sequences, such as the pattern of "I am from Iraq" followed by "I speak Arabic."
- Handwritten character recognition: Convolutional Neural Networks (CNNs) can be employed to accurately classify and distinguish various languages, such as Chinese, Arabic, and Russian, even when they exhibit different writing styles. The input to the system is a picture of a character, which is then analyzed by dividing it into smaller portions. The system identifies certain points within these sections that can be connected or overlapped with other points. This analysis helps define the overall shape of the character.
- X-ray image analysis: Convolutional neural network (CNN) models may analyze a picture of a specific body component, such as the knee, and identify potential tumor locations within the image by leveraging prior processed images that are comparable, using CNN networks. Additionally, they can be utilized to identify the specific region of an X-ray image that exhibits a tumor or other irregularities, such as broken bones.
- Cancer detection: Convolutional Neural Networks (CNNs) have been employed for the purpose of identifying cancerous cells in medical pictures, including mammograms and CT scans. CNN models utilize patient images to do comparisons with a database of images that possess similar properties. They detect the presence of indicators in a picture that suggest the presence of cancerous cells or cell damage caused by genetic and environmental causes, such as smoking.
- Visual question answering: Convolutional neural networks (CNNs) have the ability to process a picture as input and generate a response in plain language to any inquiry related to that image.
- Image captioning: Convolutional Neural Network (CNN) models are utilized to produce concise descriptions of the content depicted in photographs. This can be achieved by amalgamating various photos, including those sourced from social media platforms, with pre-existing ones. CNN models have the capability to process a series of images and produce one or many phrases that provide a description of the contents of each image.
- Biometric authentication: CNN has been employed to verify the identity of users by recognizing specific physical attributes linked to a person's facial features. CNN models have the ability to be trained on photos or movies containing humans, and produce an output vector that accurately detects distinct facial characteristics, such as the distance between the eyes, the shape of the nose, and the curve of the lips. Convolutional Neural Networks (CNNs) have the capability to identify the presence of blinking in an image and analyze the general structure of facial images that consist of several frames.
- Document classification: Convolutional Neural Networks (CNN) have the capability to utilize both textual and visual information to enhance their understanding of the content of a given document. CNN models provide the capability to classify documents into certain categories by analyzing the content of the text, such as articles pertaining to sports or politics. Additionally, they can be utilized to condense publications by extracting essential information and identifying pivotal aspects inside the text.
- CNN has been employed for the purpose of segmenting 3D medical pictures, including MRI scans. CNN models may extract a specific section of an image from a three-dimensional scan and identify various tissue types within that section by leveraging the knowledge gained by analyzing similar pictures using CNN networks[13].

9. CONCLUSION

Convolutional neural networks (CNNs) can outperform humans in visual identification tasks, but they still struggle with visual aberrations like glare and noise, which humans can handle effectively. Similar to other deep learning methodologies, Convolutional Neural Networks (CNN) heavily rely on the magnitude and caliber of the training data. Ongoing research is focused on enhancing the capabilities of Convolutional Neural Networks (CNNs) by incorporating features such as active attention and online memory. This will enable CNNs to more efficiently analyze new items.

Conflicts Of Interest

The author declares no conflict of interest in relation to the research presented in the paper.

Funding

The author's paper explicitly states that no funding was received from any institution or sponsor.

Acknowledgment

The author would like to express gratitude to the institution for their invaluable support throughout this research project.

References

[1] M. Akcayol and M. A. M. El-Moursy, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," MDPI, vol. 12, no. 5, p. 91, May 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/5/91

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (2016). [Online]. Available: https://deeplearningbook.org/

[3] M. Nielsen, Neural Networks and Deep Learning (2015). [Online]. Available: https://www.coursera.org/specializations/deep-learning

[4] F. Chollet, Deep Learning with Python (2018). [Online]. Available: https://www.deeplearning.ai/

[5] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (2019). [Online]. Available: https://www.tensorflow.org/tutorials

[6] Y. Bengio, R. Tachet des Combes, and J.-P. Vert, Deep Learning Fundamentals (2020). [Online]. Available: https://pytorch.org/tutorials/

[7] M. E. Saravanan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," National Institutes of Health, January 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33816053/

[8] Y. LeCun et al., "A Convolutional Neural Network for Image Recognition," in Advances in Neural Information Processing Systems 6, J. D. Cowan, G. Tesauro, and J. Alspector, Eds. Morgan-Kaufmann, 1998, pp. 99-106.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097-1105.

[10] Y. Goldberg, "Understanding Convolutional Neural Networks for NLP," arXiv preprint arXiv:1603.01219, 2016.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[12] A. Karpathy, "CS231n Convolutional Neural Networks for Visual Recognition," 2016. [Online]. Available: http://cs231n.stanford.edu/2016/

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.