2025 Volume: 01 Issue: 07 www.ijrpa.com ISSN 2456-9995

Review Article

International Journal Research Publication Analysis

Page: 01-09

A FOUNDATIONAL OVERVIEW OF CONVOLUTIONAL NEURAL NETWORKS

[*1]Ms. A. Kamatchi, [2]Dr. V. Maniraj

[1]Research Scholar, Department of Computer Science, A.V.V.M.Sri Pushpam College(Autonomous),Poondi,Thanjavur(Dt), Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu.

[2]Associate Professor, Research Supervisor, Head of the Department, Department of Computer Science, A.V.V.M.Sri Pushpam College (Autonomous),Poondi,Thanjavur(Dt),Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu.

Article Received: 26 October 2025

Article Revised: 14 November 2025

Published on: 05 December 2025

***Corresponding Author: Ms. A. Kamatchi**

Research Scholar, Department of Computer Science, A.V.V.M.Sri Pushpam College(Autonomous),Poondi,Thanjavur(Dt), Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu.

DOI: https://doi-doi.org/101555/ijrpa.3249

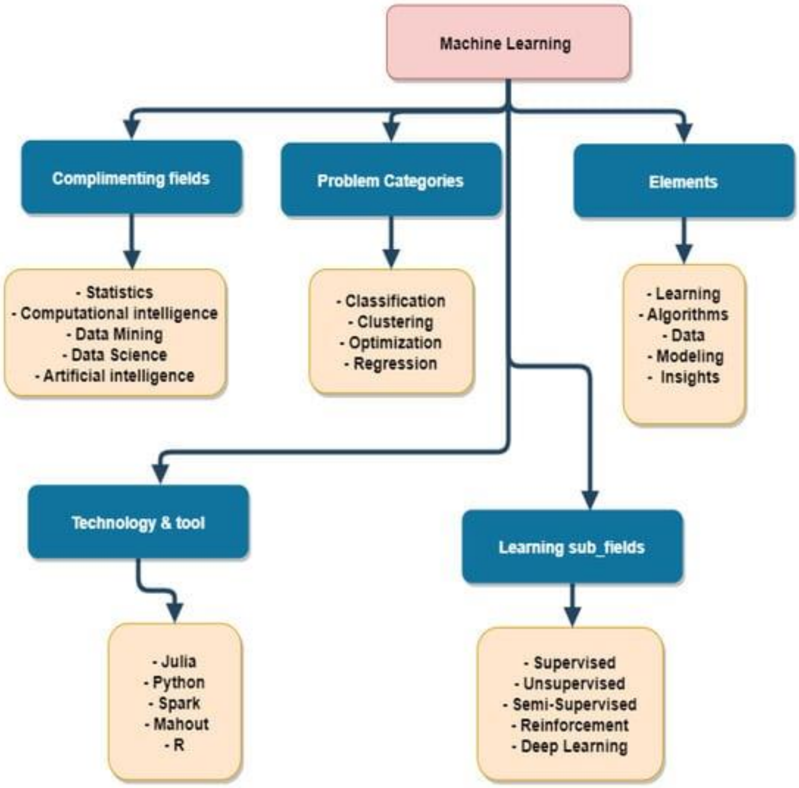
ABSTRACT

Convolutional Neural Networks (CNNs) are a fundamental class of neural networks widely employed in tasks such as image recognition and classification. They have diverse applications, including object detection, image processing, computer vision, and facial recognition. CNNs take images as input and automatically learn a hierarchy of features for classification, eliminating the need for manually engineered features. This is achieved by constructing multiple layers of feature maps, where each layer is generated by convolving the input with learned filters. Through this hierarchical approach, deeper layers are capable of capturing increasingly complex features that are robust to variations in position and distortion. The primary aim of this study is to provide a comprehensive understanding of CNNs, highlighting existing research gaps while examining their core components, functions, and other essential considerations.

INDEX TERMS: Intelligent systems, representation learning, predictive modeling, feature-learning neural network, AI-driven solutions, image labeling, Labeled data learning.

INTRODUCTION

In recent years, the adoption of machine learning (ML) has grown rapidly, with applications spanning research and real-world tasks such as text mining, spam filtering, video recommendation, image classification, and multimedia content retrieval . Among the various ML techniques, deep learning (DL) has emerged as a widely used approach in these areas. DL operates within the broader scope of ML and artificial intelligence (AI), effectively serving as a branch of AI that models information processing in a manner similar to the human brain. Traditional neural networks, from which DL evolved, have been outperformed by these advanced architectures. Furthermore, deep learning combines transformations and graph-based methods to build multi-layered learning models. Examining the subfields of learning shows that DL, as a subset of ML, is particularly focused on designing algorithms that emulate human cognition and problem-solving.



Convolutional Neural Networks (CNNs) are highly valuable in medical imaging due to their ability to accurately identify tumors and other anomalies in X-ray and MRI scans. By analyzing images of body parts, such as the lungs, CNN models can pinpoint potential tumor sites and detect abnormalities like bone fractures in X-ray images. This capability is achieved by learning from previously processed and labeled medical images, enabling the network to recognize patterns indicative of various conditions.

Survey Methodology

I conducted a comprehensive analysis of key research articles published between 2017 and 2022, with particular attention to studies from 2017, 2018, and 2019, along with selected papers from 2021 and 2022. The review focused primarily on publications from leading publishers, including IEEE, Elsevier, MDPI, ACM, and Springer, in addition to several papers from ArXiv. In total, over 60 papers covering various deep learning (DL) topics were examined, including 14 from 2017, 12 from 2018, 19 from 2019, and 15 from the years 2020 to 2022. This demonstrates that the review emphasizes recent developments in DL and convolutional neural networks (CNNs). The selected studies were systematically analyzed to achieve the following objectives:

1. Identify and describe DL and CNN techniques and network types;
2. Highlight challenges in CNN and propose possible solutions;
3. Summarize and explain CNN architectures;
4. Assess the applications of CNN.

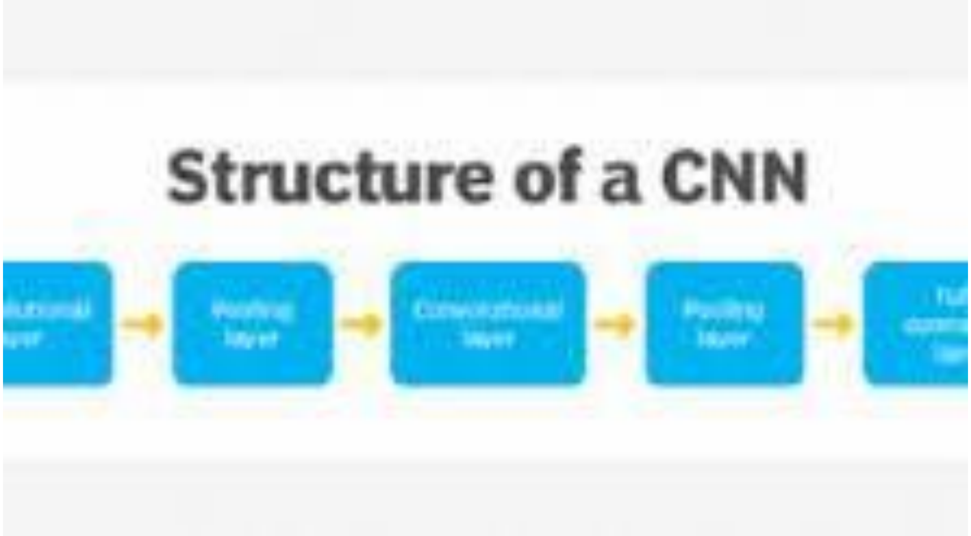
The most frequently used search terms for this review included “Deep Learning,” “Machine Learning,” “Convolutional Neural Network,” “CNN Architectures,” “CNN detection/classification/segmentation,” and “CNN Overfitting.”

Convolutional Neural Network (CNN or ConvNet)

Convolutional Neural Networks (CNNs) are a type of artificial intelligence system built on multi-layer neural networks, capable of recognizing, classifying, detecting, and segmenting objects within images. Often referred to as ConvNets, CNNs are a widely used discriminative deep learning architecture that can learn features directly from input data without requiring manual feature extraction. These networks are commonly applied in tasks such as visual recognition, medical imaging, image segmentation, natural language processing (NLP), and numerous other domains, as they are particularly well-suited to processing 2D structures . Compared to traditional neural networks, CNNs are more efficient because they automatically extract relevant features from the input, reducing the need for human intervention.

CNN Fundamentals

Understanding the various CNN components and their applications is critical to comprehending the advancements in CNN architecture. Figure_ displays several CNN parts.



CNN Layers

A typical CNN consists of four main types of layers:

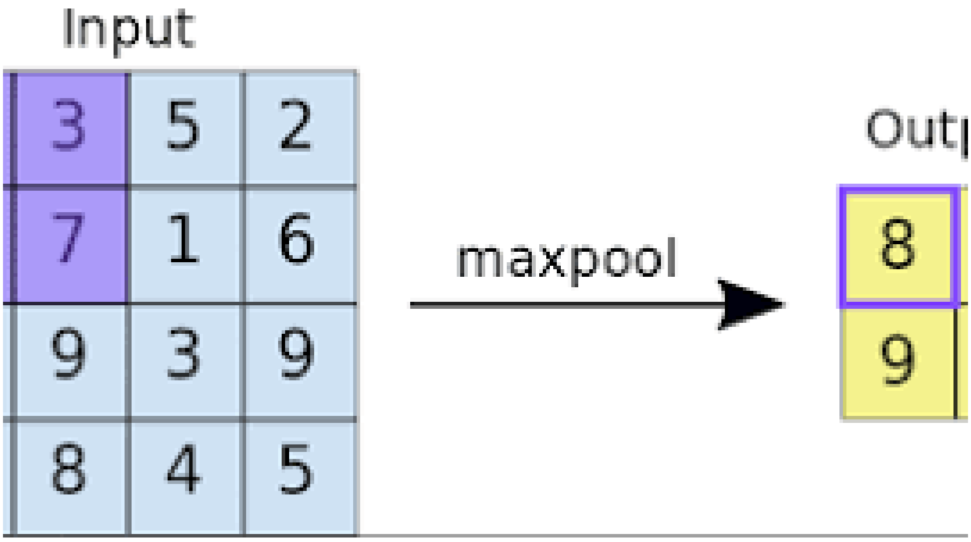
- **Convolutional layer** → Feature extraction layer
- **Pooling layer** → Subsampling or downsampling layer
- **Activation layer** → Non-linear transformation layer
- **Fully connected layer** → Dense layer

Input Image

The fundamental units of a digital image are called pixels, which represent the visual data in binary form. Each pixel, arranged sequentially in a matrix-like structure, has a value ranging from 0 to 255 that determines its brightness and color . When humans first observe an image, the brain processes a large amount of visual information almost instantaneously. Similarly, CNN layers are trained to identify simple features such as edges and curves in the initial layers, and progressively learn more complex structures, including faces and objects, in deeper layers. This hierarchical learning process allows CNNs to provide computers with a form of visual perception.

Convolutional Layer

The convolutional layer is a fundamental component of a CNN. It consists of a set of filters, or kernels, that are applied to the input data to extract important features. Each kernel has a specific width, height, and associated weights, which initially are assigned randomly and are gradually adjusted during training to capture meaningful patterns from the data. By convolving the input image (represented as an N-dimensional array) with these kernels, a feature map is generated.



3	5	2
7	1	6
9	3	9
8	4	5

A kernel is essentially a small matrix of discrete values, each associated with a weight. The initial weights are randomly chosen, and during the training process, these weights are updated so that the kernel learns to detect relevant features. Kernels allow operations in a high-dimensional implicit feature space without explicitly calculating coordinates; instead, the inner product of all data pairs in the feature space is computed. This “kernel trick” enables a linear model to behave non-linearly.

Before convolution begins, the input format must be prepared. Unlike classic neural networks that process vectorized data, CNNs accept multi-channeled images—RGB images have three channels, whereas grayscale images have only one. For example, consider a 4 × 4 grayscale image with a 2 × 2 kernel initialized with random weights. The kernel slides across the image both horizontally and vertically, computing the dot product at each position by multiplying corresponding elements and summing the results to produce a single scalar. This process is repeated across the entire image until all positions are covered .

The main image (K), the filter (L).
The output matrix is based on the equation
$(K - L + 1)$,
(1)
$4 - 2 + 1 = 3$, so the output then 3×3

In fact, the values of the dot product indicate the feature map of the output. Figure 3 visually represents the primary calculations performed at each stage. In this diagram, the smaller square (2 × 2) represents the kernel, while the larger square (4 × 4) represents the input picture. A product is then presented as a number after multiplying by both, and this sum provides an input value for the output feature map.

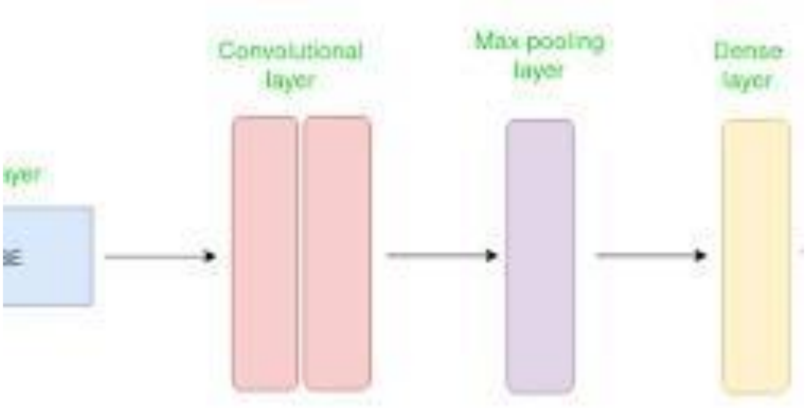
In the previous example, the kernel uses a stride of 1, which determines how many steps it moves horizontally or vertically across the input image, and no padding is applied. However, the stride value can be adjusted as needed. Increasing the stride reduces the dimensions of the resulting feature map, which can help control computational complexity.

Padding, on the other hand, affects the size of the input image and, consequently, the size of the feature map. By adding padding, the input image is effectively enlarged, allowing the kernel to cover border regions more completely and preserve spatial dimensions. Each filter in a CNN is designed to detect a specific feature; if the filter does not find a matching pattern in a region of the image, it does not activate. This mechanism enables CNNs to learn and optimize filters that are most effective for representing different objects in the data.

Padding: A limitation of the convolution operation is that information near the edges of an image can be lost, as the filter does not fully cover these regions. A simple and effective way to address this issue is to apply zero padding, which adds extra pixels with a value of zero around the image boundaries. In addition to preserving edge information, zero padding can also be used to control the dimensions of the output feature map.

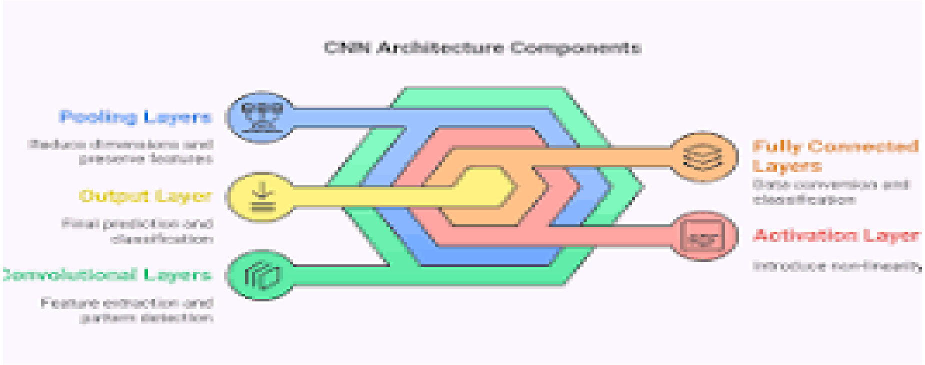
Features of CNNs: Thanks to the shared weight mechanism, CNNs are inherently translationally invariant, allowing the model to recognize features regardless of their position in the input image. During training, filters that are initially assigned random values gradually learn to detect relevant patterns, such as edges, as illustrated in Figure. However, it is important to recognize that using shared weights may not be suitable when assessing the spatial importance of specific regions in the input, as it assumes uniform significance across the entire image.

Training: A CNN model is trained using a dataset consisting of images along with their corresponding labels, which may include classes, bounding boxes, or masks. The training process relies on backpropagation, where the error between the predicted output and the actual output is calculated for each layer. This error is then used to update the weights of the neurons in the network, allowing the model to gradually learn and improve its performance.



Different Types of CNN Architectures

Classification: Image classification plays a vital role in processing multimedia data within the Internet of Things (IoT). This process analyzes input images to determine whether a particular condition or feature, such as a disease, is present. Techniques for image classification and recognition are widely applied in artificial intelligence, particularly in applications such as image retrieval, real-time object tracking, and medical imaging analysis. In recent years, deep learning approaches have gained significant attention for improving the accuracy and efficiency of these tasks.


CNN Architecture Components
Pooling Layers
Reduce dimensions and preserve features
Output Layer
Final prediction and classification
onvolutional Layers
Feature extraction and pattern detection
Fully Connected Layers
Data conversion and classification
Activation Layer
Introduce non-linearity

Detection

The difficult computer vision task of object detection involves anticipating both the location of the objects in the image and the kind of objects that were found. Beginners may find it difficult to differentiate between various related computer vision tasks.

For instance, the distinctions between object localization and object detection might be difficult to understand, even though all three tasks may be collectively referred to as object recognition. Image categorization, by contrast, is straightforward.

Segmentation

Segmentation: As the term suggests, image segmentation involves dividing an image into multiple meaningful regions, assigning an object label to each pixel. There are two primary types of segmentation: semantic segmentation and instance segmentation. In semantic segmentation, all objects of the same class are labeled identically, whereas in instance segmentation, each individual object is assigned a unique label, allowing the distinction of separate instances within the same category.

Future Directions

The classification performance of a CNN—measured in terms of accuracy, misclassification rate, precision, and recall—is strongly affected by the network's design choices, including the number of convolutional and pooling layers, the size and number of filters, the stride length, and the placement of pooling layers. Training a CNN also demands substantial computational resources, particularly GPUs, due to the intensive process of repeatedly testing and optimizing different combinations of parameters to achieve optimal results.

Popular Applications

Biometric Detection: Identity verification can be performed using unique biological traits. Biometric features such as hand geometry, retinal patterns, iris structures, and even DNA allow for the distinct identification of individuals. Object detection techniques rely on comparing input data against predefined templates to make accurate identifications. Surveillance systems, including CCTV cameras, are often used to monitor environments and capture any suspicious activity, with object detection playing a key role in tracking potential criminal behavior.

CONCLUSIONS

This paper provides a structured and comprehensive overview of deep learning, a core component of both artificial intelligence and data science. It begins with the history of artificial neural networks and progresses to modern deep learning techniques and their advancements across various domains. Key methodologies are explored, along with multi-dimensional modeling of deep neural networks. To organize this information, a taxonomy is presented that highlights different deep learning tasks and their diverse applications.

The review also considers both supervised learning with deep networks and unsupervised learning using generative models, as well as hybrid approaches that can be applied in real-world scenarios depending on the specific problem. Finally, the paper addresses several critical challenges in convolutional neural networks (CNNs), examining how various parameters influence network performance. The convolutional layer, which constitutes the core of a CNN, accounts for most of the computation, and the overall performance is affected by the number of layers, although deeper networks require more time for training and testing.

REFERENCES

1. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications, and Research Directions. *SN Comput. Sci.* **2021**, *2*, 1–21. [Google Scholar] [CrossRef]
2. Du, K.-L.; Swamy, M.N.S. Fundamentals of Machine Learning. *Neural Netw. Stat. Learn.* **2019**, 21–63. [Google Scholar] [CrossRef]
3. ZZhao, Q.; Zheng, P.; Xu, S.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Networks Learn. Syst.* **2019**, *30*, 3212–3232. [Google Scholar] [CrossRef]
4. Indrakumari, R.; Poongodi, T.; Singh, K. Introduction to Deep Learning. *EAI/Springer Innov. Commun. Comput.* **2021**, 1–22. [Google Scholar] [CrossRef]
5. AI vs Machine Learning vs Deep Learning|Edureka. Available online: https://www.edureka.co/blog/ai-vs-machine-learning-vs-deep-learning/ (accessed on 11 August 2022).
6. Cintra, R.J.; Duffner, S.; Garcia, C.; Leite, A. Low-complexity approximate convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5981–5992. [Google Scholar] [CrossRef]
7. Rusk, N. Deep learning. *Nat. Methods* **2017**, *13*, 35. [Google Scholar] [CrossRef]