International Journal of Engineering Technology & Advanced Learning
ISSN: 3107-4227, Vol. 01 (01), March 2025, pp. 73-81

Advancing Object Detection with Deep Learning: Applications in Image Processing and Computer Vision

Irfan Khan[1], Virendra Khatarkar[1], Shital Gupta[1], Sana Khan[1], Anshu Tiwari[1]

[1]*Department of Computer Engineering, Bansal Institute of Science and Technology, Bhopal, Madhya Pradesh, India.*

Abstract

Image processing and computer vision are rapidly growing fields that enable machines to understand and interpret visual data, mimicking the capabilities of human vision. With the aid of superior learning algorithms and the presence of big data, deep learning has substantially increased the predicting power of computer systems. The integration of machine learning with complex applications such as object identification, image recognition, self-driving cars, and drug discovery has become both feasible and practicable. Researchers from all fields of science have been interested in deep learning techniques as a way to use their skills to address challenges because of their better and more dependable performance. A fascinating element of this technology that will also be covered is the reuse of information in deep learning.

Keywords: Machine learning, Deep learning, Image processing, Computer Vision.

1. Introduction

Deep neural networks have been studied by scientists since 1979, but they became the emerging machine learning research area in the 2000's when, in an unsupervised way, they decreased the dimension of data [1]. It was noticed by researchers in 2012 when it was announced as the winner of the ImageNet competition. CNN (Convolutional neural networks) is used to extract visual information from massive datasets and then classify the images [2]. Since then, in deep learning, diverse research has been conducted worldwide to solve complex issues.

Deep learning is a subset of machine learning techniques that aim to automatically learn the essential aspects of a dataset. Deep neural networks, unlike ordinary neural networks, usually have more than one hidden layer. DNN's hierarchical design allows you to learn features at multiple levels of abstraction. The earliest stories teach basic features, which are subsequently accumulated in the deepest layers to build up the highest-level concepts. It is also used in a wide range of data types, including text, images, and audio [3].

The recent research in machine learning, enhanced parallel processing capabilities of hardware and graphic processing units, and the significant size availability of training data are the key reasons for the popularity of deep learning. This means deep-learning algorithms may use complex, nonlinear composition methods to autonomously learn hierarchically and distributed features efficiently using labeled and unlabeled data.

2. Related Work

Deep learning involves a wide range of learning through data, just like machine learning. DNN (deep neural network) hierarchical architecture may be applied in a range of ways to address a variety of problems. Deep learning may often be divided into three types.

2.1. Deep Supervised Learning

Regression and pattern classifications are made possible by these networks' discriminative ability. The network tries to distinguish the data objects, including multiple classes, from the labels that are supplied with the data. The network can connect input to desired output in both classification and regression tasks. The three most well-known supervised learning architectures are "deep neural networks, convolutional neural networks, and recurrent neural networks." DNN's layers are organized into a hierarchy by the neurons that make up each layer. The output of one layer feeds into the inputs of the next, and so on. Every next layer discovers more intricate patterns in the supplied data [4]. Deeper layers often acquire the highest-level abstractions within data, whereas lower layers typically learn low-level characteristics.

E-mail ID: i.khan@bistbpl.in

DNNs are the feed forward neural networks to the most layers, as seen in Figure 1, making them simple in terms of structure. CNN is a renowned supervised learning architecture that was initially developed for the analysis of visual data, such as photos and videos. However, they have demonstrated to be quite helpful for practically any sort of data, such as visual [7], audio [8], or even textual [9]. “Convolutional, pooling and fully connected layers are the three main types of layers found in CNNs. With the use of filters or kernels whose coefficients are changed during the training phase, convolution layers seek to learn important characteristics that may occur in the data. To create a feature map where the position of features is highlighted with a greater activation value, each filter is individually combined over input” [5]. Like DNNs, CNN's lower layers extract basic characteristics, and kernels learn increasingly sophisticated features as the network gets deeper. The pooling layers decrease the feature maps' dimensionality and also start introducing some new features.

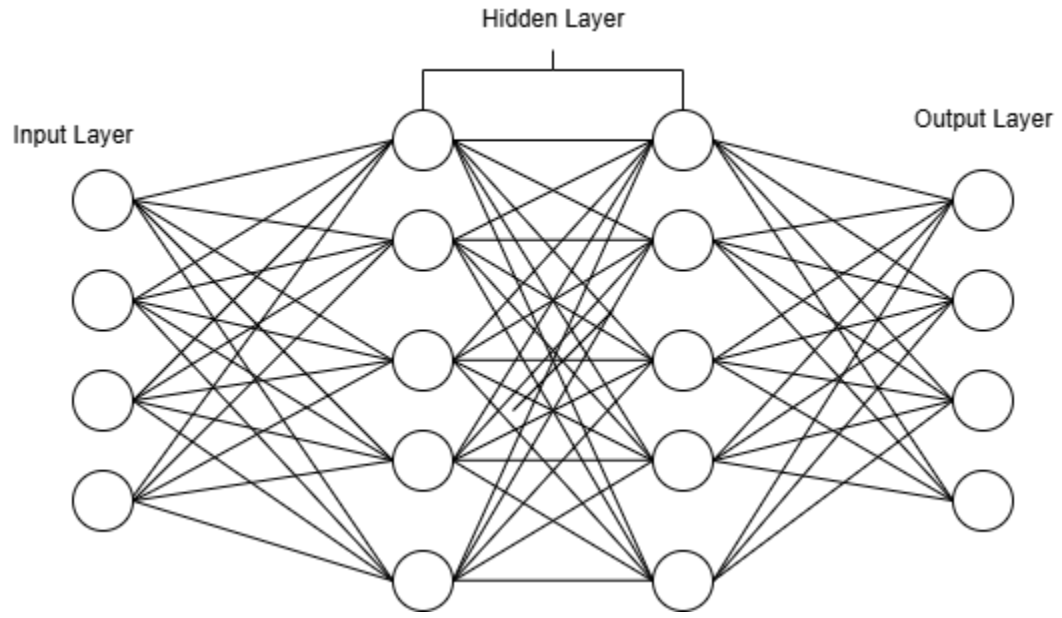

Figure 1. Deep Neural Network

Network's level of translation invariance. The method for extracting network features uses the convolutional and pooling layer, which finds local features in the input. The global characteristics are then obtained by combining the local features at the fully connected layers [6].

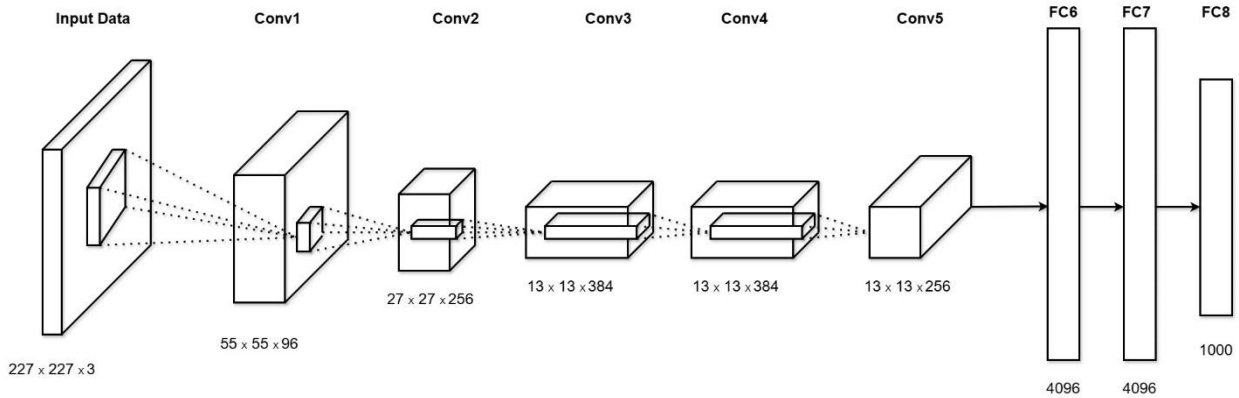

Figure 2. Alexnet CNN

CNNs were first introduced in 1979, but they only became popular in 2012 after CNNs decisively defeated all other networks in the renowned ImageNet competition. According to Figure 2, the AlexNet network has five convolutional layers, three layers are pooling layers, and three fully connected layers.

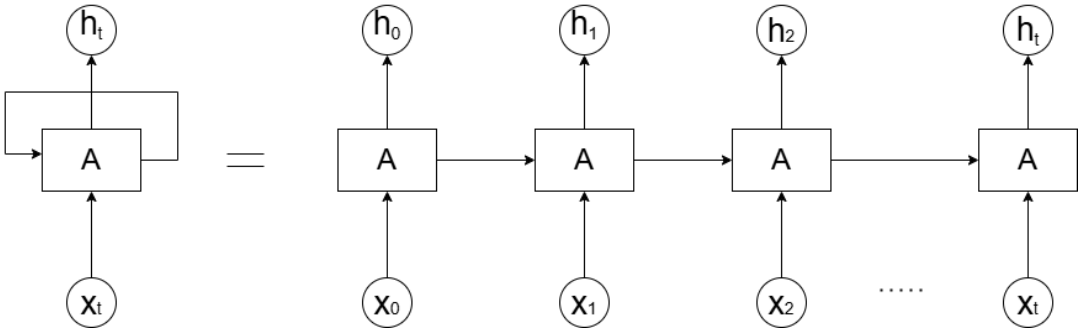

Figure 3. Recurrent neural network reconnect unit

The powerful architectures DNN and CNN are equally effective at evaluating non-sequential data, however, they are ineffective at detecting the patterns in the dataset. RNN is a new family of architectures that was created for this reason [10]. The RNN's each unit have recurrent connections, allowing the network to hold information for a long period of time. In order for the network to retain knowledge over a longer period of time, every unit in the RNN comprises recurrent connections. RNNs can now detect patterns in sequential data, including sounds, movies, and text. The long short-term memory (LSTM) network, a more contemporary and sophisticated form of RNN, enhances the pattern recognition capabilities of RNNs [11]. Figure 3 depicts the architectures of RNN and LSTM units.

2.2 Deep Unsupervised Learning

Unsupervised learning refers to learning strategies where target class labels and other task-specific supervision information are not provided during the learning phase. Deep Boltzmann machines (DBM) and deep auto encoders (DAE) are two of the most popular techniques for unsupervised learning [12, 13].

Deep bottleneck networks and auto encoders both have two parts. The data set is first compressed to relatively short-length representations. From such a short representation, the second half then is utilized to reproduce that original input. The auto encoder tries to create a precise representation throughout training in order to make it easier to retrieve the original data with minimum loss. In this approach, it unsupervised picks up important properties from the training data. This auto encoder's compact representation is frequently utilized as an input vector of the high-dimensional input that may be used for a variety of tasks including clustering, indexing, and searching as well as dimension reduction or feature embedding [4]. These encoder and decoder parts of a typical auto encoder are depicted in Figure 4.

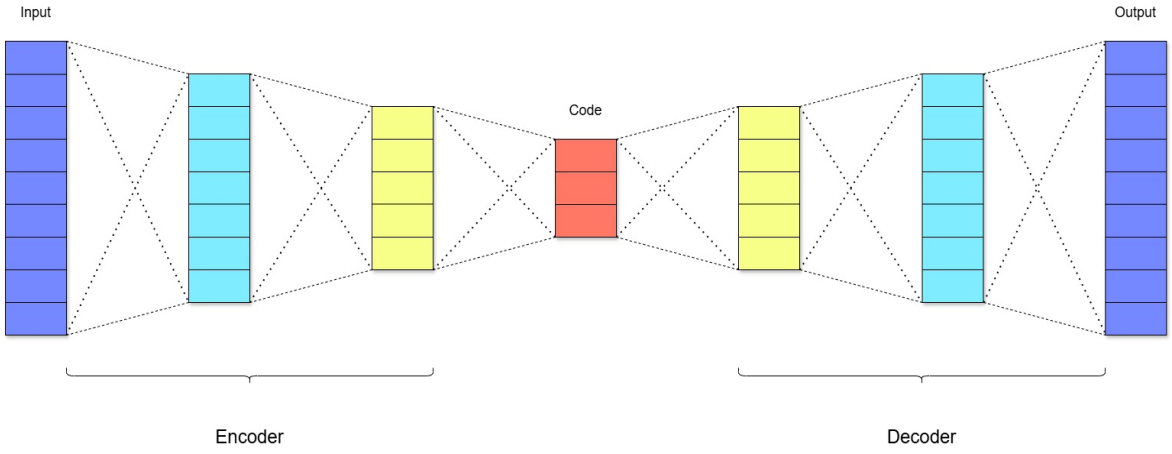

Figure 4. Architecture of auto encoder presenting the encoder and decoder

2.3 Hybrid Technique

Approach In the hybrid technique, the aim could be discrimination, which frequently receives considerable assistance from the results of generative or unsupervised deep networks. It could be done through better regularizing and optimizing deep networks for supervised learning. If you want to study the basic parameters for a subsequent supervised learning job, for instance, you may utilize a lot of unlabeled data in such an unsupervised learning method.

The strategy may also be applied when values in any deep generative or unsupervised deep network are estimated using unsupervised learning and discriminative criteria for supervised learning [4].

3. Applications to Image and Vision

One of the very first areas to benefit from deep learning breakthroughs was computer vision. With its help, computers are now capable of performing image recognition, object detection, and image segmentation. In 1989, the first notable achievement was released when a convolutional neural network was utilized to recognize handwritten numerals in postal mail [14]. Despite the system's high reliability and accuracy, no significant achievement was made until 2012. In the ImageNet competition, AlexNet succeeded over the other team by a wide majority. Researchers from all across the world have since been interested in it, and the years that followed have seen significant advancement.

Recent results in the ImageNet competition showed that CNN with residual connections achieved the human level of accuracy for image classifications & that it even beat this when many networks were included in the ensemble. [15]. It's only been able to employ computer vision in essential applications like self-driving vehicles and illness detection because of efforts over the previous five to six years, which have greatly improved the technology. Deep learning-enabled computer vision has attained specialist accuracy as in the identification of skin cancer [16], chest X-ray diagnosis [18], and disease detection utilizing multiple scans [17]

3.1 Process of Object Detection

In digital images and videos, the technique for object detection involves finding occurrences of semantic items for a particular class (like persons, cars, or birds). A typical method for object detection framework entails the generation for a significant collection of candidates which are then classified by CNN features. The technique described in [19], for instance, selective search [20] is used to generate object proposals, CNN features can be extracted for each proposal, and an SVM classifier is used to determine if the windows contain the object or not. The idea of Region with CNN features as presented in [19] has served as the foundation for a large number of works. The Regions with CNN paradigm-based approaches typically possess high detection accuracy ([21, 22]); still, there are numerous techniques that aim to enhance the performance of Regions to CNN approaches, a few of which find approximate object stances but frequently struggle to pinpoint the precise position of the object [23]. To do this, such approaches frequently adopt a joint object detection, semantic segmentation strategy [24, 25], which typically yields good outcomes.

Most publications in general on object detection by deep learning are using a variant of CNNs, the instance [18,26,27] where in a novel pooling layer and newly learning approach are given, [28] weakly-supervised cascaded CNNs, and [29] which apply CNNs in a cascaded fashion. There are, however, very few attempts at object recognition using other deep learning models.

3.2. Process of Face Recognition

The most popular use of computer vision is face recognition. Multiple face detection methods according to the mining of the handcrafted characteristics have been suggested [30–33]. In these systems, feature extraction extracts the features from such a face to create a low-dimensional representation, on which a classifier bases its predictions. Thanks to their feature learning and transformations invariant characteristics, CNNs changed the face recognition field. The first study to use CNNs for face recognition was [34], and the most advanced methods today include light CNNs [35] and VGG Face Descriptor [36]. Convolutional DBN demonstrated excellent face verification performance in [37].

Furthermore, CNNs serve as the foundation for both FaceNet [38] by Google and DeepFace [39] by Facebook. A face is 3D-modeled using DeepFace [39] and is aligned to look like a front face. Then, the input is passed through a convolution-pooling convolution filter, then now three local connecting layers, and finally two completely connecting layers, which are utilized for creating final predictions. For all that Deep Face has significant performance rates, the representation of it is difficult to understand since the same person's faces are not always grouped together throughout the learning process. Face-Net defines the triple loss function on the representations, causing the learning process to learn to cluster similar people's face representations. The foundation of OpenFace is made up of CNNs as well [40].

3.3. Action and Activity Recognition

Researchers have focused a lot of their effort on the problem of recognizing human actions and activity [41, 42]. In the recent several years, there have been numerous proposals for studies on deep learning-based human activity identification [43]. [44] Describes the application of deep learning for complicated event recognition and detection in

video sequences. Saliency maps are used mostly to locate and detect events, and after that, the previously trained features were subjected to deep learning to identify the key frame that correlates to the underlying events. Similar to the method of [46] to classify events from large video datasets, the author of [45] successfully uses a CNN-based strategy for activity recognition in volleyball; in [47] the authors apply the same method. Based on information from smartphone sensors, an activity recognition model called CNN is applied. The deep CNN model created by the authors of [48] includes a radius margin constraint in regularized terms, significantly boosting the CNN's ability to generalize for the classification of activities.

In [49], the authors examine the potential application of CNN as a features extraction model for close-grained activity; there they find out the method that orders to be able deep features did learn from the ImageNet in SVM classifier will be preferential due to the difficulty of large high intra - class variances, small intra - class variances, and a lack of training samples for each activity. The issue is split into two tasks: first, the most useful attributes for known things are predicted; second, the various components are integrated using an AND/OR graph structure. Aside from several data modalities, numerous works combine multiple model types. By combining videos and sensor datasets with a dual CNN and Long-Term and Short-Term Memory architecture, the authors of [50] offer a multi-modal, multi-stream deep learning system to solve the egotistical activity detection issue.

3.4 Datasets Categorization

Various datasets with a wide range of content depending on the application situation were used to assess the applicability of deep learning algorithms. The primary application domain is (natural) images, irrespective of the researched scenario. The list below includes a brief overview of the conventional and modern datasets that were used for benchmarking.

- Grayscale image, most popular dataset for grayscale images is MNIST [51], which includes NIST & perturbed NIST as well as other versions. The application circumstance is the identification of numerical handwriting.
- Natural Images in RGB. Images of items falling within the 101/256 categories may be found in the Caltech RGB image collections [52], such as the Caltech Silhouettes and Caltech 101/Caltech 256. Multiple classes of thousands of 32 32-color images make up the CIFAR datasets [53]. The COIL datasets [54] contain a variety of images of various things taken in a 360-degree rotation.
- Hyper spectral images. For instance, hyper spectral images may be found in AVIRIS sensor-based datasets [56] and SCIEN hyper spectral image data [55].
- Facial Images: The Audience benchmark dataset [57] is perhaps used to identify age and gender using facial image data as well as many other visual features. Another often-used dataset is face recognition in unrestricted contexts [58].
- Video Streams. The WR dataset [59, 60], which comprises a sequence of Seven kinds of industrial jobs, may be utilized for the video-based activities recognition in the assembly lines [61]. YouTube-8M [62] is a dataset that includes 4,800 various Knowledge Graphs items and Eight million different YouTube clip URLs with video-level labels.

4. Convolution Neural Network Layers

CNNs' design makes it possible for an excellent picture recognition method. The dataset from the image data is classified to use this architecture [63]. The following areas in the CNN architecture are highlighted with several layers and sections enabling image-based identification within the IoT environment. For such an IoT image from the sensing element, our system performs a conditioning procedure as illustrated in figure 5.

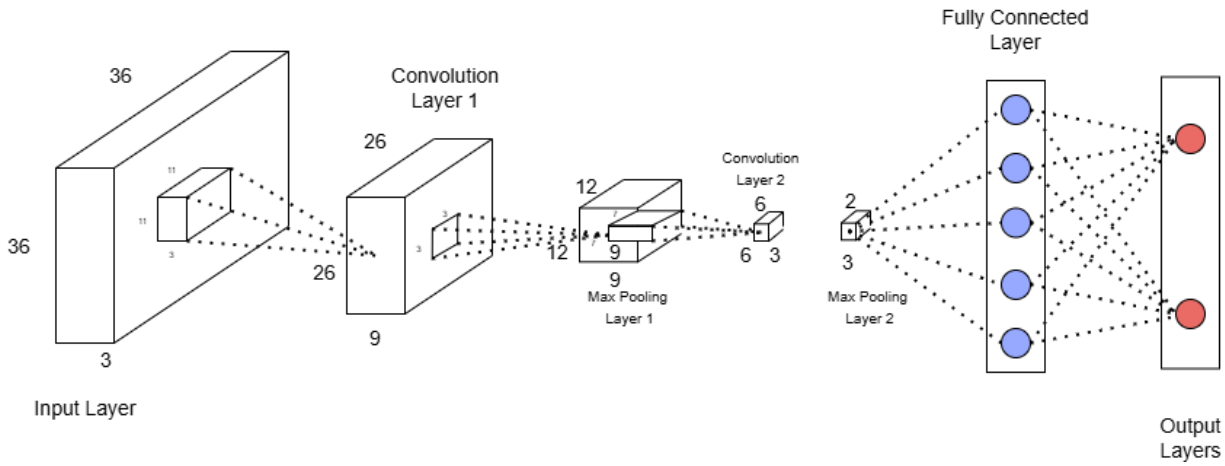

Figure 5. Convolutional Neural Network

4.1 Convolution Layer

To evaluate image processing, it is possible to set the filter size as 100, and the weight variables are taken to be 10k in size. Numerous neurons in the organization are totally coupled to the architecture. To spread weights among the local connections, this convolution layer involves decreasing the parameters [64, 66]. The convolution process is defined as;

$$ul + 1 = vl + 1ul + bl + 1 \quad (1)$$

Where u is the output layer of l. Also, v and b are weight vectors and bias items. This description is providing output data through many convolution kernels with filled pixels at the boundary of the image.

4.2 Pooling layer

The pooling layers are often utilized to drastically decrease the number of training parameters for the main process to minimize the size of the feature map. This layer, which is employed for size reduction using down-sampling principles, is situated in the middle of the neural network's structure [67-69]. To make the feature map smaller, the pooling layer uses the max pooling principle. For such feature maps from the pooling layer, the input data sizes are dynamic and move over step sizes. The parameters are set using variable mode for the feature map output in this structure.

4.3 Post-processing

In this framework's suggested post-processing, there are two functions: 1. Loss function 2. Loss function and the initialization of weight the sample set includes N samples with k^{th} samples & function representations model with anticipated value within the output section. The loss function will compute with the true value of the k^{th} samples as follows;

$$L = \sum l(yk, Yk)\ N\ k = 1 \quad (2)$$

Initiation of weight using the activation function just at the origin value, the weight initialization is employed. The following computations are made for the weight initialization function of this single-layer convolution:

$$y = v1u1 + \cdots + vnk\ unk + b \quad (3)$$

Where n-k is the dimension of the layer for the input.

5. Conclusion and Future Scope

Deep learning has brought significant advancements in how computers process and understand images and videos. It has enabled machines to recognize objects, detect patterns, and even assist in critical tasks like diagnosing diseases and driving cars. Techniques like CNNs and RNNs have shown impressive results in handling complex data, making deep learning valuable for solving real-world challenges across industries.

Looking ahead, deep learning can become even more powerful. Researchers are working on making these models faster, smaller, and more energy-efficient for use in everyday devices. These systems need to be more transparent so

people can understand and trust their decisions, especially in sensitive fields like healthcare. Combining deep learning with other technologies, such as augmented reality and smart sensors, could unlock new possibilities, from smarter homes to safer transportation. With better data and stronger computers, the future of deep learning in image and vision applications is auspicious.

References

[1]. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science. 2006 Jul 28;313(5786):504-7.
[2]. Lai X, Lin S, Zou J, Li M, Huang J, Liu Z, Li D, Fu H. Kansei engineering for the intelligent connected vehicle functions: An online and offline data mining approach. Advanced Engineering Informatics. 2024 Aug 1;61:102467.
[3]. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015 May;521(7553):436-44.
[4]. Abd Halim MA, Rahman SA. Uncovering Patterns in Online Database Usage at UiTM Negeri Sembilan: A Data Mining Approach. Journal of Information and Knowledge Management. 2024 Oct 1;14(2):1-1.
[5]. Gui J, Chen T, Zhang J, Cao Q, Sun Z, Luo H, Tao D. A Survey on Self-supervised Learning: Algorithms, Applications, and Future Trends. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024 Jun 17.
[6]. Ahmad J, Muhammad K, Bakshi S, Baik SW (2018a) Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets. Future Gener Comput Syst 81:314–330.
[7]. Badshah AM, Ahmad J, Rahim N, Baik SW (2017) Speech emotion recognition from spectrograms with deep convolutional neural network. In: 2017 international conference on platform technology and service (PlatCon), 2017. IEEE, pp 1–5.
[8]. Yang X, Song Z, King I, Xu Z. A survey on deep semi-supervised learning. IEEE Transactions on Knowledge and Data Engineering. 2022 Nov 8;35(9):8934-54.
[9]. Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. arXiv preprint arXiv:14092329.
[10]. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2018) Action recognition in video sequences using deep Bi-directional LSTM with CNN features IEEE. Access 6:1155–1166.
[11]. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507.
[12]. Liu X, Yoo C, Xing F, Oh H, El Fakhri G, Kang JW, Woo J. Deep unsupervised domain adaptation: A review of recent advances and perspectives. APSIPA Transactions on Signal and Information Processing. 2022 Aug 15;11(1).
[13]. Kumari S, Singh P. Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives. Computers in Biology and Medicine. 2024 Mar 1;170:107912.
[14]. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp 770–778.
[15]. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115
[16]. Litjens G et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88.
[17]. Rajpurkar P et al (2017) Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:171105225.
[18]. W. Ouyang, X. Zeng, X. Wang et al., "DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 7, pp. 1320–1334, 2017.
[19]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14), pp. 580– 587, Columbus, Ohio, USA, June 2014.
[20]. J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, 2013.
[21]. R. Girshick, "Fast R-CNN," in Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15), pp. 1440–1448, December 2015.
[22]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.
[23]. J. Hosang, R. Benenson, and B. Schiele, "How good are detection proposals, really?" in Proceedings of the 25th British Machine Vision Conference, BMVC 2014, gbr, September 2014.
[24]. B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, "Simultane- ´ous detection and segmentation," in Computer Vision—ECCV 2014, vol. 8695 of Lecture Notes in Computer Science, pp. 297– 312, Springer, 2014.
[25]. J. Dong, Q. Chen, S. Yan, and A. Yuille, "Towards unified object detection and semantic segmentation," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface, vol. 8693, no. 5, pp. 299–314, 2014.
[26]. J. Liu, N. Lay, Z. Wei et al., "Colitis detection on abdominal CT scans by rich feature hierarchies," in Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis, vol. 9785 of Proceedings of SPIE, San Diego, Calif, USA, February 2016.
[27]. G. Luo, R. An, K. Wang, S. Dong, and H. Zhang, "A Deep Learning Network for Right Ventricle Segmentation in Short:Axis MRI," in Proceedings of the 2016 Computing in Cardiology Conference.

[28]. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. V. Gool, "Weakly Supervised Cascaded Convolutional Networks," in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5131–5139, Honolulu, HI, July 2017.
[29]. T. Chen, S. Lu, and J. Fan, "S-CNN: Subcategory-aware convolutional networks for object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[30]. D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification," in Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13), pp. 3025– 3032, June 2013.
[31]. X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13), pp. 3208–3215, December 2013.
[32]. T. Berg and P. N. Belhumeur, "Tom-vs-Pete classifiers and identity-preserving alignment for face verification," in Proceedings of the 23rd British Machine Vision Conference (BMVC '12), pp. 1–11, September 2012.
[33]. D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: a joint formulation," in Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part III, vol. 7574 of Lecture Notes in Computer Science, pp. 566–579, Springer, Berlin, Germany, 2012.
[34]. S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," IEEE Transactions on Neural Networks and Learning Systems, vol. 8, no. 1, pp. 98–113, 1997.
[35]. X. Wu, R. He, Z. Sun, and T. Tan, A light CNN for deep face representation with noisy labels.
[36]. O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in Proceedings of the British Machine Vision Conference 2015, pp. 41.1-41.12, Swansea.
[37]. G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12), pp. 2518– 2525, June 2012.
[38]. F. Schrof, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15), pp. 815–823, IEEE, Boston, Mass, USA, June 2015.
[39]. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14), pp. 1701–1708, Columbus, Ohio, USA, June 2014.
[40]. B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: a general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, 2016.
[41]. S. Voulodimos, D. I. Kosmopoulos, N. D. Doulamis, and T. A. Varvarigou, "A top-down event-driven approach for concurrent activity recognition," Multimedia Tools and Applications, vol. 69, no. 2, pp. 293–311, 2014.
[42]. S. Voulodimos, N. D. Doulamis, D. I. Kosmopoulos, and T. A. Varvarigou, "Improving multi-camera activity recognition by employing neural network based readjustment," Applied Artificial Intelligence, vol. 26, no. 1-2, pp. 97–118, 2012.
[43]. K. Makantasis, A. Doulamis, N. Doulamis, and K. Psychas, "Deep learning based human behavior recognition in industrial workflows," in Proceedings of the 23rd IEEE International Conference on Image Processing, ICIP 2016, pp. 1609–1613, September 2016.
[44]. Gan, N. Wang, Y. Yang, D.-Y. Yeung, and A. G. Hauptmann, "DevNet: A Deep Event Network for multimedia event detection and evidence recounting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, pp. 2568–2577, USA, June 2015.
[45]. T. Kautz, B. H. Groh, J. Hannink, U. Jensen, H. Strubberg, and B. M. Eskofer, "Activity recognition in beach volleyball using a DEEp Convolutional Neural NETwork: leveraging the potential of DEEp Learning in sports," Data Mining and Knowledge Discovery, vol. 31, no. 6, pp. 1678–1705, 2017.
[46]. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F.-F. Li, "Large-scale video classification with convolutional neural networks," in Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, (CVPR '14), pp. 1725– 1732, Columbus, OH, USA, June 2014.
[47]. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," Expert Systems with Applications, vol. 59, pp. 235–244, 2016.
[48]. L. Lin, K. Wang, W. Zuo, M. Wang, J. Luo, and L. Zhang, "A deep structured model with radius-margin bound for 3D human activity recognition," International Journal of Computer Vision, vol. 118, no. 2, pp. 256–273, 2016.
[49]. S. Cao and R. Nevatia, "Exploring deep learning based solutions in fine grained activity recognition in the wild," in Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 384–389, Cancun, December 2016.
[50]. J. Shao, C. C. Loy, K. Kang, and X. Wang, "Crowded Scene Understanding by Deeply Learned Volumetric Slices," IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 3, pp. 613–623, 2017.
[51]. Y. LeCun, L. Bottou, Y. Bengio, and P. Hafner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2323, 1998.
[52]. L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 594–611, 2006.
[53]. Krizhevsky and G. Hinton, Learning multiple layers of features from tiny images, 2009.
[54]. S. A. Nene, S. K. Nayar, and H. Murase, Columbia object image library (coil-20), 1996.

[55]. T. Skauli and J. Farrell, “A collection of hyperspectral images for imaging systems research,” in Proceedings of the Digital Photography IX, USA, February 2013.
[56]. M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, “220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3,” Datasets, 2015.
[57]. E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of unfiltered faces,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.
[58]. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Tech. Rep University of Massachusetts, Amherst, 2007.
[59]. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-Ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462–3471, Honolulu, HI, May 2017.
[60]. Sef, L. Lu, A. Barbu, H. Roth, H.-C. Shin, and R. M. Summers, “Leveraging mid-level semantic boundary cues for automated lymph node detection,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface, vol. 9350, pp. 53– 61, 2015.
[61]. Voulodimos, D. Kosmopoulos, G. Vasileiou et al., “A dataset for workflow recognition in industrial scenes,” in Proceedings of the 2011 18th IEEE International Conference on Image Processing, ICIP 2011, pp. 3249–3252, Belgium, September 2011.
[62]. Voulodimos, D. Kosmopoulos, G. Vasileiou et al., “A threefold dataset for activity and workflow recognition in complex industrial environments,” IEEE MultiMedia, vol. 19, no. 3, pp. 42–52, 2012.
[63]. Hsia, C. H., Yen, S. C., & Jang, J. H. (2019). An intelligent iot-based vision system for nighttime vehicle detection and energy saving. Sensors & Materials, 31(6(1)), 1803-1814.
[64]. Tripathi, Milan. "Analysis of Convolutional Neural Network based Image Classification Techniques." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 100-117.
[65]. Rao, Anushree Janardhan, Chaithra Bekal, Y. R. Manoj, R. Rakshitha, and N. Poornima. "Automatic Detection of Crop Diseases and Smart Irrigation Using IoT and Image Processing."In Innovative Data Communication Technologies and Application, pp. 363-374. Springer, Singapore, 2021.
[66]. Durairaj, M., and J. Hirudhaya Mary Asha. "The Internet of Things (IoT) Routing Security—A Study." In International Conference on Communication, Computing and Electronics Systems, pp. 603-612. Springer, Singapore, 2020.
[67]. Shanmugapriya, T., K. Kousalya, J. Rajeshkumar, and M. Nandhini. "Wireless Sensor Networks Security Issues, Attacks and Challenges: A Survey." In International conference on Computer Networks, Big data and IoT, pp. 1-12. Springer, Cham, 2019.
[68]. Yadav, Vicky, and Rejo Mathew. "Analysis and Review of Cloud Based Encryption Methods." In International conference on Computer Networks, Big data and IoT, pp. 169-176. Springer, Cham, 2019.
[69]. Paul, Vedant, and Rejo Mathew. "Data Storage Security Issues in Cloud Computing." In International conference on Computer Networks, Big data and IoT, pp. 177-187. Springer, Cham, 2019.