P-ISSN 1991-8941
E-ISSN 2706-6703
Vol. 19, No. 2 (2025)

Journal of University of Anbar for Pure Science
(JUAPS)

journal homepage: juaps.uoanbar.edu.iq

Open Access

Review Paper

Accuracy and efficiency trade-offs in deep learning approaches for object recognition: A comparative study

Shahad F. Mohammed[1*], **Khalid Shaker**[2]

[1]Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq

[2]Department of Information Technology ,College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq

ARTICLE INFO

Received: 15/11/2024
Accepted: 13/03/2025
Available online: 19/11/2025
December Issue
10.37652/juaps.2025.155224.1336

Cite @ JUAPS

Corresponding author
Shahad F. Mohammed
shh22c1006@uoanbar.edu.iq

ABSTRACT

Object recognition is a research area in Computer Vision and Image Processing that covers several topics, including face recognition, gesture recognition, human gait recognition, and traffic road signs recognition, among others. It plays a vital role in several real-time applications such as video surveillance, traffic analysis, security systems, and content-based image retrieval. It is the task in which an object can be recognized and labelled within an image. It aims to bring the visual perception capabilities of human beings into machines and computers. Artificial Intelligence (AI) has a great interest in this field and is involved in most, if not all, of the various fields of life. Deep Learning (a new area of Machine Learning) is one of the AI techniques that is generated from Artificial Neural Network (ANN). Recently, it has gained popularity due to its competitive results in improving the efficiency of real-time decision-making. For this reason, it has been widely used in many fields, including object recognition. Therefore, this review provides insight into previous studies on how deep learning methods have been applied to object recognition, among various datasets. It classifies the deep learning methods along with different targeted objects, their contributions, the challenges they faced, and how the results were gained. These aspects will be discussed to introduce the researchers to more general knowledge about the recent techniques applied in this field. Practical experiments have proven the efficiency of Convolutional Neural Networks (CNNs) in object recognition tasks, achieving high-accuracy results within an acceptable timeframe.

Keywords: *Classification, Convolutional neural network, Deep learning, Object recognition*

1 INTRODUCTION

Digital technologies are essential to modern life to fulfil user demands [1]. Object recognition has received the attention of researchers in computer vision and image processing [2]. It is the process of identifying, classifying, and giving appropriate labels for items in images [3]. In various fields such as autonomous driving, medical imaging, virtual reality, augmented reality (AR/VR), AI-driven robotics, and others, object recognition plays a crucial role. Along with that, the use of Deep Learning to tackle visual problems is expanding [4]. Object detection aims to identify visual items of specific classes using bounding boxes, classifying them into their appropriate categories [5–7]. Based on their features, objects in images/videos are categorized in object recognition, which could be done using machine learning, deep learning, and image processing techniques [8]. The process of object recognition entails getting the distinctive data and feeding it into a classifier [9]. This approach gains precise recognition of objects by learning feature values [10]. It includes real-time scanning and neural network training, based on probability values, and the identification model

to classify objects [11]. Object recognition encounters difficulties, including shifting illumination, occlusions, and various viewpoints [12]. Neural networks are used by object recognition devices to track objects precisely [13]. It has been noticed that accurate object recognition in images can be achieved through applying Deep Learning techniques [14]. With the advancement of Deep Convolutional Neural Networks (DCNNs) and increased CPU power, CNNs are used to classify, detect, and segment objects in images [15].

2 DEEP LEARNING ARCHITECTURE

Deep Learning, a subgroup of Machine Learning, employs layered neural network techniques like Deep Neural Networks (DNN) and Convolutional Neural Networks (CNNs) to analyze data representations, aiding in fields like image analysis and pattern recognition that have applications for resolving complicated issues in domains like biology, medicine, and others [16, 17]. Deep Learning is an Artificial Intelligence function that imitates how the human brain absorbs information and forms patterns to be used in decision-making [18]. There are three main categories of deep learning: generative deep architectures (unsupervised), discriminative deep architectures (supervised), and hybrid deep architectures [19]. The classification of the deep learning approaches/architectures is illustrated in Figure 1.

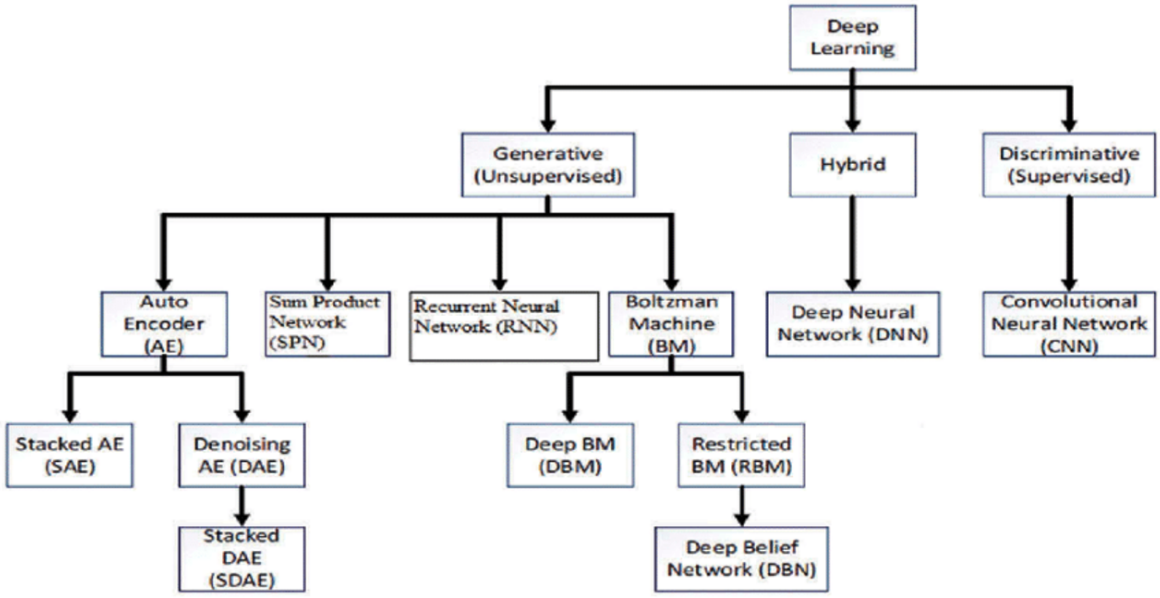

Fig. 1 Taxonomy of deep learning techniques [20]

2.1 Deep networks for unsupervised or generative learning

Deep learning generative architecture is also referred to as unsupervised learning, which uses unlabelled data to train models to extract valuable information from large and complicated datasets [21]. It has shown high effectiveness in natural language processing, image and speech analysis, and the de novo production of drug-like compounds with desired characteristics [22].

2.1.1 Autoencoder (AE)

Autoencoder (AE) is one of the Artificial Neural Networks (ANNs) that is used for unsupervised learning tasks when the unlabelled data are not needed, obtained for feature learning and input data reconstruction with the aim of acquiring an "informative" data representation [23, 24]. It discovers the ideal parameters needed to rebuild its output in the most similar way to its input. It uses the Backpropagation algorithm, and the target values are

made to equal the inputs and attempts to learn how to approximate the identity function [25, 26]. AEs consist of three or more layers in the neural network (NN):

1. An input layer (to have adequate coding (spectra in speech or image pixels, for example) 1.A hidden layer (when the number of hidden layers is greater than or equal to 1) usually has fewer dimensions than the input layer and is named as an under-complete or sparse autoencoder.

2. An output layer (that matches the input layer dimensions) [26].

AEs reduce dimensionality for complex high-dimensional data. They are helpful with limited resources applications due to data compression [25]. Additionally, they perform exceptionally well in abnormality identification by measuring the reconstruction error [24].

2.1.1.1 Vanilla auto-encoder Rumelhart (1985), who first presented the idea of AE in a research study [24]. The most basic type of auto-encoder is called "vanilla". A vanilla auto-encoder is composed of an input and output layer, with one or more hidden layers. Figure 2 shows the general structure of a vanilla auto-encoder when the input data to the input layer is denoted by X, the data in the hidden layer is denoted by Z, and the reconstructed output data within the output layer is symbolized by X' [25, 27]. This was employed efficiently for unsupervised models, especially feature extraction and dimensionality reduction [28].

2.1.1.2 Stack auto-encoder (SAE) Stacking multiple layers of auto-encoders produces the stacked auto-encoders (SAEs), which are sophisticated neural network structures that enhance representation learning and feature extraction.

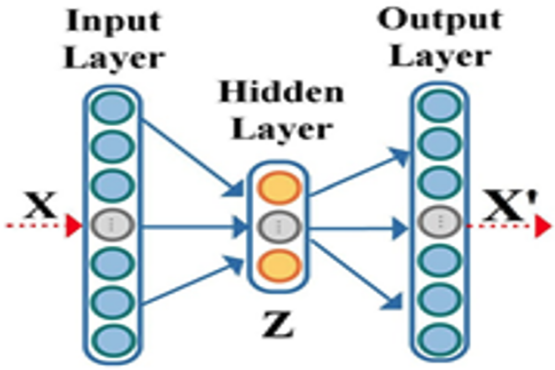

Fig. 2 The Auto Encoder Structure [25]

SAEs are increasingly popular in industrial process monitoring and anomaly detection due to their multilayer structure, where each layer trains a sequence of auto-encoders, where the output of one AE will be the input to the next one, firstly the initial layer catches the fundamental features after that the following layers obtains the other ones [29], this means that the SAE training process is obtained by a layer-wise approach [30, 31].

2.1.1.3 Auto-encoder hyper-parameters Initializing several autoencoder hyperparameters is required before starting the training process. Some of these are established before training and stay constant, whereas others can be dynamically adjusted throughout training to improve the model's execution. The process of choosing and modifying these parameters is necessary because it directly affects the efficiency of the AE's outcomes. These hyperparameters include [25]:

- Activation function.
- Number of neurons in each layer.
- Optimization algorithm.
- Number of epochs.
- Learning rate.
- Size of latent space.

- Number of hidden layers.
- Batch size.

In the field of computer vision, autoencoders have been utilized to perform many tasks, such as object recognition. Liu et al [32] utilized the Context-LGM (Contextual Latent Generative Model) and Variational auto-encoder (VAE) for context-aware object recognition, achieving an accuracy of 80.27% using the Data Science Bowl (DSB) 2017 dataset. A multi-modal adversarial auto-encoder structure is presented by Nitsch et al. [33] to transfer information from images to a point cloud for object recognition applications. They used the RGB-D object dataset with an accuracy result of 55.7% for images and 64.59% for point clouds, while using the 3D MNIST dataset with an accuracy result of 92.2% for images and 80.4% for point clouds. Xiong et al. [34] employed the feature transfer learning and a stacked autoencoder to recognize different operating conditions in several diesel engine types, concentrating on the reconstruction of vibration signals. These studies clarify the significant effect of auto-encoder techniques on the precise object recognition outcomes in various fields.

2.1.2 Boltzmann machine (bm)

Due to their interconnected binary units, Boltzmann Machines (BMs) are strong graphical models based on statistical machines that are utilized in unsupervised learning. BMS was suggested by Hinton and Sejnowski. They have been used in a variety of domains, such as quantum computing and biology [35, 36]. BM is a network of symmetrically connected, neuron-like units that choose whether to be on or off (0 or 1) in a random manner [18]. Boltzmann machines are undirected networks, allowing values to be multiplied by weights through Gibbs' sampling procedure [37]. The learning task in the Boltzmann Machine involves identifying ideal parameters for faithfully depicting the data distribution, requiring certain essential points to be achieved. To gain this learning task, the following essential points should be achieved [38]:

- Parameter optimization
- Learning algorithm
- Training process
- Regularization technique
- Empirical evaluation

Boltzmann machines (BMs) are split into two classes: Restricted Boltzmann Machine (RBM) and Deep Boltzmann Machine (DBM), which comprise many layers of RBM. These RBMs are referred to as Deep Belief Networks (DBN) when they are piled on top of one another [36, 39].

2.1.2.1 Restricted Boltzmann Machine (RBM) BMs take a long time for training due to exponential link growth with the nodes increasing by 1. RBMs, as presented by Smolensky [36, 37], are typically less expensive than Boltzmann Machines due to their more effective training techniques and simpler architecture [40]. The requirements that change a BM to an RBM are:

- The hidden nodes might not be linked to one another.
- The visible nodes might not be linked to one another [37], which is the opposite of the BM.

The training process of RBM can be unsteady, especially with low-dimensional clustering datasets. Methods such as biased Monte Carlo sampling are suggested to improve model quality and training efficiency [41].

2.1.2.2 Deep Boltzmann Machine (DBM) RBMs may capture hierarchical data representations by stacking numerous of them. Because of its multi-layer structure, DBMs can represent more complex relationships between variables [42]. The layered structure enhances the generalization skills and facilitates better processing of high-dimensional data [43]. Figure 3 shows the general structure of Boltzmann, Restricted Boltzmann, and Deep Restricted Boltzmann Machine.

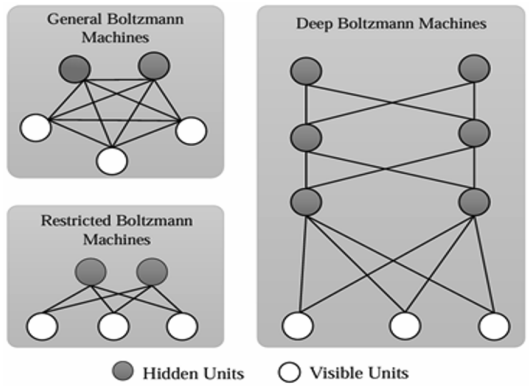

Fig. 3 A visual comparison between BM, RMB, and DMB [44]

Boltzmann Machine, including its various architectures, has been used in object recognition. One of the studies in this field is being used in more complicated object recognition applications. Zorzi et al. [45] utilized RBMs to show the scaling capability and reached more than 90% accuracy ratio on CIFAR-10, even with the steady performance in comparison with other techniques like CNN. Wan et al. [46] proposed Deep Belief Network (DBN) models with fine-tuning utilizing supervised learning, reached successful outcomes on various object recognition criteria, such as accuracies of 97% on the SVHN dataset and 91% on the CIFAR-10 dataset. When Deep Boltzmann Machines (DBMs) perform less well than CNNs in object recognition tasks due to training challenges and high computational costs, they are excellent at capturing complex data structures [47].

2.1.3 Recurrent neural network (RNN)

RNN is an Artificial Neural Network designed to identify patterns in data sequences, including time series, speech recognition, or natural language. The recurrence enables RNNs to maintain context and generate well-informed predictions according to the previous input data [48]. RNNs have a directed graph structure, with self-looping links. This means that the output of a hidden node at the current time is transmitted to the corresponding one for the next time interval, enabling them to store data for a longer period [48, 49]. Figure 4 shows the general structure of the Recurrent Neural Network (RNN).

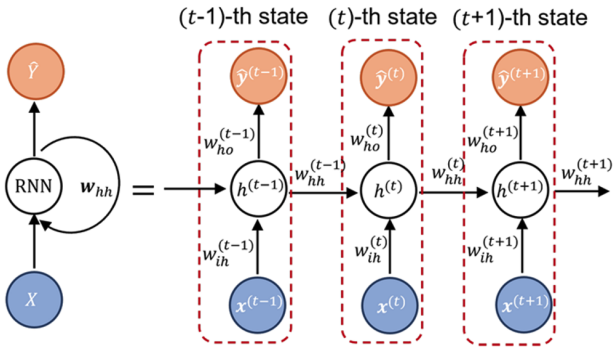

Fig. 4 General structure of RNN, in terms of time, the unrolled form is on the right, with regard to time, of the left side, where recurrent feedback is represented by the directed loop [49]

The short-term memory challenge of a simple RNN training process limits them from keeping data for long sequences [50]. To avoid that, more sophisticated RNNs have been developed, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) [48], bidirectional LSTM, bidirectional GRU, and Bayesian RNN [50, 51].

2.1.3.1 Long Short-Term Memory (LSTM) Long Short-Term Memory (LSTM) is a specific kind of RNN that was introduced to deal with issues like gradient vanishing and explosion, for efficient sequential data management [52]. An LSTM Network works by feeding its unit with the current input at a given time step, as well as the output from the previous time step. After that, the output is transmitted to the subsequent time step. For categorization purposes, the last hidden layer of the last time step and occasionally all hidden layers are commonly used [53].

Three primary gates that the LSTM consists of for managing the data flow as follows:

1. **Input Gate:** It determines the recent data based on the previous and current input, using the tanh function to generate a vector. The

input data is filtered by the sigmoid activation function.

2. **Forget Gate:** It decides how and which data should be removed (forgotten) from the previous cell state by utilizing the sigmoid activation function. If the result is 0, this implies "completely forget"; otherwise, if it is 1, it implies "completely retain". Allowing LSTM to concentrate on the crucial information for the present prediction.

3. **Output Gate:** It regulates the information flow from the cell state to the following hidden state, by performing a sigmoid function and multiplying by tanh to guarantee proper scaling of the output, deciding the following hidden state to be transmitted.

LSTMs work efficiently for tasks such as short-term load forecasting because these gates cooperate to allow LSTMs to preserve long-term dependencies in data [54]. Figure 5-a shows the general structure of the LSTM, and Figure 5-b shows the internal architecture of the LSTM.

The object recognition using RNN has been of interest in recent studies. A deep complementary feature classification network was developed by Chen et al. [55] for planet disease recognition with an accuracy of 93.46% which is higher than the baseline model by 7.2%. Using the Snapshot Wisconsin dataset and labelled files from the wildlife program of the British Columbia Ministry of Transportation and Infrastructure. This research indicates the use of LSTM networks for object identification tasks, particularly in the topic of wildlife monitoring. RNNs, the Long Short-Term Memory (LSTM) networks, have shown promising results in object recognition when combined with other deep learning techniques, particularly for temporal context tasks [56].

2.1.4 Sum-product network (SPN)

Sum-product networks (SPNs) are sophisticated probabilistic models developed to improve inference and learning by utilizing a rooted acyclic-directed graph topology. They are effective with a range of tasks, including natural language processing and computer vision, because of their ability to combine the benefits of probabilistic graphical models with deep learning [57]. SPNs consist of two types of nodes:

1. **Sum nodes:** A weight is assigned to each child node, signifying its contribution. These nodes sum their child nodes' output, which may be other sum or product nodes. Consequently, mixtures of distributions can be efficiently represented.

2. **Product nodes:** These nodes are responsible for the multiplication of their child nodes' outputs, to capture the combined distribution of independent variables.

In addition to the leaf nodes that describe the variables or data features that have been observed. SPNs can model complicated probability distributions effectively and enable inference in polynomial time accurately, due to the hierarchical organization of these nodes [58]. Figure 6 presents an example of applying a sum-product network on 2 Boolean variables Y1 and Y2 [59]. Here are additional literature reviews on the methods of unsupervised (generative) deep learning applied to object recognition systems, as shown in Table 1. Zhang et al. [60] utilized Sparse Auto-encoder with softmax classifier on the Minist Dataset, improving the recognition accuracy to 0.985. This is a high accuracy with a recall rate improved by 4% compared to other models. However, it faces Local optimization issues, learning time-consuming, gradient elimination, and parameter initialization issues. Another study by Rahal et al. [61] applied Hidden Markov Models (HMMS) with Deep Sparse Auto-encoder (SAE) on P-KHATT, APTI, IFN/ENIT, and MNIST datasets. The study enhances feature extraction, dictionary learning, codebook generation, and recognition performance, achieving 99.95% accuracy for P-KHATT and 99.40% accuracy for IFN/ENIT using a 13 by 13-pixel patch. However, it has limitations because of information loss, frequent

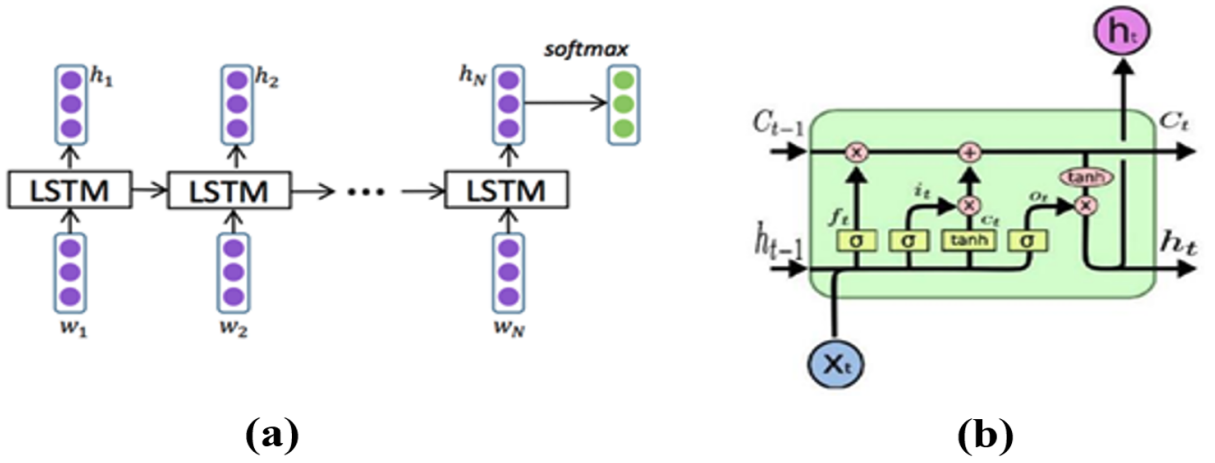

Fig. 5 (a): The high-level LSTM model architecture [51], and (b): The LSTM internal architecture [55]

Descriptors, a big codebook, and spatial layout issues. Kodepogu et al. [62] employed the CNN, RNN, and LSTM with the Deep Auto-encoder (DAE) on the GTSDB dataset.

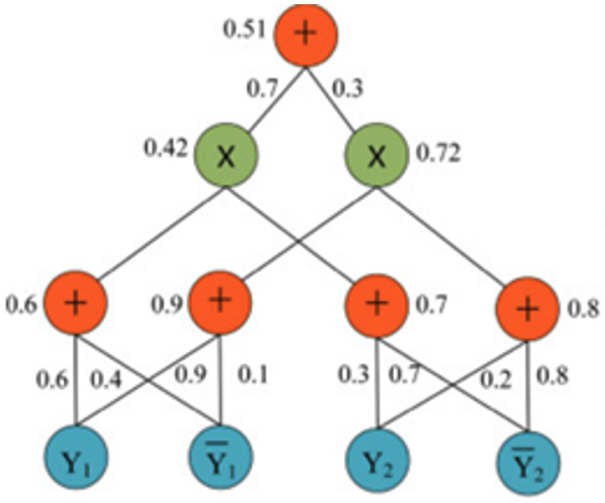

Fig. 6 Applying SPN on 2 Boolean variables Y1 and Y2 [59]

The study enhances feature representation, is efficient in capturing essential features from input images for future classification applications, and improves robustness for traffic sign recognition. However, it still faces potential issues in handling diverse real-world situations and the complexity of multiple classifiers' integrity, which can reduce the overall efficiency and speed of the recognition system. While Yu et al. [63] suggested employing the stacked denoising auto-encoders (SDAE) and manifold regularization on the WM-811k dataset. The SDAE improves wafer map defect recognition by efficiently learning noise-resistant features, achieving 98% accuracy and better generalization compared to conventional approaches. Although the training cost increased due to potential issues like excessive computing complexity and overfitting to the training set. Huang et al. [64] suggested BlazePose architecture with LSTM and YOLO v4 as a classifier on 243 videos showing 27 different actions, with 5 distinct activities that are all object based, achieving an accuracy of 95.91% in identifying actions and 97.68% in object recognition on 243 videos involving 27 distinct actions and 5 object-based activities, but significant occlusion and appearance changes can significantly impact the accuracy of landmark detection. Muzahid et al. [65] suggested using the Hand-crafted features based on ORB and SIFT with the K-NN, decision tree, and random forest (RF) classifiers that are tested on a dataset of 8000 samples from 100 class objects. Although achieving an accuracy of 85.6%, using hand-crafted features may not perform as well as deep learning approaches in more complex cases. Lu

et al. [66] proposed using a Progressive Conditional Generative Adversarial Network (GAN) with a CNN as a classifier on the ModeNet40 dataset, which improved the precision of object recognition with an accuracy of 95.2%. This faces generalization challenges across various datasets and overfitting issues due to its reliance on synthetic data. Ismael et al. [67] suggested employing the GAN-LSTM on the CARRADA dataset, which achieved an object classification accuracy enhancement of 7.86% compared to other approaches; however, it faces challenges with sparse point clouds and multipath issues associated with millimeter-wave radar. As summarized in Table 1 .

Table 1 Unsupervised (generative) deep learning methods applied to object recognition systems

Method(s)	Classifier	Dataset	Achievement	Limitation	Ref.
Auto-encoder (Variational auto-encoder)	Contextual posterior Transformer	Data Science Bowl (DSB) 2017	An accuracy result of 80.27%; effective modeling of object–context relation; hierarchical contextual features; robustness against environmental factors.	Dependence on object–context relation; computational complexity; limited generalization to diverse contexts; annotation challenges.	[32]
Auto-encoder	Multi-modal adversarial auto-encoder (AE)	RGB-D object; 3D MNIST	Study reduces labeling effort using the RGB-D object dataset (55.7% accuracy) and achieves 92.2% accuracy on 3D MNIST, demonstrating successful transfer learning and feature disentanglement.	3D point-cloud data labeling is time-consuming; transferring abstract features across modalities may reduce performance; strong dependence on data distribution.	[33]
Restricted Boltzmann Machine (RBM)	CNNs	CIFAR-10	Shows scalability, reaching more than 90% accuracy.	Performance peaks compared with some other techniques such as CNN.	[45]
Deep Belief Network (DBN)	CNN with DropConnect regularization	SVHN and CIFAR-10	Enhanced DBN models with supervised fine-tuning; high performance on object-recognition tasks; accuracy of 97% on SVHN and 91% on CIFAR-10.	For large-scale datasets, DBNs are susceptible to overfitting and require high computational cost.	[46]
Deep Boltzmann Machine (DBM)	Convolutional Deep Belief Network (CDBN)	CIFAR-10 and NORB	DBMs capture complicated data structures; CDBN reaches about 78.9% accuracy on CIFAR-10 for object recognition.	High computational cost and training challenges may make them weaker than CNNs in object-recognition applications.	[47]
Deep Complementary Feature Classification Network integrating weakly supervised detection with DeepLabv3+ and CRF	RNN (bidirectional Gated Recurrent Unit, Bi-GRU)	PlantVillage dataset; self-collected grape-disease dataset	Accuracy of 99.21% on PlantVillage and 93.46% on the grape-disease dataset, surpassing the baseline model Xception-65.	High accuracy but longer prediction time than some models; trade-off between computational efficiency and precision.	[55]
LSTM	Random Forest (RF)	Evaluation on a real testbed	On a real testbed, achieves 98.1% accuracy in object-recognition tasks.	Requires substantial training overhead compared with some other techniques.	[56]
Sparse auto-encoder	Softmax classifier	MNIST dataset	High recognition accuracy (rate 0.985), improving classification and recall by 4% compared with other models.	Local-optimization issues; time-consuming training; gradient vanishing; parameter-initialization problems.	[60]
Hidden Markov Models (HMMs) with Deep Sparse Auto-encoder (SAE)	Hidden Markov Models (HMMs)	P-KHATT; APTI; IFN/ENIT; MNIST	Improved feature extraction, dictionary learning, codebook generation, and recognition; 99.95% accuracy for P-KHATT and 99.40% for IFN/ENIT using a 13×13 pixel patch.	Information loss; frequent descriptors; large codebook; spatial-layout issues.	[61]
Deep auto-encoder (DAE)	CNN, RNN, and LSTM with DAE	GTSDB	Improved feature representation and efficient capture of essential features from input images; enhanced robustness for traffic-sign recognition.	Handling diverse real-world situations and integrating multiple classifiers can be complex and may reduce overall efficiency and speed.	[62]

Continued on next page

Table 1 (continued).

Method(s)	Classifier	Dataset	Achievement	Limitation	Ref.
Stacked denoising auto-encoders with manifold regularization	SDAE	WM-811k	SDAE improves wafer-map defect recognition by learning noise-resistant features, achieving 98% accuracy and better generalization than conventional methods.	Weighted cost and training issues such as high computational complexity and overfitting to the training set.	[63]
BlazePose architecture	LSTM and YOLO v4	Private dataset	Accuracy of 95.91% in action recognition and 97.68% in object recognition on 243 videos with 27 actions and 5 object-based activities.	Significant occlusion and appearance changes can strongly affect landmark-detection accuracy.	[64]
Hand-crafted features based on ORB and SIFT	K-NN, decision tree, and random forest (RF)	8000 samples from 100 object classes	Enhanced accuracy rate of 85.6%.	Hand-crafted features may perform worse than deep-learning approaches in more complex cases.	[65]
GAN	CNN	ModelNet40	Progressive conditional GAN-based augmentation improves object-recognition precision to 95.2% accuracy.	Generalization challenges across datasets and overfitting risks due to reliance on synthetic data.	[66]
GAN-LSTM	GAN-LSTM	CARRADA	Object-classification accuracy improved by 7.86% compared with other approaches.	Challenged by sparse point clouds and multipath issues in millimeter-wave radar.	[67]

2.2 Deep networks for supervised or discriminative learning

Deep Networks for supervised or discriminative learning are models that can learn from labelled data to predict or classify data based on the input features. These are multi-layer networks that use nonlinear functions to transform the data. Famous architectural designs involve Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). DNNs have a hierarchical structure that makes them suitable for complicated feature extraction. CNNs are excellent in spatial data processing and a perfect choice for tasks that include images. Meanwhile, RNNs are efficient with sequential data, such as natural language processing and time series. Their efficiency comes from their ability to recognize complex patterns in data [68]. Figure 7 shows the general structure of each of these architectural designs.

2.2.1 Convolutional neural network (CNN)

A Convolutional Neural Network, also known as a ConvNet or CNN, is inspired by the biological retina. It is a multi-layer deep learning technique used for object detection and recognition, neural language processing, image classification, medical image analysis, and speech recognition. They excel in visual tasks, much more efficiently than most traditional approaches. CNNs extract local features from higher layers of an image and combine them into more complex features at lower layers. CNNs are computationally costly and have a long training time on large datasets, which is why they are trained on GPUs. There are three key features of the ConvNet: Subsampling (pooling), Weight Sharing, and the Local Receptive Field [69]. The following figure shows an example of the ConvNet general architecture. Figure 8 shows an example of the CNN layers.

The CNN utilization for object recognition tasks has notable advancements through different datasets and applications. Recent research demonstrates the CNNs' efficiency in recognition precision enhancement in various environments. For example, the CNNs' efficiency is proved through benchmark datasets such as MSRC, Caltech 101, and Pascal VOC 2012, achieving recognition accuracies of 92.25%, 91.91%, and 93.50%, respectively [70]. Common CNN architectures, including ResNet, AlexNet, GoogleNet, and VGG16, are widely used for object recognition tasks. Also, MobileNet V2 obtained an accuracy ratio of 92.80% on a fruit dataset [71]. While Raj et al. [72] utilized the CNN for human activity recognition with an accuracy ratio of 97.20% on the WISDM dataset. Instead of emphasizing the recognition of objects, it focuses

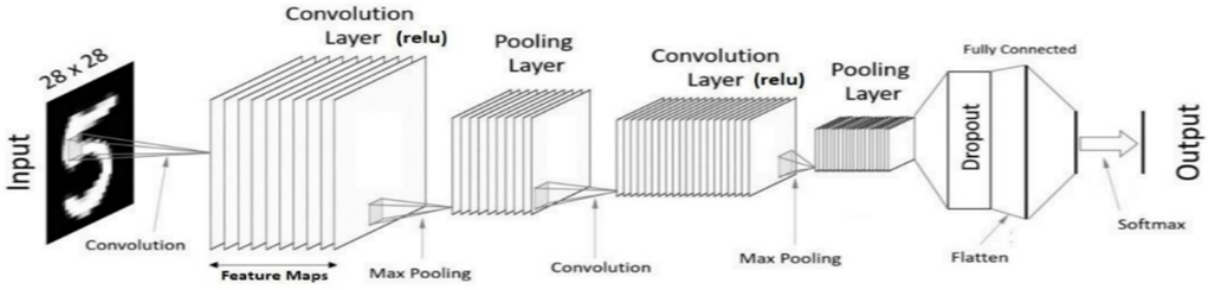

Fig. 7 The Inner structure of the several architectures of the supervised (discriminative) Network models [68]

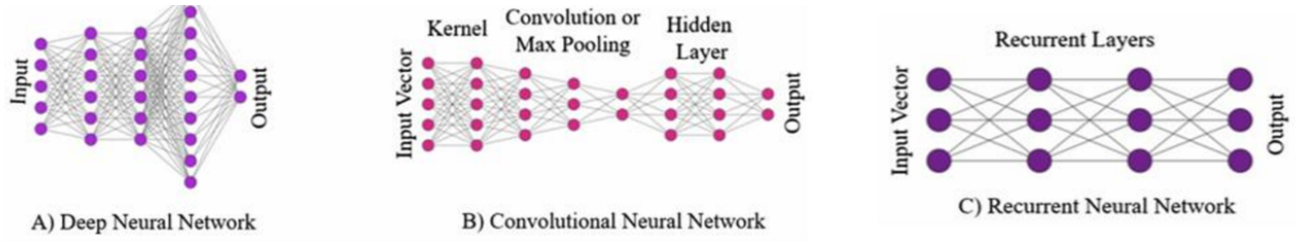

Fig. 8 An example of the general architecture of a ConvNet [70]

on the efficiency of the CNN in classification and recognition tasks. Jadhav et al. [73] used the pre-trained CNN (VGG16) with an accuracy of 91.25% for object detection and recognition. The accuracy result was 89.9% of handwritten Hindi characters, but it is still limited due to the computational resource requirements and the dataset complexity. While Bui et al. [74] employed the pre-trained CNN on Pascal VOC2007, achieving a satisfactory accuracy result of 76.7%, facing challenges related to memory constraints, and lessened precision from fewer convolutional layers. Wang et al. [75] suggested using AlexNet-RNN with a softmax classifier on the W-RGBD dataset, achieving 3% higher recognition accuracy in comparison to other methods, and low computational cost. However, during the transfer process, it faces data shortage and high computational complexity. Yaseen et al. [76] employed Cross-SplitNet with CNN as a classifier on the COCO 17, VOC 12, and VOC 07 datasets, achieving a precision (AP) of 0.819 at 22.9 frames per second (FPS) and a 1.9% increase in mean average precision (mAP) on the COCO dataset. It faces some issues in detecting densely stacked objects. Sarker et al. [77] suggested using the CNN on the Yale dataset and a self-generated video dataset, achieving a recognition accuracy of 97%, which is a notable enhancement over the conventional models, but still faces some challenges, like the absence of online training feedback techniques and the excessive computational costs, as summarized in Table 2.

Table 2 Supervised (discriminative) deep learning methods applied to object recognition systems

Method(s)	Classifier	Dataset	Achievement(s)	Limitation	Ref.
K-means clustering, region-based segmentation, SIFT, KAZE, and BRISK for feature extraction	1-D CNN	MSRC-v2, Caltech101, and Pascal VOC 2012 datasets	The method achieved high accuracy in segmentation and feature fusion, with strong precision, recall, and F1 scores, demonstrating superior object-recognition performance compared with other methods.	It encounters issues with object merging and occlusion in complex environments; segmentation still struggles with detailed backgrounds and has notable computational complexity.	[70]

Table 2 (continued).

Method(s)	Classifier	Dataset	Achievement(s)	Limitation	Ref.
CNN	MobileNet V2	Aggregated dataset for fruits	Achieved a high accuracy of 92.80% in object recognition.	—	[71]
CNN	CNN	WISDM	Achieved accuracy of 97.20% for human activity recognition.	Limited dataset size and relatively high computational complexity.	[72]
CNN (VGG16)	CNN	ImageNet	For object detection, accuracy is 91.25%; for object recognition of handwritten Hindi characters, accuracy is 89.9%.	High computational-resource requirements and dataset complexity.	[73]
CNN	CNN	ImageNet (training) and Pascal VOC2007 (testing)	Obtained a satisfactory object-recognition accuracy of 76.7%.	Faces memory-constraint issues and reduced precision when fewer convolutional layers are used.	[74]
AlexNet–RNN	Softmax classifier	W-RGBD dataset	About 3% higher recognition accuracy than other methods, with low computational cost.	During transfer, it faces data-shortage problems and high computational complexity.	[75]
Cross-SplitNet	CNN	COCO 17; VOC 12; VOC 07	Achieved precision (AP) of 0.819 at 22.9 frames per second, with a mean average precision (mAP) increase of 1.9% on the COCO dataset.	Because of missing features, it has difficulty detecting densely stacked objects.	[76]
CNN	CNN	Yale and a self-generated video dataset	Obtained recognition accuracy of 97%, a notable improvement over conventional models.	Lacks online training feedback and has high computational cost.	[77]

2.3 Hybrid deep networks

Generative models are flexible when they can be trained on both labelled and unlabelled data, while the discriminative models can't be trained on the unlabelled data but excel in supervised tasks. Hybrid networks have the ability to train both models simultaneously within a single structure, thereby gaining advantages from both. There are three types of Hybrid Deep Network models:

- **Hybrid Model 1:** a combination of various generative or discriminative models, to extract features that are stronger and more significant. For example, they can be CNN+LSTM, AE+GAN, and so on.
- **Hybrid Model 2:** a combination of generative models followed by discriminative models For example, they can be DBN+MLP, GAN+CNN, AE+CNN, and so on.
- **Hybrid Model 3:** a combination of various generative or discriminative models followed by a non-deep learning classifier. Examples could be AE+SVM, CNN+SVM, and so on [78].

Recently, there has been considerable interest in the topic of object recognition using Hybrid Deep Neural Networks. In the field of video motion recognition, Mihanpour et al. [79] introduced CoReHAR, which is a Hybrid Deep Network that integrates CNN with RNN, obtaining recognition accuracy of 95% on the UCF101 dataset. Nagarajan et al. [80] also introduced a hybrid model consisting of GAN and DCNN for object recognition, trained with HAAVO. The obtained accuracy measures were 0.940 testing accuracy, 0.946 precision, and 0.953 recall. Balasubaramanian et al. [81] used a Multi-channel CNN with a Sparse Auto-encoder (SAE) and the softmax classifier on the Minist handwritten digit library, the MIT face database, and the Oxford-17 flowers dataset, reaching a recognition accuracy of 0.985, but with a long training time. Shiri et al. [82] employed a combination of PCA with Depth-wise Separable CNN on a dataset of augmented images, improving recognition accuracy to achieve an accuracy of 94.16% and an F1 score of 96.009%. However, a large amount of data is required for training. Denton et al. [83] combined the CNN with the RNN and employed this combination on the Fruit-360 dataset, achieving a recognition accuracy of 89.18%, but it faces the high computational complexity challenge. as summarized in Table 3.

Table 3 Hybrid deep learning methods applied to object recognition systems

Method(s)	Classifier	Dataset	Achievement(s)	Limitation	Ref.
CoReHAR	CNN + RNN	UCF101	Improved performance in human action recognition across diverse scenarios, achieving 95% accuracy.	High computational requirements; sensitivity to camera motion, object occlusion, and background noise.	[79]
HAAVO	GAN + DCNN	Not mentioned	Achieved accuracy values of 0.940 (testing accuracy), 0.946 (precision), and 0.953 (recall).	Faces issues in distance estimation and multi-class object recognition for persons with visual impairments.	[80]
Multi-channel CNN + Sparse auto-encoder (SAE)	Softmax	MNIST, MIT, and Oxford-17 datasets	Highest recognition rate reached 0.985.	Long training time.	[81]
PCA + depth-wise separable CNN	PCA + depth-wise separable CNN	Augmented image dataset	Improved recognition accuracy, achieving 94.16% accuracy and 96.009% F1 score.	Requires large amounts of training data.	[82]
CNN + RNN	CNN + RNN	Fruit-360	Recognition accuracy of 89.18%.	High computational complexity.	[83]

3 POPULAR OBJECT RECOGNITION DATASETS FOR DEEP LEARNING

Researchers still acquire many data types for public libraries and their personal studies. Here are the most popular object recognition datasets used in Deep Learning research are illustrated.

3.1 The imagenet dataset

The ImageNet dataset was introduced by academics at Stanford and Princeton universities. With more than 14 million images classified into approximately 20,000 categories. It is one of the largest human-annotated image datasets ever designed. It was established to help in the research and development of visual object recognition. The growth of online search engines and digital image sharing led to its construction. This dataset utilizes basic text-based queries and is generated from WordNet6, a sizable collection of English phrases to search for images associated with WordNet proposals via keywords. For human annotation, it uses the Amazon Mechanical Turk (AMT) crowdworker platform. In 2010, the ImageNet Large Scale Visual Recognition Challenge was founded with a concentration on object recognition and localization in images. A neural network-driven machine learning model let Alex Krizhevsky and his University of Toronto team win the competition in 2012. Neural networks turned into the most common machine learning approach again after ImageNet's 2012 win revolutionized the area of deep learning by employing neural networks in the competition [83]. The ImageNet dataset faces challenges such as the computational costs required for training large models, overfitting due to its large size, and variability in object recognition tasks [84].

3.2 Coco dataset (common objects in context)

Microsoft presented the MS COCO dataset in 2015 [85]. It is a big image dataset that comprises over 330,000 images and 2.5 million labelled objects across 80 classes. It is crucial for computer vision tasks such as recognition, captioning, and segmentation, offering comprehensive annotations for model training and performance evaluation of object detection models [86]. The COCO dataset encounters some issues with incomplete object labelling, trouble identifying tiny objects, insufficient annotations, and unbalanced classes [87].

3.3 Pascal voc (visual object classes) dataset

The PASCAL VOC dataset is an important evaluation resource in computer vision, especially for object detection and classification tasks. According to the last update on this dataset, which occurred in 2012. It includes 11,530 images with 27,450 annotated objects across 20 classes, which contain standard daily life objects such as people, animals, and vehicles. Bounding boxes and class labels are appended to each image, offering ground truth data for model evaluation and training [88]. It faces some challenges, including accurately annotating image classes and efficiently analysing algorithm

performance [89]. The PASCAL VOC dataset has many versions (Table 4), with the 2012 version becoming a popular release.

Table 4 PASCAL VOC dataset statistics (2005–2012)

Dataset statistics	VOC 2005	VOC 2006	VOC 2007	VOC 2008	VOC 2009	VOC 2010	VOC 2011	VOC 2012
Number of images	1578	5304	9963	4340	7054	10103	11530	11530
Number of annotated images	1578	5304	9963	4340	7054	10103	11530	11530
Object categories	4	10	20	20	20	20	20	20
Object annotation statistics	2209 ann. obj.	4754 ann. obj.	24640 ann. obj.	10363 ann. obj.	17218 ann. obj. 3211 segm.	23374 ann. obj. 4203 segm.	27450 ann. obj. 5034 segm.	27450 ann. obj. 6929 segm.

3.4 Open image dataset

The Open Image dataset is an appropriate option for large-scale object detection, image classification, unfiled annotation, and visual relationship detection in deep learning. Kuznetsova et al. [90] presented the Open Image dataset, which solves issues such as unbalanced data distribution and multi-labels to enhance the execution. It consists of 9 million images, with annotations for object detection included in approximately 600 classes of objects. The images come with a Creative Commons Attribution license, which allows sharing and modification, and is free from design constraints. They were collected from Flickr without using specified class names or tags. It faces challenges with highly unbalanced label distribution, label-related problems, and the complexity of variant object detection environments [90, 91].

3.5 The visual genome dataset

The Visual Genome dataset, a combination of objects, images, relationships, and attributes, connects computer vision and natural language processing, aiding in machine learning model training and visual concept analysis. It includes more than 108,000 images, associated with 5.4 million area descriptions, and 2.8 million objects connected by 1.5 million relationships [92]. To build a human-like system, graphs and scenes construction, image captioning, and visual question answering [93]. The comprehensive annotations in the dataset allow for the recognition of patterns and connections that are usually overlooked in conventional datasets, which improves the usability of AI models [94]. This dataset encounters some challenges, including ensuring accuracy and homogeneity of annotations across images [95].

Many other datasets including BDD, DOTA, KITTI, Caltech, Davis, SUN RGB-D, and so on.

4 FRAMEWORKS OF DEEP LEARNING IMPLEMENTATION

This section presents the common frameworks for modularized deep learning approaches implementation, concentrating on distribution, optimization, and infrastructure support. The most common are as follows:

- NVIDIA cuDNN
- Deeplearning-4j
- Tensor-flow
- Torch/PyTorch
- Theano
- Cognitive network toolkit (CNTK)
- Keras
- Caffe
- MX-net
- DIGITS

5 CONCLUSION

This paper provides an overview of deep learning concepts and reviews the latest papers on Deep Learning in the field of object recognition, exploring various popular architectures that are used in some selected object recognition applications. In particular, three classes of deep learning architectures, namely the Generative (unsupervised), Discriminative (supervised), and Hybrid deep architecture, are explained with their approaches in detail. These three categories show a comprehensive, flexible, and reliable performance in solving many issues. For instance, the unsupervised architecture categories are the Auto-encoder, Boltzmann Machines, Recurrent Neural Networks, and Sum-Product Networks. They explained in this paper, with the relevant literature for each of them, which are used for object recognition. Then we've emphasized the most common object recognition datasets used in deep learning, and after that, we've pointed out the most widely used deep learning frameworks for implementation. With the association of the comparative results of the related works. When supervised learning algorithms handle labelled data, and unsupervised learning techniques handle unlabelled data, they are utilized in such cases since it is difficult to acquire labelled data when handling large amounts of data, and they fail to perform satisfactorily in these situations. Datasets for object recognition are crucial because they provide the varied labelled examples needed for training and testing models effectively. Deep learning methods are used for feature extraction or reducing the complexity of features. According to past studies, CNN proved a better performance in categorization than AE and BM, and therefore gets used more commonly. Amazon's product suggestions and Facebook's automatic tagging algorithms demonstrate how CNN handles images faster and more efficiently. Finally, the discussed techniques obtained a high accuracy level automatically. Future work suggests combining CNN with some other methods, like AE, using feature extraction and selection as a hybrid approach for enhancing object recognition accuracy.

ACKNOWLEDGEMENT

N/A

FUNDING SOURCE

No funds received.

DATA AVAILABILITY

N/A

DECLARATIONS

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to publish

All authors consent to the publication of this work.

Ethical approval

N/A

REFERENCES

[1] Reijers W. Digital hermeneutics: philosophical investigations in new media and technologies. AI & SOCIETY. 2020;38(6):2351–2354. 10.1007/s00146-020-01006-x

[2] Bansal M, Kumar M, Kumar M. 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. Multimedia Tools and Applications. 2021;80(12):18839–18857. 10.1007/s11042-021-10646-0

[3] Bansal M, Kumar M, Kumar M. 2D Object Recognition Techniques: State-of-the-Art Work. Archives of Computational Methods in Engineering. 2020;28(3):1147–1161. 10.1007/s11831-020-09409-1

[4] Qi S, Ning X, Yang G, Zhang L, Long P, Cai W, et al. Review of multi-view 3D object recognition methods based on

deep learning. Displays. 2021;69:102053. 10.1016/j.displa.2021.102053

[5] Kaur R, Singh S. A comprehensive review of object detection with deep learning. Digital Signal Processing. 2023;132:103812. 10.1016/j.dsp.2022.103812

[6] Liu Y, Sun P, Wergeles N, Shang Y. A survey and performance evaluation of deep learning methods for small object detection. Expert Systems with Applications. 2021;172:114602. 10.1016/j.eswa.2021.114602

[7] Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B. A survey of modern deep learning based object detection models. Digital Signal Processing. 2022;126:103514. 10.1016/j.dsp.2022.103514

[8] Chavan C, Hembade S, Jadhav G, Komalwad P, Rawat P. Computer Vision Application Analysis based on Object Detection. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 2023;07(04). 10.55041/ijsrem19015

[9] Wang W, Lai Q, Fu H, Shen J, Ling H, Yang R. Salient object detection in the deep learning era: An in-depth survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022;44(6):3239-59. 10.1109/TPAMI.2021.3051099

[10] Park HY, Hyun CH. Object recognition method and the device thereof. Korean patent KR101961462B1; 2019.

[11] Liu T, Liu X, Tang X, Zhang Z. Automatic object recognition method and system thereof, shopping device and storage medium. US patent 10,872,227 (published earlier as US 2019/0303650 A1). United States Patent and Trademark Office; 2020. Assignee: BOE Technology Group Co., Ltd.

[12] Hashimoto D, Takeyasu S, Hirano K. Object recognition apparatus. US patent application US20190370978A1; 2019.

[13] Railkar Y, Nasikkar A, Pawar S, Patil P, Pise R. Object Detection and Recognition System Using Deep Learning Method. In: 2023 IEEE 8th International Conference for Convergence in Technology (I2CT). IEEE; 2023. p. 1–6. 10.1109/i2ct57861.2023.10126316

[14] Ruiz Sarmiento JR, Monroy J, Moreno FA, González-Jiménez J. Tutorial para el reconocimiento de objetos basado en características empleando herramientas Python. In: Actas de las XXXIX Jornadas de Automática, Badajoz, 5-7 de septiembre de 2018. Universidade da Coruña. Servizo de Publicacións; 2020. p. 998–1005. 10.17979/spudc.9788497497565.0998

[15] UCUZAL H, BALIKCI CICEK AGI, ARSLAN AGAK, COLAK C. A Web-Based Application for Identifying Objects In Images: Object Recognition Software. In: 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE; 2019. p. 1–5. 10.1109/ismsit.2019.8932735

[16] Rossi U. Review. Brumal Revista de investigación sobre lo Fantástico. 2023;11(1):347–351. 10.5565/rev/brumal.1002

[17] Zhan T. DL 101: Basic introduction to deep learning with its application in biomedical related fields. Statistics in Medicine. 2022;41(26):5365–5378. 10.1002/sim.9564

[18] Abdul lateef AA, Al-Janabi STF, Al-Khateeb B. Survey on intrusion detection systems based on deep learning. Periodicals of Engineering and Natural Sciences (PEN). 2019;7(3):1074. 10.21533/pen.v7i3.635

[19] Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing. 2014;3(1). 10.1017/atsip.2013.9

[20] Selvaganapathy S, Nivaashini M, Natarajan H. Deep belief network based detection and categorization of malicious URLs. Information Security Journal: A

Global Perspective. 2018;27(3):145–161. 10.1080/19393555.2018.1456577

[21] Kenig M, Lahini Y. Unsupervised generalization of correlated quantum dynamics on disordered lattices. Physical Review A. 2023;107(1). 10.1103/physreva.107.012430

[22] Chen Y, Wang Z, Wang L, Wang J, Li P, Cao D, et al. Deep generative model for drug design from protein target sequence. Journal of Cheminformatics. 2023;15(1). 10.1186/s13321-023-00702-2

[23] Berahmand K, Daneshfar F, Salehi ES, Li Y, Xu Y. Autoencoders and their applications in machine learning: a survey. Artificial Intelligence Review. 2024;57(2). 10.1007/s10462-023-10662-6

[24] Bank D, Koenigstein N, Giryes R. In: Autoencoders. Springer International Publishing; 2023. p. 353–374. 10.1007/978-3-031-24628-9_16

[25] Berahmand K, Daneshfar F, Salehi ES, Li Y, Xu Y. Autoencoders and their applications in machine learning: a survey. Artificial Intelligence Review. 2024;57(2). 10.1007/s10462-023-10662-6

[26] Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. IEEE transactions on emerging topics in computational intelligence. 2018;2(1):41-50

[27] Zhang Y, Zhang E, Chen W. Deep neural network for halftone image classification based on sparse auto-encoder. Engineering Applications of Artificial Intelligence. 2016;50:245–255. 10.1016/j.engappai.2016.01.032

[28] Liu Y, Ponce C, Brunton SL, Kutz JN. Multiresolution convolutional autoencoders. Journal of Computational Physics. 2023;474:111801. 10.1016/j.jcp.2022.111801

[29] WANG J, ZHANG H, MIAO Q. An attention graph stacked autoencoder for anomaly detection of electro-mechanical actuator using spatio-temporal multivariate signals. Chinese Journal of Aeronautics. 2024;37(9):506–520. 10.1016/j.cja.2024.03.024

[30] Hoang DT, Kang HJ. A survey on Deep Learning based bearing fault diagnosis. Neurocomputing. 2019;335:327–335. 10.1016/j.neucom.2018.06.078

[31] Hinton GE, Osindero S, Teh YW. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation. 2006;18(7):1527–1554. 10.1162/neco.2006.18.7.1527

[32] Liu M, Sun X, Zhang F, Yu Y, Wang Y. Context-LGM: leveraging object-context relation for context-aware object recognition. arXiv preprint arXiv:211004042. 2021

[33] Nitsch J, Nieto J, Siegwart R, Schmidt M, Cadena C. Learning Common and Transferable Feature Representations for Multi-Modal Data. In: 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE; 2020. p. 1601–1607. 10.1109/iv47402.2020.9304669

[34] Xiong G, Ma W, Zhao N, Zhang J, Jiang Z, Mao Z. Multi-Type Diesel Engines Operating Condition Recognition Method Based on Stacked Auto-Encoder and Feature Transfer Learning. IEEE Access. 2021;9:31043–31052. 10.1109/access.2021.3057399

[35] Ventura E, Cocco S, Monasson R, Zamponi F. Unlearning regularization for Boltzmann machines. Machine Learning: Science and Technology. 2024;5(2):025078. 10.1088/2632-2153/ad5a5f

[36] Zhang X, Chen J. Deep learning based intelligent intrusion detection. In: 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN). IEEE; 2017. p. 1133–1137. 10.1109/iccsn.2017.8230287

[37] Harshvardhan G, Gourisaria MK, Rautaray SS, Pandey M. UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system. Journal of King Saud University - Computer and Information Sciences. 2022;34(8):6400–6413. 10.1016/j.jksuci.2021.01.017

[38] Ventura E, Cocco S, Monasson R, Zamponi F. Unlearning regularization for Boltzmann machines. Machine Learning: Science and Technology. 2024;5(2):025078. 10.1088/2632-2153/ad5a5f

[39] Senthilnath J, Nagaraj G, Sumanth Simha C, Kulkarni S, Thapa M, Indiramma M, et al. DRBM-ClustNet: A Deep Restricted Boltzmann–Kohonen Architecture for Data Clustering. IEEE Transactions on Neural Networks and Learning Systems. 2024;35(2):2560–2574. 10.1109/tnnls.2022.3190439

[40] Wang X, Chu J, Yu H, Gong Z, Li T. Self-supervised Gaussian Restricted Boltzmann Machine via joint contrastive representation and contrastive divergence. Knowledge-Based Systems. 2024;299:112121. 10.1016/j.knosys.2024.112121

[41] Béreux N, Decelle A, Furtlehner C, Seoane B. Learning a restricted Boltzmann machine using biased Monte Carlo sampling. SciPost Physics. 2023;14(3). 10.21468/scipostphys.14.3.032

[42] Vera M, Rey Vega L, Piantanida P. Information flow in Deep Restricted Boltzmann Machines: An analysis of mutual information between inputs and outputs. Neurocomputing. 2022;507:235–246. 10.1016/j.neucom.2022.08.014

[43] Wang Q, Gao X, Wan K, Hu Z. Generative and discriminative infinite restricted Boltzmann machine training. International Journal of Intelligent Systems. 2022;37(10):7857–7887. 10.1002/int.22908

[44] Bozcan I, Oymak Y, Alemdar IZ, Kalkan S. What is (Missing or Wrong) in the Scene? A Hybrid Deep Boltzmann Machine for Contextualized Scene Modeling. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2018. p. 1–6. 10.1109/icra.2018.8460828

[45] Zorzi M, Testolin A, Stoianov I. Modeling language and cognition with deep unsupervised learning: A tutorial overview. Frontiers in Psychology. 2013;04. 10.3389/fpsyg.2013.00515

[46] Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R. Regularization of neural networks using dropconnect. In: International conference on machine learning. PMLR; 2013. p. 1058-66

[47] Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09. ACM; 2009. p. 609–616. 10.1145/1553374.1553453

[48] Takahashi N, Yamakawa T, Minetoma Y, Nishi T, Migita T. Design of continuous-time recurrent neural networks with piecewise-linear activation function for generation of prescribed sequences of bipolar vectors. Neural Networks. 2023;164:588–605. 10.1016/j.neunet.2023.05.013

[49] Gajamannage K, Jayathilake D, Park Y, Bollt E. Recurrent neural networks for dynamical systems: Applications to ordinary differential equations, collective motion, and hydrological modeling. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2023;33(1)

[50] Apaydin H, Feizi H, Sattari MT, Colak MS, Shamshirband S, Chau KW. Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. Water. 2020;12(5):1500. 10.3390/w12051500

[51] Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. arXiv preprint arXiv:230517473. 2023

[52] Kang Q, Yu D, Cheong KH, Wang Z. Deterministic convergence analysis for regularized long short-term memory and its application to regression and multi-classification problems. Engineering Applications of Artificial Intelligence. 2024;133:108444. 10.1016/j.engappai.2024.108444

[53] Minaee S, Azimi E, Abdolrashidi A. Deep-sentiment: Sentiment analysis using ensemble

of cnn and bi-lstm models. arXiv preprint arXiv:190404206. 2019

[54] Liu S, Kong Z, Huang T, Du Y, Xiang W. An ADMM-LSTM framework for short-term load forecasting. Neural Networks. 2024;173:106150. 10.1016/j.neunet.2024.106150

[55] Fang W, Chen Y, Xue Q. Survey on research of RNN-based spatio-temporal sequence prediction algorithms. Journal on Big Data. 2021;3(3):97

[56] Yue J, Miao Z, He Y, Du N. Loss Architecture Search for Few-Shot Object Recognition. Complexity. 2020;2020(1):1041962

[57] Degtyarenko I, Deriuga I, Grygoriev A, Polotskyi S, Melnyk V, Zakharchuk D, et al. Hierarchical Recurrent Neural Network for Handwritten Strokes Classification. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2021. p. 2865–2869. 10.1109/icassp39728.2021.9413412

[58] Bottcher W, Machado P, Lama N, McGinnity TM. Object recognition for robotics from tactile time series data utilising different neural network architectures. In: 2021 International Joint Conference on Neural Networks (IJCNN). IEEE; 2021. p. 1–8. 10.1109/ijcnn52387.2021.9533388

[59] Xia R, Zhang Y, Liu X, Yang B. A survey of sum–product networks structural learning. Neural Networks. 2023;164:645–666. 10.1016/j.neunet.2023.05.010

[60] Zhang S, Cheng Q, Chen D, Zhang H. Image target recognition model of multi-channel structure convolutional neural network training automatic encoder. IEEE Access. 2020;8:113090-103

[61] Rahal N, Tounsi M, Hussain A, Alimi AM. Deep Sparse Auto-Encoder Features Learning for Arabic Text Recognition. IEEE Access. 2021;9:18569–18584. 10.1109/access.2021.3053618

[62] Kodepogu K, Manjeti V, Divya M, Anusha M, Jaswanth M, Kumar K. Traffic Sign Detection and Recognition by Using Auto Encoder Deep Learning Classification Models. Traffic. 2023;13(4)

[63] Yu J. Enhanced Stacked Denoising Autoencoder-Based Feature Learning for Recognition of Wafer Map Defects. IEEE Transactions on Semiconductor Manufacturing. 2019;32(4):613–624. 10.1109/tsm.2019.2940334

[64] Huang YP, Kshetrimayum S, Chiang CT. Object-Based Hybrid Deep Learning Technique for Recognition of Sequential Actions. IEEE Access. 2023;11:67385–67399. 10.1109/access.2023.3291395

[65] Muzahid AAM, Wanggen W, Sohel F, Bennamoun M, Hou L, Ullah H. Progressive conditional GAN-based augmentation for 3D object recognition. Neurocomputing. 2021;460:20–30. 10.1016/j.neucom.2021.06.091

[66] lu g, he z, zhong y, han y. Enhanced radar for object recognition based on GANs. In: Lei T, editor. 5th International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2023). SPIE; 2023. p. 160. 10.1117/12.2689832

[67] M Abdulkareem I, K AL-Shammri F, A Khalid NA, A Omran N. Proposed Approach for Object Detection and Recognition by Deep Learning Models Using Data Augmentation. International Journal of Online and Biomedical Engineering (iJOE). 2024;20(05):31–43. 10.3991/ijoe.v20i05.47171

[68] Wani M, Bhat F, Afzal S, Khan A. Advances in Deep Learning. Singapore. Springer; 2019. 10.1007/978-981-13-6794-6

[69] Shyam R. Convolutional neural network and its architectures. Journal of Computer Technology & Applications. 2021;12(2):6-14

[70] Naseer A, Mudawi NA, Abdelhaq M, Alonazi M, Alazeb A, Algarni A, et al. CNN-Based Object Detection via Segmentation Capabilities in Outdoor Natural Scenes. IEEE

Access. 2024;12:84984–85000. 10.1109/access.2024.3413848

[71] Surve Y, Pudari K, Bedade S, Masanam BD, Bhalerao K, Mhatre P. Comparative Analysis of Various CNN Architectures in Recognizing Objects in a Classification System. In: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). IEEE; 2024. p. 1–5. 10.1109/i2ct61223.2024.10544049

[72] Raj R, Kos A. An improved human activity recognition technique based on convolutional neural network. Scientific Reports. 2023;13(1). 10.1038/s41598-023-49739-1

[73] Jadhav J, Attar M, Patil S, Beg S. Object recognition using CNN. International Journal of Advance Research, Ideas and Innovations in Technology. 2018;04(02):1987–1991.

[74] Bui HM, Lech M, Cheng E, Neville K, Burnett IS. Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure. IEEE Access. 2016;4:10059–10066. 10.1109/access.2016.2639543

[75] Wang SY, Qu Z, Li CJ. A Dense-Aware Cross-splitNet for Object Detection and Recognition. IEEE Transactions on Circuits and Systems for Video Technology. 2023;33(5):2290–2301. 10.1109/tcsvt.2022.3221658

[76] Usman Yaseen M, Anjum A, Fortino G, Liotta A, Hussain A. Cloud based scalable object recognition from video streams using orientation fusion and convolutional neural networks. Pattern Recognition. 2022;121:108207. 10.1016/j.patcog.2021.108207

[77] Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN computer science. 2021. 10.20944/preprints202108.0060.v1

[78] Taspinar YS, Selek M. Object Recognition with Hybrid Deep Learning Methods and Testing on Embedded Systems. International Journal of Intelligent Systems and Applications in Engineering. 2020;8(2):71–77. 10.18201/ijisae.2020261587

[79] Mihanpour A, Rashti MJ, Alavi SE. CoReHAR: A Hybrid Deep Network for Video Action Recognition. International Journal of Web Research. 2020;3(1):1-10

[80] Nagarajan A, M P G. Hybrid Optimization-Enabled Deep Learning for Indoor Object Detection and Distance Estimation to Assist Visually Impaired Persons. Advances in Engineering Software. 2023;176:103362. 10.1016/j.advengsoft.2022.103362

[81] Balasubaramanian S, Cyriac R, Roshan S, Maruthamuthu Paramasivam K, Chellanthara Jose B. An effective stacked autoencoder based depth separable convolutional neural network model for face mask detection. Array. 2023;19:100294. 10.1016/j.array.2023.100294

[82] Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. arXiv preprint arXiv:230517473. 2023

[83] Denton E, Hanna A, Amironesei R, Smart A, Nicole H. On the genealogy of machine learning datasets: A critical history of ImageNet. Big Data & Society. 2021;8(2). 10.1177/20539517211035955

[84] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6):84–90. 10.1145/3065386

[85] Zimmermann E, Szeto J, Pasquero J, Ratle F. Benchmarking a Benchmark: How Reliable is MS-COCO?; 2023. Available from: https://arxiv.org/abs/2311.02709

[86] Mao X, Chen Y, Zhu Y, Chen D, Su H, Zhang R, et al. COCO-O: A Benchmark for Object Detectors under Natural Distribution Shifts. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2023. p. 6339-50

[87] Everingham M, Van Gool L, Williams C, Winn J, Zisserman A. The PASCAL Visual Object Classes (VOC) challenge. International Journal

of Computer Vision; 2010. 10.1007/s11263-009-0275-4

[88] Paniego S, Sharma V, Cañas JM. Open Source Assessment of Deep Learning Visual Object Detection. Sensors. 2022;22(12):4575. 10.3390/s22124575

[89] Everingham M, Eslami SMA, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision. 2014;111(1):98–136. 10.1007/s11263-014-0733-5

[90] Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. International journal of computer vision. 2020;128(7):1956-81

[91] Pan C, Peng J, Bu X, Zhang Z. Large-scale object detection in the wild with imbalanced data distribution, and multi-labels. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024;46(12):9255-71

[92] Perkel JM. A graphics toolkit for visualizing genome data. Nature. 2022;608(7923):636–637. 10.1038/d41586-022-02191-z

[93] Xu W, Zhong Q, Lin D, Li G, Cao G. CoolBox: A flexible toolkit for visual analysis of genomics data. BMC bioinformatics. 2021. 10.1101/2021.04.15.439923

[94] Goldman M, Craft B, Zhu J, Haussler D. Abstract 5039: Visualization and analysis of cancer genomics data using UCSC Xena. Cancer Research. 2022;82. 10.1158/1538-7445.AM2022-5039

[95] Pearce TM, Nikiforova MN, Roy S. Interactive Browser-Based Genomics Data Visualization Tools for Translational and Clinical Laboratory Applications. The Journal of Molecular Diagnostics. 2019;21(6):985-93. 10.1016/j.jmoldx.2019.06.005

How to cite this article

Mohammed SF, Shaker K. Accuracy and efficiency trade-offs in deep learning approaches for object recognition: A comparative study. Journal of University of Anbar for Pure Science. 2025; 19(2):151-170. doi:10.37652/juaps.2025.155224.1336