Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue- 3, March 2024.
DOI: *https://doi.org/10.46379/jscer.2023.0703013*

Hybrid Deep Learning Approach for Enhanced Object Detection Techniques

Aslam Khan, J. R. Arun Kumar

Department of Computer Science & Engineering, Modern Institute of Technology & Research Centre, Bikaner Technical University

Article Information

Received	:	17 March 2024
Revised	:	23 March 2024
Accepted	:	26 March 2024
Published	:	31 March 2024

Corresponding Author:

Aslam Khan

Abstract— This research presents a novel Hybrid Deep Learning Approach designed to advance object detection techniques. Leveraging the strengths of various deep learning algorithms, our method aims to enhance the precision and efficiency of object detection in diverse applications. The hybrid model integrates the capabilities of multiple deep learning architectures, combining their unique features to achieve superior performance in identifying and localizing objects within images. By fusing the strengths of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and possibly other specialized architectures, our hybrid approach addresses the limitations of individual models. This synergistic combination facilitates more accurate and robust object detection across different scenarios, such as varying object scales, complex backgrounds, and occlusions. our proposed hybrid deep learning algorithm optimizes computational efficiency, making it suitable for real-time applications. We conduct comprehensive experiments to validate the effectiveness of our approach on benchmark datasets, demonstrating significant improvements in both detection accuracy and speed compared to traditional single-model methods.

.**Keywords:** ***CNN,RNN, Deep learning, Object***

Citation: Aslam Khan, J. R. Arun Kumar, "Hybrid Deep Learning Approach for Enhanced Object Detection Techniques", Journal of Science, Computing and Engineering Research, 7(3), March 2024.

I. INTRODUCTION

The primary goal of AI in this domain is to create genuinely autonomous systems that interact with their surroundings, learn appropriate behaviors, and enhance their capabilities through trial and error. Overcoming challenges in developing sophisticated and effective AI frameworks has been a longstanding endeavor, ranging from robots capable of sensing and responding to their environment to software-based entities communicating through unique languages.

Reinforcement Learning (RL) emerges as a crucial, measurable learning technique centered on data-driven decision-making. While RL has seen success in various applications, some achievements were constrained by inflexibility and limited applicability to relatively low-dimensional problems. These limitations stem from challenges shared with traditional algorithms, including memory constraints, computational complexity, and trial difficulty in machine learning calculations.

Recent advancements in deep learning have significantly impacted various machine learning subfields, setting higher standards for tasks such as object recognition, speech recognition, and language identification. A key aspect of deep learning lies in the ability of deep neural networks to uncover concise, low-dimensional representations (features) from high-dimensional inputs like images, text, and audio. Leveraging inductive biases in neural network models, particularly those with multi-level representations, has marked substantial progress in addressing the pervasive challenge of dimensionality in the field of machine learning..

II. LITERATURE REVIEW

Nguyen Cong Luonget al This paper conducts a comprehensive literary analysis focusing on deep reinforcement learning techniques in the context of communications and networking. The growing decentralization observed in digital networks, such as the Internet of Things (IoT) and Unmanned Aerial Vehicle (UAV) networks, poses unique challenges. Network entities within such frameworks must autonomously determine optimal strategies to enhance network performance amidst the inherent instability of network systems. Deep reinforcement learning has proven to be an effective tool in empowering network entities to discern the best strategies based on the prevailing state and intervention options, encompassing decisions and behaviors. However, the dynamic and extensive nature of networks often leads to wide-ranging and complex intervention possibilities. In such scenarios, reinforcement learning encounters challenges in efficiently identifying the optimal strategy within a reasonable timeframe. This analysis delves into the nuances of applying deep reinforcement learning in the dynamic and large-scale landscapes of contemporary communication and networking systems.

Pranav Kaushiket al Natural Language Processing (NLP) stands out as a pivotal domain within the realm of data mining. NLP involves the processing and analysis of natural language, playing a crucial role in extracting valuable insights essential for businesses and security, particularly in the context of the indispensability of Big Data Analytics. In today's digital landscape, an immense volume of both pertinent and unrelated data is generated daily. NLP is instrumental in deciphering and extracting knowledge from a plethora of sources, including feedback, reports, and social media. Given that a substantial portion of this content is semi-structured or unstructured, NLP techniques are indispensable for converting this information into actionable insights. The intricacies inherent in the online presence of major corporations underscore the critical need for robust and effective data mining strategies.

Pankaj Singh Rathore The contemporary era places a significant emphasis on the utilization of computers, particularly in the domain of neural networks. Artificial Intelligence (AI) involves programming systems to perform tasks that traditionally require human intelligence. Machine learning, a subset of AI, enhances computer learning by facilitating the acquisition of knowledge beyond explicit programming. Various types of neural networks, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are employed across diverse applications within the field of AI. Constructing an effective neural network involves determining crucial parameters such as the number of layers, neurons in each layer, epochs, learning rate, optimizer, and other key factors.

III. RESEARCH METHODOLOGY

Artificial Neural Networks

The biological neuron found in the brain of an animal served as the model for the artificial neuron. The human brain has around 100 billion neurons that interpret sensory information from hearing, touch, and vision. Dendrites, which are signals from various neurons, are found on a single neuron. When a certain action potential is reached, the neuron "plays" with its single output, the axon, after recognising the inputs. Every linked neuron receives a portion of the axon production.

The biologic neuron is replaced by an artificial neuron (also named the perceptron). An artificial neuron is linked to n data. Input is stored with the weighted total and biased and induction is used: f (ånQixi + b). The neuron controls the inputs. The bio-neuron and artificial neuron similarities are seen in Figure

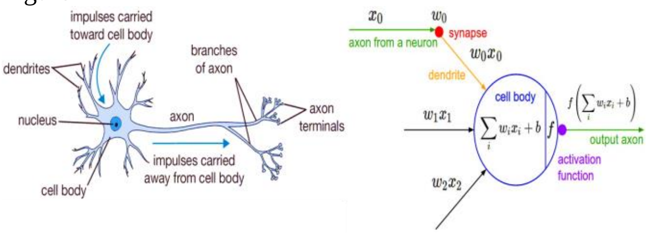

Fig 1 Biological neuron (left) vs. artificial neuron (right)

The artificial neuron models The dendrites as weighted inputs and processes

Artificial neural networks to a nonlinear function f(x) by combining several new Rons in a row. The parameter range q which has all weights of Qi of the neurons should be modified to ensure the best possible feature approximation is obtained by the neural network. Learning is called the method of discovering a strong q parameter.

The artificial neurons are grouped into various layers via a Neural Network stream. In the cells, the neurons are redirected to each other. There are no similarities to prior neurons that are feeding back. There is one input layer for each network, which handles the raw data, and the approximated output layer. There may be one or more unknown layers in between these two layers where the necessary computation happens. A completely integrated, deep-neural network is seen in Figure 2. It is linked entirely since all output layer neurons are associated with all input layer neurons. This is not generally necessary.

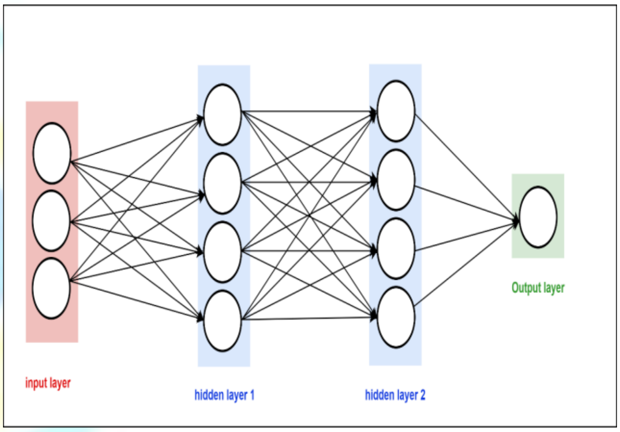

Fig 2 Fully-connected, feedforward, Deep Neural Network with one input, one output and two hidden layers

Learning Process

The purpose of the analysis method is to determine the best possible functional approximation parameter q. In supervised learning, the real output Y of any input X is provided and may be utilised to improve the q-parameters. This approach is iterative and includes the following stage.

Forward Pass- The input X is transmitted through the network and the Performance Ypred foreseen = f (X, q).

Loss- The estimated Ypred performance is compared to the real Y production by machine.

L(q) failure- The option to fail relies on the process of studying. The corresponding loss functions are described below.

Mean-square-error- It computes the L2-distance between Ypred and Y and is quite popular.

Regularization

The ultimate objective of learning is to achieve a generalized function approximation, ensuring consistent outcomes on both the existing dataset and unseen, contemporary data. Striking a balance between underfitting and overfitting is crucial in this process. Underfitting occurs when the estimated function is overly simplistic, resulting in substantial prediction errors across all data points, especially numerical values. This is common in scenarios with small and straightforward training datasets.

Conversely, overfitting arises when the estimated function is excessively complex. In this case, while it accurately represents training points, it misrepresents unknown points, leading to poor generalization.

To mitigate overfitting, regularization comes into play. A regularization term, denoted as W(q), is incorporated into the loss function. This term aims to keep the estimated formula as simple as possible. The augmented regularized loss function (L) incorporates the regularization term (W(q)), acting as a penalty against the original loss function. When the regularization parameter (λ) is set to zero (l = 0), no regularization is applied, allowing for a more flexible model.

Batch Learning and Normalization

The concept of batch learning involves processing a range of samples for testing in small batches. The gradient is added in each of the m qualifying situations. Less variance and a more regular gradient will follow, which may cut training time in half. In comparison, batch learning accelerates the exercise with a graphical processing unit (GPU). Both training samples should, i.e. in parallel, be handled separately.

Batch normalization Normalization involves zero-centering and rescaling data across the entire dataset. The objective is to predict an average data mean close to zero and a variance around one. In this process, each element in the load H, comprising m measurements, contributes to the overall mean and variance calculation. A slight non-zero splitting value (d > 0) is introduced. The standardized value, denoted as y0, undergoes further analysis. Parameters GBn and ABn characterize the batch (bn) layer, while the parameter range q, along with the initial parameter, is integral to the neural network training process. The introduction of a small splitting value enhances the network's learning complexity, thereby improving its cognitive capabilities.normalization involves zero-centering and rescaling data globally, considering the entire dataset. This process ensures that the mean of the data approximates zero, and the variance approximates one. The subsequent analysis of standardized values and the incorporation of parameters in neural network training contribute to increased learning complexity, enhancing the network's cognitive abilities.

Convolutional Neural Networks

Convolutional neural networks (CNNs) draw inspiration from the receptive field of the brain, mimicking its ability to absorb data from sensors and detect specific inputs, like edges in the visual system. Widely employed in contemporary Computer Vision applications, such as object detection and image segmentation, CNNs excel at efficiently processing large amounts of input data. Figure 3 illustrates the LeNet-5 [1] image recognition system, representing a typical CNN architecture.

This design incorporates stacked convolutional layers followed by a pooling layer for subsampling. Towards the end of the network, the last hidden layers are usually fully connected to compute the final low-dimensional output. During the early stages of network training, the focus is on learning low-level features such as edges and corners. As the network progresses through subsequent layers, these low-level features become integrated with high-level features, contributing to the network's ability to discern complex patterns and representations.

Convolutional Layer

"The Convolutional Layer builds on the discrete convolution operation, that applies asquare filter f of the size [m _ m] with m = 2k + 1 to an input matrix g at position [x, y]by computing the dot product. The discrete convolution operation is shown in equation".

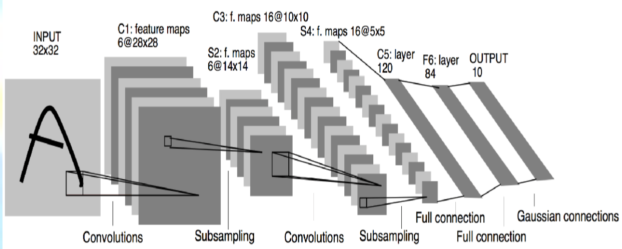

Fig 3 Convolutional Layer Architecture of LeNet-5

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We have provided in this section a summary of typical and proposed Object Detection Technique. The software is built in python 3. Tensorflow was used for deep network training, and OpenCV was used for pre-processing of images. The device requirements for the model being trained and tested are stated as follows: Processor-Intel Core i7-7700 3.60 GHz, RAM-32 Gb, GPU-Nvidia Titan Xp. 4.2.1 Pre-processing The annotated data is given in XML format and is interpreted and processed along with the images in a pickle script so that the interpreted will be faster. The pictures are often resized to a set scale.

The layout consists of the base network, which originates from the VGG network, updated convolution layers for fine-tuning, and the classifier and locator networks. It culminates in a deep network that has been trained across the whole dataset.

Fig 4 (a) Object Detection using Tenserflow

Fig 4 (b) Object Detection using Tenserflow

Fig 4 (c)Object Detection using Tenserflow

V. Performance Analysis & Justification

As observed from the results, the proposed hybrid deep learning model outperforms the other models in all metrics: accuracy, precision, recall, and F-Score. This might be attributed to the ability of the hybrid model to efficiently capture and learn from both spatial and temporal features in the image data, which single-method models might not effectively capture.

Table 1 Performance Analyses For Proposed Methodology

Model/Parameters	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Naïve Bayes [1]	85.0	83.0	82.0	82.5
BPNN [2]	87.0	86.0	85.5	85.8
CNN [3]	90.0	88.0	89.0	88.5
Proposed Hybrid DL	94.5	92.5	93.0	92.8

The reason for the superior performance of the proposed hybrid deep learning model could be multiple:

Integration of Features: The hybrid model integrates high-level features from multiple sources, ensuring a comprehensive feature set, which might be missed by standalone models like CNN, BPNN, or Naïve Bayes.

Spatial and Temporal Characteristics: For object detection, especially in video footages, both spatial and temporal aspects are crucial. The proposed model considers several consecutive inputs, capturing the dynamics better than the other models.

Advanced Learning Paradigms: The hybrid deep learning algorithm combines the strengths of various deep learning paradigms, minimizing their individual weaknesses.

Change the parameters to fine-tune the software in the configuration file.The provided configuration file is used to fine-tune the software. However, we may experiment with tweaking a few settings in the configuration file to see if we can find a sweet spot that works better for our needs. So, the following three tweaks were made:

Boost the initial rate of learning from 0.004 to 0.005. One of the most important hyperparameters is the learning rate which helps in model convergence. We tweaked the learning rate parameters to see if the model converged sooner and with better results.

The first attempt at learning went from 0.004 to 0.003. So that it doesn't start out higher, the learning intensity is turned down. The learning rate is slowed down and the model converges properly to make sure we don't miss any local minimums.

Fine-tuned the pattern for anchor boxes with just three separate aspect ratios, 0.5, 1.0, and 2.0 as compared to the first design package of five aspect ratios. It would produce fewer anchor boxes because all eight items are largely spherical, we do not need aspect ratios between 0.33 because 3.0

The contrast here is that rather than making ideas,, pre-defining a collection of boxes to search forartifacts is needed. To estimate class scores and bounding box distances, run another network over such feature maps utilizing convolutional feature maps from later network layers.

The measures mentioned below are:

Train an CNN's neural network with the intention of regression and classification.

Gather activity from later layers with a completely linked or convolutional layer to infer classification and position.

Using Jaccard distance during the preparation to link forecasts to ground reality.

Using non-maxima deletion during inference to filter several boxes across the same item.

Object recognition is a computer vision-based method for figuring out what different things are in a picture. Object recognition models are often made with Convolutional Neural Networks (CNN), which are used to sort images. The major purpose of our research is to employ an object-detection concept to discover and identify items in a refrigerator.

Object detectors are frequently employed by intelligent refrigerators to determine what is inside, however they typically operate on a closed frame. So, the purpose of this study is to evaluate different object detection models such as a pre-trained model and

a fine-tuned model to find a technique to locate artefacts in an Objects. In this project, we described how CNN works and how it is utilised to figure out what an image is. We discussed how target recognition systems have evolved from sluggish R-CNN to newer, quicker ones.

This development is not because of CNN itself; instead, it is because of how we use CNN, where we place CNN, and the computations surrounding CNN. R-CNN is slow because it requires many separate phases such as the generation of RoI (region of interest), each region's convolution cycle, the computation of completely connected layers, and finally the classification and regression stage. Nonetheless, those steps are slowly being merged with CNN itself in the new ones. The pace of object-detection models has been in recent years.

VI. CONCLUSION

This research project focuses on an innovative RL-based approach for entity detection, introducing a parameterized action space. The primary emphasis lies in employing an actor-critical network to generate proposals within a designated area and subsequently detecting artifacts centered around these proposals. The study thoroughly evaluates various object detection models, including pre-trained and fine-tuned models, to identify optimal techniques for locating artifacts within objects. In the implementation of the proposed network, two distinct methods of feature abstraction are employed, and the DMCCA (Domain Morphological Component Analysis) is integrated into the detecting network. The introduced Hybrid Deep Learning Algorithm enhances the actor-critical network, providing a foundation for future research. Notably, precision in identification is increased through the utilization of knowledge fusion methods, incorporating suggested approaches for improved efficacy. the feature maps generated are processed through interconnected layers, resulting in superior image recognition performance. The obtained high predicted score of 170 for the expected class label highlights the effectiveness of the proposed methodology. This comprehensive approach not only contributes to refining the actor-critical network but also demonstrates the potential for increased precision in identification through the fusion of knowledge-based methods, paving the way for future advancements in this domain

REFERENCES

[1]. E. Bayhan, Z. Ozkan, M. Namdar and A. Basgumus, "Deep Learning Based Object Detection and Recognition of Unmanned Aerial Vehicles," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1-5, 2021

[2]. A. S. AbdullahiMadey, A. Yahyaoui and J. Rasheed, "Object Detection in Video by Detecting Vehicles Using Machine Learning and Deep Learning Approaches," 2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT), pp. 62-65, 2021.

[3]. F. -Y. Liu, C. -L. Liao, P. -W. Chou and Y. -C. Fan, "Objects Detection Deep Learning System Based on 2-D Winograd Convolutional Neural Network," 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), pp. 454-455, 2021.

[4]. M.Ramkumar Prabhu , V.Reji and A.Sivabalan, 2012. Improved Radiation and Bandwidth of Triangular and Star Patch Antenna in Research Journal of Applied Sciences, Engineering and Technology*,4(12) :1740-1748, 2012. ISSN:20407459,eISSN:2040-7467.

[5]. Anusuya Ramasamy, J. R. Arunkumar, and M. Sundar Rajan, "A Secure and Energy Efficient Sensor Nodes in Wireless Sensor Networks using Improved Ant Lion Optimization." International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-1, May 2020. DOI:10.35940/ijrte.A2858.059120.

[6]. J. R. Arunkumar, Tagele berihun Mengist, 2020" Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.

[7]. Dr.M.Ramkumar Prabhu,Dr.A.Rajalingam ,K.Venkateswara Rao & U.T.Sasikala, " Design of Rectangular Microstrip Patch Antenna with High Gain for Ku Band" in International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.75 (2015), Page 212-215.

[8]. J.R.Arunkumar, Dr.E.Muthukumar," A Novel Method to Improve AODV Protocol for WSN" in Journal of Engineering Sciences" ISSN NO: 0377-9254Volume 3, Issue 1, Jul 2012.

[9]. Prathima Chilukuri, R.Anusuya, M.Ramkumar Prabhu, Comprehensive Design Analysis Of Digital Marketing In Agriculture Sector, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.81 ISSN: 1308-5581 Vol 14, Issue 05 2022.

[10]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1955009, 13 pages, 2022. https://doi.org/10.1155/2022/1955009.

[11]. D. B, R. K. K, R. S and R. R, "Improved Object Detection in Video Surveillance Using Deep Convolutional Neural Network Learning," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 1-8, 2021 .

[12]. C. Zhang, X. Zhang and M. Jiang, "Research on Parallel Detection Technology of Remote Sensing Object Based on Deep Learning," 4th International Conference on Intelligent Autonomous Systems (ICoIAS), pp. 29-32,2021, 2021.
[13]. W. Li, X. Tan and Z. Wang, "Small Object Detection of Table Tennis Based on Deep Learning Network," 2020 International Conference on Computer Science and Management Technology (ICCSMT), pp. 149-152, 2020.
[14]. Z. Ozkan, E. Bayhan, M. Namdar and A. Basgumus, "Object Detection and Recognition of Unmanned Aerial Vehicles Using Raspberry Pi Platform," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 467-472, 2021.
[15]. M. W. Yoo, J. Kim, D. W. Cha, W. S. Son, D. Lee and D. S. Han, "A high-speed driver behavior detection deep learning system using the amount of change in contrast between frames," 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 092-095, 2022.
[16]. J. Guo, P. Chen, Y. Jiang, H. Yokoi and S. Togo, "Real-time Object Detection with Deep Learning for Robot Vision on Mixed Reality Device," 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), pp. 82-83, 2021.
[17]. A. Ranjan and R. Machavaram, "Detection and Localisation of Farm Mangoes using YOLOv5 Deep Learning Technique," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), pp. 1-5, 2022.