iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58240

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Review - Construction Safety Management using Deep Learning and Computer Vision

Dr. Kalaivani D[1], Muhammad Hisham Gundagi[2], Shourya Biswas[3], Praneeth V[4], Mohammed Yousuf[5]
Department of ISE, New Horizon College of Engineering, Bangalore, India

Abstract: ***SafeZone is a deep learning and computer vision model that is aimed at improving the safety of construction workers and reducing potential hazards at construction sites. Three aspects are taken into consideration while designing the model – whether workers are wearing safety gear, potential accident zones at the site, and check for the quality of materials given to the workers. This project is intended to automate the task of identifying lack of safety and potential hazards at construction sites and subsequent implementation of safety measures to eliminate the chance of accidents. A vast majority of industrial work-sites are often never automated - and instead, manually supervised with minor application of object-based sensor technology. The project will attempt to bridge this gap with the help of Convolutional Neural Networks - which allows for a streamlined, utilitarian and flexible suite of tools that help supervisors and on-site working personnel in keeping their heads on a swivel.***
Keywords: Convolutional Neural Network, YOLO, Object detection, Personal Protective Equipment, Support Vector Machine, Non-Maximum Supression.

I. INTRODUCTION

In 2020-23 alone, around 700 workers have met their end in industrial accidents, with 213 sustaining injuries [1]. Such numbers for a growing nation are often not uncommon, as occupational accidents in developing countries come at a higher rate, and with a less covered labor force due to state legislative affairs.

In the developing world, workers often face unregulated and unprotected exposure to known hazards, such as silica and asbestos. Decades ago, these minerals were the source of problems commonly faced by millions of workers in the developed world. However, the difference now is that developing country workers are being exposed when widespread knowledge is available about the risks and effective preventive measures.

In the construction industry, workers' unsafe behaviours and acts contributed to nearly 80% of accidents. Previous studies indicate that safety performance can be significantly improved through identifying, monitoring, analyzing and modifying unsafe behaviors on construction sites. Behavior-based safety (BBS) research has been an emerging and promising trend in safety research. [13][17]

Computer vision is an interdisciplinary area that deals with how computers can provide enriched information to support and achieve a high-level understanding by digital images or videos. In recent years, it has attracted a growing research interest in the construction industry due to its potential for automated and continuous monitoring for many construction fields, e.g. occupational safety, productivity, and quality. Computer vision can process and analyse the obtained images or videos in a timely manner and provide a rich set of information (e.g. site conditions, locations and behaviours of project entities) about a construction scene to provide the accurate and comprehensive understanding of construction activities. [2][13][16]

Neural networks are a subset of machine learning, and they are at the heart of deep learning algorithms. They are composed of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network. [13][14]

II. LITERATURE REVIEW

Seo et al [2] presented a study on computer vision aiding in construction safety and enhancement and prevention of accident causes. Citing unsafe conditions and unsafe acts as the primary reasons for the loss of lives or injuries of such. Several snippets on application are also given.

Kamoona, Gostar [3] presented an essential study on the utility of Random Finite Set-Based Anomaly Detection algorithms for use in industrial sites, by using incomplete data and providing accurate statistics.The proposed solution of anomaly detection and deviations - while not an unfounded technique - had rarely been implement prior to such a degree.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Megha Nain et al 2021, [4] outlined the prime cases and needs for deep learning involved in sensor technology around industrial zones - as well as going in depth with SURF and BRIEF. Semantic and instance segmentation utilization as well as standard CNN proposals have made this study a proper case in deep learning for industrial safety.

Zhao, Chen et al [5] introduces the utility of YOLOv3 for an enhanced, and reliable real-time PSPE wear detection framework. A mid-point between RCNN and SSD frameworks used in image and video recognition software, they believe a good median between deftness and reliability has been found.

Han, Zeng et al [6] solidifies on the usage of YOLO detectors by involving cases of bounding-box regression techniques and dataset operations on targeted objects.

Gallo, Di Rienzo et al [7] mentions the utilization of DNN and visual photogrammetry to lift objects that the neural network has been trained with in order to check and verify if they are worn, or not.

Jaekyu and Sangkyub Lee [8] theorize on the causations and trajectories that lead to fatal errors and risk-factor cases in industrial locales undergoing hazardous conditions. Analyzing said factors and involving them into object recognition models to detect trajectories of risk before they occur is something that is detailed well.

Lee, Khan and Park [9] dictate about the grounded theory method - a qualitative analysis technique to gather, process and output data from a collection of events and theories regarding real life experiences. A statistical analysis done via computer code.

Guo, Zou [10] prod at the utility of RCFN convolutional neural networks – while touching on the points regarding the increasing scale of industrialization versus worker health and safety, backed by legislation and state laws. They lay out tables and options on what can be tracked, what can be taught to be tracked, and what can accurately be primed to be picked up on.

Krizhevsky, Sutskever [11] detailed on the process that deep convolutional neural networks operate on to filter and refine output. Utilizing ReLU Non-Linearity, training on multiple GPUs for a vast array of output classes, and overlap pooling – to be trained in the ImageNet LSVRC-2010 contest.

Ker, Wang [12] opines on the utility of neural networks and soft computing in order to detect malignant illnesses and trace pathogens cleanly and effectively. A medical usage.

Shobha, T., & Anandhi, R. J. [13] enunciated the usage of differential analysis algorithms in the field of neural networks and classifier designs.

Kalaivani, D, Deshmukhi H. [14] outlined the practicalities of using a dynamic neighbour discovery scheduling algorithm to be used in WSNs in order to uplift them, make them more accurate – as well as ensure they are flexibly usable. As well as walking us through dynamic neighbour sampling.

R. Ghatkamble, P. B. D and P. K. Pareek [15] dictate the usage of YOLO in waste management system – an important algorithm used in this very project. Implementation and integration with physically involved metrics.

Vandana, C.P., & Ajeeth A. Chikkamannur, [16] detail on the content of feature studies – their importance in neural networks, and delve deeper into their experimental studies and results.

Krishnaveni, A., [17] lays out the uses of PIR and IR sensors – along with electronically two-toned systems for bank safety. The crucial aspect of this paper is its emphasis on the reliance on IoT systems.

III. PROPOSED METHOD

A. Principle of YOLO architecture

YOLOv3 presents an impressive network structure that translates target detection into a regression problem. It directly provides bounding box coordinates and classification categories for targets across multiple image locations, ensuring swift detection, meeting real-time demands. Let's delve into the various facets of this algorithm. [5][15]

Operating as a supervised learning algorithm, YOLOv3 divides a given image into S × S grids. Each grid predicts B bounding boxes and their associated category confidence if an object's center lies within it. This approach excels in detecting small or overlapping objects. A bounding box encompasses five essential parameters (x, y, w, h, C) denoting center position, width, height, and confidence level, respectively. The confidence score reflects the accuracy of the bounding box containing the object, calculated using IoU and object presence probability.[14][15]

The confidence calculation is represented as:

$$C = \Pr(object) \times IoU_{pred}^{truth}$$

where $\Pr(object)$ represents whether the object is contained in the grid. If the bounding box contains the object, then $\Pr(object) = 1$; otherwise, $\Pr(object) = 0$. where the Intersection over Union (IoU) measures the accuracy of the contained objects.

$$IoU = \frac{area(ground\ truth) \cap area(prediction\ box)}{area(ground\ truth) \cup area(prediction\ box)}$$

The final confidence, denoted as c, integrates object classification probability, object presence, and IoU accuracy.

$$c = Pr(class_i|object) \times Pr(object) \times IoU_{pred}^{truth}$$

YOLOv3 employs the darknet53 network for feature extraction, enhanced by residual units for better model convergence. This network architecture facilitates expanded layering, thereby improving feature extraction.

Introducing a 1 × 1 convolution kernel within the residual module reduces convolutional operation channel count, thereby minimizing network parameters, reducing overall weight, and computational load.

Distinct from its predecessors, YOLOv3 operates on three-scale feature maps, significantly enhancing small target detection capabilities.

During the detection phase, YOLOv3 utilizes a full convolution approach, offering flexibility in input image size and reducing computation by replacing fully connected layers with convolutional layers. This method significantly reduces computational requirements. The output of YOLOv3 is a tensor with $S \times S \times ((4 + 1) \times B \times C'$ dimensions, where $S \times S$ is the number of input images divided into grid cells; B stands for the number of predicted bounding boxes per grid cell, and C' signifies the categories of detected objects. Taking the Visual Object Classes (VOC) dataset, containing 20 object types, as an example, the network generates 96 prediction results, which may not accurately reflect the real scenario. To refine these predictions, a technique called non-maximum suppression is employed: Initially, a threshold is set for the confidence score of each bounding box. Boxes with confidence scores below this threshold are discarded. Remaining bounding boxes are sorted based on their confidence scores, and the box with the highest score is selected. Each remaining bounding box's Intersection over Union (IoU) with the highest-scored box is calculated. If the IoU surpasses a certain threshold, indicating significant overlap, the box is removed. The process continues by selecting the highest-scoring box among the unprocessed ones and repeating the threshold filtering and IoU calculation steps. This method iteratively refines the predictions by retaining high-confidence bounding boxes while eliminating redundant or overlapping detections, ensuring more accurate object localization and classification.

B. Safety Conditions Considered

There are different conditions wherein the safety of the workers is being compromised. Computer vision could absolutely help us in automated monitoring of such acts, thereby assuring the overall safety in construction. Below are the safety connditions being taken into consideration.

1) *Personal Protective Equipment (PPE):* Personal protective equipment, abbreviated as "PPE", is an equipment for the protection against personal injury and illnesses at the workplace. These can be used to overcome various workplace hazards. The figure below illustrates the various components of PPE.

Figure 1. Personal Protective Equipment (PPE)

2) *Falling from Heights:* Usage of safety gear like body harness, and body nets to prevent injuries by fall. CCTV framework can be implemented to monitor hazardous locations to check for harness.
3) *Body Posture:* The body posture of the workers is crutial for preventing accidents, especially in hazardous zones, such as scaffolding framework. Improper body position and posture can increase the possibility of incidents.
4) *Trench Collapse:* Trench collapse leads to lots of fatalities every year. With the usage of protective equipment while entering any deep and big trenches could avoid accidents cases.
5) *Quality of Materials:* All construction tools and transport equipment should be of proper design and in good form, without any defects, in order to ensure the safety of the workers as well as freshly built structures.

C. Compliance Monitoring

1) *Failiur to use PPE:* Wearing proper PPE kits would greatly reduce the probability of occurance of accidents. Yu Zhao et al., [5] presents a model based on deep learning for real time detection of safety vests and helmets, developed on YOLO architecture. 2 approaches were used: (I) Object detection using YOLOv3 algorithm and (II) Using Kalman filter and Hungarian algorithm for pedestrian tracking.
2) *Hazardous zones:* The presence of hazardous zones at construction sites, such as excavation and trenches, moving parts of machinery, and working at heights, is inevitable. As such, the introduction of safety measures in accident prone zones is a must to prevent worker injuries and casualties.
3) *Transport and Handling of Materials:* The proper design and maintenance of transport, earth-moving, and material handling equipment are crucial. Only well-trained workers should operate this equipment to prevent accidents.
4) *Electrical Safety:* Regular checks for all electrical tools are essential to detect and address any faults promptly. Any identified fault should be immediately removed and reported. Consider implementing a Safety Alarm System at various electrical points to promptly receive and respond to fault notifications.

IV. EXPERIMENTAL APPROACH

In this section, we comprehensively describe the overall development approach of the object recognition models for construction site safety management, including the development environment, the safety conditions being considered, the procedure for collecting the image data for training, the image pre-processing technique, the structure of the image datasets, the backbone model, and the development of the construction worker safety management model.

A. Data Collection

Data collection is an extremely crucial phase in Computer Vision. The construction image data are collected through construction site CCTV and surveillance cameras, which are installed at strategic locations where the worker density and possibility of accidents is high. The cameras record images and videos on-site for verification of PPE compliance. Depth based Cameras examine hazardous acts and conditions for construction workers.[15]

B. Method

Image classification and detection commonly employ Support Vector Machines (SVM), Deep Learning (DL) leveraging Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These DL methods, particularly utilizing CNNs and RNNs, excel in autonomously extracting image features, contributing significantly to accurate classification and detection processes.

When it comes to real-time detection, several top-performing algorithms stand out, notably YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), RetinaNet, and R-FCN (Region-based Fully Convolutional Network). Among these, YOLO-V3 emerges as the fastest and most accurate, surpassing competitors like SSD and R-FCN according to recent studies [68].

The distinct advantage of YOLO-V3 lies in its lower computational load compared to architectures like R-CNN. This reduced computational complexity significantly enhances YOLO-V3's speed in detecting construction entities, making it a preferred choice for real-time applications in construction safety.

YOLO employs a pivotal technique known as Non-Maximum Suppression (NMS) during the output phase. NMS serves to eliminate redundancy by refining bounding boxes. Boxes with lower confidence levels but a high degree of overlap are systematically pruned, ensuring that each object is represented by a single bounding box, enhancing the accuracy and efficiency of object localization and recognition.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

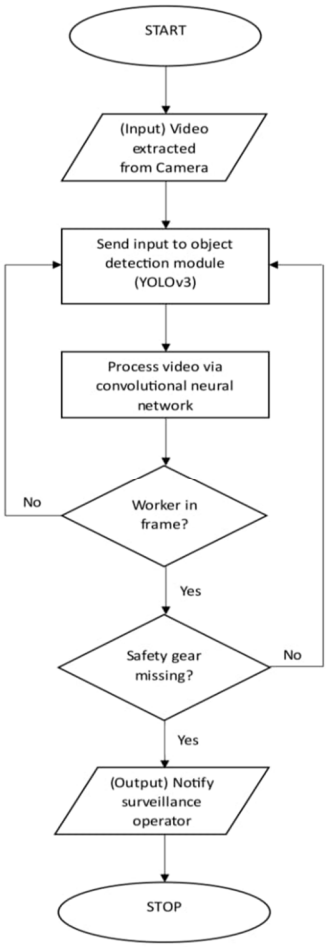


Figure 2. Overall architecture of the safety monitoring system

C. Evaluation Indices

Precision, recall and F1 serves as an indicator of a model's performance, representing the ratio of correctly identified "true" objects among those classified as true by the object recognition model. Precision quantifies the proportion of actual true objects within the model's positive classifications, often referred to as Positive Predictive Value (PPV). Precision is calculated using the below equation:

$$Precision = \frac{TP}{TP + FP}$$

Recall represents the ratio of correctly predicted true objects to all actual true objects. It measures the proportion of actual true objects correctly identified by the model among all the true objects present in the original construction site images. Recall is calculated using the below equation:

$$Recall = \frac{TP}{TP + FN}$$

The F1 score, defined as the harmonic mean of precision and recall, serves as a balanced measure for evaluating a model's performance, especially in scenarios with imbalanced labeled data. It offers a more accurate assessment by considering both precision and recall simultaneously.

$$F1\,Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

The Intersection over Union (IoU) serves as an evaluation metric for assessing object recognition accuracy. It quantifies the degree of overlap between the predicted object's location (bounding box) and the labeled ground truth.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

An IoU value closer to 1 signifies an accurate prediction by the object recognition model. The IoU is calculated by determining the area of overlap between the predicted bounding box and the ground-truth bounding box, then dividing it by the area of union, which represents the combined area encompassed by both bounding boxes. Typically, an IoU threshold of 0.5 is set as a standard measure. This threshold determines whether a prediction is considered a true positive or not, affecting performance metrics like precision and recall. Higher IoU thresholds demand more precise predictions, evaluating model performance rigorously.

V. DISCUSSION

Object Detection serves as a fundamental technique in Computer Vision, especially bolstered by the integration of Deep Learning, amplifying its impact. However, challenges persist within detection and tracking methods in Computer Vision, particularly in scenarios with poor lighting, blurriness, and occlusion. Additionally, the variance in execution speed due to computer hardware poses a significant concern. Limited dataset sizes lead researchers to rely on data augmentation techniques, often using small samples, which complicates comparison and metric evaluation, like precision and recall.

Privacy concerns also pose challenges, making compliance management complex. This paper presents a theoretical exploration of safety and compliance measures. Future research could implement these theories to yield comprehensive practical results. Integrating Ontology and Computer Vision techniques could bridge the semantic gap in upcoming research. Ontology, known for its computer-oriented and logic-based attributes, aids in modeling domain knowledge through explicit definitions of classes, instances, and relationships.

Developing a framework involving Ontology and Computer Vision techniques could establish spatial relationships between objects, detectable through Computer Vision processes. While Deep Learning excels in feature extraction, its limitation lies in identifying causality. Safety monitoring, for instance, requires not only identifying individuals and operational conditions but also evaluating their interactions for a safer workplace. Future research could delve into analyzing these interactions, presenting an area ripe for exploration. Moreover, novel algorithms are needed for concurrent identification of multiple tasks, representing another avenue for future research and development.

VI. CONCLUSION AND FUTURE ENHANCEMENT

The pursuit of progress often dominates, yet a safe and healthy workforce remains essential for advancement. Within civil engineering, Computer Vision techniques have emerged as crucial tools for enhancing safety and compliance monitoring. This paper delves into methods leveraging Computer Vision frameworks and Deep Learning to efficiently and progressively monitor safety and compliance in construction environments. This safety management framework not only enhances productivity but also ensures safety for workers and newly constructed structures. The paper reviews various Computer Vision and Deep Learning techniques while also addressing challenges and potential areas for future research. The ultimate goal is to cultivate a healthier and more progressive environment in construction by employing suitable safety and compliance techniques.

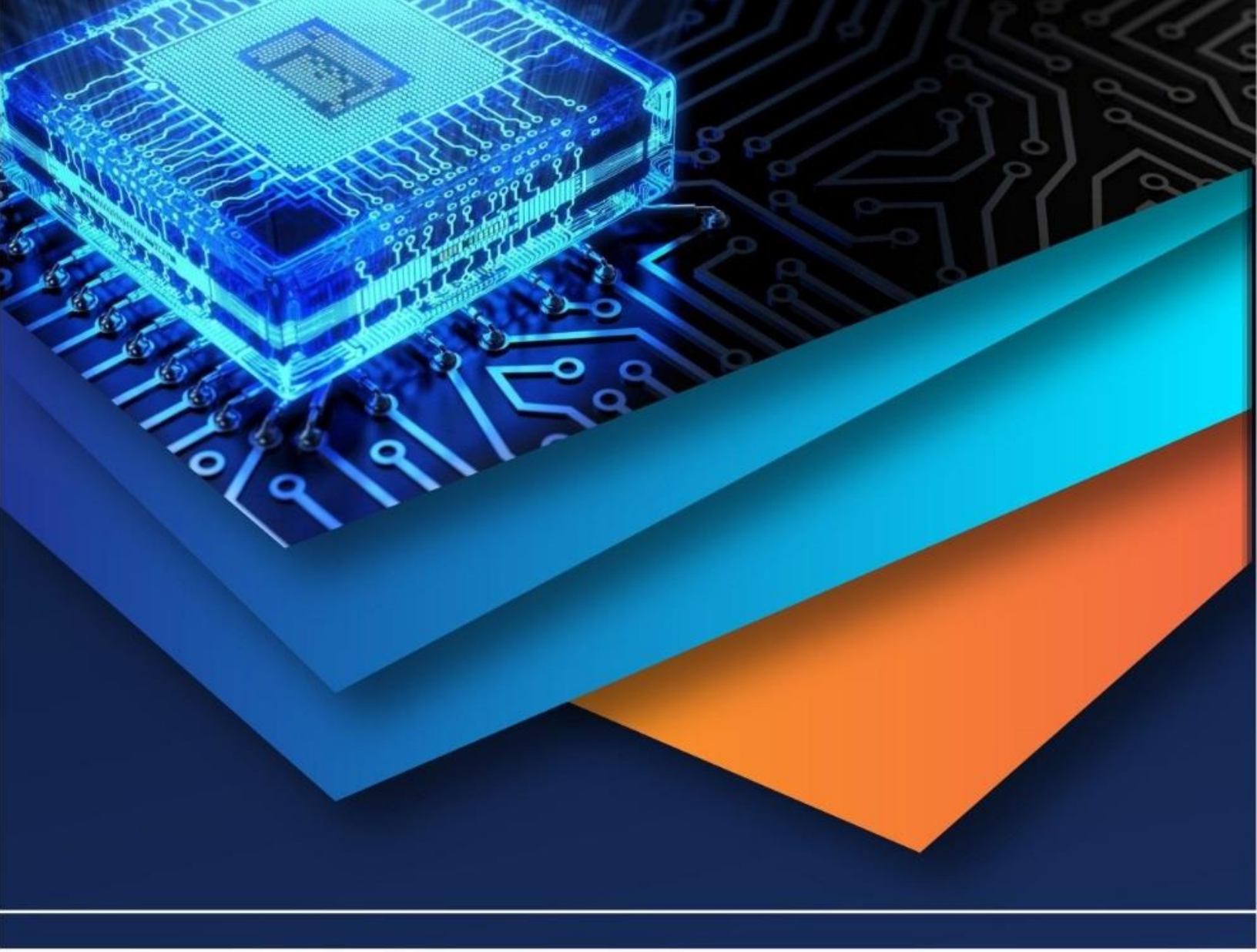
However there are some potential future enhancements. Data can be integrated from various sources like LiDAR, thermal imaging, or radar with visual data from Computer Vision. Fusion of multi-modal data can enhance object detection accuracy, especially in challenging conditions like low light or adverse weather. Implementing methods for incremental learning to adapt models continuously to new data from the construction site. This allows models to learn and incorporate new object classes or changes in the environment over time without retraining from scratch. Develop intuitive interfaces that facilitate easy interaction with the model's outputs, enabling non-technical users at construction sites to understand and act upon safety recommendations or alerts effectively. These enhancements aim to push the boundaries of the project by leveraging emerging technologies, ensuring model robustness, real-time applicability, and a heightened focus on safety and compliance within construction environments.

REFERENCES

[1] The Indian Express, “700 Workers died in industrial accidents between 2020-23, Assembly told”, Avinash Nair, September 20, 20230
[2] Joon-Oh Seo, Sang-Uk Han, Sang-Hyun Lee, Hyoungkwan Kim, “Computer vision techniques for construction safety and health monitoring.”, Advanced Engineering Informatics, 2014-2015
[3] Ammar Mansoor Kamoona, Amirali Khodadadian Gostar, Ruwan Tennakoon, Alireza Bab-Hadiashar, David Accadia, Joshua Thrope, Reza Hoseinnezad, “Random Finite Set-Based Anomaly Detection for Safety Monitoring in Construction Sites”, 1109/ACCESS.2019.2932137, 2019
[4] Megha Nain et al, “Safety and Compliance Management System Using Computer Vision and Deep Learning”, 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012013
[5] Yu Zhao, Quan Chen, Wengang Cao, Jie Yang, Jian Xiong, Guan Gul, “Deep Learning for Risk Detection and Trajectory Tracking at Construction Sites”. Special Section on AI-Driven Big Data Processing, IEEE 2019

[6] Kun Han, Xiangdong Zhen, “Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features”, 10.1109/ACCESS.2021.3138407, 2021
[7] Gionatan Gallo, Francesco Di Rienzo, Federico Garzelli, Pietro Ducange, Carlo Vallat, " A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge”, 10.1109/ACCESS.2022.3215148, 2022
[8] Jaekyu Lee and Sangyub Lee, “Construction Site Safety Management: A Computer Vision and Deep Learning Approach”, Sensors 2023, 23, 944
[9] Doyeop Lee, Numan Khan & Chansik Park, “Rigorous analysis of safety rules for vision intelligence-based monitoring at construction jobsites”, International Journal of Construction Management, 23:10, 1768-1778, DOI: 10.1080/15623599.2021.2007453
[10] Brian Guo, Yang Zou, “A review of the applications of Computer Vision to construction health and safety”, University of Canterbury, New Zealand; University of Auckland, New Zealand, 2018
[11] Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, University of Toronto, 2012
[12] Justin Ker, Lipo Wang, Jai Rao, Tchoyosin Lim, “Deep Learning Applications in Medical Image Analysis”, special section on soft computing techniques for image analysis in the medical industry current trends, challenges and solutions
[13] Shobha, T., & Anandhi, R. J.(2020). Robust Classifier Design with Ensemble Neural Network using Differential Evolution. International Journal of Engineering Trends and Technology,ISSN: 2231-5381 ,174-181
[14] Kalaivani, D, DESHMUKHI, H. (2023). ENERGY EFFICIENT DYNAMIC NEIGHBOR DISCOVERY SCHEDULE ALGORITHM FOR WIRELESS SENSOR NETWORKS. Journal of Theoretical and Applied Information Technology, 101(10).
[15] R. Ghatkamble, P. B. D and P. K. Pareek, "YOLO Network Based Intelligent Municipal Waste Management in Internet of Things," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-10, doi: 10.1109/ICERECT56837.2022.10060062.
[16] Vandana, C.P., & Ajeeth A. Chikkamannur, (2020). Feature Selection: An empirical Study. International Journal on Emerging Trends in Technology (IJETT), 2(69), (165-170).
[17] MKM Ms. Krishnaveni A, “IoT based Bank Locker Security System with Face Recognition”, Grenze International Journal of Engineering and Technology 9 (1), 2446-2452.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)