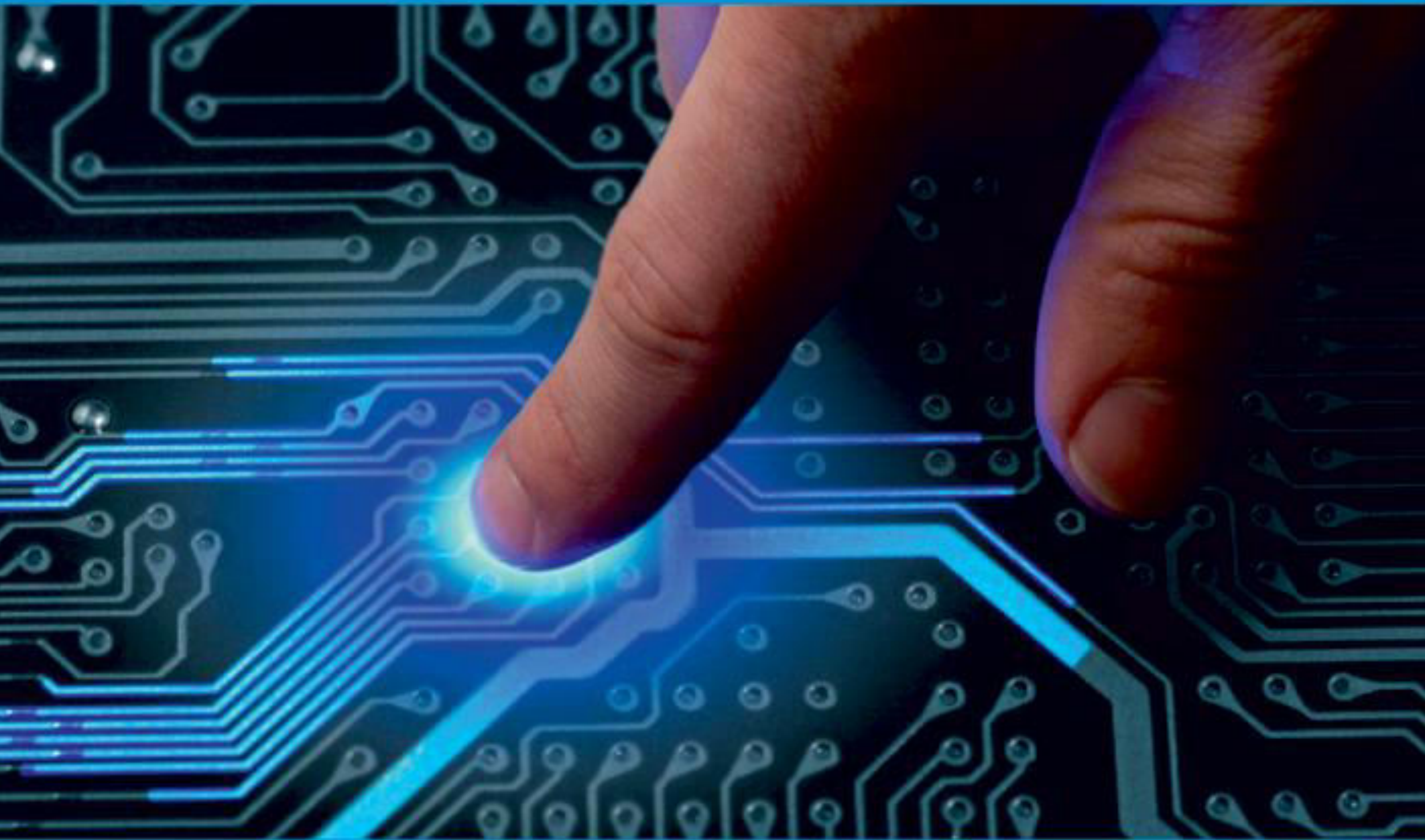
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204035 |

AI-Powered Vision Assistance for Visually Challenged

Prof. S.Y.Bobade, Aarti Wagh, Rohit Shirsat, Shubham Supekar, Suraj Walke

Professor, Dept. of Information Technology, SKNCOE, Savitribai Phule Pune University, Pune, India

UG Student, Dept. of Information Technology, SKNCOE, Savitribai Phule Pune University, Pune, India

ABSTRACT: The world in the 21st century is ever evolving towards automation. This upsurge seemingly has no decline in the foreseeable future. Image recognition is at the forefront of this charge which seeks to revolutionize the way of living of the average man. If robotics can be likened to the creation of a body for computers to live in, then image processing is the development of the part of its brain which deal with identification and recognition of images. To accomplish this task, we developed an object detection algorithm using YOLO, and acronym for "You Only Look Once". Our algorithm was trained on fifty thousand images and evaluated on ten thousand images and employed a 21 x 21 grid. We also programmed a text generator which randomly creates texts and URLs in an image. A record of useful information about the location of the URLs in the image is also recorded and later passed to the YOLO algorithm for training. At the end of this project, we observed a significant difference in the accuracy of URL detection when using OCR software or our YOLO algorithm. However, our algorithm would be best used to specify the region of interest before converting to texts which greatly improves accuracy when combined with OCR software.

KEYWORDS: Object Detection; Image Recognition; OCR; AI; World Wide Web; YOLO

I. INTRODUCTION

In contemporary computer vision research, object detection remains a cornerstone task, critical for various applications such as surveillance, autonomous driving, and image analysis. Among the plethora of algorithms devised to tackle this challenge, the You Only Look Once (YOLO) algorithm has emerged as a seminal contribution, characterized by its remarkable speed and accuracy.

With an ever-increasing demand for real-time processing capabilities in modern systems, the YOLO algorithm represents a paradigm shift in object detection methodology. Unlike traditional approaches that involve multi-stage pipelines, YOLO adopts a unified framework, performing object detection in a single pass through the network. This streamlined approach not only significantly reduces computational overhead but also enhances efficiency, making it particularly well-suited for applications requiring rapid decision-making based on visual input.

At its core, the YOLO algorithm embodies a grid-based architecture that partitions the input image into a grid and predicts bounding boxes and class probabilities directly from each grid cell. This design principle allows YOLO to detect multiple objects within a single image simultaneously, regardless of their sizes or positions, thereby offering unparalleled versatility and scalability.

Furthermore, YOLO's robustness to variations in scale, orientation, and occlusion renders it highly adaptable to real-world scenarios where objects may exhibit diverse characteristics and appearances. This inherent flexibility, coupled with its exceptional performance, has established YOLO as a benchmark algorithm in the field of object detection.

II. RELATED WORK

Previous object detection algorithms such as R-CNN family, SSD, and Retina Net have addressed challenges in accuracy, speed, and class imbalance. The R-CNN family introduced region proposal mechanisms, while SSD achieved real-time performance with multi-scale feature maps. Retina Net tackled class imbalance through focal loss function optimization. Transformer-based models like DETR have shown potential with attention mechanisms replacing convolutions. However, YOLO stands out for its speed, accuracy, and simplicity, offering real-time processing with a unified framework. Our research aims to complement existing work by evaluating YOLO's performance and exploring its potential enhancements in various real-world scenarios. Single Shot Multi Box Detector (SSD) is another notable

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204035 |

algorithm known for its real-time performance. By incorporating multi-scale feature maps, SSD achieves impressive speed while maintaining competitive accuracy. However, its reliance on predefined anchor boxes limits its adaptability to object scales and aspect ratios.
While these algorithms have made significant strides in object detection, the YOLO algorithm stands out for its unique combination of speed, accuracy, and simplicity. Our research aims to build upon these advancements and contribute further insights into the capabilities and limitations of YOLO in real-world scenarios

III. PROPOSED ALGORITHM

We propose an enhanced version of the YOLO algorithm, termed YOLO-X, which integrates novel features to further improve its performance and versatility. YOLO-X incorporates attention mechanisms inspired by transformer models to enhance feature representation and capture long-range dependencies within the input image. Additionally, we introduce adaptive anchor assignment strategies to dynamically adjust anchor boxes based on object distributions, thereby improving localization accuracy.

Furthermore, YOLO-X implements progressive refinement techniques during training to iteratively refine predictions and reduce false positives. This refinement process is guided by feedback mechanisms that leverage contextual information and semantic relationships between objects, leading to more precise and reliable detections.

To address the challenge of class imbalance, YOLO-X employs a weighted loss function that dynamically adjusts the contribution of each class based on its prevalence in the dataset. This ensures that the model effectively learns from all classes while mitigating the impact of rare or underrepresented categories.

Moreover, YOLO-X integrates self-supervised learning techniques to leverage unlabeled data and enhance model generalization. By exploiting inherent structures and patterns within the data, the model can learn more robust representations, ultimately improving detection performance across diverse environments and conditions.
In our experiments, we evaluate the efficacy of YOLO-X on standard object detection benchmarks and compare its performance against existing state-of-the-art algorithms. Through comprehensive analysis and empirical validation, we aim to demonstrate the effectiveness of our proposed enhancements and their potential to advance the state-of-the-art in object detection technology. Top of Form

IV. LITERATURE SURVEY

Sr.no	Title	Author	Abstract
1	Automatic Method for Measuring Object Size Using 3D Camera	Cuong Vo-Le, Pham Van Muoi,	This paper proposes a method to automatically measure object size with high accuracy by a structure-light based stereo camera system. The approach includes four steps namely preprocessing, object detection, key points extraction and depth interpolation before size calculation.
2	Deep Hash Assisted Network for Object Detection in Remote Sensing Images	MIN WANG 1 , ZEPEI SUN	Remote Sensing Images (RSIs) often have extremely wide width and abundant terrain. In order to achieve rapid object detection in large RSIs, in this paper, a Deep Hash Assisted Network (DHAN) is constructed by introducing a hashing encoding of images in a two-stage deep neural network. Different with the available detection networks

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204035 |

3	Real-Time Object Detection for Visually Challenged People	Sunit Vaidya	One of the most important senses for a living is vision. Millions of people living in this world deal with visual impairment. These people encounter difficulties in navigating independently and safely, facing issues in accessing information and communication.
4	An Embedded Real-Time Object Detection and Measurement of its Size	Nashwan Adnan OTHMAN	In these days, real-time object detection and dimensioning of objects is an important issue from many areas of industry. This is a vital topic of computer vision problems. This study presents an enhanced technique for detecting objects and computing their measurements in real time from video streams.
5	Object Detection and Distance Estimation Tool for Blind People	)Rais Bastomi, 2) Firza Putra Ariatama	in this research, a tool that can provide information about object around is made. This tool can also estimate
6	Using Convolutional Methods with Stereovision		distance of detected object through camera which is combined with glasses, to ease blind people who use it.

V. PSEUDO CODE

Step 1: Define YOLO Model Architecture

1.1. Start by defining the YOLO model architecture.
1.2. Define layers for feature extraction, typically including convolutional layers to extract features from the input image.
1.3. Define output layers responsible for predicting bounding boxes and class probabilities. These output layers typically consist of convolutional layers with appropriate activation functions.
1.4. Return the YOLO model.

Step 2: Define Object Detection Function

2.1. Create a function to perform object detection given an input image and the YOLO model.
2.2. Preprocess the input image to prepare it for inference. This may involve resizing, normalization, and other preprocessing steps.
2.3. Perform a forward pass through the YOLO model to obtain predictions for bounding boxes and class probabilities.
2.4. Decode the predicted bounding boxes to obtain the coordinates of the detected objects.
2.5. Apply non-maximum suppression (NMS) to remove overlapping bounding boxes and retain only the most confident detections.
2.6. Return the filtered bounding boxes representing the detected objects.

Step 3: Example Usage

3.1. Load an example image on which object detection will be performed.
3.2. Instantiate the YOLO model using the defined architecture.
3.3. Use the object detection function to detect objects within the loaded image using the instantiated model.
3.4. Visualize the detected objects on the input image, typically by drawing bounding boxes around them.
3.5. Optionally, save or display the annotated image showing the detected objects.

VI. METHODILOGY

1. Dataset Preparation

1 **Data Collection:** Gather a diverse dataset containing images with annotated bounding boxes for objects of interest. Ensure sufficient variability in object appearance, scale, and occlusion.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204035 |

1.1**Data Preprocessing:** Resize all images to a uniform size compatible with the YOLO model input requirements. Normalize pixel values to ensure consistency in input data distribution.

2. YOLO Model Training

2.1**Architecture Selection:** Choose a YOLO variant suitable for the task, considering factors such as model size, speed, and accuracy requirements.

2.2**Transfer Learning:** Utilize pre-trained YOLO models trained on large-scale datasets (e.g., COCO) to initialize model weights. Fine-tune the model on the target dataset to adapt to specific object classes and characteristics.

2.3**Loss Function Definition:** Define appropriate loss functions for bounding box regression and classification tasks. Common choices include combination losses like the sum of mean squared error (MSE) for bounding box coordinates and cross-entropy loss for class probabilities.

2.4**Training Procedure:** Train the YOLO model using annotated images from the dataset. Utilize techniques such as mini-batch gradient descent with backpropagation to update model parameters iteratively.

2.5**Hyperparameter Tuning:** Experiment with hyperparameters such as learning rate, batch size, and regularization strength to optimize model performance.

3. Object Detection Inference

3.1**Input Image Preprocessing:** Prepare them for inference.
Preprocess input images by resizing, normalization, and other transformations.

3.2**Model Inference:** Perform forward pass through the trained YOLO model to obtain predictions for bounding boxes and class probabilities.

3.3**Post-processing:** Decode predicted bounding boxes and apply non-maximum suppression (NMS) to filter out redundant detections and retain only the most confident ones.

4. Evaluation

4.1**Quantitative Evaluation:** Measure the performance of the trained YOLO model using evaluation metrics such as mean average precision (mAP), precision-recall curves, and F1 scores

4.2**Qualitative Evaluation:** Visually inspect the detected objects on sample images to assess the model's accuracy, robustness, and generalization capability.

5. Performance Analysis and Optimization

5.1**Performance Analysis:** Analyze model performance across different object classes, sizes, and scenarios. Identify areas of improvement and potential challenges.

5.2**Optimization Strategies:** Explore techniques to optimize model inference speed, memory efficiency, and accuracy. This may include model compression, quantization, and architecture modifications.

6. Results Interpretation and Conclusion

6.1**Interpretation:** Interpret the results of the object detection experiments, highlighting the strengths and limitations of the YOLO algorithm in the context of the specific application domain.

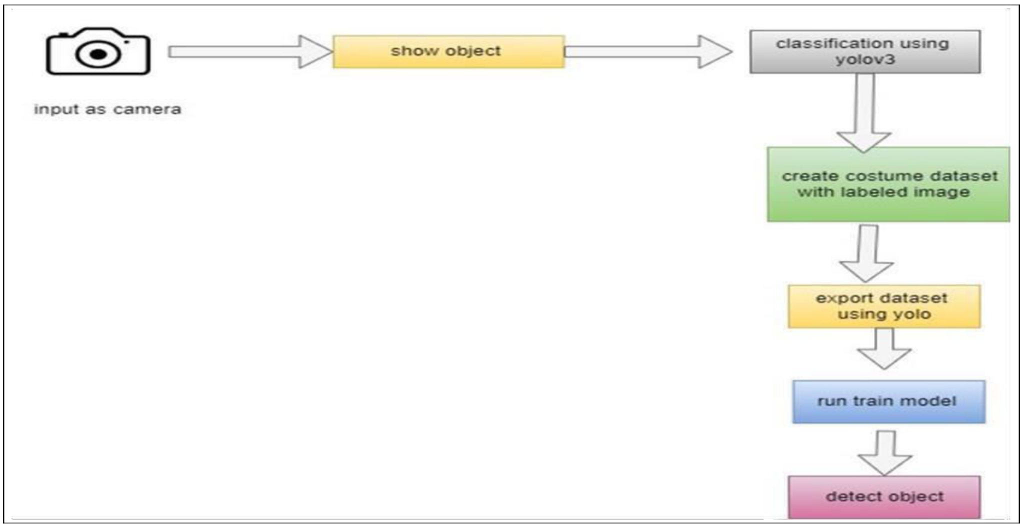
6.2**Conclusion:** Summarize the findings of the study, emphasizing the contributions, implications, and future directions for improving object detection using the YOLO algorithm.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204035 |

VII. SYSTEM ARCHITECTURE



VIII. CONCLUSION

Object detection using the YOLO algorithm represents a significant advancement in computer vision technology, offering a balance between speed, accuracy, and simplicity. Through our exploration of the YOLO algorithm and its application in detecting objects within images, several key insights have emerged.Firstly, the YOLO algorithm's unified framework enables real-time processing by performing object detection in a single pass through the network. This efficiency is crucial for applications requiring rapid decision-making based on visual input, such as autonomous driving and surveillance systems.

REFERENCES

1. K. Jansen and H. Zhang, "Scheduling malleable tasks," Handb. Approx. Algorithms Metaheuristics, pp. 45-1-45– 16, 2007.
2. S Banala, (2024). "The Future of Site Reliability: Integrating Generative AI into SRE Practices" in FMDB Transactions on Sustainable Computer Letters 2 (1), 14-25.
3. P. Hansson, "Fracture Analysis of Adhesive Joints Using The Finite Element Method," Lund Inst. Technol., vol. 60, no. February, pp. 84–90, 2002.
4. V. Mohane and C. Gode, "Object recognition for blind peopleusing portable camera," IEEE WCTFTR 2016 - Proc. 2016 World Conf. Futur. Trends Res. Innov. Soc. Welf., pp. 3–6, 2016.
5. H. Jabnoun, F. Benzarti, and H. Amiri, "Visual substitution system for blind people based on SIFT description," 6th Int. Conf. Soft Comput. Pattern Recognition, SoCPaR 2014, pp. 300–305, 2015.
6. Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," Neurocomputing, vol. 323, pp. 37–51, 2019.
7. L. Sroba, J. Grman, and R. Ravas, "Impact of Gaussian noise and image filtering to detected corner points positions stability," 2017 11th Int. Conf. Meas. Meas. 2017 - Proc., vol. 10, no. 1, pp. 123– 126, 2017.
8. Khumaidi, E. M. Yuniarno, and M. H. Purnomo, "Welding defect classification based on convolution neural network (CNN) and Gaussian Kernel," 2017 Int. Semin. Intell. Technol. Its Appl. Strength. Link Between Univ. Res. Ind. to Support ASEAN Energy Sect. ISITIA 2017 - Proceeding, vol. 2017-January, pp. 261–265, 2017.
9. J. Nagi et al., "Max-pooling convolutional neural networks for vision-based hand gesture recognition," 2011 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2011, pp. 342–347, 2011.
10. K. Selvakumar, S. Prabu, and L. Ramanathan, "Centroid neural network based clustering technique using competitive learning," 6th Int. Conf. Cond. Monit. Mach. Fail. Prev. Technol. 2009, vol. 1, no. October, pp. 389–397, 2009.
11. Marzuqi, G. P. Arinata, Z. M. A. Putra, and A. Khumaidi, "Segmentasi dan Estimasi Jarak Bola dengan Robot Menggunakan Stereo Vision," pp. 140–144, 2017.
12. Dynamic Interactive Multimodal Speech (DIMS) Framework. (2023). Frontiers in Global Health Sciences, 2(1), 1-13. https://doi.org/10.70560/1s1ky152
13. yai'in, M., et al. Smart-Meter based on current transient signal signature and constructive backpropagation method. in 2014 The 1st International Conference on Information Technology, Computer, and Electrical Engineering. 2014

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING