International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AI DRIVEN REAL TIME PADESTRIAN DETECTION TO PREVENT TRAFFIC ACCIDENTS

[1]D .shirisha, [2]Mr.A. sudheer

[1]M.Tech Scholar in Department of CSE Sri Indu College of Engineering And Technology. Hyderabad
[2]Assistant .Professor in Department of CSE Sri Indu College of Engineering And Technology. Hyderabad

Abstract

Object detection and recognition have become important research areas in the fields of Artificial Intelligence, Computer Vision, and Deep Learning due to their wide range of real-world applications such as autonomous vehicles, surveillance systems, medical imaging, industrial automation, and smart city monitoring. Traditional object detection systems mainly rely on handcrafted feature extraction methods and classical machine learning algorithms, which often fail under complex environmental conditions such as poor lighting, background noise, occlusion, and scale variation. Recent advancements in Deep Learning and Convolutional Neural Networks (CNNs) have significantly improved the accuracy and efficiency of object detection systems. This research presents an intelligent object detection and recognition framework using Convolutional Neural Networks (CNNs), AlexNet architecture, Support Vector Machine (SVM), Non-Maximum Suppression (NMS), Intersection over Union (IOU), and backpropagation algorithms. The proposed system integrates deep feature extraction and classification mechanisms to identify objects accurately from images and video streams. CNN-based feature extraction automatically learns discriminative visual features while AlexNet improves image representation learning. SVM classifiers are used for efficient object classification and recognition. NMS eliminates duplicate detection boxes, and IOU is used for bounding box evaluation and localization accuracy. The proposed framework improves detection performance and reduces false positives under challenging scenarios. Experimental analysis demonstrates that the system achieves higher mean Average Precision (mAP), lower miss rates, and improved object localization accuracy compared with traditional object detection approaches such as HOG+SVM and DPM methods. The research highlights the importance of deep learning architectures in developing scalable and intelligent object recognition systems for real-time applications.

Keywords— Object Detection, Convolutional Neural Networks, AlexNet, Support Vector Machine, Deep Learning, Computer Vision, Non-Maximum Suppression, Intersection over Union, Artificial Intelligence, Image Recognition.

I. INTRODUCTION

Object detection and recognition are fundamental tasks in computer vision and artificial intelligence applications. The primary objective of object detection systems is to identify and localize objects within digital images and video streams accurately. These technologies are widely used in surveillance systems, autonomous vehicles, industrial automation, robotics, healthcare systems, facial recognition, and smart traffic management applications.

Traditional object detection methods mainly rely on handcrafted feature extraction techniques such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Haar features, and classical machine learning classifiers. Although these methods perform reasonably well in controlled environments, they fail under challenging conditions such as occlusion, illumination changes, background clutter, and object scale variation. Manual feature engineering also limits the adaptability and scalability of traditional object detection systems.

Recent advancements in Artificial Intelligence (AI) and Deep Learning (DL) have significantly transformed the field of object detection. Deep learning models, especially Convolutional Neural Networks (CNNs), can automatically learn hierarchical feature representations directly from raw image data. CNN-based architectures have achieved remarkable performance in image classification, segmentation, and object detection tasks.

Krizhevsky et al. introduced AlexNet, a deep convolutional neural network architecture that achieved groundbreaking performance in image classification tasks on the ImageNet dataset [1]. AlexNet demonstrated the effectiveness of deep neural networks in extracting complex image features automatically.

Girshick et al. proposed the Region-Based Convolutional Neural Network (R-CNN) framework, which integrated region proposals and CNN-based feature extraction for object detection tasks [2]. Fast R-CNN and Faster R-CNN further improved detection speed and localization accuracy [3], [4].

Redmon et al. introduced YOLO (You Only Look Once), a real-time object detection framework capable of performing fast and accurate object detection from images and videos [5]. SSD (Single Shot Detector) also improved real-time object detection efficiency

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

by performing object localization and classification simultaneously [6].

The proposed research introduces an intelligent object detection framework using CNN, AlexNet, Support Vector Machine (SVM), Non-Maximum Suppression (NMS), IOU evaluation metrics, and backpropagation algorithms. The framework combines deep feature extraction and machine learning classification mechanisms to improve object detection accuracy and localization performance.

The system processes digital images and video streams, extracts visual features using CNN architectures, classifies detected objects using SVM classifiers, and evaluates localization accuracy using IOU metrics. NMS is used to eliminate duplicate bounding boxes and improve final detection quality.

The proposed framework aims to improve detection accuracy, reduce miss rates, support real-time processing, and provide scalable object recognition capabilities for intelligent surveillance and monitoring applications.

II. LITERATURE SURVEY

Several researchers have contributed significantly to the field of deep learning-based object detection and image recognition systems. Traditional computer vision methods focused on handcrafted feature extraction and classical machine learning techniques for object classification tasks.

Dalal and Triggs proposed Histogram of Oriented Gradients (HOG) feature descriptors combined with Support Vector Machine classifiers for human detection applications [7]. Their work became one of the earliest successful approaches for object recognition.

Krizhevsky et al. introduced AlexNet, a deep convolutional neural network architecture consisting of multiple convolutional layers, pooling layers, and fully connected layers [1]. AlexNet achieved state-of-the-art performance in image classification and demonstrated the effectiveness of GPU-based deep learning training.

Girshick et al. proposed R-CNN, which combined region proposal algorithms with CNN-based feature extraction for object detection [2]. The approach improved detection accuracy significantly compared with traditional feature-based methods.

Fast R-CNN and Faster R-CNN architectures further improved object detection speed and localization efficiency by integrating region proposal networks and end-to-end training mechanisms [3], [4].

Redmon et al. introduced YOLO, a unified real-time object detection framework that performs object localization and classification

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

simultaneously [5]. YOLO significantly improved detection speed and enabled real-time applications such as autonomous driving and video surveillance.

Liu et al. proposed SSD (Single Shot MultiBox Detector), which improved detection speed and multi-scale object localization performance [6].

Several researchers also explored CNN and SVM hybrid architectures for image classification tasks. CNN models automatically learn deep visual features, while SVM classifiers improve decision boundaries for object classification.

Non-Maximum Suppression (NMS) algorithms are widely used for eliminating duplicate detection boxes and improving localization accuracy. IOU (Intersection over Union) metrics are commonly used for evaluating bounding box overlap and object localization quality.

Although existing deep learning models achieve high accuracy, many systems still face challenges such as computational complexity, real-time processing limitations, object occlusion, and small object detection difficulties.

III. EXISTING SYSTEM

Existing object detection systems mainly rely on traditional computer vision algorithms and machine learning techniques. Earlier systems used handcrafted feature extraction methods such as HOG, Haar features, and SIFT combined with classifiers such as SVM and AdaBoost.

These systems required manual feature engineering and performed poorly under varying environmental conditions. Traditional methods struggled with:

- Illumination changes
- Scale variations
- Object occlusion
- Background noise
- Complex scenes
- Real-time processing limitations

Region proposal methods such as Selective Search improved object localization but introduced computational overhead. Classical CNN-based systems improved recognition accuracy but often suffered from slow detection speed and duplicate bounding box generation.

Existing systems also faced challenges such as:

- Increased false positive detections
- Poor localization accuracy
- High training complexity
- Limited adaptability
- Reduced detection performance in crowded environments

These limitations motivated the development of intelligent deep learning frameworks capable of improving object recognition and localization performance.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IV. PROBLEM STATEMENT

Traditional object detection systems fail to provide accurate and efficient object recognition under challenging real-world conditions such as poor lighting, background clutter, object occlusion, and crowded environments. Existing systems require extensive manual feature engineering and often suffer from high false positive rates, localization errors, and computational complexity.

There is a need for an intelligent object detection framework capable of automatically learning visual features, accurately localizing objects, reducing duplicate detections, and supporting real-time surveillance and monitoring applications. The proposed research aims to develop a CNN-based object detection system integrating AlexNet, SVM, IOU, and NMS mechanisms for improved recognition performance and localization accuracy.

V. PROPOSED SYSTEM

The proposed system introduces an intelligent object detection and recognition framework using Convolutional Neural Networks (CNNs), AlexNet architecture, Support Vector Machine classifiers, Non-Maximum Suppression (NMS), and IOU evaluation mechanisms.

The framework processes input images and video streams for object localization and recognition tasks. Initially, the system performs image preprocessing operations such as resizing, normalization, and noise reduction. CNN architectures extract deep visual features automatically from input images.

AlexNet architecture is used for feature representation learning due to its deep convolutional structure and efficient image classification performance. The extracted features are passed to SVM classifiers for object categorization and recognition.

The system generates bounding boxes around detected objects and evaluates localization accuracy using Intersection over Union (IOU) metrics.

□genui□{“math_block_widget_always_prefetch_v2”:{“content”:“IOU=\frac{\text{Area of Overlap}}{\text{Area of Union}}”}}□

Non-Maximum Suppression (NMS) eliminates duplicate detection boxes and retains only the highest confidence detections.

The proposed framework provides several advantages:

- Improved object detection accuracy
- Reduced false positives
- Better localization performance
- Real-time detection support
- Automatic feature extraction
- Scalability for large datasets
- Efficient surveillance monitoring

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

VI. METHODOLOGY

The proposed methodology consists of several stages for performing object detection and recognition efficiently.

Initially, image datasets containing multiple object categories are collected and preprocessed. Image preprocessing operations include resizing, normalization, grayscale conversion, and noise reduction.

Convolutional Neural Networks extract deep visual features automatically from images. CNN layers perform convolution, activation, pooling, and feature learning operations to capture hierarchical image patterns.

AlexNet architecture is used for feature extraction and deep representation learning. The network contains multiple convolutional layers, pooling layers, ReLU activation functions, and fully connected layers for efficient image analysis.

Backpropagation algorithms update neural network weights during training to minimize classification loss and improve recognition performance.

Support Vector Machine classifiers are used for object classification based on extracted CNN features. SVM improves classification boundaries and enhances recognition accuracy.

Bounding boxes are generated for detected objects. Intersection over Union (IOU) metrics evaluate localization quality and bounding box overlap.

□genui□{“math_block_widget_always_prefetch_v2”:{“content”:“mAP=\frac{1}{N}\sum_{i=1}^{N}AP_i”}}□

Non-Maximum Suppression removes redundant bounding boxes and retains optimal detections based on confidence scores.

The proposed framework supports image-based and video-based object detection operations for surveillance and intelligent monitoring applications.

VII. IMPLEMENTATION

The proposed system is implemented using Python programming language and deep learning frameworks such as TensorFlow, Keras, and OpenCV. Jupyter Notebook and Visual Studio Code are used as development environments.

The implementation begins with dataset preparation and image preprocessing operations. Images are resized and normalized before being passed into CNN architectures.

AlexNet architecture is implemented using convolutional layers, pooling layers, ReLU activations, dropout layers, and fully connected layers. CNN models are trained using labeled image datasets for object recognition tasks.

Support Vector Machine classifiers are integrated with CNN feature extraction layers for efficient classification performance. The training process uses backpropagation algorithms and gradient optimization techniques to improve network accuracy.

The system also implements:

- Bounding box generation
- IOU evaluation
- Non-Maximum Suppression
- Real-time video detection
- Performance visualization
- Detection confidence analysis

OpenCV is used for image acquisition, video frame processing, and real-time object visualization. The implementation supports scalable image and video processing for intelligent surveillance applications.

VIII. RESULTS

The proposed object detection framework demonstrated significant improvements in recognition accuracy and localization performance compared with traditional object detection systems.

Experimental analysis showed that CNN-based feature extraction improved detection efficiency under varying environmental conditions such as poor lighting, background clutter, and object occlusion. AlexNet successfully learned discriminative visual features from image datasets and improved classification performance.

The integration of SVM classifiers improved object recognition boundaries and reduced false positive detections. Non-Maximum Suppression effectively eliminated duplicate bounding boxes and improved localization quality.

IOU evaluation metrics demonstrated higher localization accuracy and improved overlap performance. The system achieved higher mean Average Precision (mAP) and lower miss rates compared with HOG+SVM and DPM-based methods.

The proposed framework also supported real-time object detection operations suitable for:

- Intelligent surveillance systems
- Autonomous vehicles
- Smart traffic monitoring
- Security applications
- Crowd analysis systems

The results confirmed that deep learning-based object detection systems significantly outperform traditional computer vision methods in terms of accuracy, scalability, and robustness.

IX. CONCLUSION

This research presented an intelligent object detection and recognition framework using Convolutional Neural Networks, AlexNet architecture, Support Vector Machine classifiers, Non-Maximum Suppression, and IOU evaluation techniques.

The proposed system successfully integrated deep feature extraction and machine learning classification mechanisms for improving object detection accuracy and localization performance. CNN architectures automatically learned hierarchical visual features while AlexNet improved representation learning capabilities.

The integration of SVM classifiers enhanced object recognition performance and reduced classification errors. IOU metrics and NMS mechanisms improved bounding box localization and eliminated redundant detections.

Experimental analysis demonstrated that the proposed framework achieved higher mean Average Precision and lower miss rates compared with traditional object detection approaches such as HOG+SVM and DPM methods.

The research highlights the effectiveness of deep learning technologies in developing scalable and intelligent object detection systems suitable for surveillance, autonomous driving, smart cities, and security monitoring applications.

Future enhancements such as YOLO, SSD, EfficientNet, Vision Transformers, multi-object tracking systems, cloud-based surveillance integration, and embedded AI deployment using Raspberry Pi and NVIDIA Jetson Nano can further improve real-time object detection performance and scalability.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in *Proceedings of CVPR*, 2014.

[3] R. Girshick, “Fast R-CNN,” in *Proceedings of ICCV*, 2015.

[4] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified Real-Time Object Detection,” in *Proceedings of CVPR*, 2016.

[6] W. Liu et al., “SSD: Single Shot MultiBox Detector,” in *European Conference on Computer Vision (ECCV)*, 2016.

[7] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *Proceedings of CVPR*, 2005.

[8] K. He et al., “Deep Residual Learning for Image Recognition,” in *Proceedings of CVPR*, 2016.

.

[9] C. Szegedy et al., “Going Deeper with Convolutions,” in *Proceedings of CVPR*, 2015.

[10] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *International Conference on Learning Representations (ICLR)*, 2021