ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse611032021.pdf
https://doi.org/10.30534/ijatcse/2021/611032021

An Analysis on Object Recognition Using Convolutional Neural Networks

Mamoona Saleem[1], Salman Afsar[2], Ahmed Mateen[3], Arslan Zaheer[4], Muhammad Tariq[5], Muhammad Asim Raza[6]
[1, 2,3,5,6]Department of Computer Science, University of Agriculture Faisalabad.
[4]National University of Computer and Emerging Sciences, Faisalabad
Corresponding Author Email: ahmedmatin@hotmail.com

ABSTRACT

The global development and progress in scientific paraphernalia and technology is the fundamental reason for the rapid increase in the data volume. Several significant techniques have been introduced for image processing and object detection owing to this advancement. The promising features and transfer learning of Convolutional Neural Network (CNN) have gained much attention around the globe by researchers as well as computer vision society, as a result of which, several remarkable breakthroughs were achieved. This paper comprehensively reviews the data classification, history as well as architecture of CNN and well-known techniques by their boons and absurdities. Finally, a discussion for implementation of CNN over object detection for effectual results based on their critical analysis and performances is presented.

Key words: Neural Networks, Object Recognition, Convolutional

1. INTRODUCTION

The volume of image data has been highly increased with the rapid advancements in mobile internet and social media, as human beings cannot process efficiently such large volume of data. So, it is expected to handle such tasks automatically with the aid of automated process. With the better understanding of image processing technology, extensive recognition of image and exact identification the object target of the image become more and more significant [1]. The people are widely concerned about the classification of images along with obtaining the semantic object category and image location [2], that's the reason that object detection technology has taken wide attraction globally [3]. Object detection technology tends towards the detection of target objects with the theoretical concepts and methods of pattern recognition and image processing, concluding the semantic group objects, and mark targeted position of target object in image [4].

It's a quite challenging taskto recognize image using computer technology automatically. Noise disturbance, complex background, low resolution, attitude and scale changes and other factors impacting the object detection performance seriously. The conventional methods for object detection were not as robust as convolutional neural networks, as it is not robust to illumination change, thus lacking generalization abilities. Object recognition was quite in-efficient during 2010-2012 in PASCAL VOC challenge[5], with small gains through the establishment of collection systems and enlists variants using traditional methods. Due to this reason, a variety of techniques proposed to elevate the performance of recognition object. Convolutional neural network (CNN) among they successful model about deep learning, having strong ability of sequential learning features, and the recent research proves that the feature extraction done by CNN has a stronger and reliable ability of discrimination and generalization than hand-crafted features.

The CNN has attained quite a great success in many areas of computer vision. An impressive outcome on the ImageNet by joining the dataset of Large-Scale Visual Recognition Challenge (ILSVRC) along with Le Cun's technique and finally the last fine-tuning techniques to obtain good learning. The obtained outcome popularised CNN as it achieved the error rate of 15 percent vs. 26 percent of traditional methodology which is overwhelming contribution to the growth of efficient object finding techniques[6].

In 2014, Zisserman and Simonyan [7] find impact of changing with depth of an CNN on localization as well as organization accuracy in ImageNet challenge, that recovers then futuristic by utilizingCNN's layers deeper in 16 and 19. A architecture of 16-layer CNN contains five layers of pooling (2x2 neighborhood-max-pooling), fully-connected of three layers and 13 convolutional layer's (with 3x3 liters). Concealed coatings contain rectified (ReLu) activations. A Fully-connected layers scales channels 4096 to SoftMax 1000 outputs and which can be systemizewith the help of dropout.

In 2016 conqueror about the object finding group in the ImageNet objective challenge the also based on CNN. This technique has used a amalgamation of CRAFT region scheme generation [8], CNN gated bi-directional[9], breakthrough generation, as well as assembling the clustering.This work has been used so far for the object detection and recognition purposes?

2. RESEARCH METHOD

The main progress after division is furthermore extraction, representation and the recognition of human actions is unbelievable. When the highlights may contain the data on existence, they are space time volumes (STV) at that point and when it is possible for us to include discreet Fourier changes (DFT) picture contours which spatially contain the image power variant. For the extraction of highlights, STV and DFT are all pictured, though they feel distressed. Since

the nearby highlights are stronger against clamour and obstruction.

2.1 Identification of the need for review

In this phase, we look for existing SLRs on Object recognition from Image using convolutional neural networks. In recent times, many papers are published on Object recognition in many disciplines (Image using convolutional neural networks Based Techniques). This thing shows that the Image using convolutional neural networks is becoming popular in upcoming object recognition based on CNN in this table 1.

Table 1: Electronic Databases

Identifier	Database	URL
ED1	IEEE	http://ieeexplore.ieee.org/
ED2	ACM	http://dl.acm.org/
ED3	Science Direct	http://sciencedirect.com/
ED4	Springer Link	http://link.springer.com/

2.2 Research Questions

In this review paper, classification of the available image processing techniques has been elaborated and also detailed literature review of object detection feature of image processing has been done (for review of using convolutional neural networks). There are a couple of important questions which will be answered in this review paper. Each inquiry has appropriate inspiration to demonstrate the need behind the inquiry in this SLR is these inquiries feature of Object recognition from Image using convolutional neural networks.

1) What are the standard image classification techniques?
2) What are other artificial architectures neural network has been used so far for the object detection and recognition purposes?
3) Why convolutional neural networks are preferred over other artificial neural network architectures?
4) What is the basic structure of convolutional neural networks (CNN)?
5) What are the advantages, dis-advantages, variations and solutions to previous work proposed in each method of convolutional neural networks?
6) Last, what is the best convolutional neural network architecture so far considering performance as benchmark?

To facilitate an organized review about convolutional neural networks of image processing, its necessary about develop a better understanding of hierarchical classification techniques which has been described in section 3 below. On the basis of this taxonomy, literature review of artificial neural networks, based on object detection, is elaborated in section 4 , the major purpose is to provide a broad overview of recent work done including the advances and defects in each method which related to artificial neural networks. In next Section 5, detailed review about the convolutional neural network has done. Comparison at various convolutional neural network methods has done about conclude which is the best approach so far.

2.3 Data Sources

Four electronic databases that are given in Table 1 are considered as the primary data sources for the extraction of relevant studies. Moreover, Google scholar had been considered as the source for external studies. Alt-though, results from Google scholar contains duplicate's that already had been extracted from the other four electronic databases. So, after removing those du-plicate results only unique results are considered for primary studies.

Search Terms

The given search terms had been extracted from the major terms given which are given in relevant literature studies and primary question. To find relevant outcomes from the electronic databases listed in Table 1, following search terms were defined. The PICO format of the search terms has been shown in which defines proper categories of the search terms. Given search terms are combined using the conjunction *(AND)* and disjunction *(OR)* operators. So, after combing the search terms an automated search string has been engendered which is given below:

((("Document Title": " machine learning *" OR ", artificial neural networks" OR" regional proposal") AND ("*Object recognition*" OR "convolutional neural networks ") AND ("Document Title": "Computer vision "OR "object Detection")

In addition, some other keywords had been used to filter out the results from databases; these keywords contain broadcasting, aspects, feature, and features. For extraction of articles from Google Scholar, the search string used is given below:

"Object recognition from Image using convolutional neural networks" This search string is reflected as a generic search string and results from the first four pages of Google scholar has been extracted which contains 20 results for the external category.

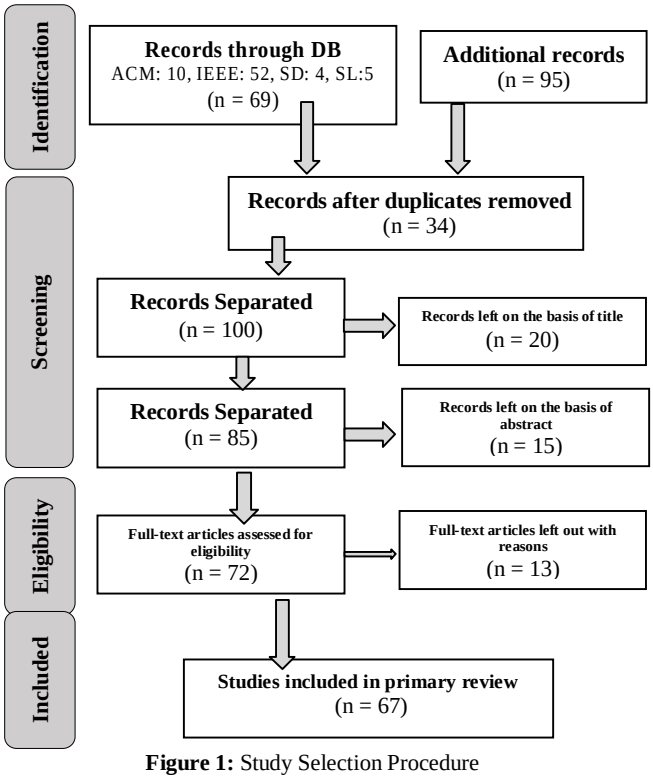

Figure 1: Study Selection Procedure

2.4 Study Selection Procedure

The study selection procedure of this systematic collected works review is visualized in Figure 1. This process of study selection consists of 3 phases. Each of these phases is defined below:

Identification

In this phase, results had been extracted from the electronica databases which are listed in Table 1. These results contain duplicates, as they are the straight-ahead outcomes that are extracted from the search string.

Screening

In this stage, three sorts of filtration happen on the separated outcomes from above stage. Right off the bat, copies of the outcomes are expelled that are extricated from each electronic database. After this, results channel based on title. In this progression, the title of the considerable number of results are analyzed and just those outcomes are incorporated which have a title significant to the point of SLR.

After these steps, only those results are extracted for the full study that has their title as well as abstract relevancy with the topic of our SLR.

Eligibility Control

In this stage, chose considers from the above stage had been full content prepared and just significant articles had been chosen for essential investigation and further handling of our SLR. A legitimate clarification of incorporation and rejection criteria is characterized in the segment underneath which unequivocally clarifies the criteria and study determination technique based on these criteria.

2.5Inclusion & Exclusion Criteria

The criteria of consideration and prohibition are utilized to control the qualification for the choice of essential examination from full-content perusing of articles that had been chosen after the screening stage. It has been seen that just such articles are chosen for the essential examination that has legitimate approvals of procedures which are characterized for extraction of proposal results. Just such Articles that are recovered from companion checked on diaries and gatherings had been chosen in the last investigation. For the choice of an article, it must be essential that it ought to incorporate curiosity of procedures and different strategies. Articles from 1988 to 2018 will be incorporated for study. A few articles that give suggestion procedures of mixed media things and tunes and so forth can be incorporated into the essential investigation. Also, chief select just such articles which give proposal procedures dependent on Object recognition from Image using convolutional neural networks.

2.6Image Processing Algorithms Classification

Conventional Schemes such as the Parzen windows and Bayesian discriminant based on arithmetical pattern recognition were famous till early 1990s. Seeing that, ANNs gained plenty attention and are notably used as a substitute for clustering techniques and classical pattern recognizing. Any other attractive trainable machine for object recognition is non-parametric feed-forward ANNs, which can discriminate several textures if provided with triumphing features. Presently, the usage of ANNs in picture processing will increase the aforesaid conventional applications. To address with low-level image processing duties along with image development and noise suppression a part of feed-forward ANNs and SOMs have they used so far. In order to deal with image processing of the low level Hopfield ANNs had been provided as a tool for finding out an appropriate method *to* deal with complicated (NP-whole) optimization glitches. therefore, they come to be the correct options to conventional optimization the image processing algorithms and they could be formulated as optimization issues.

Distinct issues addressed within the subject of digital image processing could be described into what we have chosen about the name of image processing chain (see Fig. 2).

1) Preprocessing. It has been done initially before applying any other image processing operation. The reason behind is scale-up or scale-down the image as per requirements.

2) Data reduction & feature extraction. Extracting them specific portion or component from an image as per requirements is called data reduction or feature extraction. Extracted features usually have fewer pixels as compare to the original image.

3) Segmentation. Division of image into several regions that are connected with each other on the basis of some specific criteria. Example is the image operation applied on textures to produce some

*4) Object detection and recognition.*Observing the exact location i.e. orientation, position, and scales the object within image.

5) Image understanding. It can addresses about the specific arrangement of object and its in-depth analysis.

Optimization strategies aren't seen as an isolated step in the whole process. however, it can be considered as a set of strategies, which help the other steps in following figure 2.

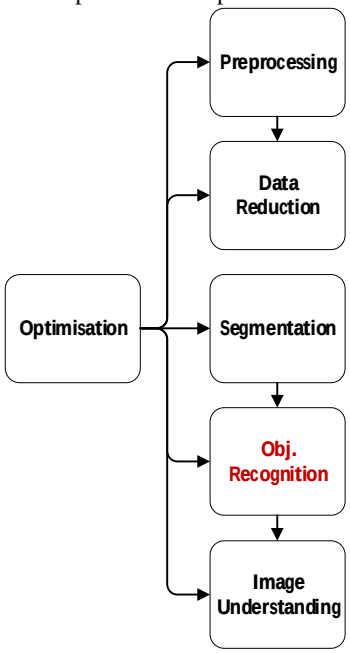

Figure 2: Architectural process diagram.

In figure 3, besides, task achieved using algorithm, its can be processing competencies are in part regulate via the distraction degree about the enter records. Following abstraction stages has been distinguished based on numbering.

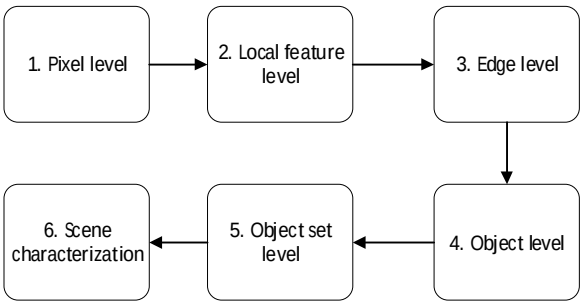

Figure 3: Abstraction Level of Image Processing Chain

Image Processing Based on Neural Networks

In this section, we will be review various object detection systematic techniques strut on artificial neural networks before moving to convolutional neural networks (CNN)[10] that contains the review of feature based other artificial neural networks too for object detection.

Object recognition

Object recognition is among the important issues based on computer vision and is particularly a complicated problem to accomplish. In lots of respects, object recognition is pretty a similar other to computer vision functionality, it as consists of creating an package that's forms a regular behavior to deformation and adjustments in the perspective in addition to lighting fixtures. A prime factor that makes object recognition a distinct issue is it includes each and classifying and finding regions of an image [11].

Through late 2000s, predominant answers for object recognition is to make use of feature descriptors, which includes scale-invariant function transform (SIFT) [12]which evolved through David Lowe(1999) along with histograms orientated gradients (HOG) [13]which became popularized by 2005. By 2010s, technology has been moved towards the usage of convolutional neural networks[14].

Pixel-based object detection

Several ANN methods have been introduced for object recognition based on pixel data. Strategies using the weight sharing algorithms[15], the recurrent networks developed in 1980, the ART network is the contribution of Grossberg, mixture-of-experts used to divide the space problem into homogeneous regions, fuzzy ANNs consists of fuzzy logics and ANNs method, bi-directional auto-associative memories (BAM) is a sub-class of recurrent neural network which was brought by the Neocognitron is a hierarchical, multilayered ANN delivered with the aid of piecewise-linear neural classifiers based at the Kohonen learning vector quantization (LQV2) , higher-order ANNs and Hopfield ANNs. A hardware has been designed to cope with object detection problem based on ANNs: the RAM network suited particularly to implement WSI and the optical implementations. ultimately, Self-Organizing Feature Maps (SoMs) being used to extract features from pixel-based data.

Various architectures of network have been introduced to deal with the object variations position and orientation. An approach has been taken that is proportional to illumination change. In order to gain appropriate results and for better classification, the distinctionis required between invariant recognition in 2D&3D images. A novel approach has been introduced for object recognition knows as what-and-where filter that was proportional to the 2D translations, scale, and in-plane rotational. Its can be combination of what (filter bank) along the other with where (an invariant module). Several other methodologies need to learn via explicitly training. Objects statistical intensity model was built by Egmont-Peterson and Arts. A two-stage ANN approach so far for recognition of nodules chest radiographs was developed by Penedo et al. Nodule sub-images were partly used to trained these ANNs.

Rare cases like object detection occlusion or the existence of more than one objects inside the image are processed by a classifier that is rarely been taken into consideration. McQuiod developed an experimental structure which is capable for multiple object detection simultaneously within an image[9].

Convolutional Neural Network

Before starting discussion on various aspects of convolutional neural networks, it is necessary to mention here other artificial neural-network approaches based on feature-based object detection including: feed-forward ANNs, Hopfield ANNs, a fuzzy-ANN and RAM-based ANNs. SOMs are often executedin order to perform feature extraction before object recognition, even if SOMs are trained to perform object classification.

Compared to the pixel-based approach described in the previous section, neural architectures are developed on a smaller scale for feature-based object recognition. It reflects the point that foremost focus to develop or select them most optimal structures for them recognition tasks. Various feature-based approaches have a common pathway that discrepancies in rotation and scale are crumbled to the structures, e.g., statistical moments. It is also noted that a certain measure of noise always influences the computed structures, as a result of which the recognition performance is deteriorated. Therefore, the subsequent classifier performs the major task of filtering out the noise and distortions due to these features. Additionally, whenever a large object is to be detected and densely sampled, feature extraction should be performed. Contrarily, a neural classifier is comprised of so many parameters that a qualified generalization will be obstructed.

The major issue in solving the computer-vision, image processing-based issues using a conventional neural network is that a usual normal image comprises of a large quantity of information. A monochrome low-resolution image i.e. (620x480) contains about 297-600 pixels. A general assumption leads to a conclusion that if every pixel of this picture is allocated a separate inter-linked, 297-600 weights required for each neuron. Must need 2,073,600 weights for a full HD image (1920x1080) andIf the pictures provided are polychrome, then quantity weights about increases according to them color channels (typically three). Thus, it can be seen that the general sum of free factors can be in the

link rapidly turn into particularly enormous as the doppelgänger dimensions' rise. Moreover, big data effects overfitting and in-efficient enactment.

Moreover, many patterns finding methods involves that explanation is translationally constant. It's can wasteful for the separately train of the neurons observing that the similar systematic technique can be the right-bottom corner and left-top crook of an image. A fully-connected neural network failed to notice this structure and therefore failed in this case, thus in such scenarios.

Generalized Architectural Overview

As CNN's are feedforward networks, so the information flow only occurs in one direction, from input to output. CNN's are biologically inspired just like an Artificial Neural Networks (ANN). Visual Cortex in brain comprises of alternating layers of simple and complex cells (Hubel & Wiesel), motivates their architecture. CNN came in various variations but in general, it consists of convolutional and pooling (or subsampling) layers, which are grouped into components. These connected layers follow these components. Components are often bonded together at the top of each other, thus forming a deep model in figure 4.

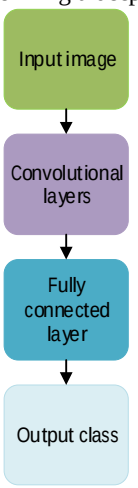

Figure 4:CNN Image Classification Pipeline [16]

Convolutional Layer

This layer acts as a feature extractor, themain responsibility is to learn the feature representations of their input images. Convolutional layers comprise of neurons which are arranged in feature maps having a sympathetic field, which can be connected to the neighborhood about the neurons of the backward layers through the set of trainable weights. Inputs are convolution using learned about the weights of the compute a featured map, obtained convolved results sent via non-linear activation. Neurons inside a feature map have weights that constrained to be equal; but, diverse article maps inside the similar layers may have different weights In order to compute the *k*th output feature map,$Y_k$can be computed as:

$$Y_k = f\,(Wk * x) \qquad (1)$$

Where x is the input image; Convolutional filter belongs to kth feature map is represented by W_k. the * sign used to calculate the product of the filter model at each point of the input image.

Pooling Layers

The main responsibility of pooling layers is to decrease the spatial resolution of feature maps. Basically, network deep layers require less information about features exact spatial locations, at the same time more filter matrices are requiredin order to recognize multiple high-level patterns. The depth of data volume can be increased by decreasing the height and width of data-volume that could be helpful in keeping the computation time at a reasonable level.

Data volume size can be reduced by adding a pooling layer after a convolutional layer. The layer down-samples the activation maps. The main issue with pooling layer that it may destroy the information about spatial relationships b/w subparts of patterns. They can do by adding a max pooling layer after infinite and convolutional layer [17].

Fully Connected Layers

Multiple pooling and convolutional layers are bonded together at the top of each order in the other to extract more abstract article moving concluded the network. Every neuron in one layer connects to another neuron in each layers using full connected layer phenomenon . It is in standard of the same as the outmoded multi-layer perceptron neural network.

Training

Artificial Neural Networks uses learning algorithms to adjust their free parameters (i.e., the biases and weights) in order to get desired network output. Commonly used algorithm for this purpose is backpropagation [18]. It proposes the solution to iterate weights to reach at constant point. In old version, gradient descent method has been used for optimization purpose. However, in modern literature, gradient descent is undoubtedly a time-consuming technique and found unreliable to for minimization of errors.

R-CNN

In this section, we will study various methods that joinCNN's with regional proposal classification along with how are they generated [19] (also called as Region of interest RoI)

Overview/Description

R-CNN forward computing comprises of various stages, as shown in fig. 5. After taking images as the input, at first, the region of interest (RoI) has been generated [20]. Generated RoIs are category-independent bounding boxes and they have a high ratio of comprising the targeted object. A distinct method called Selective Search has been used for making the RoIs (see a reference for appropriate data).

It has been used to determine features from every region proposal also called Region of Interest (RoI). The sub-image restricted inside the bounding-box has been used toward parallel the input size of the image in CNN and then passed further near the network. Once network can be successfully extracted structures from the input, the structures are then passed to support vector machines (SVM) thats gives them final classification of vectors.

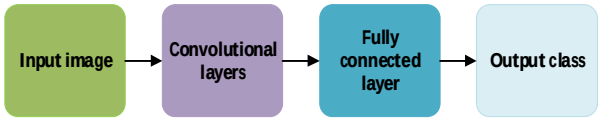

Figure 5: Stages of R-CNN Forward Computation.

In the above figure 5 this design has been prepared in multiple phases, begins off advanced with the convolution network. When CNN learning has been finished, help to vector machines (SVM) are then passed further to CNN features. Eventually, the learning procedure begins for region proposal.

Issues

R-CNN is quite a significant architecture because it contributed toward provision of a first working solution for object recognition using CNN's. As R-CNN is among the first one, it has a number of defects that have been addressed by later research.

In 2015, Girshick pointed out 3 main problems of R-CNN in his paper:

- The training comprises of couple of steps, as shown in fig. 5
- Secondly, the training takes quite a long time which is making it quite expensive for practical operational work. Also, for both Support Vector Machine as well as region proposal training, features have been taken and stored on disk from each region proposal. A long period of computation is required along quite a large number of disk space.
- Last one and the quite important, processing of object detection algorithm is quite slow and time consuming for the image even using the GPU. Forward computing is one of the reason along with only one Regional Proposal can be generated at a time. In case of multiple RoIs, system is not capable of handling them concurrently thus due to which an overlapping is occurred in results, which make the method non-reliable and in-efficient.

Fast R-CNN

This architecture was, moving toward more practicality, this architectural method for object recognition focused towards performing the forward permit about the CNN consists of complete image, as a substitute for the isolation Regional Proposal.

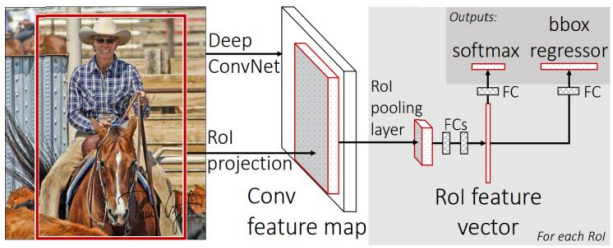

Figure 6: (Fast) Region-based Convolutional Networks (R-CNN).

General description

Figure 6 explains the basic architecture of fast R-CNN. This (Fast R-CNN) technique receipts entire image by way of an input along with identification of Region of Interest (RoI). Image has been processing using various layers comprises of convolutional layer as well as max pooling layer in order to generate a convolutional article map. Next step is to extract an article vector from each Region of Interest (RoI) using article map. Then, feature vector will be provided as an input to the fully-connected layers, that further acts as an input to further two output layers, that aresoftmax layer, responsible for producing probability estimation and the later known as bounding box (responsible for refinements of initial candidate boxes)[21].

Classification evaluation

As described via the authors, fast R-CNN takes quite a short span of time in-order to classify an image in comparison to conventional R-CNN, nearly takes not more than a second time using state-of-art Graphical Processing Unit. The motive at the back of is because each RoI has used same feature map.

As recognition time decreases, overall computation time starts depending on the response time of region proposal generation method. Due to this, Region of Interest production forms a computational barrier. In case of multiple RoIs, the evaluation of time span consumes on the convolutional layer dominates the period expended estimating the fully-connected layer. Time consumption canister be decreased by compressing the connected layers implementing truncated singular value decomposition but the drawback is it will result into a bit of issue in accuracy but overall performance gain with respect to time will be more then 30 % of original time.

Training

As described [16], Fast R-CNN is quite more efficient to use for training purposes rather than R-CNN, through almost nine-fold decrease popular training time-span. The can be whole network (including the fully connected layers and RoI layer) could be trained by using the two algorithms that are back-propagation and stochastic ascent descent. Usually, pre-trained algorithm can be chosen as a starting opinion and at that point refined to next-level. Mini-batches of N-images is used for training. Each mini-batch image is responsible for providing a sample to R/N RoIs. If the connexion over amalgamation with the. ground-truth container is above 0.5, then RoI samples are consigned to the class. Beside this, left over RoI belongs to the background class.

As per classification, computational and memory usage have been shared for the RoI from the same image. The original image has been tossed horizontally with a probability about 0.5 for data-augmentation. Using a multi-task loss function, the bounding box regressors and the soft-max classifiers are fine-tuned together, both are considered to be the accurate class of the experimented RoI and offset about sample bounding container from accurate bounding box.

3. MAIN RESULTS

They can be main Three results support the Fast R-CNN contributions:

- State-of-the-art mAP on VOC07, 2010, and 2012
- Fast training and testing matched to R-CNN, SPPnet

- Fine-tuning convolution layers in VGG16 improves mAP

3.1 Faster R-CNN

A Faster R-CNN network is presented and trained for RoI production and its detection. Primarily, training has been done separately on networks. Afterward, the next step consists of the combining and refining of networks. While refining process, certain layers are kept fixed due to which many layers get trained in return.

A single image is provided as an input to the trained network. Feature maps generated from the image using shared fully convolutional layers. Region Proposal Network(RPN) taking feature maps as an input. RPN is considered as a "weak" detector, only responsible for detection Whether there is an object and generate a bounding box proposal Object. But in fact, the convolutional layer is "powerful" enough to detect, locate, and classify objects simultaneously [22]. RPN then produces the region proposals, which are fed as an input along with article maps into the final recognition layers. Detection of the layers containRoI pooling layer ot the as a result of which final classifications.

A foremost reason for using shared convolutional layers is the negligible computational cost of region proposals. Moreover, computing region proposals using CNN has the extra benefit of its dependency on GPU as compared to old-fashioned RoI generation methods (i.e. Selective Search), which were applied using CPU.

3.2 SSD

The Single Shot MultiBox Detector (SSD)[23] is a unified framework for the object detection with a single network. It often promotes integrated detection. Neither, it produces any proposal nor it takes part in resampling the image segments. However, the process of object detection in SSD is carried out through a single pass of CNN[24].

It somehow resembles a descending window method, in which procedure is instantiated with an avoidance of the regular bounding values. Diverse scales and aspect ratios can be included in it. The calculations for the object predictions depends on the boxes, which includes offset parameters and hence, responsible for predicting the correctness of bounding box covering the object as compared to a defaulted box[25].
This algorithm is based on several scales which uses the feature maps from different convolutional layers as a response to the classifier[26]. As this method generates quite a large number about the classifier, boxes filters the boxes up to a suppression stage, as a result of which the boxes below a certain threshold are eliminated.

4. CONCLUSION

This paper has presented a detailed review of the fundamental context for CNN implementation and classification for the object detection phenomenon based on its pros. and cons. Moreover, the constraints of conventional neural networks in image recognition are demonstrated. The paper also describes the advantages of the CNN as an effective solution to numerous computer vision problems. Afterward, the evolution about the convolutional object detection from the R-CNN can be recently introduced optimal techniques is demonstrated. This review not only describes the advancement in the structure of the CNN but its implementation and computational effectiveness are also illustrated. In fact, the paper describes the comprehensive review of techniques to overcome computational bottlenecks, integration of different phases into the CNN, improvement in response time, automatic error correction and its optimization over time lapse.

The Future Work

The exact study of the designed techniques, implementation and real-time hardware tests can be a future trend for the researchers. For further research response, time reduction on commercial computers would be a thought-provoking topic. Due to hardware cost, the claim of real-time performance by various techniques is not achieved yet. It can be achieved by implementing these techniques on real-time hardware. So that more and more applications can be commercialized for consumers. Hardware cost and size reduction can also be an optimal approach towards future trends. Moreover, according to aforementioned suggestion, an approach can be acknowledged for producing a complete convolutional or neural system, which can be able to learn inherent features for an object classification automatically. Correspondingly, this system should also be capable of differentiating the object can be since a convinced part of a scene. However, this systemic technique can be implemented by integrating geometric interference and CNN. Yet, it depends on time and scope that which research direction the research takes.

REFERENCES

[1] C. Szegedy, A. Toshev, and D. Erhan, **Advances in Neural Information Processing Systems,***Deep Neural Networks for object detection*, vol. 26, pp. 2553–2561, 2013.

[2] K. Q. Huang, W. Q. Ren, and T. N. Tan, **A review on image object classification and detection**, *Chinese Journal of Computers*, 2014.

[3] F. Xie, M. Zhang, J. Zhao, J. Yang, Y. Liu, and X. Yuan, **A Robust License Plate Detection and Character Recognition Algorithm Based on a Combined Feature Extraction Model and BPNN**, *Journal of Advanced Transportation*, vol. 2018, 2018, doi: 10.1155/2018/6737314.

[4] O. Russakovsky, J. Denf, S. Hao, K. Jonathan, S. Satheesh,M. Sean and H. Zhiheng, **ImageNet Large Scale Visual Recognition Challenge**, *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[5] D. Hoiem, D. S.K., and J. H. Hays, **Pascal VOC 2008 Challenge**, *World Literature Today*, 2009.

[6] T. Y. Lin *et al.*, **Microsoft COCO: Common Objects in Context**, *Springer International Publishing*, vol. 8693, pp. 740–755, 2014.

[7] K. Simonyan and A. Zisserman, **Very deep convolutional networks for large-scale image recognition**, *arXiv preprint arXiv*, p. 1409.1556, 2014.

[8] B. Yang, J. Yan, Z. Lei, and S. Z. Li, **Craft objects from images**, *In Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition, pp. 6043–6051, 2016.

[9] X. Zeng, Zeng, Xingyu, W. Ouyang, Y. Junjie, Li, Hongsheng, X. Tong, K. Wang, Y. Liu,**Crafting gbd-net for object detection**, *arXiv preprint arXiv*, pp. 1610–2579, 2016.

[10] M. Egmont-Petersen, U. Schreiner, and S. C. Tromp, **Detection of leukocytes in contact with the vessel wall from in vivo microscope recordings using a neural network**, *IEEE Trans. Biomed. Eng*, vol. 47, no. 7, pp. 941–951, 2000.

[11] R. Girshick, **Fast r-cnn**, *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

[12] D. G. Lowe, **Object recognition from local scale-invariant features**, *In Computer vision, 1999. The proceedings of the seventh IEEE international conference*, vol. 2, pp. 1150–1157, 1999.

[13] N. Dalal and B. Triggs, **Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition**, *CVPR, IEEE Computer Society Conference on (2005)*, vol. 1, pp. 886–893, 2005.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, **Rich feature hierarchies for accurate object detection and semantic segmentation**, *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[15] S. Ren, K. He, R. Girshick, and J. Sun, **Towards real-time object detection with region proposal networks**, *In Advances in neural information processing systems*, pp. 91–99.

[16] X. Xinfeng, D. Du, L. Qian, Y. Liang, W. Tang, L. O. Zhong, L. Mian, P. H. Huynh and R. S. Monggoh. **Exploiting Sparsity to Accelerate Fully Connected Layers of CNN-Based Applications on Mobile SoCs**, *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 17, no. 2, p. 37, 2018.

[17] J. Wang and X. Xiaolong, **Non-local neural networks**, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, no. 3, 2018.

[18] C. M. Bishop, **Pattern Recognition and Machine Learning (Information Science and Statistics)**, *Springer-Verlag New York, Inc., Secaucus, NJ, USA*, 2006.

[19] A. Krizhevsky, I. . Sutskever, and G. E. Hinton, **Imagenet classification with deep convolutional neural networks**, *In Advances in neural information processing systems*, pp. 1097–1105, 2012.

[20] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders, **Segmentation as selective search for object recognition**, *In Computer Vision (ICCV), IEEE International Conference*, pp. 1879–1886, 2011.

[21] Z. He, D. Liang, S. Zhang, X. . Huang, and S. Hu, **Traffic-sign detection and classification in the wild**, *onference on Computer Vision and Pattern Recognition (CVPR)*.

[22] Ren and Shaoqing, **Faster R-CNN: towards real-time object detection with region proposal networks**, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 6, pp. 1137–1149, 2017.

[23] W. Liu, D. Anguelov, E. Dumitru, C. Szegedy, R. Scott, F. Cheng-Yang and C. B. Alexander**Ssd: Single shot multibox detector**, *In European Conference on Computer Vision Springer*, pp. 21–37, 2016.

[24] M. Srivastava, S. Mishra, H. Singh, and H. Singh, **HIERARCHAL CLASSIFICATION OF SATELLITE**, vol. 7, no. 19, pp. 2524–2529, 2020.

[25] T. A. Al-asadi and M. A. Almaamory, **HIDING FINGERPRINT IN IRIS IMAGE BASED ON**, vol. 7, no. 19, pp. 3902–3909, 2020.

[26] B. Mallikeswari and D. P. Sripriya, **A DOUBLE FILTERING-DENSITY ALGORITHM FOR ENHANCING IMAGES OF DIFFERENT TYPES**, *JCR*, vol. 7, no. 13, pp. 2394–5125, 2020, doi: https://doi.org/10.1109/cvpr.2016.232.