International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.184

Utilizing Machine Learning Techniques such as Object Detection and Segmentation to Facilitate Workplace Safety

Senthil Kumari P
Asst prof - Dept. of artificial intelligence and data science engineering - PSNA college of engineering and technology
Dindigul, India
psk045@gmial.com

Surya Prakash S
UG student-Dept. of artificial intelligence and data science engineering - PSNA college of engineering and technology
Dindigul, India
suryaprakashpro@gmail.com

Vignesh Karthick S
UG studenet - Dept. of artificial intelligence and data science engineering - PSNA college of engineering and technology
Dindigul, India
vigneshmsdhoni1929@gmail.com

Sheik Abdulla M
UG student - Dept. of artificial intelligence and data science engineering -PSNA college of engineering and technology
Dindigul, India
sheikabdullah7447@gmail.com

Nishok Krishna N G
UG studenet -Dept. of artificial intelligence and data science engineering - PSNA college of engineering and technology
Dindigul, India
ngnishokkrishna@gmail.com

Abstract— The study proposes an efficient method that employs the power of "machine learning" to make the "workplace safer" through the use of "computer vision" or image processing. It employs the "object detection" and "image segmentation" algorithms, which have been trained to detect workplace dangers such as an employee not wearing the safety equipment properly, and then the object detected is passed to the image segmentation part of the algorithm, which is used to highlight the object that causes the abnormality in the workplace environment, "Mask Region-Based Convolutional Neural Network" is an efficient algorithm to detect objects, classify, and to produce pixel-wise mask. If the system detects an abnormality, it sends a warning via the cloud to an application where the administrator can take action to remedy the problem as soon as possible. It also alerts nearby people to solve the situation. It also makes use of an existing surveillance camera, making this system relatively simple and cheap to set up. Thus, this system takes advantage of the capabilities of machine learning and computer vision to improve worker safety and avoid accidents by alerting the employee who is not taking proper safety measures in the workplace atmosphere.

Keywords— Machine Learning, Image Segmentation, Object Detection, Mask Region-Based Convolutional Neural Network, Workplace Safety, Computer Vision.

I. INTRODUCTION

Machine learning has huge potential for use in workplace safety. Machine learning might be used to discover potential workplace safety issues, predict when accidents are likely to occur, and develop new strategies for preventing mishaps. Machine learning may be used to improve workplace safety by automating risk assessment and detection operations. By recognizing and monitoring things and people in an environment, object detection and segmentation can help to improve workplace security. An object detection system can be used to scan an environment for possible dangers. Security professionals can be informed when a possible danger is discovered and take the necessary action by recognizing and monitoring items and individuals in a monitored area.

Artificial intelligence is being used to improve workplace safety in a variety of ways. AI tools are employed to collect, process, and store real-time data and safety information. Automatic categorization algorithms are used by safety analysts to detect accidents or events (or possible accidents) in a certain location or with specific personnel. These algorithms may also be used to identify possible risks by analyzing certain patterns of behavior, such as those that signal exhaustion or stress.

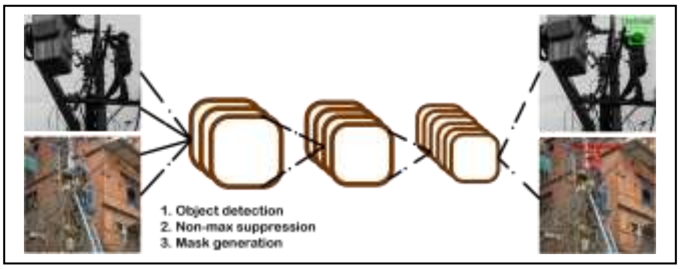

Fig. 1 Overview

A. *Need For Ml In Workplace Safety*

When it comes to sensitive situations, workplace safety is extremely important for the following example: -

- Dangerous chemical spills in the chemical industry
- On construction site employees not wearing safety helmets
- In the lab researchers not wearing safety jackets and glasses.

Human surveillance isn't even practical in some places, like: -

- Nuclear power plants
- In vast places where a lot of labor is required for surveillance, humans watching an environment for a long time causes many errors.

On the other hand, machine algorithms can be used for long-term surveillance with higher accuracy.

B. *Objective Of The Project*

The goal of workplace safety is to keep employees safe while they are at work. This includes the following: -

- Provide a safe working environment.
- Ensuring that employees are wearing proper safety equipment.
- Preventing the organization from being negatively impacted.
- Preventing harm to employees.
- Protecting organizational property due to accidents.
- Eliminating the death rate due to workplace accidents.

II. LITERATURE SURVEY

Annotated pictures are required for training in a deep learning-based workplace safety strategy. Engineers have difficulties when annotating photos with annotations indicating violations of safety norms. The true-negative rate of majority vote-based crowdsourcing annotation is low.

Artificial intelligence is increasingly being utilized to detect safety infractions in public spaces. A Bayesian network model is utilized to predict the likelihood of an accident occurring. The model predicts the likelihood of future incidents using data from previous accidents [1]. AI can detect trends that people may miss, such as harmful working conditions or defective equipment. To effectively detect safety violations in public spaces using the input data, a Bayesian network model may significantly raise the genuine negative rate of annotation.

This strategy can be utilized in our model to determine the probability of the predicted bounding box, in the model we utilize this data to perform non-max suppression to discover the best-fit bounding box for a specific item.

The main challenge with picture segmentation is that training time on a high-quality image dataset is long. This technique addresses this issue by stacking the entire problem into four layers.

The split/merge technique is used to suggest a four-layer picture segmentation procedure. The watershed algorithm is used to divide a picture into many sections in the first layer [2]. In the second layer, a co-evolutionary mechanism is used to produce final segment centers by combining related main areas. In the third layer, a meta-heuristic method connects the remaining areas to their corresponding determined centers using two operators. An evolutionary method is utilized in the last layer to integrate the generated comparable and nearby areas.

This model is used to comprehend the divide and concur technique, which swiftly generates predictions on the input image. This method also provides predictions on a wide range of images, including greyscale, colorful, and even textured images.

Due to picture speckles and pixel similarities, image segmentation is extremely challenging in crowded tiny images. This suggested method employs a Convolutional neural network (CNN) to build a region-based segmentation that is effective at recognizing segmental events in pictures with a high degree of pixel similarity.

Convolutional neural networks are fed images as input. In architecture, image input is given as a matrix, and the input enables regions to be analyzed by filtering and shifting scanning the whole matrix [3]. This is how the working time and savings from convolution computations are calculated. The pooling strategy is used to mix images while the convolution procedure is used to scan the image. The final stage produces Rectifier Linear Negative Interference that Interferes with the Image (ReLU). Completely Connected Layer. The prediction layer (FCN) is reached, and an estimate of the item or area is generated.

The region-based object detection is a fundamental building block for developing a Mask region-based convolutional neural network. The main difference between these architectures is that it loses a lot of information during the pooling process, resulting in the generation of harsh edges on the feature map so we use RoI Align for the pooling process which does not produce any data loss with the help of non-quantized strides.

Image segmentation may be carried out using a variety of approaches with varying degrees of accuracy and precision. This study compares techniques used for Image Segmentation Based on Edge Detection.

The proposed image segmentation model is based on edge detection and a convolutional neural network.

The suggested approach employs edge detection to determine the borders of various objects in an input picture. It then uses a convolutional neural network to categorize the various items based on their borders and appearance [4]. The architecture of the network is made up of convolutional layers, a pre-trained model, and two fully-connected layers with a SoftMax layer in between.

This study compares the impact of different thresholds on the canny operator. As a result, these tests receive varied results. Canny outperforms the other approaches, and the low and high thresholds can affect edge accuracy.

Object detection and image segmentation in real-time with high-resolution resources is difficult in complicated crowded contexts. This suggested methodology employs an ideal strategy to generate a class, bounding box, and mask for the identified item in real-time on extremely high-resolution resources.

A precise mask region convolutional neural network (precise Mask R-CNN) is proposed for object recognition and instance segmentation in high-resolution images. This model makes use of RoI Align, which employs a non-quantized stride to provide lossless feature map pooling [5]. Each time an item appears in the image, this model creates segmentation masks and bounding boxes for that particular occurrence.

This model is used to identify infractions in the workplace, such as determining if an individual is wearing a helmet correctly on a construction site. The algorithm's object recognition component aids in locating the anomaly-causing item, and its mask creation component is used to draw attention to the specific anomaly-causing event.

III. PROPOSED METHODOLOGY

Workplace security with humans will be labor-intensive, yet fault tolerance will be high. This proposed system uses data science and machine learning methods to monitor its surroundings for abnormalities using object detection and image segmentation algorithms. The initial stage in image segmentation is object detection; at this point, anomalies are also discovered, and a mask must be constructed; the major goal of instantaneous image segmentation is to emphasize the event that triggers the anomaly and the segmentation is carried out with the architecture of Mask Region-Based Convolutional Neural Network which takes input frame detected by the object detection phase of the algorithm.

A. Object Detection Phase

The video stream from the surveillance camera is split into individual frames and then resized to fit the input frame size supported by the constructed Mask Region-Based Convolutional Neural Network. The parsed frames are passed through a series of convolutional layers to generate a feature map for those input frames. Using the feature map, object detection can be carried out using the Region Proposal Network (RPN), Then it is passed to the RoI Align then to CNN architecture to classify the object, and regression to define the boundary of the detected object. It is qualified if the predicted boundary box and the actual boundary box have more than 50% value on Intersection Over Union (IOU) during the training phase. Non-max suppression is employed because the RPN will predict several boundary boxes for a single object. Non-max suppression looks at the output accuracy for each bounding box for an object and selects the box with the highest accuracy as the resultant bounding box for that specific class.

B. Segmentation Phase

The segmentation section takes the feature maps and passes them to the RoI Align for pooling, which resizes the feature map created by the Region Proposal Network (RPN) to a fixed-size feature map before passing them to the fully connected convolution layer for image segmentation. The segmentation algorithm is a binary classification algorithm that takes each object in the feature map and generates a binary mask for the object and the background. Finally, the object detection step of the algorithm returns the class and the boundary box, and the mask is applied on top of the object.

C. Advantage of the proposed system over the existing system

The proposed system can be implemented in a workplace for cost-effective monitoring, reducing the amount of labor required for surveillance. Some existing methodologies only detect abnormal workplace instances, but the proposed system highlights the instances with the mask generation part of the algorithm and generates real-time alerts.

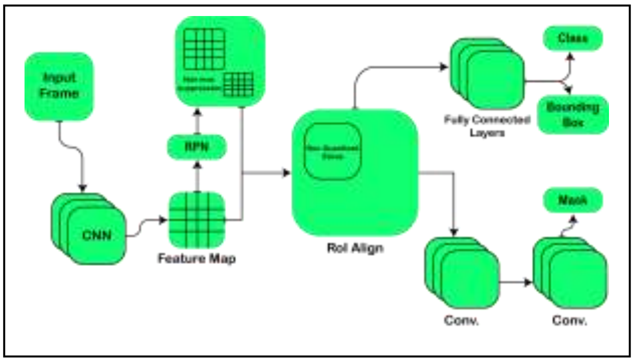

Fig. 2. System Architecture

IV. SYSTEM STRUCTURE

The proposed system takes the camera feed to manipulate each frame to the desired format then pass to the machine learning model to classify and segment the image frame finally classification of alerts and warnings is passed to the dashboard of the application.

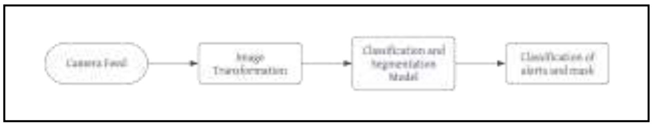

Fig. 3. General WorkFlow

A. Admin Dashboard

The system administrator may construct a new host network for their organization and then configure all of the surveillance cameras. Once configured, the administrator can watch live forecasts and can receive alerts & warnings.

B. Employees Dashboard

Each employee in an organization may connect to their organizational network to receive live feeds, the location of the anomaly that causes events, alerts, and warnings.

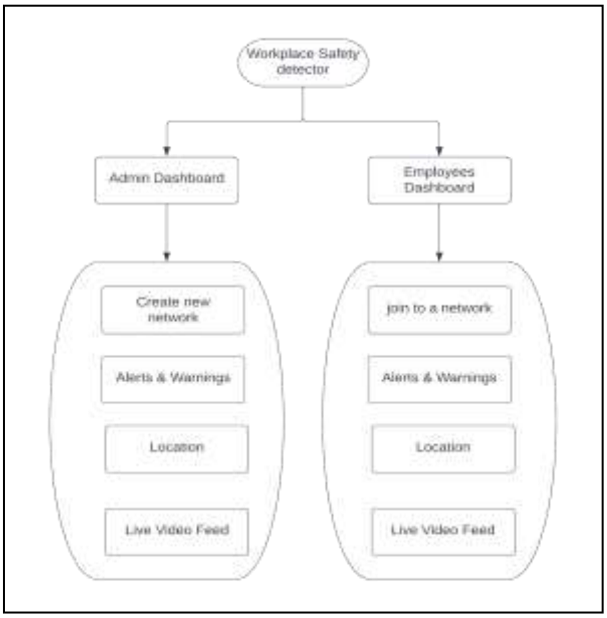

Fig. 4. Application Dashboard

C. *Mask RCNN To Generate Alerts and Mask*

The processed image from the surveillance camera is fed into the Mask Region-Based Convolutional Neural Network for the generation of the mask, boundary box, and classification of alerts, warnings, and suggestions and to forecast prediction.

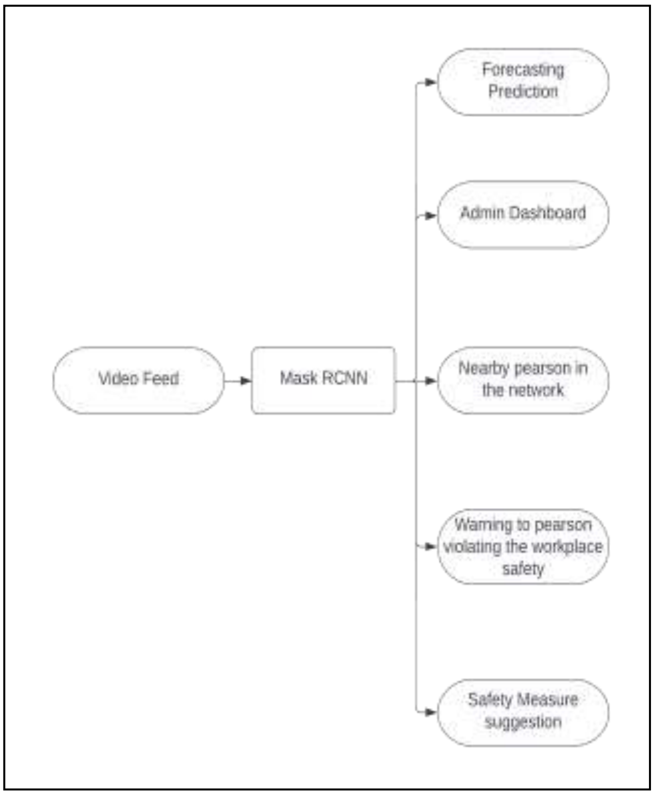

Fig. 5. Algorithm's Outcome

V. Module Description And Performance

The methodology proposes several layers to provide great modularity

- Image Transformation
- Object Detection
    - Feature Map Generation
    - Region Proposed Network
    - Non-max Suppression
    - Classification
    - Regression
- Image Segmentation
    - RoI Align
    - FC Layer (Mask Generation)
- Desktop Application (Server)
- Mobile Application (Users in the network)

D. *Image Transformation*

The process of extracting, resizing, and recoloring a picture that is in the desired output format for a trained model is known as image transformation.

The following processes are done in the image transformation module: -

- The video feed from the surveillance camera is collected and the frame rate is trimmed to the desired level
- The FPS trimmed video then pass for extraction of each frame in the video.
- The extracted frames or images are converted to greyscale by normalizing the RGB values between 0 to 255
- The grayscale image is then resized to match the required input resolution by the trained model.
- The reshaped images are then flattened to load into a data loader.
- Then the appropriate images with their paired annotation

E. *Object Detection*

Detecting instances of semantic objects of a specific class (such as people, buildings, or cars) in digital images and videos is the task of object detection, a branch of computer science related to computer vision and image processing. Well-researched object detection domains include pedestrian detection and face detection. Application areas for object identification in computer vision include picture retrieval and video surveillance.

The following process is carried out in the object detection module: -

- Feature Map Generation:

Apply filters or feature detectors to the input picture using the activation function to build feature maps or activation maps. Feature detectors or filters aid in the identification of various characteristics in a picture, such as edges, vertical lines, horizontal lines, bends, and so on.

- Region Proposed Network:

A Region Proposal Network, or RPN, is a fully convolutional network that predicts object limits and scores at each place at the same time. The RPN is fully trained to create high-quality region suggestions. RPNs are intended to anticipate region proposals with a broad range of sizes and aspect ratios efficiently. Anchor boxes are used as references in RPNs at various scales and aspect ratios. The technique is conceptualized as a pyramid of regression references that avoids enumerating pictures or filters with different scales or aspect ratios.

- Non-max suppression:

Non-Max Suppression (NMS) is a computer vision approach that is employed in a variety of jobs. It is a type of algorithm that selects one entity (e.g., bounding boxes) from a set of overlapping entities. To achieve the desired outcomes, we may pick the selection criteria. Most typically, the requirements are some type of probability number and some form of overlap measure (e.g., Intersection over Union).

- Classification

The pooled feature map is used to classify the object that which class it belongs.

- Regression

The detected object will have a bounding box this was detected by the regression.

F. *Image Segmentation*

Image segmentation is splitting a digital image into several image segments, also known as image regions or objects, in digital image processing and computer vision (sets of pixels). The purpose of picture segmentation is to simplify and/or transform an image's representation into something more relevant and easier to examine. Image segmentation is commonly used to find objects and boundaries (lines, curves, and so on) in pictures. Picture segmentation is the process of labeling every pixel in an image so that pixels with the same label have specific properties.

G. *Desktop Application*

The desktop application is used to process the input feed in real-time with the proposed machine learning algorithm and push the prediction to the user end (Mobile application) through the cloud.

H. *Mobile Application*

The mobile app fetches information from the cloud database and storage to display the alerts, and live video feed and to specify the location where the anomaly is detected

I. *Performance*

epoch	loss	val_loss
1	0.074437419	0.062248869
2	0.051469764	0.031099439
3	0.029761078	0.036314331
4	0.0263856	0.022122867
5	0.022803274	0.023090442
6	0.020320808	0.018060902
7	0.017557671	0.006110405
8	0.01474829	0.018123651
9	0.017121169	0.015766486
10	0.015634849	0.015485728

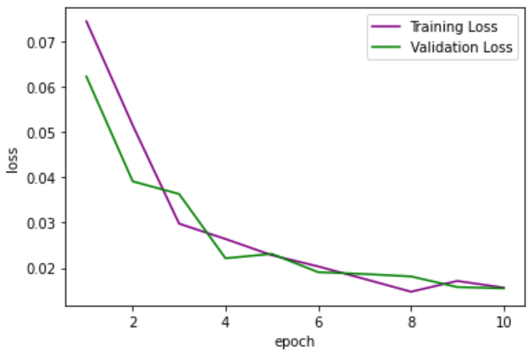

Fig. 6. Graph Epoch Vs Loss

VI. Conclusion

AI technology can assist to improve safety by automating procedures and delivering alarms, when possible, risks are recognized; but, in bigger areas, human-based surveillance may not be effective so machine learning algorithms can be utilized to rapidly discover patterns and trends in workplace safety incidents, assisting in the addresses these issues of potential safety dangers. This could result in the development of preventative measures and the implementation of improved safety regulations.

References

[1] Y. Wang, P.C. Liao, C. Zhang, Y.i. Ren, X. Sun, P. Tang Crowdsourced reliable labeling of safety-rule violations on images of complex construction scenes for advanced vision-based workplace safety Adv. Eng. Inf., vol. 42 p. 101001, 2019, 10.1016/ j.aei.2019.10100.

[2] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.

[3] R. Yayla and B. Şen, "Region-based Segmentation of Terrain Fields in SAR Images," *2020 28th Signal Processing and Communications Applications Conference (SIU)*, 2020, pp. 1-4, doi: 10.1109/SIU49456.2020.9302432.

[4] W. Zhou, X. Du and S. Wang, "Techniques for Image Segmentation Based on Edge Detection," 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), pp. 400-403, 2021, doi: 10.1109/CEI52496.2021.9574569.

[5] H. Su, S. Wei, M. Yan, C. Wang, J. Shi and X. Zhang, "Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 1454-1457, doi: 10.1109/IGARSS.2019.8898573.

[6] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Autom. Constr., vol. 35, pp. 131-141, 2013, https://doi.org/10.1016/ j.autcon.2013.05.001.

[7] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[8] Shaoqing Ren et al., "Faster R-CNN: towards real-time object detection with region proposal networks", International Conference on Neural Information Processing Systems, pp. 91-99, 2015.

[9] Kaiming He et al., "Mask R-CNN", IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 99, pp. 1-1, 2017.

[10] JeongbinHwang et al., "Development of training image database using web crawling for vision-based site monitoring", vol. 135, 2022

[11] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp. 1440-1448, 2015, doi: 10.1109/ICCV.2015.169.

[12] V. Tiwari, C. Pandey, A. Dwivedi and V. Yadav, "Image Classification Using Deep Neural Network," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 730-733, 2020, doi: 10.1109/ICACCCN51052.2020.9362804.

[13] Wang Liwei, Li Jiming, Zhou Guomin and Yang Dongyong, "Application of depth transfer learning in hyperspectral image classification[J/OL]", Computer engineering and application, pp. 1-8, 01 2019.

[14] J. Wang, Y. Wu and Y. Mei, "Insulator instance segmentation based on deep learning network Mask RCNN," 2022 China International Conference on Electricity Distribution (CICED), Changsha, China, pp. 1424-1427, 2022, doi: 10.1109/CICED56215.2022.9929201.

[15] A. Shihavuddin, M. K. Kiron, M. I. Islam, M. H. Maruf, R. H. Ashique and S. M. Kabir, "Cascaded 3-Stage Nuclei Segmentation Using U-net, Faster-RCNN and SegNet for Higher Precision," 2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, pp. 1-5, 2021, doi: 10.1109/STI53101.2021.9732598.