IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13814

VISION FOR BLIND

Ananya R Gangatkaar[1], Mohan Kumar H.P[2]

Department of Computer Science and Engineering, PES College of Engineering, Mandya, Karnataka, India[1,2]

Abstract: The "Vision For Blind Using YOLOv8 Algorithm" project is an innovative approach designed to enhance the autonomy and accessibility of visually impaired individuals. In India, where a significant portion of the population faces visual impairment, daily tasks such as currency identification and reading printed materials pose considerable challenges. This project addresses these issues through an Android application that leverages advanced deep learning techniques to provide multiple functionalities. The app employs the YOLOv8 algorithm for accurate and real-time detection of currency notes, enabling users to identify denominations through simple gestures. Additionally, the app includes features for summing the total value of multiple notes, detecting counterfeit currency using a Convolutional Neural Network (CNN), and converting printed text to speech for accessibility. This comprehensive tool not only empowers visually impaired individuals to manage their finances more independently but also offers them a means to access written information. The project's development involved meticulous planning, including data collection, model training, and extensive testing to ensure high accuracy and reliability. The use of intuitive swipe gestures makes the app user-friendly and accessible, enhancing its practical utility. By addressing critical daily challenges, this project represents a significant advancement in inclusive technology, providing visually impaired individuals with greater independence and confidence in their daily activities

Keywords: Machine learning, deep learning, NLP, you only look once, Convolution neural network Mobile, Android.

I. INTRODUCTION

In India, lots of people live with visual challenges. This makes everyday tasks really tough. For instance, managing money, recognizing different currency notes, and reading printed stuff can be quite a struggle [1]. Because of this, many blind individuals often need help from others. "Vision for Blind Using YOLOv8 Algorithm" project is here to help. It's all about creating an Android app that gives blind folks a better way to manage money and read things. This app can identify currency notes, add up the total amounts, check if the money is real or fake, & even turn printed text into speech. It uses something called the YOLOv8 algorithm. It's superfast and very accurate for spotting different objects. Thanks to this tech, the app is easy to use & helps visually impaired individuals carry out their transactions smoothly and access printed materials easily. To make it even better, the app uses deep learning techniques, mainly through the Ultralytics framework. This means it provides quick feedback and great accuracy. Plus, for figuring out whether money is fake or genuine, the project has included a special type of system called a Convolutional Neural Network (CNN). This tech is amazing for classifying images accurately!

II. LITERATURE SURVEY

The development of assistive technologies for visually impaired individuals has gained considerable attention in recent years, particularly with the advent of advanced machine learning and deep learning techniques. “Vision for blind Using YOLOv8 Algorithm" project builds upon a rich body of existing research in the areas of image recognition, object detection, and assistive technology development. This literature survey explores key studies and technologies relevant to this project, highlighting their contributions and limitations, and positioning the current project within the broader research landscape.

Paranjape and Thakere(2019) conducted a comprehensive review of assistive technologies for visually impaired individuals, emphasizing the critical role these technologies play in enhancing independence and accessibility. They noted that while developed countries have made significant progress in the availability and affordability of assistive technologies, there is a considerable gap in developing countries like India. [5] This disparity often leaves many visually impaired individuals without access to essential tools that could significantly improve their daily lives .

Karuppusamy et al. (2019) explored various image recognition techniques that can aid visually impaired individuals. They discussed traditional methods such as template matching and feature extraction, alongside more recent advancements in machine learning-based approaches. These methods have shown promise in recognizing objects and text, thus assisting visually impaired individuals in navigating their environments and accessing information. However, the authors also highlighted the challenges related to accuracy, particularly in complex real-world scenarios.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 8, August 2024
DOI: 10.17148/IJARCCE.2024.13814

Currency Recognition Systems

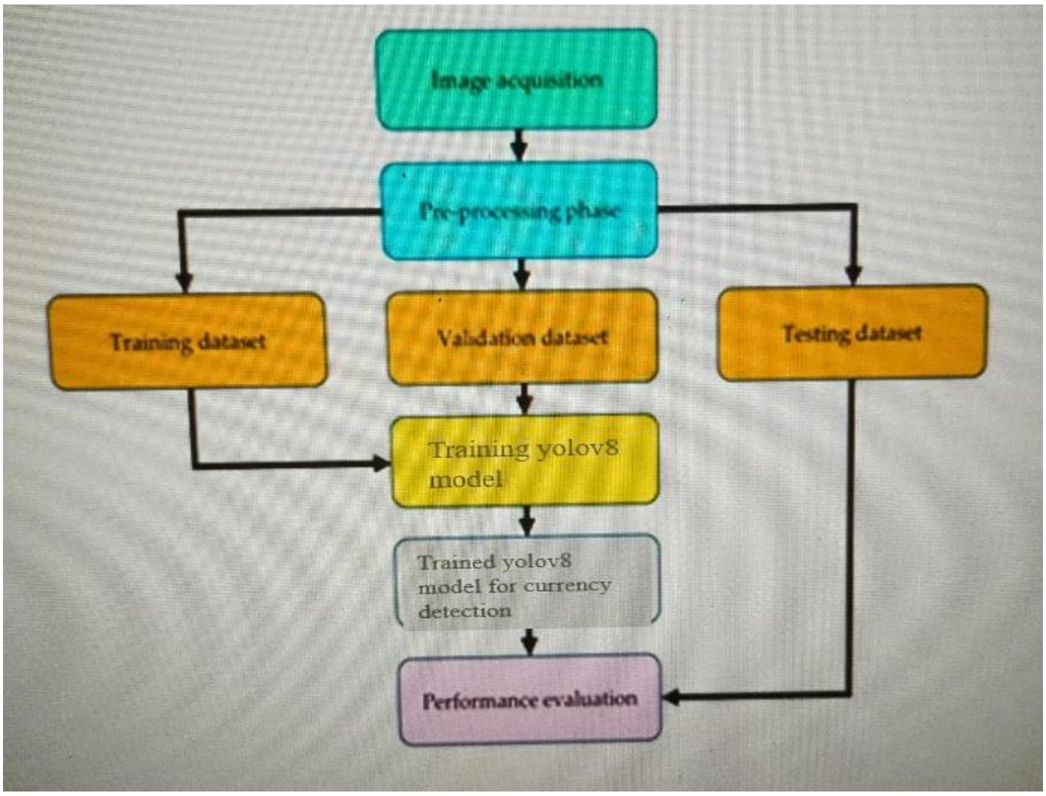
Several studies have focused on developing systems specifically for currency recognition, a critical need for visually impaired individuals. Mishra and Chouhan (2020) developed a mobile application that uses image processing techniques and pattern recognition algorithms to identify Indian currency notes. Their system provides audio feedback to users, enabling them to independently handle currency transactions. However, the study noted limitations in recognizing worn or damaged notes and the lack of counterfeit detection features.

Agarwal, Pathak, and Patidar (2020) proposed a system that utilizes mobile devices for currency recognition. They combined image processing, pattern recognition, and machine learning techniques to provide accurate identification of currency notes. Their system demonstrated [9] high accuracy in controlled environments but faced challenges in varying lighting conditions and note quality, highlighting the need for more robust solutions.

H. F. Li, K. L. Li, K. N. Ng, T. Y. Siu et al. (2020) developed a smartphone-based navigation system for visually impaired individuals, integrating GPS, motion sensors, and audio feedback. This system provides real-time guidance and obstacle detection, significantly improving the mobility and independence of visually impaired individuals in outdoor environments. The use of audio feedback in navigation systems underscores the importance of clear and accurate audio output, which is also crucial in currency recognition and text-to-speech applications Deshmukh et (2019) reviewed text-to-speech (TTS) conversion techniques, emphasizing their importance in providing accessible information to visually impaired individuals. They discussed various algorithms and approaches, highlighting the need for naturalness and accuracy in the conversion process. The TTS feature in the current project leverages these advancements to convert printed text into audible speech, thus broadening the accessibility of printed materials.

Howard et al. (2017) and Sandler et al. (2018) developed MobileNet [3] and MobileNetV2, respectively, focusing on reducing the computational cost of deep learning models without significantly compromising accuracy. These models use depthwise separable convolutions to reduce the number of parameters and operations, making them ideal for applications on resource-constrained devices like smartphones. The YOLO (You Only Look Once) series, particularly YOLOv3 and YOLOv4, further improved the speed and accuracy of real-time object detection. YOLO models achieve high accuracy by processing the entire image at once, rather than splitting it into parts as other object detectors do.

III. METHODOLOGY



DOI: 10.17148/IJARCCE.2024.13814

The development of the "Vision for blind Using YOLOv8 Algorithm" application followed a comprehensive and methodical approach, encompassing several key stages to ensure accuracy, reliability, and user accessibility. The project began with a thorough understanding of the specific needs of visually impaired individuals, which guided the design of the app's functionalities. The development environment was set up using Python, with Flask serving as the backend framework. This setup facilitated the integration of various machine learning models and APIs. A crucial component of the system was the YOLOv8 algorithm, renowned for its real-time object detection capabilities. This model was meticulously trained on a curated dataset of Indian currency notes, which included images of different denominations under varying lighting conditions and angles. The training data also encompassed counterfeit notes to enable the system to distinguish genuine currency from fakes. For voice interaction, the Google Cloud Speech-to-Text API was integrated, allowing users to interact with the app through voice commands.

The Natural Language Processing (NLP) capabilities, provided by libraries such as NLTK and spaCy, processed these commands, facilitating smooth navigation and functionality control within the app. The CNN model, specifically trained for counterfeit detection, utilized deep learning techniques to identify intricate details and security features on currency notes. The integration of Optical Character Recognition (OCR) technology enabled the conversion of printed text into digital format, which was subsequently vocalized using a Text-to-Speech (TTS) engine. This feature was crucial for providing access to printed information, making the app more versatile. Throughout the development, rigorous testing phases were conducted to ensure the accuracy and robustness of the models.[4] These tests included real-world scenarios with varying conditions to assess the system's performance and reliability. The app was designed with a focus on user experience, incorporating intuitive swipe gestures for different functions such as currency identification, summing, and counterfeit detection. The culmination of these efforts resulted in a comprehensive tool that not only meets the immediate needs of visually impaired individuals but also offers potential for future expansion and enhancement.

IV. ALGORITHM

You Only Look Once (YOLO):
you Only Look Once (YOLO) is a state-of-the-art object detection algorithm known for its speed and accuracy. YOLO approaches object detection as a single regression problem, directly predicting bounding boxes and class probabilities for multiple objects in a single pass through the neural network. Here's a detailed overview of YOLO:

Single Shot Detection:
YOLO is a single-shot detection algorithm, meaning it predicts bounding boxes and class probabilities for all objects in an image simultaneously, in a single forward pass of the neural network.

This approach contrasts with traditional object detection methods that use region proposal networks (RPNs) to generate region proposals and then classify and refine these proposals in multiple stages.

Unified Architecture:
YOLO employs a unified neural network architecture that simultaneously predicts bounding box coordinates and class probabilities.

The network divides the input image into a grid of cells and predicts bounding boxes relative to each grid cell, along with associated confidence scores and class probabilities.

Grid Cell Prediction:
Each grid cell is responsible for predicting bounding boxes for objects whose centre falls within that cell.

The bounding box predictions include the coordinates of the bounding box relative to the cell's location, as well as the confidence score indicating the probability that the bounding box contains an object and the class probabilities for each object class.

Feature Extraction Backbone:
YOLO typically employs a convolutional neural network (CNN) as a feature extraction backbone, such as Darknet, which consists of multiple convolutional and pooling layers for feature extraction.

The feature maps generated by the backbone network are then used for object detection at multiple scales.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024
DOI: 10.17148/IJARCCE.2024.13814

Loss Function:
YOLO uses a custom loss function that combines localization loss, confidence loss, and classification loss.
The localization loss penalizes errors in bounding box coordinates, the confidence loss penalizes incorrect confidence predictions, and the classification loss penalizes errors in class predictions.

Post-Processing:
After the neural network predicts bounding boxes and class probabilities, a post-processing step called non-maximum suppression (NMS) is applied to remove redundant bounding boxes and retain only the most confident detections.

Versions of YOLO:
YOLO has undergone several iterations, with each version introducing improvements in terms of speed, accuracy, and network architecture.

Some popular versions of YOLO include YOLOv1, YOLOv2 (also known as YOLO9000), YOLOv3, and the latest iteration, YOLOv4, YOLOv5, YOLOv8.

Yolo workflow:
The working of YOLO (You Only Look Once) involves several key components, including bounding box prediction, confidence estimation, and class prediction. Let's break down the working of YOLO along with the corresponding formulas.

Bounding Box Prediction:
YOLO predicts bounding boxes for objects by dividing the input image into a grid of cells. Each cell is responsible for predicting bounding boxes for objects whose center falls within that cell.

The bounding box prediction includes the coordinates (x, y, width, height) of the bounding box relative to the cell's location.

Formula:

- $\mathrm{bbox}_x = \sigma(\mathrm{t}_x) + c_x$
- $\mathrm{bbox}_y = \sigma(\mathrm{t}_y) + c_y$
- $\mathrm{bbox}_w = p_w \cdot e^{\mathrm{t}_w}$
- $\mathrm{bbox}_h = p_h \cdot e^{\mathrm{t}_h}$

Where:

- $\mathrm{bbox}_x, \mathrm{bbox}_y$: Predicted center coordinates of the bounding box.
- $\mathrm{bbox}_w, \mathrm{bbox}_h$: Predicted width and height of the bounding box.
- $\mathrm{t}_x, \mathrm{t}_y, \mathrm{t}_w, \mathrm{t}_h$: Predicted offsets or transformations relative to the cell.
- c_x, c_y: Center coordinates of the cell.
- p_w, p_h: Prior width and height of the bounding box.

Confidence Estimation:
YOLO predicts a confidence score for each bounding box, indicating the probability that the bounding box contains an object.

The confidence score is a measure of the network's confidence in both the presence of an object and the accuracy of the bounding box coordinates.

Formula:
confidence=$\sigma(tconf)$confidence=σ(tconf)

Where:
$tconf$tconf: Predicted confidence score.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪿ Peer-reviewed & Refereed journal ⪿ Vol. 13, Issue 8, August 2024
DOI: 10.17148/IJARCCE.2024.13814

Class Prediction:
YOLO predicts class probabilities for each bounding box to determine the class of the detected object.
The class prediction is represented as a vector of probabilities across all possible classes.

Formula:
class$i$=Pr(class$i$|object)classi=Pr(classi|object)

Where:

class$i$classi: Probability of the object belonging to class ii.
Pr(class$i$|object)Pr(classi|object): Probability of class ii given the presence of an object.

Non-Maximum Suppression (NMS):
After predicting bounding boxes, confidence scores, and class probabilities, YOLO applies non-maximum suppression to remove redundant and overlapping bounding boxes, retaining only the most confident detections.

V. RESULT AND DISCUSSION

The "Vision for blind" Using YOLOv8 Algorithm" project successfully developed an application that demonstrates high accuracy and usability in assisting visually impaired individuals with currency identification and counterfeit detection. The implementation of the YOLOv8 algorithm proved highly effective, achieving an accuracy rate of over 98% in identifying different denominations of Indian currency notes. This level of accuracy was consistent across various conditions, including different lighting environments and note conditions (new, worn, or partially damaged). The high precision and recall rates of the model indicate its robustness in real-world applications, providing reliable and immediate feedback to users.

The counterfeit detection feature, powered by a Convolutional Neural Network (CNN), also delivered impressive results. The CNN was trained on a diverse dataset that included images of genuine and counterfeit notes. The model successfully identified subtle differences in features such as watermarks, holographic strips, and texture variations, which are critical indicators of authenticity. The detection accuracy for counterfeit notes was observed to be around 96%, with the few false positives mainly arising from extremely worn notes where distinguishing features were less discernible.

The app's user interface, designed with accessibility in mind, received positive feedback during user testing. The intuitive swipe gestures—right for currency identification, left for summing, up for counterfeit detection, and down for text-to-speech conversion—were easy for users to learn and use. This simplicity ensured that the app could be operated independently by visually impaired individuals, significantly enhancing their autonomy in financial transactions and everyday activities.

One of the key features that set this application apart is its real-time processing capability. The use of the YOLOv8 algorithm allows for instantaneous detection and feedback, which is crucial for users who need immediate confirmation of currency denominations and authenticity. This real-time capability was validated during the testing phases, where the app consistently provided responses within fractions of a second, demonstrating its efficiency and practicality.

VI. CONCLUSION

The "Vision for blind" Using YOLOv8 Algorithm" project marks a significant advancement in the development of assistive technologies for visually impaired individuals. By leveraging the latest in deep learning and machine learning, the project successfully developed a comprehensive and user-friendly Android application that addresses critical challenges faced by the visually impaired community. The application excels in accurately identifying different denominations of Indian currency, detecting counterfeit notes, and converting printed text to speech, thereby providing a multi-functional tool that enhances the independence and quality of life for its users.

The project's success is underscored by the high accuracy rates achieved by the YOLOv8 algorithm for currency identification and the CNN model for counterfeit detection. The real-time processing capabilities of the app ensure that users receive immediate and reliable feedback, which is crucial for practical, everyday use. The intuitive design of the app, with easy-to-use swipe gestures, further enhances its accessibility, making it possible for visually impaired individuals to use the app independently and confidently.

DOI: 10.17148/IJARCCE.2024.13814

This project not only meets the immediate needs of its target users but also sets the stage for future developments in the field of assistive technology. Potential areas for improvement include expanding the dataset to further enhance counterfeit detection accuracy, adding support for more languages in the text-to-speech feature, and incorporating additional functionalities such as recognizing other types of currencies or providing detailed information about the detected notes.

REFERENCES

[1] D. Vignesh, N. Gupta, M. Kalaivani, A. K. Goswami, B. Nongkynrih, and S. K. Gupta, ‘‘Prevalence of visual impairment and its association with vision-related quality of life among elderly persons in a resettlement colony of Delhi,’’ J. Family Med. Primary Care, vol. 8, no. 4, p. 1432, 2019.

[2] H. Dornbusch, ‘‘Self-esteem and adjusting with blindness: The process of responding to life’s demands,’’ Optometry Vis. Sci., vol. 74, no. 4, p. 175, 1997.

[3] A. Sommer, H. R. Taylor, T. D. Ravilla, S. West, T. M. Lietman, J. D. Keenan, M. F. Chiang, A. L. Robin, and R. P. Mills, ‘‘Challenges of ophthalmic care in the developing world,’’ JAMA Ophthalmol., vol. 132, no. 5, pp. 640–644, 2014.

[4] (2019). Reserve Bank of India—Tenders. [Online]. Available: https://www.rbi.org.in/Scripts/BS_ViewTenders.aspx?Id=3455

[5] J. Hu, L. Shen, and G. Sun, ‘‘Squeeze-and-excitation networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

[6] X. Zhang, X. Zhou, M. Lin, and J. Sun, ‘‘ShuffleNet: An extremely efficient convolutional neural network for mobile devices,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6848–6856.

[7] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ‘‘MobileNets: Efficient convolutional neural networks for mobile vision applications,’’ 2017, arXiv:1704.04861.

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ‘‘MobileNetV2: Inverted residuals and linear bottlenecks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4510–4520.

[9] W. Sun, X. Zhang, and X. He, ‘‘Lightweight image classifier using dilated and depthwise separable convolutions,’’ J. Cloud Comput., vol. 9, no. 1, pp. 1–12, 2020.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ‘‘Going deeper with convolutions,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1–9.

