International Journal of
Image Processing and Pattern Recognition

ISSN: 2456-6985
Volume 11, Issue 2, 2025
July—December
DOI (Journal): 10.37628/IJIPPR

https://journalspub.com/journal/ijippr

Review IJIPPR

Deep Learning with CNNs: A Survey of Techniques and Real-World Applications

Teena Uppal*

Abstract

Convolutional Neural Networks (CNNs) have emerged one of the most powerfully used tools in deep learning, particularly in image and pattern recognition tasks. CNNs are being revolutionized through deep learning resulting in a dominant architecture in computer vision as well as other various domains. This paper provides a comprehensive survey of CNN-based deep learning, its methodologies, regularizations, and computational efficiencies. We also address the challenges and future directions of CNN. This paper provides overview of CNNs. CNNs are differentiated into several layers: input layer, convolutional layer, activation layer, pooling layer, fully connected layer, and the output layer. Over time CNNs evolved significantly. Key methods of training CNNs include data augmentation, transfer layering, and batch normalization as well as additional techniques like dropout and early stopping to improve generalization. CNN has real world applications in security, medical imaging, traffic management, defence, agriculture and pattern recognition. This paper emphasizes the significance of CNN architecture, model evolution, training techniques, and practical use cases. Despite overall development, CNNs continue to evolve and will be even more intelligent and efficient as AI technology advances in future.

Keywords: CNN (convolutional neural networks), artificial intelligence, real world applications, natural language processing (NLP), rectified linear unit

INTRODUCTION

CNN is a part of deep learning that provides grid like data efficiency in images and time series. CNN makes people accomplish impossible tasks to make possible, such as autonomous vehicles, face recognition, intelligent medical treatment, driver less cars, and self-service supermarket, have transformed machine learning through feature interaction and representation learning. CNN plays an important role in image processing, medical diagnosis, and Natural Language Processing (NLP). It has been widely used for object recognition tasks but now it is being examined in other aspects like object tracking, text detection and recognition, pose estimation, action recognition, scene labelling, and many more. This paper provides an overview of art techniques and challenges to be addressed. This paper also provides an overview of CNN architecture, its methodologies, applications, and optimizations. Earlier model of CNN was designed to classify handwritten digits and was successful in recognizing visual patterns directly from input image even without any pre-processing, but it works for just simple patterns not for complex patterns due to lack of training data and computing power, so this operation failed. CNN is quite versatile, easy to implement, and less complex than traditional

***Author for Correspondence**
Teena Uppal
E-mail: Uppalteena9@gmail.com

Student, Department of Multimedia, BBK DAV College for Women, Amritsar, Punjab, India

Received Date: April 28, 2025
Accepted Date: August 07, 2025
Published Date: December 22, 2025

Citation: Teena Uppal. Deep Learning with CNNs: A Survey of Techniques and Real-World Applications. International Journal of Image Processing and Pattern Recognition. 2025; 11(2): 1–5p.

models especially in the field of visual analysis.

CNN ARCHITECTURE AND EVOLUTION IN MODELS

Convolutional Neural Networks (CNN) is a deep learning model designed to perform specific tasks, like recognize patterns and to process images, CNNs are fully different from traditional machine learning which require manual feature extraction, whereas CNN automatically learn important features from input of image or data and makes them highly efficient for tasks such as object detection, image recognition, and text analysis. CNNs have evolved significantly with improvements in network depth, computational efficiency and optimization techniques. So, here, we learn different layers and models of CNN.

CNN Workflow Step by Step

CNN works as a smart detective that analyzes an image in small pieces and then slowly builds understanding of the whole picture. This task takes place in different layers, and each layer has its own task. There are six layers such as input layer, convolution layer, activation layer, pooling layer, fully connected layer, and output layer. Let us discuss these layers in detail.

Input layer (Feeding the Image)

First step in CNN is to feed the image as input. Computers do not see images, like humans, so they see images with numbers. Firstly, images are converted into a grid of numbers, and each number represents the brightness of pixels. For example, a grayscale image of 32 x 32 pixels is represented as a 32 x 32 in grid of numbers and in case of a colored image (RGB), it is stored as three separate grids (Red, Green and Blue channels).

Convolution Layer (Feature Extraction)

CNN looks for small patterns in the image using filters known as kernels. These filters move across the image and detect features like curves, edges, colors, or textures. In simple words we can say that it extracts features from input data using filter-based operations. This stage of CNN does not recognize complete objects because it includes complex patterns and this stage is just for simple patterns. For example, a filter may detect horizontal edges, vertical edges and by stacking multiple filters, the network can recognize complex objects, but it will identify simple patterns only.

Activation Functions (Adding Non-Linearity)

After convolution layer extracts features, comes the role of activation function which helps introduce non-linearity. The most common activation function is ReLu (Rectified Linear unit). So, in other words, activation function helps models analyzes complex patterns and without it models would not be able to capture complex patterns. Its basic function is to add non-linearity.

Pooling Layer (Reducing Size, Keeping Important Features)

Its basic feature is to reduce the file size while preserving the important details. It helps to make network faster by reducing amount of data to process. Most common form of Pooling is max pooling where only the highest value is kept. Fully Connected layer (Final Decision-making) after several layers of convolution and pooling the extracted features are flattened into a single layer and is sent to this layer. This layer generally makes predictions. For example, if we want to train CNN to recognize each digit from 0 to 9, the final layer will have 10 neurons. And if we want to train a CNN to classify between a dog and a cat, the final layer will have two neurons.

Output Layer (Prediction)

This final layer uses an activation function, like SoftMax, which assigns probability to each category and the highest probability determines the final output. For example, if a CNN is classified between animals:

- Cow 80%.

- Cat 18%.
- Bird 2%.

The network will predict Cow as a result.

Evolution in Models

The first mathematical models of neurons were the MP model, introduce in 1943 by McCulloch and Pitts [1]. Then in late 1950 and early 1960, a single layer perceptron model was proposed by Rosenblatt which added learning ability to MP model. Then in 1986, Hinton et al. [2] proposed a multi-layer feed for network trained by error back propagation algorithm which addresses problems that single layer perceptron could not solve. Waibel et al. [3] proposed a time delay neural network in 1987 for speech recognition which was one dimensional. Then Zang [4] proposed two dimensional CNN which was called shift variant artificial neural network (SIANN). Then in 1989, Lecun et al. [5] constructed a CNN for handwritten zip code recognition. In 1990s [6], CNNs, like chaotic neural networks and general regression neural networks, were successfully proposed. Le-Net-5 was most famous [7]. In 2006, Hinton et al. [8] proposed first multi-hidden layer artificial neural networks that have excellent feature learning ability. Second-layer-wise pretraining can overcome difficulties of training deep neural networks. In 2012, the best classification results at that time were achieved using a deep CNN in the ImageNet Large Scale Visual Recognition Challenge [9, 10].

CNN Training Techniques

Training a CNN requires several techniques to ensure its efficient learning and avoiding common problems like overfitting and poor generalization. Below are some training techniques of CNNs:

Data Augmentation

CNN requires large amounts of data to generalize which can be time consuming and expensive. Data Augmentation artificially expands dataset by applying transformations to it. Basically, it increases data variability to prevent overfitting.

COMMON DATA AUGMENTATION METHODS

- *Rotation:* Randomly rotates images to make the model invariant to orientation.
- *Flipping:* Horizontally or vertically flips images to simulate real-world variations.
- *Cropping:* CNN dynamically studies spatial patterns of images and crop them.
- *Scaling & Zooming:* Increases or decreases the size of objects in an image.
- *Color Jittering:* Adjusts brightness, contrast, and saturation.
- *Adding Noise:* Introduces small distortions, making the model more robust.

Transfer Learning

Instead of training a model from basic level from start transfer learning allows models trained on previous datasets to be reused, e.g., ImageNet.

PARTS OF TRANSFER LEARNING

- *Feature Extraction:* A pre-trained CNN models convolution layers are freeze and only final classification layer is used for feature extraction.
- *Fine-Tuning:* Deeper layers of these pre-trained models are used for new datasets to adapt to tasks.

Batch Normalization

Batch normalization stabilizes and speeds up training process. It basically takes a mini batch of data, then calculates its mean and variance. Each variation is adjusted to mean value of 0 and variance 1. It allows values to adjust as needed by learning two extra parameters, gamma and beta.

Dropout

When CNN learns training data instead of general patterns overfitting happens. So, it is important to reduce overfitting here dropout helps as it randomly deactivates neurons during training so that their focus goes on multiple feature representation. It basically improves models' generalization by stopping neurons from adapting a lot.

Early Stopping

When the model performs well on training data but poorly on unseen data again the case of overfitting there also early stopping comes to rescue by preventing neurons from learning too much. Here model is rained as usual, but the only difference is that the validation loss is tracked. It benefits by saving computation time and prevents models from memorizing noise.

Learning Rate Scheduling

The main challenge is choosing the correct learning rate, if the learning rate is too high it causes unstable training and if it is too low it can cause slow convergence. Hence, it is very important to have perfect learning rate. Learning Rate Scheduling helps improve convergence by adjusting learning rate.

Types of Learning Rate Scheduling

- *Step Decay:* Training rates are reduced at fixed intervals.
- *Exponential Decay:* Learning rates are decreased exponentially.
- *Cyclic Learning Rate (CLR):* Alternates between high and low learning rates.

Gradient Clipping

When weights become extremely large sometimes gradients explode making training unstable. So, gradient clipping handles exploding gradients in deep networks by limiting maximum value of gradients. It helps in stabilizing training and prevents divergence in deep networks.

REAL WORLD APPLICATION OF CNN

CNNs are used in real life in many places like Computer Vision, Natural Language Processing, Medical Imaging, and Security and Surveillance.

Computer Vision

It helps in image classification used mainly in google photos or face book facial recognition or even your face lock password in mobile. It also helps with object detection, for example, in YOLO, SSD for autonomous vehicles, etc., image segmentation is also a part where computer vision is used in form of CNN through medical image analysis, satellite imaging, etc.

Natural Language Processing (NLP)

CNNs have applications in sentiment analysis and text classification.

CNNs extract key features from text for sentiment analysis and spam detection. It identifies things such as names, date's locations in text, etc. CNN based encoders even improve accuracy in deep learning models.

Medical Imaging

CNNs have automated analysis for medical images.

It helps identify tumors or cancer in MRI, CT scans, and X-rays. It also assists in detecting diabetic retinopathy and glaucoma. Not only this, but it also helps in detecting COVID 19 by classifying chest X-rays.

Security and Surveillance

CNNs improve accuracy in biometric and surveillance applications.

It detects suspicious activity in surveillance cameras, identifies individuals with face recognition in real time. CNN also detects any type of malware or virus found keeping the device safe.

CONCLUSIONS

CNNs have revolutionized deep learning in many ways. Some of which are discussed above. This survey has outlined training techniques of CNN, its real-world examples, evolution of models, and many more things which conclude that CNN has more advancements to be made which can make it smarter. As AI continues to evolve, CNNs still remain a fundamental component for making intelligent systems.

From the start of using CNN in handwritten digits to their modern roles in various advanced areas CNNs have both flexibility and power. As AI advances, CNNs will remain the pillar in building more intelligent systems.

REFERENCES

1. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943 Dec;5(4):115–33.
2. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986 Oct 9;323(6088):533–6.
3. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ. Phoneme recognition using time-delay neural networks. In: Backpropagation. New Jersey: Psychology Press; 2013. p. 35–61.
4. Zhang W, Tanida J, Itoh K, Ichioka Y. Shift-invariant pattern recognition neural network and its optical architecture. In: Proceedings of annual conference of the Japan Society of Applied Physics. 1988 Aug;564.
5. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989 Dec;1(4):541–51.
6. Aihara K, Takabe T, Toyoda M. Chaotic neural networks. Phys Lett A. 1990 Mar 12;144(6–7):333–40.
7. Specht DF. A general regression neural network. IEEE Trans Neural Netw. 1991 Nov 1;2(6):568–76.
8. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 2002 Aug 6;86(11):2278–324.
9. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2021 Jun 10;33(12):6999–7019.
10. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25.

