INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES

ISSN: 2322-0821(O) VOLUME 12 ISSUE 2
ISSN: 2394-9910(P) April 2024 - June 2024
www.irjeas.org

Original Article

Deep Spotter: A Deep Learning Approach For Image Detection

Kirti Sahu [1], Dr. P K Sharma [2], Manvendra Singh[3]

[1] *M. Tech Scholar, Computer Science & Engineering Department, NRI Institute of Research and Technology Bhopal*
keertisahu2015@gmail.com
[2] *Associate Professor & HOD, Computer Science & Engineering Department, NRI Institute of Research and Technology Bhopal*
principal.nirt@gmail.com
[3]*HOD, Computer Science & Engineering Department, NRI Institute of Research and Technology Bhopal*

Corresponding Author: keertisahu2015@gmail.com
DOI –10.55083/irjeas.2024.v12i02006

ABSTRACT: Current advancements in the digital era are characterized by the abundance of visual data, making it necessary to apply sophisticated methods for object recognition. Detecting one object in the image is easy; however, identifying several objects simultaneously is quite a tough challenge. The paper designs object recognition and proposes a solution for detecting multiple objects within the image. Among the many alternatives, Convolutional Neural Networks (CNNs) stand out as being particularly effective in solving this problem. Among the various methods used in CNNs, Deep Spotter is one that, based on the YOLO (You Only Look Once) architecture, efficiently detects objects in images by making full use of deep learning. Instead of treating object detection as a repurposed classifier, as most existing methods within a unified framework do, Deep Spotter casts it as a regression problem. It detects bounding boxes and determines class probabilities directly from entire images in a single sweep and bridges every individual step in the detection process up to that point in one neural network. It allows all steps to be end-to-end optimized and improves the performance of the detection.

Keywords: Convolutional Neural Networks (CNN), Optimization, Detection

1. INTRODUCTION

Nowadays, image data uploaded to social media is increasing in a rapid pace and it is beyond control of human's manual operation. For handling massive images, automatic computer systems with highly sophisticated image processing techniques such as image pattern recognition are strongly required. The computer systems detect and locate objects from images and this technique is called image recognition technology. The main purpose of image recognition is to classify objects in images and retrieve their positions with high accuracy. In practical applications, image recognition technology is extensively used in various fields from security (e.g. facial recognition system) to cloud based services (e.g. image categorization and sorting) and visual content management (e.g. stock photo database). However, there still exist some difficulties in automated image detection such as complicated backgrounds, noise and low quality images. Convolutional Neural Networks (CNNs) are promising Neural networks specifically developed for handling image data. CNNs are capable of learning hierarchical patterns and features from raw image data automatically and thus suitable for recognition tasks such as object detection. As With the evolution of technology, internet facilities are available everywhere not only on Personal Computers (PCs) but also on smart phones. Social media websites are acquiring massive amount of image data uploaded every day. It is impossible for humans to efficiently process such large amount of data, so computerized automated systems are needed for image scale analysis. Image processing technique uses methods such as pattern recognition to identify and pinpoint objects within an image.The objective of image recognition is to classify objects and determine their locations. Some of the applications of image recognition are: security, cloud based image categorization and visual content database (e.g. stock photo websites). Automated image detection techniques are limited for complex background, noisy and low resolution images. Convolutional Neural Networks (CNNs) have recently shown spectacular success in image recognition. This paper presents a Deep Spotter system which is based on CNNs for recognizing objects within images. This architecture can be implemented in any field.

2. LITERATURE REVIEW

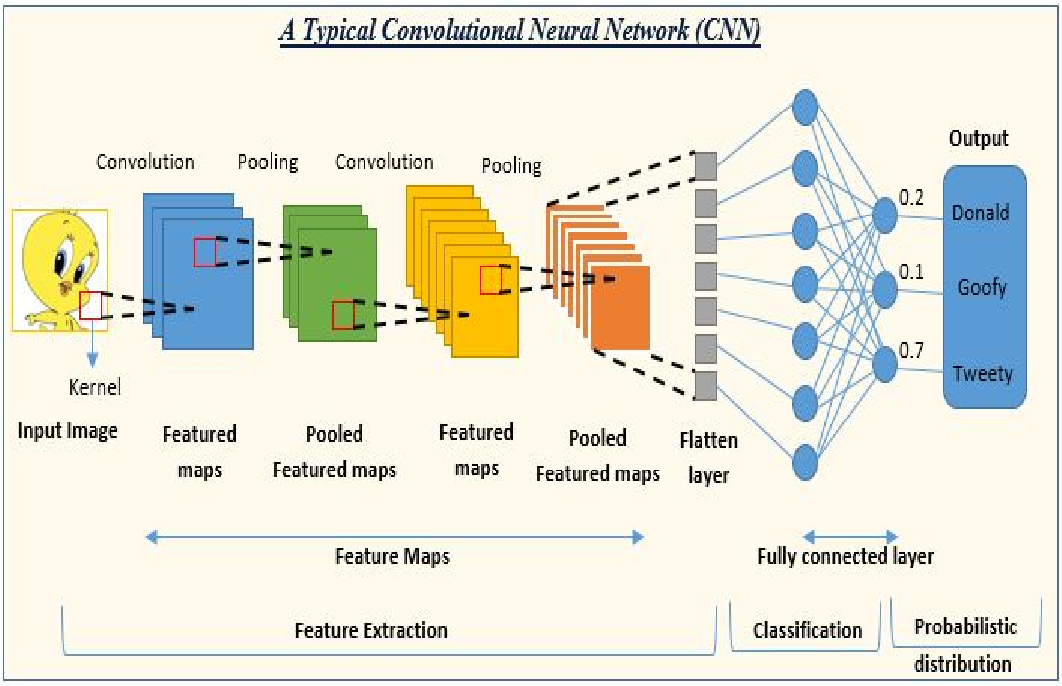
Dr. Suwarna Gothane from CMR Technical Campus, Hyderabad, outlines key limitations of the YOLO algorithm in their paper titled "A Practice for Object Detection Using YOLO Algorithm." They emphasize the significant training time of YOLO, taking about 47 seconds for each test proposal, which hinders its real-time usability. Moreover, YOLO requires substantial computational resources to store feature maps, leading to high storage demands and lengthy training periods. Additionally, the algorithm's reliance on a limited number of regions from selective search may limit its ability to detect objects effectively in complex scenes. Dr. Gothane suggests the need to replace existing systems with more efficient ones to overcome these challenges, particularly addressing time consumption and resource utilization issues. On the other hand, Asst. Prof. Shreya Kapoor, Naman Mittal, and Akarsh Vaidya from MAIT (IPU) in Delhi discuss object detection and classification using YOLO and Convolutional Neural Networks (CNN) in their paper. They stress the importance of robust and efficient object recognition methods and highlight CNN's effectiveness in feature extraction compared to traditional techniques. The paper outlines the basic architecture and steps of their image recognition model, covering image acceptance, object detection with CNN, and the use of OpenCV for this purpose. The authors also explore potential accuracy improvements and applications of object recognition in defect detection and education. Lastly, TausifDiwan, G. Anirudh, and Jitendra V. Tembhurne's paper titled "Object Detection using YOLO: Challenges, Architectural Successors, Datasets, and Applications" focuses on single-stage object detection algorithms like YOLO. The authors acknowledge YOLO's limitations, such as overlooking important aspects of object detection and providing a biased view towards YOLOs. While discussing challenges and advancements in single-stage object detection, the paper may not extensively compare YOLO with other cutting-edge methods or thoroughly explore future challenges and advancements beyond single-stage detectors. Each paper contributes valuable insights into object detection methodologies, addressing limitations, methodologies, and future prospects, which are beneficial for researchers and practitioners in this field.

3. PROBLEM DEFINITION

Due to the internet and social media, there is a deluge of digital images, and people find it difficult to comprehend them all in a manual way. Despite the availability of image recognition technologies, the automatic detection of objects in images is still challenging due to the presence of complex backgrounds, noise, and lower resolutions. Although Convolutional Neural Networks (CNNs) are effective for image recognition, there is a need for systems that can efficiently identify and locate objects in images in an automatic way. This study proposes the design of a computer vision system based on CNNs techniques and open CV library. The proposed system aims to detect objects in images in an automatic manner to ease the processing of large visual datasets. The development of robust image recognition algorithms is expected to improve computer vision and aid applications such as facial recognition for security and image classification for cloud services and visual databases.

Convolution Neural Network: Within specific network types in the field of computer vision, artificial intelligence, and machine learning, Convolutional Neural Networks are potent varieties of enhanced learning systems that are mainly applied to analyze and interpret visual data. At first, CNNs fused to perform functions such as image classification and recognition, but they are widely used now in jobs such as object identification, facial recognition, medical image identification, and automated driving.

Architecture and Layers: What actually distinguishes CNN is its configuration that aims to mimic the processes that take place in the layers of animal visual cortex. A standard CNN consists of multiple layers, each fulfilling a distinct role in extracting valuable features from input images: A typical CNN comprises several layers, each layer serves a unique role in extracting significant features from the input images.



Input Layer: While this first layer is fed raw pixel values of an image as an input.

Convolutional Layers: That is why it is worth emphasizing that the real idea behind CNN is in the convolution layers. Every layer applies a number of the learnable filters to the input data and convolves them with an appropriate mathematical operation. It is a process of developing different features (like edges, texture, and patterns) from the image inputting.

Activation Function: Finally, after convolution, ReLU or any other activation function is applied to the desired output map to inject non-linearity into the network and then maps it to the subsequent layer.

Pooling (Sub sampling) Layers: After that comes the use of pooling layers which is utilized to reduce the dimensionality of the feature maps created by the convolutional layers. Pooling is techniques aids in the reduction of the spatial dimension of the feature maps but still preserves important information.

Fully Connected Layers (Dense Layers): At the final stage of the network, dense layers are used to make decision and to classify on extracted features. These layers aggregate the learned features to deliver predictions or classifications with reference to the input database.

Application of CNN

Various fields have been revolutionized through the use of CNNs, which can handle visual data in an exceptional manner:

- **Image Classification**: The classification of objects within images is a domain where CNNs truly triumph and surpass human-level accuracy on standard datasets like Image Net.
- **Object Detection**: This task requires accurate detection and localization of multiple objects in the image through combining convolutional layers with techniques such as region proposal networks (RPNs) or anchor boxes.
- **Semantic Segmentation**: In this case, CNNs are used for pixel-level classification which means that every pixel has to be labeled with a particular class.
- **Face Recognition**: Biometric authentication is made possible due to robust face recognition systems that were developed using CNNs.
- **Medical Image Analysis**: CNNs are utilized for tasks such as tumor detection in medical imaging which is a valuable aid for healthcare providers.

YOLO (You only Look Once)

YOLO (You Only Look Once) is a new, fast way to spot objects made by Joseph Redmon and his team. YOLO changed the game by treating object spotting as one quick step. It guesses where objects are and what they are all at once with a single look

4. KEY FEATURES OF YOLO

Single Neural Network: YOLO processes the entire image at once through a single convolutional neural network (CNN), making it extremely fast and efficient.

Unified Detection: YOLO performs object detection and localization in a single step, directly predicting bounding boxes and class probabilities without requiring region proposal networks or post-processing steps.

Real-time Performance: YOLO demonstrates remarkable real-time capabilities, processing images at high speeds. exceeding 30 frames per second.

Understanding YOLOv1 Architecture YOLOv1, The original version of YOLO, introduced the foundational concepts and architecture that laid the groundwork for subsequent iterations. Let's delve into the key components and design principles of YOLOv1.

Architecture Overview

Input and Preprocessing YOLOv1 takes an input image of fixed size (e.g., 448x448 pixels) and preprocesses it by normalizing pixel values to the range [0, 1].

CNN Backbone: The input image is passed through a deep convolutional neural network (CNN) to extract hierarchical features. YOLOv1 typically uses Darknet-19, a variant of the Darknet architecture, as its backbone network.

Detection Head: The feature maps generated by the CNN are then used for predicting bounding boxes and class probabilities. YOLOv1 divides the input image into a grid of cells (e.g., 7x7 grid for a 448x448 image) and predicts bounding boxes and confidence scores at each grid cell.

Bounding Box Prediction: In YOLOv1, each grid cell predicts multiple bounding boxes (usually 2

by the computer brain, doing many guesses at the same time in just one go.

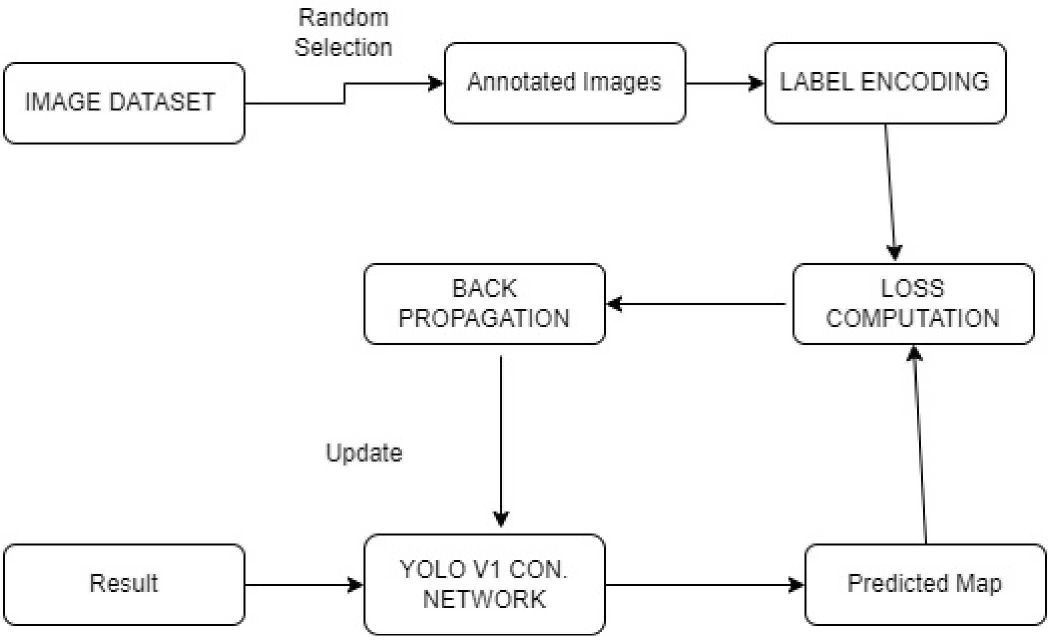
per cell), each defined by 5 parameters: (x, y, w, h, confidence). Here, (x, y) specify the center coordinates of the bounding box within the grid cell, (w, h) denote its width and height relative to the entire image, and confidence signifies the objectness score.

Class Prediction: In addition to bounding boxes, YOLOv1 predicts class probabilities for each grid cell. This is achieved using a softmax activation function applied to the class scores.

Non-max Suppression: To filter out redundant bounding boxes and improve precision, YOLOv1 applies nonmaximum suppression (NMS) based on intersection-over-union (IoU) thresholding.

Data Set and How Yolo Works: The Image Net dataset serves as a vital benchmark in computer vision, containing images classified into numerous categories. This research specifically utilizes a subset of 20 classes from the Image Net 1000 dataset due to GPU memory limitations. The selected 20 classes are as follows:

- aeroplane
- bicycle
- bird
- boat
- bottle
- bus
- car
- cat
- chair
- cow
- diningtable
- dog
- horse
- motorbike
- person
- pottedplant
- sheep
- sofa
- train
- tv-monitor



Implementation using Tensor Flow:
The implementation of the YOLO (You Only Look Once) algorithm was carried out using the Tensor Flow library, a popular open-source deep learning framework. YOLO revolutionizes object detection by treating both object detection and localization as regression problems.

Regression Approach in YOLO:
In contrast to previous object detection pipelines such as R-CNN (Region-based Convolutional Neural Network), YOLO simplifies the task by integrating detection and localization into a single neural network architecture. By treating object detection and localization as regression problems, YOLO significantly streamlines the process and avoids the complexity associated with separate components for detection and localization.

This unified approach not only simplifies the model architecture but also makes it easier to fine-tune and train. With YOLO, the entire object detection pipeline becomes more efficient and scalable, allowing for real-time performance even with limited computational resources.

How to Use:
Before running the code, it's essential to ensure that the weight file is placed in the correct directory, which is the 'test' folder. The weight file contains the pre-trained parameters of the YOLO model and is crucial for the system to make accurate predictions.

Testing Options:

- **testDB:** This feature enables testing the code with databases like Image net. It's crucial to recognize that the model has been trained exclusively on the Image net dataset. Thus, using any alternative dataset would necessitate retraining the model, a step that has not yet been undertaken.
- **testLive**: With this option, the system tests from a live webcam feed. It captures real-time video frames from the webcam and performs object detection on each frame.
- **testFile**: This option enables testing on a single image. The system takes a single input image file and performs object detection on it.

Running the Code:

To run the file, execute the following command in the parent directory:

python filename.py option
Here, 'filename.py' represents the name of the Python script file containing the YOLO implementation, and 'option' represents one of the testing options mentioned above (e.g., 'testDB', 'testLive', or 'testFile').
By specifying the appropriate testing option, users can choose the desired mode of operation for the YOLO system, whether it's testing on a database, live webcam feed, or a single image file.

Requirements:

- **Tensorflow 1.0**: TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.
- **OpenCV 2:** OpenCV 2 is a versatile and powerful open-source library widely used for image processing and computer vision

applications. It enables tasks such as face detection, object tracking, and landmark detection, among many other functionalities.

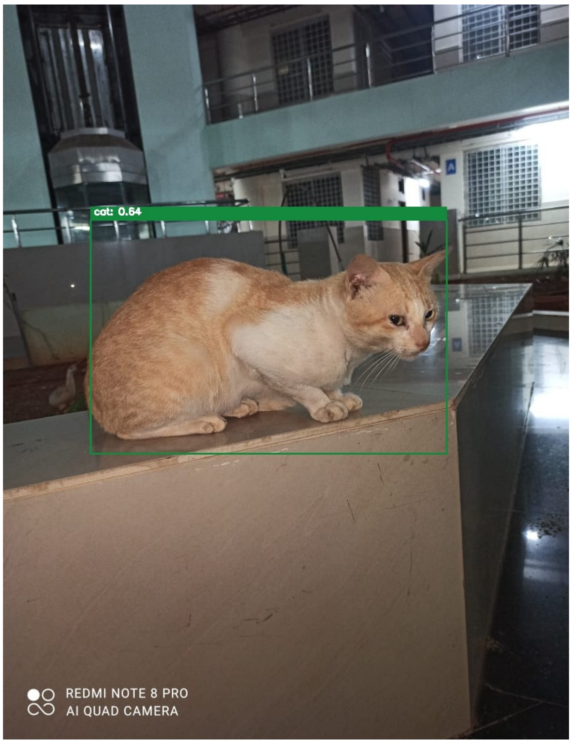
- **Python 2**: Python 2 or above version can be used.
- **Pre-trained weights**: Since, we are using pre-trained model, so after training, weight file is generated, and I have downloaded this weight file from third party sources.

5. RESULTS

I have tested my model on some of the pictures to analyze the performance of the model and also to have the look on the images where my model detection will fail. Given below are some of the results I have got after executing the model.

- In PICTURE 1, despite of objects in small size or far away, the model correctly predicts the two cycles and a person.
- PICTURE 2 is a simple picture of cat.
- In PICTURE 3, there are two persons, which the model predicts correctly despite the person face is not in front of the camera. Since the characteristics are not clearly noticeable, the model predicts less confidence score for them.
- In PICTURE 4, the model fails because it is predicting motor bike instead of cycle. It's because the characteristics are not clearly visible of the cycle.
- In PICTURE 5, model is predicting two cars correctly, but it fails to detect people on right side hand of the car, that's because of the umbrella those people have.

PICTURE 1



PICTURE 2

PICTURE 3

PICTURE 4

6. CONCLUSION AND FUTURE WORK

In our work, named "Deep Spotter," the focus is on computer vision problems and advancements, particularly in relation to object detection in images. The rise in digital visual data has created a great demand for effective object recognition. We propose to address it through a system that utilizes object detection using two techniques: CNNs (Convolutional Neural Networks) and an algorithm called YOLO [You Only Look Once] for real-time object detection.

CNNs, inspired by the human visual system, have improved the accuracy and efficiency of image recognition tasks. With the methodology that includes feature extraction, activation functions, pooling layers, and fully connected layers, CNNs can learn and even extract important features from visual data automatically.

Some tasks include object detection, image classification, and image segmentation, object recognition and object classification. Furthermore, the YOLO algorithm is unique in that it performs real-time object detection by simplifying the object detection problem to just a regression problem, which predicts bounding boxes and class probabilities from full images with only one pass through.

This approach to object detection is all-in-one in a metaphorical sense and produces increased performance and real-time inference for object detection problems. This algorithm is ideal OpenCV 2 is extensively used in real-time applications like autonomous driving, surveillance, and image analysis, providing robust tools for these complex tasks.The final results from our experiments show the proposed Deep Spotter system can perform accurately as object detection by localizing and recognizing multiple objects in images. Even though the problems of complex background, noise, and lower resolution are present, our system has shown promise in object detection.

Consequently, it is likely this research will push computer vision forward as it has advanced through the object detection and object recognition with Deep Spotter model. The proposed methodology may be applied in a variety ways to analyze pictures and beyond that will contribute to the future development of processing and analyzing visual data. The future direction of Deep Spotter is to contribute to advancing the state in object recognition by improving the model architectures, making it more robust, scalable, and customized to specific domains. In facing and addressing these challenges and new opportunities, Deep Spotter aims to advance the ongoing evolution of computer vision technologies for a wide range of real-world uses.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Department of Computer Science at NRI Institute of Research and Technology - [NIRT], Bhopal for the generous support and assistance provided during the course of this work. The department's resources and guidance have greatly contributed to the successful completion of this research project. We deeply appreciate the encouragement and collaborative environment fostered by the department, which played a crucial role in shaping our research endeavors. Thank you for the valuable support and encouragement throughout this journey

REFERENCES

[1]. Praahas Amin, Anushree B. S., Bhavana B. Shetty, Kavya K., Likitha Shetty, “Object Detection using Machine Learning Technique”, 2019 International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Volume: 06 Issue: 05, May 2019.

[2]. 6. Shubhodip Sasmal. Predictive Analytics in Data Engineering. International Research Journal of Engineering & Applied Sciences (IRJEAS). 12(1), pp. 13- 18, 2024. 10.55083/irjeas.2024.v12i01004.

[3]. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, ”You Only Look Once: Unified, Real-Time Object Detection”, 2015, University of Washington, Allen Institute for AI, Facebook AI Research.

[4]. The link for the weights generated after training the model: https://drive.google.com/file/d/0B2JbaJSrWLpza08yS2FSUnV2dlE/view?resourcekey=0 - I9VUDbm5ImKILvhvSm- Tg

[5]. Dr. SuwarnaGothane, A Practice for Object Detection Using YOLO Algorithm 2021. International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN : 2456-3307 (www.ijsrcseit.com) doi : https://doi.org/10.32628/CSEIT217249

[6]. 2. Shubhodip Sasmal. Advanced Analytics with AI in Data Engineering. International Journal of Contemporary Research in Multidisciplinary. 2024. 3(1):160-167.

[7]. Naman Mittal, Akarsh Vaidya, Asst. Prof Shrey Kapoor 2019. Object Detection and Classification Using Yol International Journal of Scientific Research & Engineering Trends

[8]. 1. Shubhodip Sasmal. Cognitive Computing in Data Engineering Applications. International Journal of Contemporary Research in Multidisciplinary. 2024. 3(1): 175-180.

[9]. 3. Shubhodip Sasmal. AI and Data Engineering: A Synergistic Approach. International Journal of Contemporary Research in Multidisciplinary. 2024. 3(1): 181-187.

[10]. TausifDiwan, G. Anirudh, Jitendra V. Tembhurne 2022. Object detection using YOLO: challenges, architectural successors, datasets and applications. doi.org/10.1007/s11042-022- 13644-/

[11]. Geeksforgeeks ,”Introduction to Convolution Neural Network”,14 Mar, 2024 https://www.geeksforgeeks.org/introduction-convolution-neural-network/

[12]. https://www.linkedin.com/pulse/why-we-prefer-convolution-neural-networks-cnn-imagedata-salunke-

[13]. Raycad Convolutional Neural Network (CNN), Nov 14,2017https://medium.com/@raycad.seedot

ech/convolutional-neural-network-cnn-8d1908c010ab 37

[14]. https://www.researchgate.net/figure/An-example-of-CNN-structure-which-consists-ofinput-layer-convolution-layers_fig7_331960063.

[15]. Shubhodip Sasmal. Streamlining Big Data Processing with Artificial Intelligence, International Research Journal of Engineering & Applied Sciences (IRJEAS). 11(3), pp. 43 - 49, 2023. 10.55083/irjeas.2023.v11i03010.

[16]. FatimazahraBelharar , Enhancing Classification Accuracy for Imbalanced Image Data Using SMOTE, 7June 2023, https://medium.com/@fatimazahra.belharar/enhancing-classification-accuracy-forimbalanced-image-data-using-smote-41737783a720

[17]. DakshaUchagaonkar , YOLO v1 working explained.. 20 June 2023, https://medium.com/@daksha.uchagaonkar/yolo-v1-working-explained-ec20750be682

Conflict of Interest Statement: *The authors declare that there is no conflict of interest regarding the publication of this paper.*

Cite this Article

Kirti Sahu, Dr. P K Sharma, Manvendra Singh. Deep Spotter: A Deep Learning Approach For Image Detection. A Comparative Analysis. International Research Journal of Engineering & Applied Sciences (IRJEAS). 12(2), pp. 3 5 - 4 3 , 2024.
https://doi.org/10.55083/irjeas.2024.v12i02006