*International Journal of Engineering Applied Sciences and Technology, 2021*
**Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 159-166**
Published Online November 2021 in IJEAST (http://www.ijeast.com)

# A REVIEW ON OBJECT DETECTION IN THE HEALTHCARE USING DEEP LEARNING TECHNIQUES

Mr. Nallanti Venkateswararao
Research Scholar,
Computer Science and Engineering,
Adikavi Nannaya University,
Rajahmundry,India

Dr. Pallipamu. Venkateswara Rao
Associate professor,
Computer Science and Engineering,
Adikavi Nannaya University,
Rajahmundry,India,

*Abstract*- **Deep learning comes under the class of machine learning algorithms which made a huge impact in areas of healthcare, such as cancer diagnosis, precision medicine, Diabetic Retinopathy, Gastrointestinal (GI) Diseases Detection, Cardiac Imaging, Tumor Detection etc. It uses several layers for feature extraction and transformation. The output of one layer is given as input to the successive layers. Deep Learning Algorithms are based on the unsupervised machine learning of multiple levels of features of the data. It uses some form of gradient descent for training. Deep learning has shown impressive performance in various fields such as image segmentation, image classification and object detection. In this paper we focused on deep learning and its tool i.e. convolution neural network in generic object detection architectures along with some modifications and tricks to improve detection performance further. Finally we have given future research directions**

*Keywords:* **Machine learning algorithm, optimization, convolution neural networks.**

## I. INTRODUCTION

The digitization of medical records increases with rapid growth of Deep learning in computer vision tasks. The digitization of electronic health records increased in the US from 2007 to 2012 [1]. Structure of Human brain and neurons of the brain are the base of Deep learning, it is the subset of Machine learning [2]. Due to the rapid development in Deep learning, it has achieved considerable results in many fields such as robotics, medicine, traffic analysis etc. It contains different layers such as input layer, hidden and output layers. Nodes of adjacent layers are connected, each edge (known as connection) is associated by a weight. Here to receive results at each unit, inputs are multiplied in respect to edge weights and summed. The sum is input to the activation function for transformation. Most of the cases activation functions are a sigmoid function, tan hyperbolic or ReLU. The deep neural network's structure is shown in the figure1 below. The sigmoid function value exists between 0 and 1, Therefore, when it is used for models the output is in between 0 and 1 . Due to the differentiability, at any two points, the sigmoid curve can measure the slope. Another function which is better than logistic sigmoid is tanh and whose range is -1 to 1.

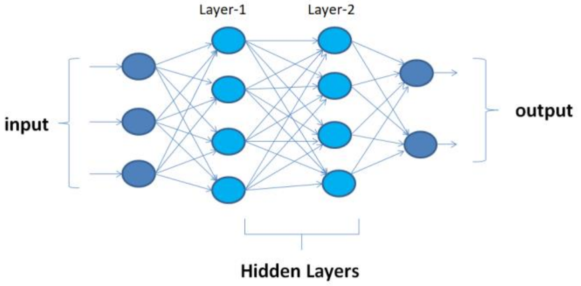

Fig 1: Deep Neural network

In classification of two classes tanh() is used frequently.The subsequent unit of the post layer is fed by the result of the output function. Sigmoid curve, sigmoid logistic are used prominently in feed forward nets.The solution of the problem is the output of the final layer. Implementation of any neural networks consists of 1.Preprocessing Dataset, 2. Convert it into a training and testing data set, 3.Train the neural network, 4. Make predictions with test data.

The paper is organized in the following sections: Introduction to Deep Learning, History of Convolutional neural networks, DNN Architectures, Training Algorithms

## II. THE ROLE OF CNN IN OBJECT DETECTION

Object Detection (OD) is a technique of computer vision based on deep learning. The rapid changes in the computer vision field make it useful in the most challenging areas such as object classification and object localization. The aim of OD technique is to find object location in the image called object localization and finding class of each object called object classification. In object detection the image is divided into M x N regions. Initially a model is prepared on a set of images for classifying into positive and negative image classes. The object detection methods are divided into region proposed based and classification based methods. In object detection CNN place important role, The object detection algorithms are Fast RCNN,FasterRCNN,HOG, R-FCN,SSD,SPP-net, YOLO.

### A. Fast R-CNN

R-CNN is an object detection model,for segmentation and localisation it uses CNN's to bottom-up region proposals.It identifies bounding boxes of object region candidates based on selective search and then gets features from each region for classification. The pitfalls in the R-CNN are overcomed by Fast R-CNN [3] to achieve speed and accuracy. The advantages it has are single training stage, higher detection quality, updating all network layers, for feature caching disk storage is not required

### B. Faster R-CNN

The object detection algorithm Faster-CNN[4] utilizes the Region Proposal Network and shares full-image Convolutional features in a cost-effective manner than R-CNN and Fast R-CNN. Faster R-CNN has two modules, the deep FCN that proposes regions and Fast R-CNN detector which uses the proposed regions. This system is a unified network for object localisation and classification and the RPN module tells the second module where to look.

### C. Histogram of Oriented Gradients (HOG)

The HOG algorithm extracts features from images data and detects objects in the field of computer vision. The HOG descriptor mainly concentrates on the object structure or object shape. In localized portions of an image it includes occurrences of gradient orientation, such as the region of interest(ROI),detection window. The HOG is simple and easy to understand.

### D. Region-based Convolutional Neural Networks (R-CNN)

The combination of region proposals with Convolution Neural Networks (CNNs) is RCNN. It is used in localizing objects with the help of a deep network. It can be trained as a robust model with a small amount of detection data. It gives high accuracy and is used to classify object proposals. without resorting to approximate techniques,it can scale to thousands of object classes.

### E. Region-based Fully Convolutional Network (R-FCN)

R-FCN is a region based detector for object detection. On the entire image, all computations are shared , unlike other costly region based detectors applied on subnetworks. Unlike Fast RCNN, RFCN has a shared and fully convolutional architecture that can yield a better result.All weights are learnable and convolutional in the algorithm, and used in classification of ROIs into object background and categories.

### F. Single Shot Detector (SSD)

Single neural network in combination with an SSD is used in real time to detect objects. At 7 frames/Seconds, accuracy of State of art is considered in object detection. Object detection methods like the Single Shot Detector region proposal are eliminated to speed up the process. SSD includes multi-scale and default boxes features to overcome the reduction in accuracy.

To match SSD with the Faster R-CNN in terms of accuracy, use the lower resolution images and boosts the speed. To handle objects of different sizes, the SSD network converges predictions from various feature maps with various resolutions.

### G. Spatial Pyramid Pooling (SPP-net)

Regardless of size of image representation, SPP-net [5] generates fixed length images.

In deformations SPP is robust, and improves the classification of all CNN based images. On the entire image, feature map computation is done only once, and for training necessary detectors, produces a fixed length image, where features are pooled in arbitrary regions. Iterative convolutional feature computations are avoided.

### H. YOLO (You Only Look Once)

YOLO is one of the object detection algorithms for localization and classification of objects in the images. Unified YOLO is extremely faster than others according to Facebook AI Research. Images can be processed at 45 frames/second in the case of the base model of YOLO. Astounding 155 frames/second can be achieved by Fast YOLO and it is double the real time detectors like mAP. In generalizing natural images to artwork in the images related domain, YOLO performs better than DPM and R-CNN.

## III. CNN ARCHITECTURE

Neuron structure in the human brain is the inspiration for Deep Learning algorithms. DL algorithms are used in images, videos of high dimension. Fully connected layers added to convolutional layers are units of a multi layer NN(Neural Network).Feature extraction is an analysis process of a convolution tool, features of images are separated and identified by it for analysis purpose. Output of the convolution process is utilized to predict the class based on extracted features in the previous stage of the image.

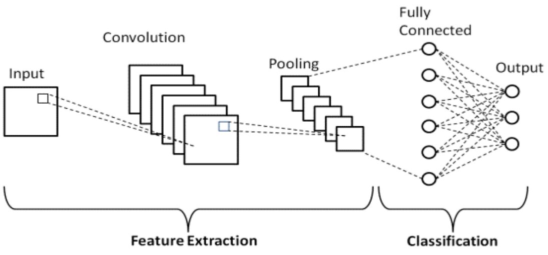

Fig 2: Architecture of CNN

### A. Convolutional Layer

The first layer of CNN is Convolutional Layer which is used for extracting the features from the input images. Here between the input image and a filter the convolution operation is performed and is applied to the entire image by sliding and finding the dot product between the filter and the input image parts. The output is termed as the Feature map which provides us with corners and edges information about the image. The feature map is given to other layers to learn remaining features of the image.

### B. Pooling Layer

Generally Pooling Layers follow Convolutional Layers. The main aim of this layer is to reduce the number of parameters and calculations to lower the computational costs. It reduces connections between layers and operates on all feature maps. There are different types of pooling operations. In max pooling the biggest element is taken from the feature map. In Average Pooling the average of the elements is found in a predefined size image. The total sum is found in Sum Pooling. These Layers usually act as a bridge between the Convolutional Layer and the FC Layer

### C. Fully connected layer

These layers are a standard Deep Neural Network, which find the predictions from the activations for classification or regression. The Fully Connected (FC) layer is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.

## IV. CNN Models

Deep CNN has a major role in image classification and recognition therefore it has become widely known. In this section various classical and modern architectures of Deep CNN are described.Major CNN models are as follows

### A. LeNet

Yann LeCun in 1989[7] proposed LeNet which is a CNN structure and It refers as LeNet-5.CNN is a kind of feed-forward neural network whose artificial neurons can perform well in large-scale image processing[Wikipedia]

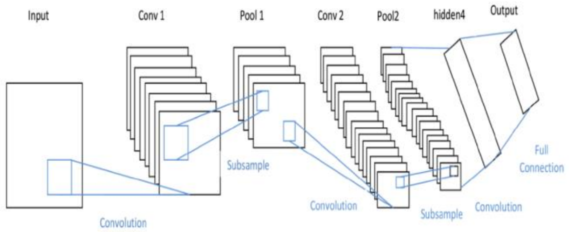

Fig 3: Architecture of LeNet [7]

### B. AlexNet model

AlexNet is CNN architecture, designed by Alex Krizhevsky , Sutskever and Geoffrey Hinton. On ImageNet data ,AlexNet [8] has an accuracy of 84.6%. It utilizes data augmentation methods such as patch extractions,image translation and horizontal reflection. It is implemented to avoid the problem of overfitting over the training data. It uses batch SGD for training for weight decay and momentum. It uses two GTX 580 GPUs and contains five convolutional layers, one max pooling, RELU as non-linearities, three Fully Connected layers and dropout.

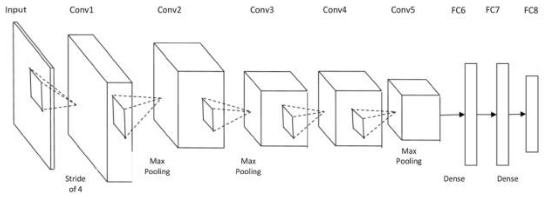

Fig 4: Architecture of AlexNet [8]

### C. ZFNet

ZFNet is a classic convolutional neural network. The design of ZFNet was motivated by visualizing intermediate feature layers and the operation of the classifier. The filter sizes are reduced and the stride of the convolutions is reduced when compared to Alexnet.

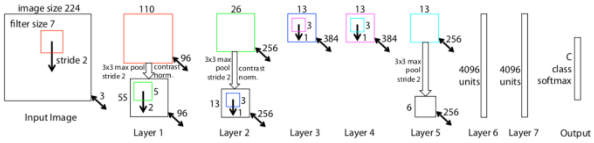

Fig 5: ZFNet architecture[9]

### D. GoogleNet Model

GoogLeNet is a deep CNN which contains 22 lyers and that’s a variant of the Inception Network.It was developed by researchers at Google. GoogleNet gives a top-5 test accuracy of 93.3% on ImageNet ILSVRC14. This model is used to enhance computer resources utilization into the model.

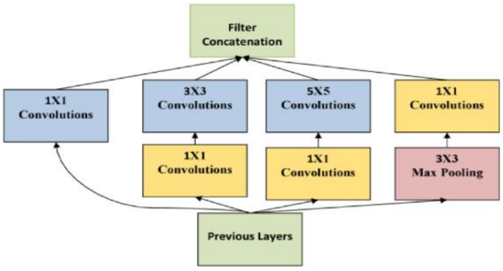

Fig 6: Architecture of GoogleNet [10]

### E. VGGNet Model

VGGNet is a CNN model architecture developed by K.Simonyan and A.Zisserman in 2014. It takes 224*224 RGB image as input and preprocess it with pixel values in the range 0-255 and subtracts the mean image values.It has 13 convolutional and 3 fully connected layer.VGG has in depth filters of smaller size (3*3) instead of having large filters.

### F. Inception model

The inception is a CNN model and has more techniques to improve both speed and accuracy.

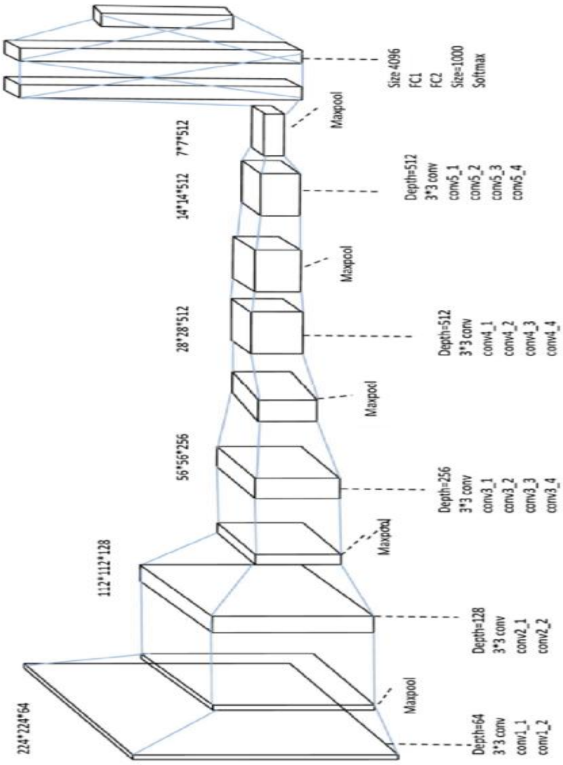

Fig 7: Architecture of VGG19 [11]

It is capable of discovering how a CNN local sparse design can be relative and secured by dense components. For dissecting end layer relationship statistics a layer by layer construction is done and these shape the units of the next layer. The prior layer units and some locale of the input image are compared and assembled into filter banks.

### G. ResNeXt model

The architecture of ResNeXt is an extension to deep residual networks, acquires the approach of replicated layers and uses the split–transform–merge procedure which uses Inception models. The ResNeXt has a group of ResNet/VGGs influenced residual blocks. The rules are as follows:the blocks split the hyper parameters and every time the spatial map is pooled by two factors. It is multi-branch and has cardinality. It gives better performance than ResNet.

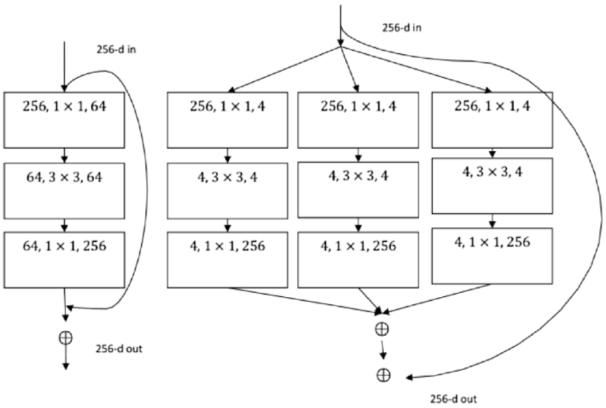

Fig 8: Architecture of ResNeXt [12]

### H. SENet Model

The SENet is CNN model and constructed With “Squeeze-and-Excitation” (SE) block which adaptively recalibrates channel-wise feature responses. This model can increase the representational power of the model concentrating on the relationship among the channels. SENets assembled groundwork on ILSVRC 2017 categorization submission and automatically diminishes the top-5 error to 2.251%. Without any computational cost it enhances the interdependencies of the channel and GPU is used to implement global average pooling.

### I. MobileNet V1/V2

Convolutional layers, which are quite expensive to compute, can be replaced by so-called depthwise separable convolutions MobileNets V1. In this model [10], the depthwise convolution replaces normal convolution followed by pointwise convolution. The depth wise separable convolutions are used to perform a solitary convolution on every color channel.

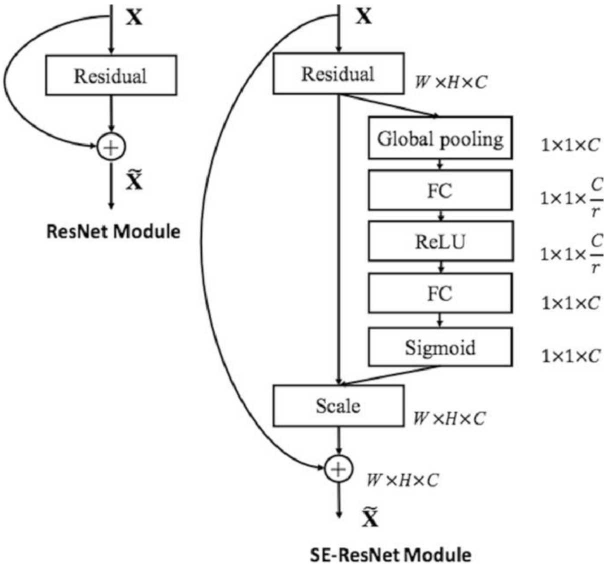

Fig 9: ResNet and SE-ResNet module

### J. DenseNet

In DenseNet, each layer passes feature-maps to all subsequent layers by getting inputs from all preceding layers and Concatenation is used. A DenseNet is a type of CNN which uses dense connections among layers, through Dense Blocks. In this architecture every layer is coupled in a feed forward manner. It has L (L + 1)/2 direct connections where L is Layer. It concatenates both output feature maps and incoming feature maps. It diminishes the vanishing gradient problem, number of parameters by using information from all previous layers.

### K. Xception model

The Xception architecture [14] is developed with depthwise distinguishable convolutions, and pointwise convolution. It is an extended version of the inception model.This architecture shows better results on ImageNet than others. The number of connections is fewer, and the model is lighter. It shows better results when compared to Inception V3.

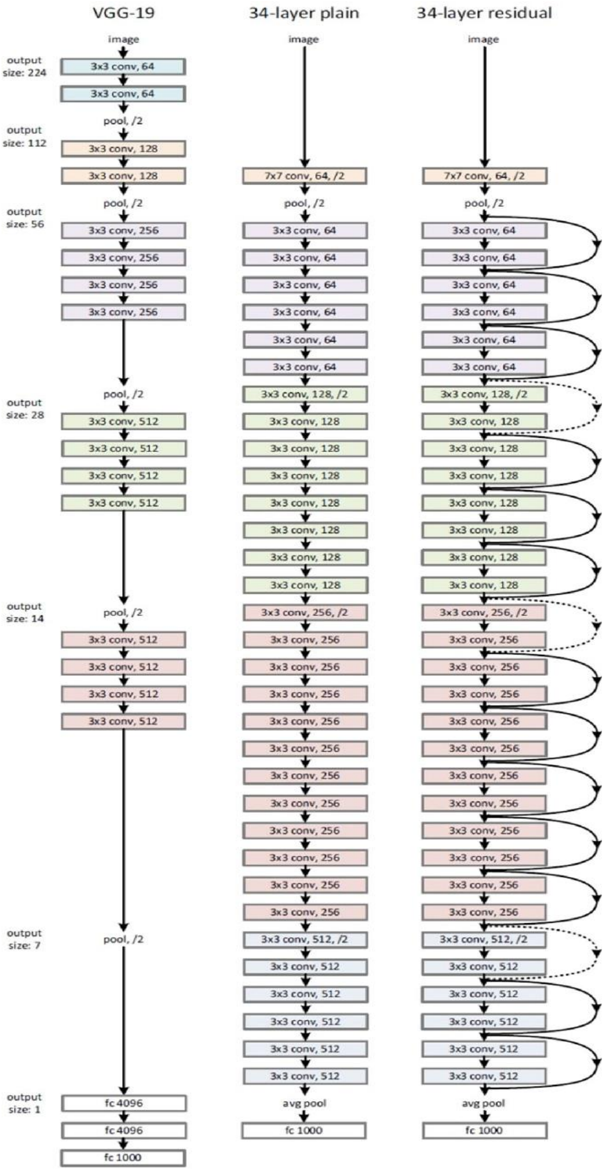

Fig 10: Architecture of ResNet[13]

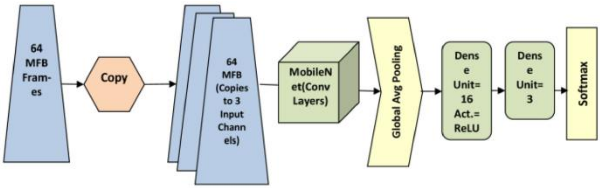

Fig 11: Architecture of MobileNet [14]

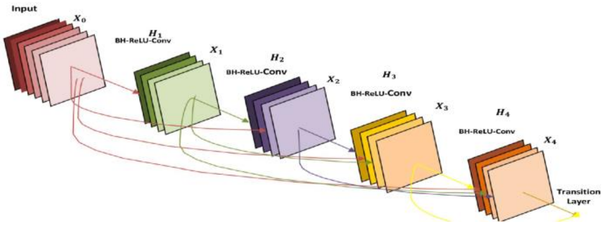

Fig 12: Architecture of DenseNet model [15]

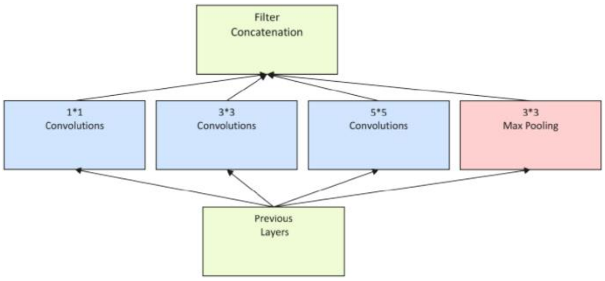

Fig 13: Architecture of Inception model [16]

### L. NAS/PNAS/ENAS

The best model on neural networks is NASNet [13] which is used for image classification and semantic segmentation. The optimization of the model can be achieved with reinforcement learning searching methods.

### M. Efficient Net

EfficientNet is CNN architecture and scaling method and with compound coefficient ϕ uniformly scales all dimensions. The EfficientNet scales depth,width and resolution with fixed scaling coefficients of the network. The compound scaling is suitable for bigger images and requires more layers and more channels for processing images.

## V. OBJECT DETECTION

It is the technique used for computer vision tasks to locate and classify existing objects in the images. The sliding window method reduces time complexities. In this method, a window with size of M × N is picked to search over the objective image. The frameworks are mainly divided into two types, One which follows the traditional object detection pipeline, generating region proposals and classifying them into different object categories. The second type is the object detection as a regression or classification problem, adopting a unified framework and getting final results.

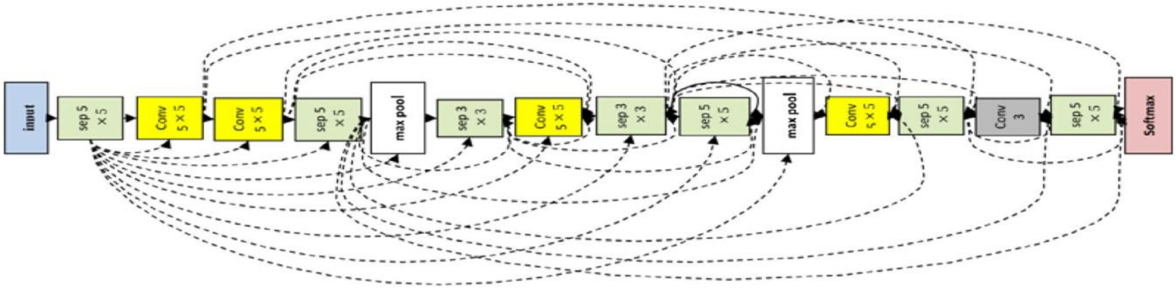

Fig 14: ENAS architecture [17]

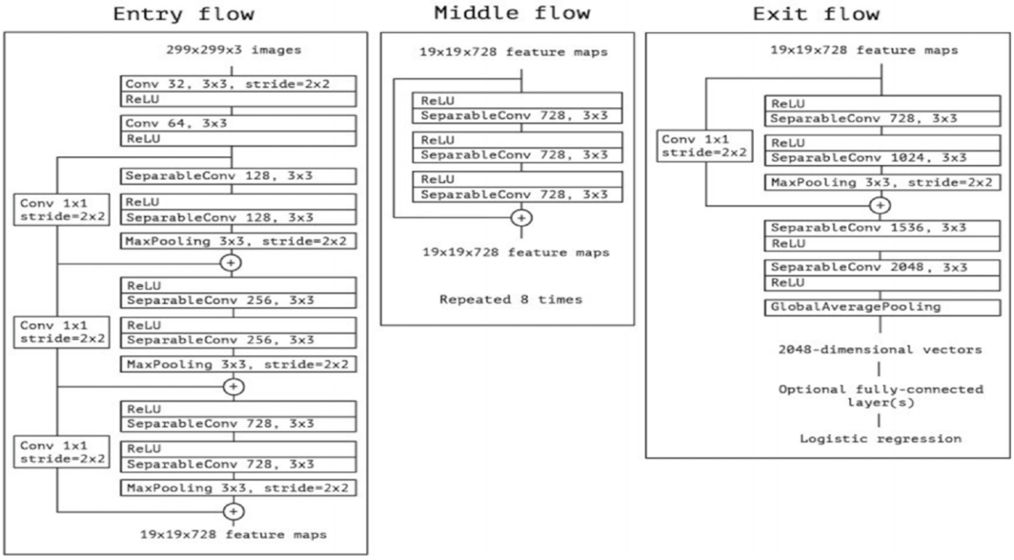

Fig 15: Architecture of Xception model [18]

## VI. CNN APPLICATIONS IN OBJECT DETECTION

This section gives a brief review of CNN applications in real time. The application of Deep learning created remarkable work in the last few years. Object detection is one of the demanding issues in many fields. The object detection is mainly used for object localization and classifying the objects in the images.

Medical face mask detection is used in public places to identify the people who do not wear face masks to stop the spreading of Covid-19 among people. The aim of Mohamed Loey[19] model is used to classify and localize the medical face mask objects in images. This model consists of ResNet-50 as a feature extractor and YOLO v2 for detection of medical face masks. The IOU is used to estimate anchor boxes and achieved 81% average precision. Arjya Das[20] presents an approach by using basic ML packages such as OpenCV,TensorFlow, Keras and Scikit-Learn for detecting face masks with an accuracy up to 95.77%. Ryumina[21] proposed a hybrid method for detecting face masks which combines visual features extracted by CNN that convey information about pixel intensity.He tested his approach on the Medical Mask Dataset , MAFA and RMFD datasets. The proposed recognises the presence/absence of a protective mask on a human's face in comparison with traditional CNNs on the MAFA and RMFD databases. Paul F. Jaeger[22] proposed a Retina U-Net, used for semantic segmentation in images and it recaptures discarded supervision signals by complementing object detection with an auxiliary task. Bhavneet Kaur[23] proposed a model salient object detection algorithm for brain image analysis by combining background and foreground connectivity, achieved the accuracy of 97.07%.Ling Dai model DeepDR[24] which is a transfer learning network to evaluate retinal image quality, retinal lesions, and Diabetic Retinopathy(DR) grades. Sehrish Qummar [25] to train an ensemble of five deep CNN models to encode the rich features and improve the classification for different stages of DR. Shanaka Ramesh Gunasekara etc proposed a model to classify and segment brain tumors using a T1 weighted MRI sequence. It consists of a CNN , Faster R-CNN,Chan–Vese algorithms for classification, tumor localization, and precise tumor segmentation respectively. The proposed model of H.N.T.K.Kaldera[26] used a CNN, for classification problem and Faster R-CNN for segmentation problem with lower computations with a higher accuracy level. This system shows average accuracy of 94%. To overcome human errors Yakub[27] proposed a model by Faster R-CNN with Region Proposal Network (RPN) and VGG-16 architecture for detecting brain tumors with an average efficiency of 77.60%.

## VII. CONCLUSION

In this review paper Deep Learning methods in object detection are discussed.It includes the history of CNN in object detection along with CNN Models, CNN Architecture and finally . The applications of deep learning in facemask detection,Diabetic retinopathy detection are introduced. Deep learning-based object detection methods have achieved huge progress, so it will remain an active research area in the healthcare domain. The future work is to design new models or update existing models to achieve better results in the field of object detection.

## VIII. REFERENCES

[1] C. J. Hsiao, E. Hing, and J. Ashman, “Trends in electronic health record system use among office-based physicians: United states, 2007-2012,”Natl Health Stat Report, no. 75, pp. 1–18, 2014.

[2] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436 (2015)

[3] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[4] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 1 Sept. 2015, doi: 10.1109/TPAMI.2015.2389824.

[6] Lee, J., Bang, J., Yang, S.I.: Object detection with sliding window in images including multiple similar objects. In: 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp. 803–806 (2017)

[7] Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)

[8] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

[9] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, pp. 818–833. Springer, Cham (2014)

[10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)

[11] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[12] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)

[13] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[14] Hussain, M., Haque, M.A.: Swishnet: a fast convolutional neural network for speech, music and noise classification and segmentation. arXiv preprint arXiv :1812.00149 (2018)

[15] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,pp. 4700–4708 (2017)

[16] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper withconvolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)

[17] Pham, H., Guan, M.Y., Zoph, B., Le, Q.V., Dean, J.: Efficient neural architecture search via parameter sharing. arXiv preprint arXiv :1802.03268 (2018)

[18] Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)

[19] Mohamed Loey a, Gunasekaran Manogaran b,c, Mohamed Hamed N. Taha d, Nour Eldeen M. Khalifa: Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, Elsevier: 102600 (2021)

[20] A. Das, M. Wasif Ansari and R. Basak, "Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV," *2020 IEEE 17th India Council International Conference (INDICON)*, 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342585.

[21] Ryumina, E., Ryumin, D., Ivanko, D., & Karpov, A. 2021. A Novel Method for Protective Face Mask Detection Using Convolutional Neural Networks and Image Histograms.International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIV-2/W1-2021, 177-182.

[22] Jaeger, Paul F., et al. “Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for

Medical Object Detection.” ArXiv:1811.08661 [Cs], Nov. 2018. arXiv.org, http://arxiv.org/abs/1811.08661.

[23] Bhavneet Kaur.(2018). An improved salient object detection algorithm combining background connectivity and foreground for brain image analysis. Computers & Electrical Engineering. 71. 692-703. 10.1016/j.compeleceng.2018.08.018.

[24] Dai, L., Wu, L., Li, H. et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. NatCommun 12, 3242 (2021). https://doi.org/10.1038/s41467-021-23458-5.

[25] S.Qummar et al., "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," in IEEE Access, vol. 7, pp. 150530-150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[26] H. N. T. K. Kaldera, S. R. Gunasekara and M. B. Dissanayake, "Brain tumor Classification and Segmentation using Faster R-CNN," 2019 Advances in Science and Engineering Technology International Conferences (ASET), 2019, pp. 1-6, doi: 10.1109/ICASET.2019.8714263.

[27] Y.Bhanothu, A. Kamalakannan and G. Rajamanickam, "Detection and Classification of Brain Tumor in MRI Images using Deep Convolutional Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp.248-252.