Neuroscience Informatics 2 (2022) 100070

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

NaaS

Localization of stroke lesion in MRI images using object detection techniques: A comprehensive review

Sangeeta Rani [a], Bhupesh Kumar Singh [b], Deepika Koundal [c], Vijay Anant Athavale [d,*]

[a] Computer Engineering Department, B. S. Anangpuria Institute of Technology & Management, Faridabad, India
[b] B. S. Anangpuria Institute of Technology & Management, Faridabad, India
[c] School of computer Science, UPES - University of Petroleum and Energy Studies, Dehradun, India
[d] Walchand Institute of Technology, Solapur, Maharashtra, India

ARTICLE INFO

Article history:
Received 16 September 2021
Received in revised form 3 April 2022
Accepted 7 April 2022

Keywords:
Lesion
Localization
RCNN (Region-based Convolutional Neural-Network)
Fast RCNN (Region – based Convolutional Neural-Network)
Selective search
YOLO (You Only Look Once)

ABSTRACT

Stroke is one of the lethal diseases that has significant negative impact on an individual's life. To diagnose stroke, MRI images play an important role. A large number of images are being produced day by day such as MRI (Medical Resonance Imaging), CT (Computed Tomography) X-Ray images and many more. Machine Learning algorithms are less efficient and time-consuming in localization of such medical images. Object detection using deep learning can reduce the efforts and time required in screening and evaluation of these images. In the proposed paper, several approaches such as RCNN (Region-based Convolutional Neural-Network), Fast R-CNN (Fast Region-based Convolutional Neural Network), Faster R-CNN (Faster Region-based Convolutional Neural Network with Region proposal Network), YOLO (You Only Look Once), SSD (Single-Shot Multibox Detector) and Efficient-Det are listed which can be used for stroke localization and classification. Comparison of RCNN, Fast R-CNN, Faster R-CNN, YOLO, SSD and Efficient-Det with accuracy are also present in this paper. A Chart of the Data Set available for object detection is also considered in this paper. By The maP (Mean-Average Precision) and the accuracy of every single method, it is identified that the speed and accuracy need to poise.

1. Introduction

Stroke has been one of the lethal diseases in the world which harm an individual's life. Stroke cases are being reported highly in lower strata of society and the younger generation. This has caused the scientific community to be more interested in the clinical diagnosis and treatment of the disease. MRI images are used in diagnosing stroke especially ischemic stroke. The diagnosis of these images is being done by human experts all over the world causing extended time and efforts which leads to the extension in the waiting period of the patients. In most cases, immediate results are required due to the deadly nature of the disease. The heavy workloads of the doctors and radiologists make the situation more difficult. The wrong diagnosis is always the possibility due to immense pressure. Intelligent agents have increased the hopes for auto-diagnosis of MRI and the most sought-after solution in the medical arena. As technology is changing its landscape, the application of computer vision is also growing. Computer vision plays a significant role in the agriculture field, medical field, automobile industry, and many more. To ease the working of computer vision, object detection has an important role [1]. A lot of research is going on in the area of object detection, such as anomaly detection, eCommerce, the health sector, and many more. The main focus of the review is object detection because visual detection is more prone to error and also causes noise [2]. The researchers have analysed methods such as RCNN, Fast RCNN, YOLO, SSD and EfficientDet that can be used in stroke detection. Stroke, the world's second most deadly diseases, has historically been one of the leading causes of harm to people's lives and health. It is marked by a high morbidity, disability rate, death rate, and recurrence rate, and it places a great burden on society and patients' families. At the moment, the incidence of stroke is rapidly increasing in low-income and younger populations [3]. The quantitative analysis of brain MRI images is critical in the diagnosis and treatment of stroke. With the growing relevance of medical imaging in clinical diagnosis, MRI has become a key foundation for stroke diagnosis and therapy, particularly for ischemic stroke, which is difficult to identify from CT scans as compared to hemorrhagic stroke [4]. Deep learning technology, illustrated by a convolutional neural network (CNN), systematically pulls characteristic values from a big

* Corresponding author.
E-mail addresses: sangeeta.rani@faculty.anangpuria.com (S. Rani), bhupesh.singh@anangpuria.com (B.K. Singh), koundal@gmail.com (D. Koundal), vijay.athavale@gmail.com (V.A. Athavale).

https://doi.org/10.1016/j.neuri.2022.100070

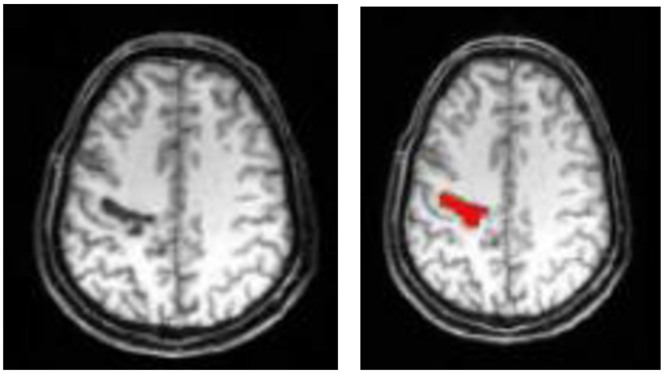

Fig. 1. (a) Image classification; (b) Object detection [10].

dataset in order to acquire more sophisticated conceptual features for classification, detection, and segmentation, enabling intelligent MRI interpretation.

In the classification, the model predicts the class label of the image. The classification approach only classifies the image but it does not locate the portion of the object in an image. So, this problem can be solved using object detection [5]. Kronberg et al. [6] employed the winning algorithm from the 2018 Brain Tumor Segmentation Challenge, trained on the BraTS 2020 dataset, to segregate necrotic core, peritumoral edema, and augmenting tumour. They concluded that the average segmentation accuracy ranges from 0.476 for T1 alone to 0.751 for the whole collection of sequences. Sarkar et al. [7] examined and evaluated their appropriateness to track mental sadness using EEG data, four neural network-based deep learning architectures, namely MLP, CNN, RNN, RNN with LSTM, their result shows that the RNN model scored the greatest accuracy among Neural Network-Based Deep Learning approaches, with 97.50 percent in the Training Set and 96.50 percent in the Testing Set. Phase. Kharouba et al. [8] collected data on consecutive patients with ELVO who were treated with thrombectomy in a prospective study. The results of their study show that the number of passes required to accomplish recanalization, as well as age, stroke severity, and collateral and reperfusion status, all had an influence on survival and functional result after 90 days. Radiologists may construct any contrast-weighted picture using specific SyMRI software by adjusting acquisition parameters like as repetition time, echo duration, and inversion time. Furthermore, automated brain tissue segmentation, volumetry, and myelin assessment are all possible. In everyday clinical paediatric imaging, the SyMRI technique allows for a quicker scan duration [9].

In the case of object detection, the technique uses the bounding box (Clip Window) around the object to detect where the object is shown in Fig. 1(a) and Fig. 1(b). Several approaches have been introduced for object detection as discussed in Table 1.

The researchers have surveyed deep learning-based approaches that can be used for object detection. The Researchers have discussed the motivation of the research, a detailed methodology of techniques, a summary of the review, and concludes with a conclusion.

2. Motivation

A significant number of medical pictures are created in a short period of time due to advancements in medical imaging and neuroscience. As a result, a strategy for reducing clinical efforts is necessary. Object detection is a technique for recognising and locating all known items in a given environment. The data from the object detector may be utilised to navigate around any barriers in the area. Automatic tooth detection, pedestrian walking, industrial inspection, quality inspection, traffic analysis, food product inspection, book identification on the shelf, and medical analysis are just a few examples of where object detection is employed. Because human involvement in any endeavour takes time. Because humans are more prone to making mistakes, the goal of this study topic is to develop a system that can recognise things more precisely.

3. Data set available

There is a various dataset available for object detection (Table 2).

4. Methodology

There are various methodologies available for object detection. Some of the techniques use Machine-Learning (ML) in which separate approaches are required for feature extraction, and separate approaches are required for classification, but in deep learning, the same algorithm can be used for feature extraction and classification. And some of the techniques use Deep-Learning (DL) The Researchers have presented a detailed explanation of some of the deep-learning-based object detection.

4.1. RCNN (Region-based Convolutional Neural-Network)

A Network named Region-based Convolutional Neural network was introduced that is based on Convolutional Neural Network and detect objects more accurately [13]. Selective search approach is used to find out region proposals. Based on regions, features are extracted and then SVM classifier is used to classify the objects.

R-CNN depends on three stages:

a. Region Proposal
b. Feature Extraction
c. Classification

a. Region Proposal: In R- CNN, a selective- search method is used for region proposal. In this approach, initially, the input-image is divided into many regions. Then based on the similarity between CNN regions, the regions are merged. It means it creates a cluster of similar regions. This process is repeated until the object is located. Finally, it establishes the bounding box on the located image.

b. Feature Extraction: The cropped portion of the identified region is taken as input to the feature extraction. Then, the cropped image is resized to pass through CNN to extract useful features. Here the object is divided into two classes background class and foreground class. For this, IOU is calculated.

$$\textit{If IOU} \geq 0.5\ [\textit{foregroundclass}] \tag{1}$$

$$\textit{IOU} < 0.5\ [\textit{backgroundclass}] \tag{2}$$

c. Classification: Here, the input to the classification is feature representation. Then SVM is used for classification purposes, which is used to predict the label of the object is located. The license plate-detection-system is proposed using RCNN [41]. The proposed license plate detection improves the detection accuracy by the semantics of region proposal techniques (Fig. 2).

4.2. Fast RCNN

In RCNN, the same network worked on the same pixels multiple times. If the image has m*n pixels, it will be given to the same network again and again, due to that performance of RCNN was not so good. To solve this problem, Fast-RCNN was introduced

Table 1
Approaches for Object Detection.

Approaches	Year	Description	DataSet
AlexNet [11]	2012	A CNN based approach that is used for Image Classification.	ImageNet
OverFeat [12]	2013	The multiscale approach is used for classification, localization, and detection.	ILSVRC 2013
R-CNN [13]	2014	Convolutional Neural Network based approach which uses selective search to reduce number of bounding box.	PASCAL VOC 2010-12 and on ILSVRC2013.
M.R. - CNN & S – CNN [14]	2015	A system that detects objects of multiple regions and uses semantic segmentation.	PASCAL VOC2007 and PASCAL VOC2012
DeepID – Net [15]	2015	Detect the object by modelling the deformation of the object. It means that, except detecting the entire object directly, it is also crucial to detect object parts that can then assist in detecting the entire object.	ImageNet
Fast R-CNN [16]	2015	Fast RCNN (Fast Region-based Convolutional Neural-Network) method is used to increase detection accuracy as training and testing depend on a deep convolutional network.	VOC 2007, VOC 2010, VOC 2012
CRAFT [17]	2016	A better approach for object detection where the focus is on reducing background regions so that objects can be detected more accurately.	PASCAL VOC07/12 and ILSVRC.
R-FCN [18]	2016	R-FCN (region-based, fully convolutional-networks) use two-stage for object detection.	PASCAL VOC2007 & 8012, MS – COCO
Faster R-CNN [19]	2016	A single Convolutional Network is used for both region proposal generation and Object Detection tasks.	PASCALVOC 2007, PASCALVOC 2012, MS COCO
Yolo [20]	2016	This approach is used to detect multiple objects, and as it looks at the image only once so, it is called you only look one.	PASCAL VOC 2007, PASCAL VOC 2012
SSD [21]	2016	To detect multiple objects in the image, only one shot is required in SSD.	PASCALVOC, COCO, and ILSVRC
YOLOv2 & YOLO 9000 [22]	2017	It is a modified version of YOLO that has the capability of detecting about 9000 categories.	PASCAL VOC and COCO
YOLOv3 [23]	2018	A Modified version of YOLO that is three-time faster than SSD and as accurate as SSD.	M.S. – COCO Dataset
Objects as Points [24]	2019	Center Point-based approach called which is more straightforward, faster, and more accurate than bounding box based detectors	MS-COCO Dataset
YOLOv4 [25]	2020	A Modified version of YOLO that is fast operating and more accurate object detector,	ImageNet and MS COCO Dataset.
P.P. – YOLO [26]	2020	A Modified version of YOLOv3 is the most effective and efficient object detector in accuracy and speed.	MS COCO Dataset.
EfficientDet [27]	2020	A new family of single-stage object detection which is scalable detection architecture and provides better accuracy and efficiency	MS COCO Dataset

Table 2
List of available Datasets.

Dataset	Description	Instances	Year
Caltech 101 [28]	Pictures of objects.	9146	2004
Caltech-256 [28]	This dataset is an improvement of Caltech-101 Dataset, which contains larger category images.	30,607	2009
CIFAR-10 Dataset [29]	Contain about ten classes of objects.	60,000	2009
CIFAR-100 Dataset [29]	Contain a large number of images used for object classification.	60,000	2009
ImageNet [30]	Vast database that contains more than 14 million images and used for object detection.	14,197,122	2009 (2014)
PASCAL VOC Dataset [31]	This dataset has a large number of images which are used for classification purpose.	500,000	2010
SUN Database [32]	It contains an extensive collection of fully annotated images.	1,30,519	2010
German-Traffic-Sign Detection Benchmark Dataset [33]	A database of German roads' traffic sign images.	900	2013
Berkeley 3-D Object Dataset [34]	This dataset contains a total of 849 images in which 50 different object classes are labelled.	849	2014
(COCO) Microsoft Common Objects-in-Context [35]	An object detection dataset that is used to classify about 80 classes.	2,500,000	2015
Fashion-MNIST [36]	This database contains images of fashion products.	60,000	2017
CINIC-10 Dataset [37]	CINIC-10 is an improvement of CIFAR-10, which is used to fill the benchmark in the ImageNet and CIFAR-10 dataset.	270,000	2018
Open LORIS-Object [38]	(Open LORIS-Object) Lifelong/Continual Robotic-Vision- dataset focuses on improving the familiar objects' continuous learning capability in the home scenario.	1,106,424 RBG-D images	2019
ISLES Dataset [39]	Annotated diffusion-weighted perfusion and diffusion brain MRI	2017	2019
Open Images [40]	An extensive database of images contains images with rich annotations.	9.2 million	2020
ImageNet [4]	Natural images	–	2021
BraTS dataset [6]	Segment necrotic core, peritumoral edema, and enhancing tumor	2020	2022

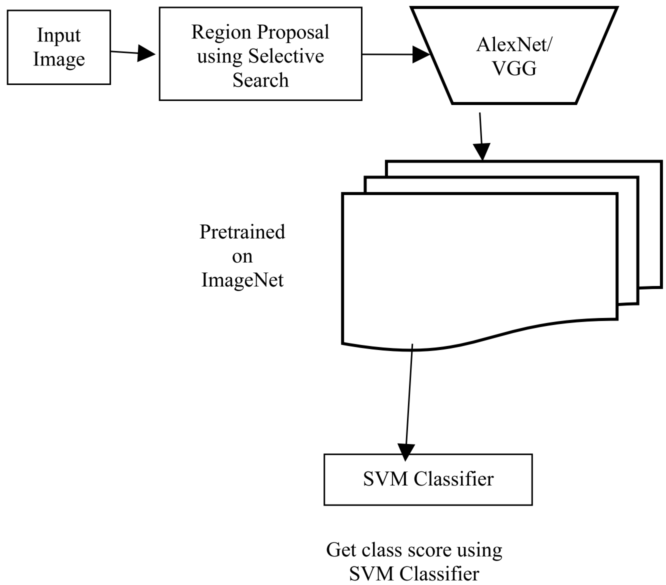

Fig. 2. A general framework of R-CNN.

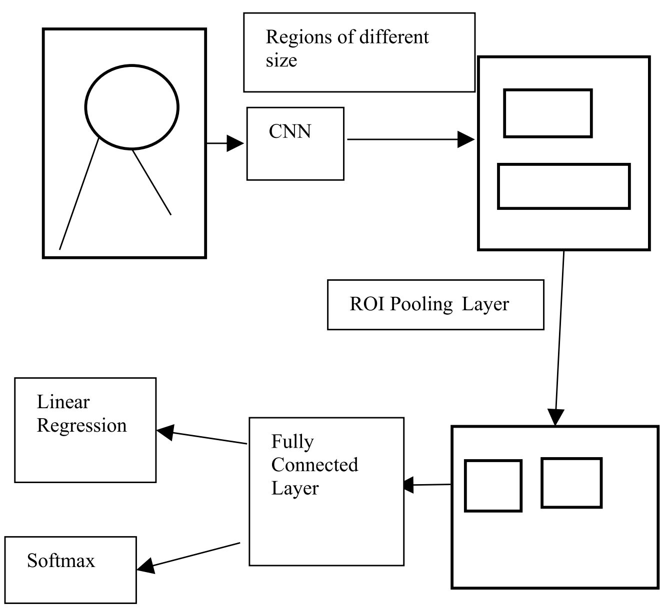

Fig. 3. A General Framework of Fast-RCNN.

[16]. Fast RCNN takes 9.5 hours of training time and 47 seconds of testing time. Detection accuracy is 70.0 (Fig. 3).

How it works:

1. First, an image is given as input to the Convolutional Neural network.
2. The network produces features which are assigned to the Region proposal algorithm.
3. The Region Proposal algorithm produces regions of different sizes. Which feed to the ROI Pooling layer.
4. ROI Pooling layer scale the obtained regions into same size regions, which are given to Fully Connected Layer.

To classify and to locate the portion of an image, output of fully connected layer is feed to linear Regression and Softmax layer.

Advantages:

1. Training and testing time is much lesser than RCNN.
2. Detection accuracy is high.

4.3. Faster R-CNN with RPN (Region-Proposal-Network)

The Region-Proposal Network (RPN) is a network that recognises an item more effectively when coupled with a fully convolutional network [19]. Based on a fully-convolutional neural network, an RPN is utilised to determine object bound and object score at each place. Fast R-CNN, which provides high-quality region recommendations for detection, uses the RPN. The experiment is carried out using the PASCAL VOC2007, 2012, and MS-COCO datasets, which offer greater accuracy with just 300 recommendations per image. R-CNN and RPN, which are faster, won first place in the ILSVRC and COCO2015 contests. The faster R CNN is used to efficiently categorise objects using a deep convolutional network. R-CNN that is faster recognises things more precisely. The system analyses oxygen levels, heart rate, body temperature, and other parameters and transmits the collected data to a central control centre using Wireless Sensor Networks, LoRa, and satellite modems [42].

Faster R CNN works similarly to the RCNN approach. However, in this approach, first, an image is feed as input to CNN, which provides a convolutional feature-map, while in the case of RCNN, Region proposals are fed to CNN. Then the region of proposals is identified from generated Convolutional feature mapping. Then ROI Pooling layer is used to form same-size regions so that it can be provided as input to a fully-connected layer. A softmax-layer is then used to predict the class of the resultant region and the offset values for the bounding box.

How faster R-CNN with RPN works:

1. An input image is taken and feed to Convolutional Network.
2. Convolutional Network generates several bounding boxes called anchor for that image.
3. These anchors are applied to RPN (Region proposal Network), which generates proposed regions of different sizes, but it creates complexity when working on different size features.
4. Now, regions proposed by RPN are applied to ROI Layer as an input, which reduces the features and maps to the same size.

When generating labels for RPN Classification to classify foreground, background, or ignore. For this, the IOU of all the bounding-boxes against all the ground-truth boxes is taken. The IOU is used to label the ROI as foreground, background, and ignore (1,0 and -1 respectively). Then ignoring the -1 label and the other two labels are used to calculate cross-entropy loss.

$$L(\{P_i\},\{t_i\}) = \frac{1}{N_{CLS}}\sum_i L_{CLS}(P_i, P_i^*) + \frac{1}{N_{REG}}\sum_{i=w,x,y,z} P_i^* L_{REG}(t_i, t_i^*) \quad (3)$$

Here
i = index of anchor
P_i = PredictedProbability of anchor i
P_i^* = ground- truth label => 1 if anchor = +Ve
0 if anchor = -Ve
t_i^* = ground truth-box is associated with positive anchor.
L_{CLS} = Classification Loss
L_{REG} = Regression Loss
w, x, y, z = two coordinates of box centre, width and height.

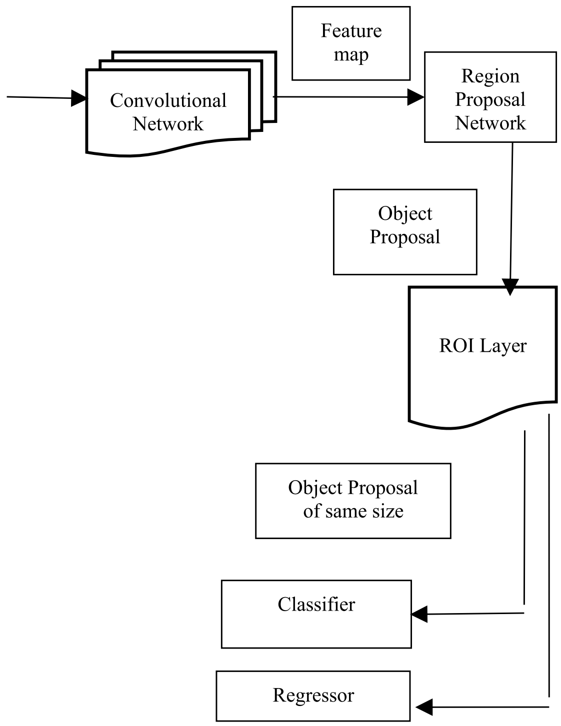

Fig. 4. Flowchart of Faster R-CNN.

5. Finally, regions produced by ROI Layer will be applied to classifiers to classify whether it is object or background; or to regressor to generate a refined bounding box (Fig. 4).

The advantages of Fast R-CNN:

a. Every time, two thousand Region-Proposals are not required to be provided to the (CNN) Convolutional Neural-Network.
b. The convolution-operation is performed only once per-image, and a feature-map is produced from it.

4.4. YOLO (You Only Look Once)

The real-time Object-Detection algorithm YOLO (You Only LookOnce) is used to forecast the object and location of objects in the picture [15]. Yolo is a rapid technique to detection because, as the name indicates, 'You Only Look Once' applies a single neural network to the picture and generates predictions using it. This method looks at the area of the image where the object's likelihood is high. The YOLO method processes pictures at 45 frames per second (FPS). At the same time, Fast YOLO processes images at a rate of 155 frames per second (FPS), while maintaining double the (maP) Mean Average Precision of existing real-time detectors.

How YOLO works

i. First of all, an input image is taken.
ii. The image is divided into S*S grids.
  a. A Bounding box is computed for each grid.
  b. For each grid, a confidence score is determined.

$$Confidence\ score = pred\,(object) * IOU^{truth}_{PRED} \tag{4}$$

Where IOU is the intersection-over-union between the observed box and the actual-box.

$$If\ Confidence\ score\ \geq 0,\ then\ there\ is\ an\ object\ present\ in\ the\ grid \tag{5}$$

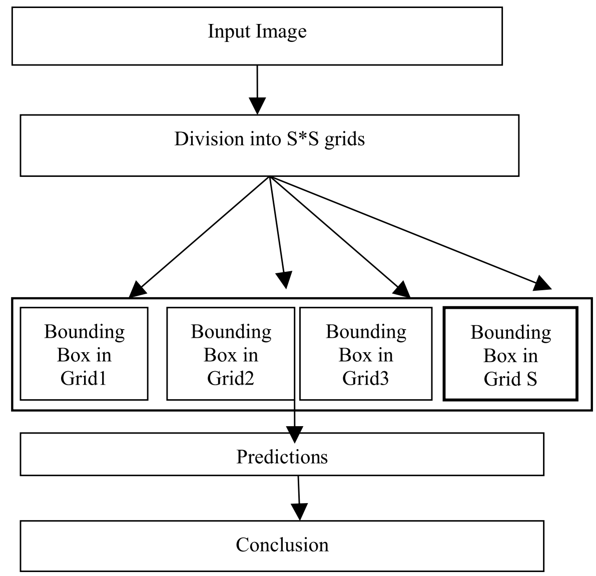

Fig. 5. Flowchart of YOLO.

$$If\ Confidence\ score\ = 0,\ then\ there\ is\ no\ object\ present\ in\ the\ grid \tag{6}$$

iii. For all of the bounding-box, network determines class probability Pr(Classi|Object)
iv. If the class probability of the bounding box > threshold value, then object can be located in an image (Fig. 5).

A model was proposed for Wood pith detection, and the model is based on YOLO [43]. For this purpose, the researchers have taken a dataset of 345 images, which is measured by detection accuracy. The YOLO model was proposed to identify objects and distinguish missing articles [44]. The researchers say that calculation is easy to fabricate and can be prepared legitimately on a total picture.

4.5. SSD- Single Shot MultiBox Detector

To detect multiple objects in the image, only one shot is required in single shot multibox detection (SSD) [21]. While in RPN, two shots (Region-Proposal and Object-Detection) are required. The network can handle various size objects. The Researcher has done experiments on datasets PASCALVOC, COCO, and ILSVRC and provides better accuracy than other methods (Fig. 6). SSD achieves 74.3% mAP (Mean Average Precision) at 59 FPS (Frames per second) for 300*300 inputs and for 512*512 input, SSD (Single Shot Multibox Detector) achieve 76.9% mAP (Mean Average Precision) at 22 FPS (Frames per second).

Loss-Function
Loss-Function in SSD is the summing up of Localization-Loss and Confidence Loss.

$$L\,(p,q,r,s) = 1/N\,(Lconf\,(p,q) + \alpha Lloc\,(p,r,s)) \tag{7}$$

(Where N = Number of Positive matches).
Lconf = Confidence Loss
Lloc = Localization Loss

The Localization loss between the Predicted-box (r) and Ground-truth-box (s) is defined as the smooth L1 Loss with qp, qy as the offset to the default bounding-box d of width w and height h.

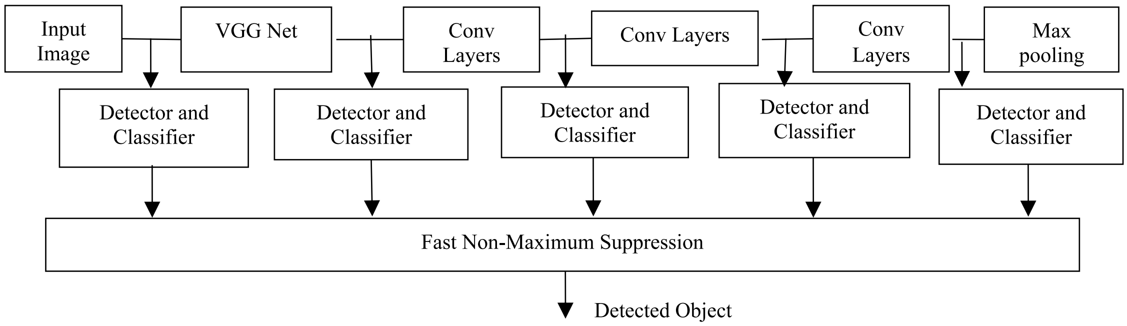

Fig. 6. Framework of SSD.

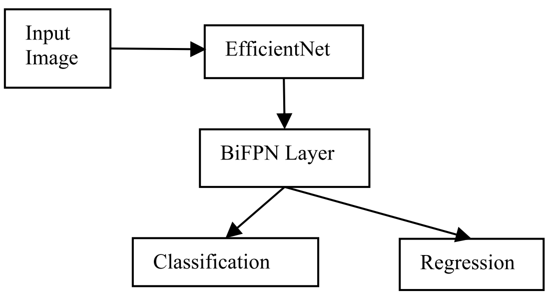

Fig. 7. A general framework of EfficientDet.

$$\mathrm{Lloc}(\mathrm{p},\mathrm{r},\mathrm{s}) = \sum_{i\in Pos}^{N} \sum_{m\in\{qp,qy,w,h\}} p^{\mathrm{k}}{}_{\mathrm{ij}}\mathrm{smooth}_{\mathrm{L1}}(\mathrm{r_i}^{\mathrm{m}} - \mathrm{s_j}^{\mathrm{m}}) \tag{8}$$

$$\mathrm{s_j}^{\mathrm{qp}} = (\mathrm{s_j}^{\mathrm{qp}} - \mathrm{d_i}^{\mathrm{qp}})/\mathrm{d_i}^{\mathrm{w}} \tag{9}$$

$$\mathrm{s_j}^{\mathrm{qy}} = (\mathrm{s_j}^{\mathrm{qy}} - \mathrm{d_i}^{\mathrm{qy}})\mathrm{d_i}^{\mathrm{h}} \tag{10}$$

$$\mathrm{s_j}^{\mathrm{w}} = \log(\mathrm{s_j}^{\mathrm{w}}/\mathrm{d_i}^{\mathrm{w}}) \tag{11}$$

$$\mathrm{s_j}^{\mathrm{h}} = \log(\mathrm{s_j}^{\mathrm{h}}/\mathrm{d_i}^{\mathrm{h}}) \tag{12}$$

$$pij^{x} = \begin{cases} 1, & if\ IOU > 0.5 between\ default\ box\ i\ and \\ & \quad ground\ truth\ box\ j\ on\ class\ x, \\ 0, & otherwise. \end{cases} \tag{13}$$

4.6. EfficientDet

A new family of single-stage object detection is scalable detection architecture and provides better accuracy and efficiency [27]. EfficientDet Object Detection method is based on EfficientNet architecture and EfficientNet architecture is a combination of Bi-FPN and combined scaling. It achieves 55.1 average precision on COCO Dataset (Fig. 7).

Bi-FPN depends on cross-scale connections and weighted feature fusion.

BI-FPN: Cross-scale Connections: Those nodes are removed with one input edge and no feature fusion. An extra edge is added from original input to an output node to fuse more features. The same layer is repeated multiple times to enable more high-level feature fusion.

Bi-FPN: Weighted feature fusion: Multiple input features are first resized to the same resolution and then sum them up as different input features are at different resolutions, so they contribute to output feature unequally. For each input, additional weight is added so that network can learn the importance of each input feature.

Compound scaling is applied to scale up the dimension of the network (Table 3).

From this research paper, new researchers can easily differentiate between different object Detection techniques in terms of their methodology adopted.

A network called Region-based Convolutional Neural Network, which is based on Convolutional Neural Network, was introduced to recognize objects more correctly. If the picture has m*n pixels, it will be fed into the same network again, resulting in poor RCNN performance. At each location, an RPN is used to calculate item bound and object score. The RPN is used by Fast R-CNN, which gives high-quality region suggestions for detection. The detection accuracy is excellent. In the RCNN technique, an image is supplied as input to CNN, which outputs a convolutional feature-map, but in the RCNN approach, Region suggestions are fed to CNN. The region of suggestions is then determined using the resulting Convolutional feature mapping. The Convolutional Network creates a number of bounding boxes.

5. Conclusion

In this run-of-the-mill life, brain stroke is becoming a major concern. Manual diagnosis takes more time and more prone to errors which may be disastrous for the patients as they need immediate results. In this technology-driven era, automation in detecting disease can give efficient and accurate results. MRI images play an important role in detecting strokes. Machine learning algorithms have been in use to evaluate medical data-set for decades. These algorithms provide less efficient and time-consuming results. Object detection using deep learning algorithms is more efficient in detecting strokes from MRI images. In the proposed paper, various object detection methodologies such as (YOLO) You only look once, (RCNN) region-based convolutional network, (SSD) Single shot multibox detector, etc., are studied. A comparison chart of all these methods is also presented. In this chart, it can be seen that the SSD methods have the highest mean average precision among all the ways. The computation time of SSD is less among all the methods. But in these methodologies, it is reviewed that when accuracy is high, it is taking more time to detect objects. So, it is concluded that there should be a balance between speed and accuracy.

Human and animal rights

The authors declare that the work described has not involved experimentation on humans or animals.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Table 3
Comparison of Various Object Detection techniques.

Method	RCNN [13]	Fast RCNN [16]	Faster RCNN with RPN [19]	YOLO(You Only Look-Once) [20]	SSD [18]	EfficientDet [27]
Advantages	For object localization and recognition, R-CNN is a Simple and straightforward approach.	A single pass is used to process images instead of 2000 region proposals, so the speed and accuracy are higher than R-CNN.	Because the convolution -operation is performed only once-per image, and a feature map is produced from it, so the Faster RCNN approach is more suitable from RCNN	In this approach, regions are not used for localization, but this method looks at the portion of the image where the object is present.	Only one shot is required to detect multiple objects in the images.	EfficienDet is a new Single Stage detector that provides better accuracy and efficiency.
Performance	66.0 Mean Average Precision	66.9 maP	73.2 maP	63.4 maP	74.3 and 76.8 maP	55.1 average Precision
Speed	0.03 frames per second	0.5 frames per second	7 frames per second	45 frame per second	46 and 19 Frame per second	410B Flops

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

References

[1] A. Andreopoulos, J.K. Tsotsos, 50 years of object recognition: directions forward, Comput. Vis. Image Underst. 117 (8) (2013) 827–891, https://doi.org/10.1016/j.cviu.2013.04.005.
[2] B.V. Baheti, S.N. Talbar, S.S. Gajre, A training-free approach for generic object detection, IETE J. Res. 68 (1) (2019) 482–495, https://doi.org/10.1080/03772063.2019.1611491.
[3] Report on stroke prevention and treatment in China Writing Group, Brief Report on Stroke Prevention and Treatment in China, 2019, Chin. J. Cerebrovascular Dis. 17 (5) (2020) 272–281.
[4] S. Zhang, S. Xu, L. Tan, H. Wang, J. Meng, Stroke lesion detection and analysis in MRI images based on deep learning, J. Healthc. Eng. 2021 (2021).
[5] A.R. Pathak, M. Pandey, S. Rautaray, Application of deep learning for object detection, Proc. Comput. Sci. 132 (Iccids) (2018), https://doi.org/10.1016/j.procs.2018.05.144.
[6] R.M. Kronberg, D. Meskelevicius, M. Sabel, M. Kollmann, C. Rubbert, I. Fischer, Optimal acquisition sequence for AI-assisted brain tumor segmentation under the constraint of largest information gain per additional MRI sequence, Neurosci. Inform. 100053 (2022).
[7] A. Sarkar, A. Singh, R. Chakraborty, A deep learning-based comparative study to track mental depression from EEG data, Neurosci. Inform. 100039 (2022).
[8] R. Kharouba, P. Gavriliuc, N.E. Yaghmour, J.M. Gomori, J.E. Cohen, R.R. Leker, Number of stentriever passes and outcome after thrombectomy in stroke, J. Neuroradiol. 46 (5) (2019) 327–330.
[9] C. Andica, A. Hagiwara, M. Hori, K. Kamagata, S. Koshino, et al., Review of synthetic MRI in pediatric brains: basic principle of MR quantification, its features, clinical applications, and limitations, J. Neuroradiol. 46 (4) (2019) 268–275.
[10] S. Zhang, S. Xu, L. Tan, H. Wang, J. Meng, Stroke lesion detection and analysis in MRI images based on deep learning, J. Healthc. Eng. (2021) 1–9, https://doi.org/10.1155/2021/5524769.
[11] A. Krizhevsky, G.E. Hinton, ImageNet classification with deep convolutional neural networks, pp. 1–9, https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf, 2012.
[12] P. Sermanet, et al., Overfeat: integrated recognition, localization and detection using convolutional networks, in: 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, Banff, Canada, https://arxiv.org/abs/1312.6229v4.
[13] R. Girshick, et al., Rich feature hierarchies for accurate object detection and semantic segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2014) 580–587, https://doi.org/10.1109/CVPR.2014.81.
[14] S. Gidaris, N. Komodakis, Object detection via a multi-region and semantic segmentation-aware U model, in: Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 2015, pp. 1134–1142.
[15] W. Ouyang, et al., DeepID-Net: deformable deep convolutional neural networks for object detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June, 2015, pp. 2403–2412.
[16] R. Girshick, Fast R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2015 International Conference on Computer Vision, ICCV 2015, 2015, pp. 1440–1448.
[17] B. Yang, et al., CRAFT objects from images, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, pp. 6043–6051.
[18] J. Dai, et al., R-FCN: object detection via region-based fully convolutional networks, Adv. Neural Inf. Process. Syst. (2016) 379–387.
[19] S. Ren, K. He, R. Girshick, Faster R-CNN: towards real-time object Ddtection with region proposal networks, pp. 1–9, https://arxiv.org/abs/1506.01497v3, 2016.
[20] J. Redmon, S. Divvala, R. Girshick, A. Farhadi You, Only Look Once: Unified, Real-Time Object Detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[21] W. Liu, et al., SSD: Single Shot Multibox Detector, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905, 2016, pp. 21–37.
[22] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 6517–6525.
[23] J. Redmon, A. Farhadi, YOLOv3: an incremental improvement, https://doi.org/10.48550/arXiv.1804.02767, 2018.
[24] X. Zhou, D. Wang, P. Krähenbühl, Objects as points, https://arxiv.org/abs/1904.07850v2, 2019.
[25] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, YOLOv4: optimal speed and accuracy of object detection, https://doi.org/10.48550/arXiv.2004.10934, 2020.
[26] W. Kurdthongmee, K. Suwannarat, Locating wood pith in a wood stem cross-sectional image using YOLO object detection, in: Proceedings - 2019 International Conference on Technologies and Applications of Artificial Intelligence, TAAI 2019, 2019.
[27] M. Tan, R. Pang, Q.V. Le, EfficientDet: scalable and efficient object detection, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2020) 10778–10787, https://doi.org/10.1109/CVPR42600.2020.01079.
[28] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in: CVPR 2004, Workshop on Generative-Model Based Vision, pp. 1–9, https://doi.org/10.1016/j.cviu.2005.09.012.
[29] A. Krizhevsky, Learning Multiple Layers of Features from Tiny Images, https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf, 2009.
[30] J. Deng, et al., ImageNet: a large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, May 2014, pp. 248–255.
[31] M. Everingham, et al., The Pascal Visual Object Classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338, https://doi.org/10.1007/s11263-009-0275-4.
[32] J. Xiao, et al., SUN database: large-scale scene recognition from abbey to zoo, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2010) 3485–3492, https://doi.org/10.1109/CVPR.2010.5539970.
[33] S. Houben, et al., Detection of traffic signs in real-world images: the German traffic sign detection benchmark, in: Proceedings of the International Joint Conference on Neural Networks, May 2014.
[34] A. Janoch, et al., A category-level 3-D object dataset: putting the Kinect to work, in: Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 1168–1174.

[35] T.Y. Lin, et al., Microsoft COCO: Common Objects in Context, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8693 (PART 5), 2014, pp. 740–755.
[36] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, pp. 1–6, available at http://arxiv.org/abs/1708.07747, 2017.
[37] L.N. Darlow, et al., CINIC-10 is not ImageNet or CIFAR-10, available at http://arxiv.org/abs/1810.03505, 2018.
[38] Q. She, et al., OpenLORIS-object: a robotic vision dataset and benchmark for lifelong deep learning, available at http://arxiv.org/abs/1911.06487, 2019.
[39] C.U. Perez Malla, M.D.C. Valdes Hernandez, M.F. Rachmadi, T. Komura, Evaluation of enhanced learning techniques for segmenting ischaemic stroke lesions in brain magnetic resonance perfusion images using a convolutional neural network scheme, Front. Neuroinform. 13 (2019) 33.
[40] A. Kuznetsova, et al., The Open Images dataset V4: unified image classification, object detection, and visual relationship detection at scale, Int. J. Comput. Vis. 128 (7) (2020) 1956–1981, https://doi.org/10.1007/s11263-020-01316-z.
[41] M.A. Rafique, W. Pedrycz, M. Jeon, Vehicle license plate detection using region-based convolutional neural networks, Soft Comput. 22 (19) (2018) 6429–6440, https://doi.org/10.1007/s00500-017-2696-2.
[42] R.K. Garg, J. Bhola, S.K. Soni, Healthcare monitoring of mountaineers by low power Wireless Sensor Networks, Inform. Med. Unlocked 27 (2021) 100775.
[43] X. Long, et al., PP-YOLO, An effective and efficient implementation of object detector, 2020, arXiv.
[44] M. R, et al., Enhanced missing object detection system using YOLO, 2020, pp. 1407–1411.