Currency Recognition Using Deep Learning

Han Juin Wong 21138830

1 Abstract.

This project presents a deep learning-based approach for recognizing the denominations of New Zealand banknotes using the YOLOv8 object detection algorithm. The model is designed to not only to understand the foundation of Convolutional Neural Network (CNN), but also assist visually impaired individuals. Leveraging the power of CNN and the latest YOLOv8 architecture, the model is trained on images of $10, $20, and $50 notes. The project highlights key advancements in object detection, such as real-time processing, single-shot prediction, and improved performance over previous YOLO versions. Lastly, The final model demonstrates high accuracy in identifying banknote denominations.

2 Introduction.

As highlighted in paper, the number of visually impaired individuals in the global population is increasing. (Chatterjee, Obaidat, Samanta, Islam, & Joseph, 2023; Tasnim, Pritha, Das, & Dey, 2021; Thomas & Meehan, 2021). A common difficulty is accurately identifying the denomination of banknotes, which can lead to financial dependency, vulnerability to fraud, or unintentional overpayment. While some current solution help solved this problem, they are often unreliable or inaccessible. With recent advancements in artificial intelligence, deep learning has emerged as a powerful tool for automating visual tasks. Convolutional Neural Networks (CNNs), have shown impressive results in image classification and object detection applications (Das, Kundu, Sazzad, & Rahman, 2023). However, CNNs are not the only option, newer architectures such as Vision Transformers, and hybrid models like EfficientDet, also offer competitive performance and unique advantages. (Bandara, 2025) Moreover, on mobile platforms, recognition systems can deliver both high accuracy and real-time performance. Traditional computer vision techniques often fall short in generalizing across varying lighting conditions, note orientations, and backgrounds. This project addresses these gaps by designing a custom banknote recognition system for New Zealand currency using YOLOv8, a state-of-the-art object detection model known for its speed and accuracy. In addition to implementing YOLOv8, the project also explores other deep learning-based approaches through comparative evaluation, with the goal of identifying the most effective and practical model for currency recognition in assistive applications.

3 Related Work.

Islam, Hasan, and Rahman (2024) uses a different approach in detecting currency. They used the vision transformer model, a state-of-the-art approach that outperforms traditional CNN models. Vision transformer is a deep learning model that suitable for image recognition. It leverage the capabilities of transformer model, which often used in Natural Language Processing. Vision Transformer process starts by dividing an image into small, fixed size patches, which are then flattened and embedded into vectors. Positional embeddings are added to retain spatial information. These patches, along with a classification token, are passed through a transformer model made up of multiple layers. The classification token gathers information from all patches and represents

the entire image. Finally, this token is passed through a classification head, typically a few layers with softmax to output the predicted class of the image.

Abass et al. (2023) uses the Efficient and Accurate Scene Text (EAST) model and Tesseract Optical Character Recognition (OCR), another novel approach to detect currency denominations. The EAST model, commonly used for text detection, is a neural network-based architecture specifically designed to identify text in images. OCR is known for accurately detecting and extracting text from images. In their approach, the EAST model is used for feature extraction and model training. It generates two outputs: scores and geometry. The scores indicate the probability that a region contains text, while the geometry provides the coordinates of the text regions. The OCR is responsible for making predictions using the trained model. It can easily process scanned text and convert it into machine-readable and editable format.

In Tong and Yan (2023) paper, the authors proposed an enhanced YOLOv5 model for detecting transparent watermarks on currency notes, a method that aligns with this research approach. Specifically, they integrated the Squeeze-and-Excitation (SE) attention block into YOLOv5 backbone to improve feature extraction. The SE module emphasizes important channels while eliminating less useful ones, thus enhancing the model's ability to identify small watermark features. The study also implemented different data augmentation techniques to improve generalization. The experiments compared multiple versions of YOLOv5 (s, m, l) with and without SE. The YOLOv5l-SE variant achieved the highest performance, with 99.9% precision and a mAP@0.5 of 99.61%, albeit at the cost of increased training time. This indicates that integrating attention mechanisms like SE can significantly improve detection accuracy.

The author developed a real-time, offline, and mobile-friendly currency recognition system for Sri Lankan banknotes to support visually impaired individuals. It uses a deep learning approach called EfficientDet-lite2, which is optimized for mobile development. (Bandara, 2025) EfficientDet-lite2 consists of three key components: the backbone, the Bidirectional Feature Pyramid Network (BiFPN), and the detection head. The backbone is responsible for extracting features from the input image. The BiFPN aggregates features from different levels of the backbone. The detection head applies small convolutional layers to the features to predict the bounding box, class scores, and objectness score.

Shoumik et al. (2022) concluded that applying transfer learning with a CNN model can improve the accuracy of bank note identification. Transfer learning involves leveraging previously learned knowledge from a pre-trained model and applying it to a new model. By building upon this existing knowledge, the new model gains a strong foundation, allowing it to start from an advanced point rather than learning from scratch.

4 Contribution.

This paper introduces a YOLOv8-based currency recognition system specifically for New Zealand banknotes, targeting the denominations of $10, $20, and $50. Unlike prior research that relied on YOLOv5 or traditional CNNs, this work adopts a more recent and optimized detection architecture to achieve improved accuracy and faster inference. A custom dataset was created using real-world images captured via smartphone, annotated to support robust model training.

In addition to model development, this study conducts a comparative analysis with other deep learning models mentioned in the related work, evaluating performance across precision, recall, mAP. The system is designed with real-time responsiveness and low computational overhead in mind, making it feasible for future

mobile deployment. This project contributes both a new dataset and experimental benchmarks to the field of assistive currency recognition, with potential application in tools for visually impaired individuals in New Zealand.

5 Methodology

5.1 Data Collection

An iPhone 15 Pro was used to capture video footage of New Zealand banknotes ($10, $20, and $50) at a resolution of 1080p and 30fps. Since both the front and back sides of the notes display the same number design, each denomination is labeled using a single label, resulting in three labels for the three different notes. To improve generalization, data augmentation was performed by introducing slight variations in angle and perspective across the captured frames. The recorded video is then imported into Python for frame extraction and resized to 640 $\times$ 640 pixels to comply for YOLOv8 input requirement . In total, 209 sample images across the three note denominations are prepared for training.

5.2 Data Annotation

Labeling was performed using LabelImg, an open-source graphical image annotation tool that supports YOLO-compatible formats. After frame extraction and resizing, each image was manually annotated by drawing bounding boxes around the visible currency notes. Each annotation includes the class index corresponding to the denomination ($10, $20, or $50) and the normalized coordinates of the bounding box (center x, center y, width, height), as required by the YOLO format. The annotations were saved in text files with the same base filename as the images, ensuring compatibility with the YOLOv8 training pipeline. To ensure that the model's performance is not compromised by the dataset, the bounding boxes were always drawn as close as possible to the classified text. For instance, Figure 1 shows a sample bounding box for a $10 New Zealand note. This manual labeling process ensured high-quality ground truth data for training and evaluation.

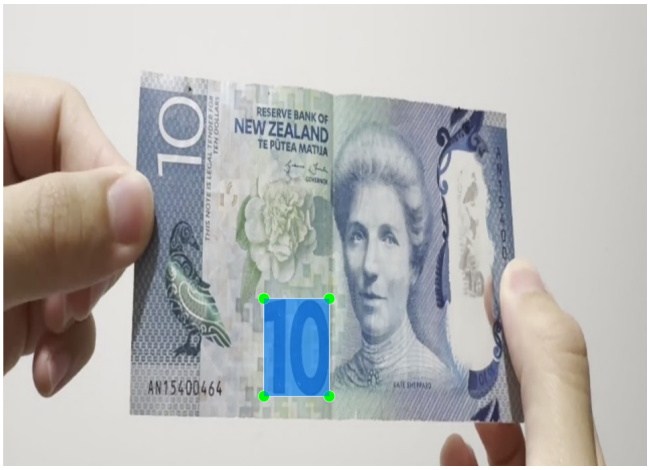

Figure 1: Annotation Sample for $10 New Zealand notes

5.3 Model Selection

To classify the denominations of the banknotes, the YOLO object detection model is used. YOLO is a single-shot detection algorithm built on top of Convolutional Neural Networks, which enables it to perform object classification in real-time. CNNs have been widely used in image classification tasks due to their strong spatial feature extraction capabilities, making them a suitable foundation for high-accuracy vision systems. Among various YOLO versions, YOLOv8 was selected for this project due to its significantly improved architecture, offering better accuracy, faster inference speed, and enhanced feature extraction compared to its predecessors such as YOLOv5. YOLOv8 is designed to be lightweight and scalable, which makes it suitable for real-time applications like currency recognition. Its ability to process the entire image in a single forward pass reduces latency, which is essential for deployment on mobile or embedded systems. The specific algorithm and architectural components of YOLO will be discussed in detail in the following algorithm section.

5.4 Workflow Diagram

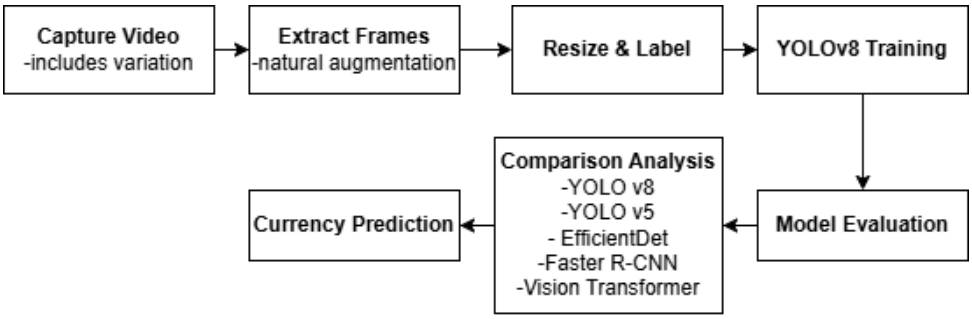

Figure 2: Workflow Diagram

Figure 2 shows the workflow pipeline of the model training and evaluation process. The pipeline begins with video capture using a smartphone, which naturally introduces variations in angle and lighting, it served as a form of data augmentation. Extracted frames are then resized to 640×640 pixels and manually labeled using the LabelImg tool in YOLO format. These annotated images are used to train the YOLOv8 object detection model. Following training, the model undergoes evaluation using standard metrics such as precision, recall, and mAP. To assess the relative performance of YOLOv8, a comparison analysis is conducted involving multiple object detection architectures, including YOLOv5 as a baseline, EfficientDet, Faster R-CNN, and Vision Transformer models. The best-performing model is then used for final currency denomination prediction. This pipeline is designed to support future deployment in real-time assistive applications for visually impaired users.

6 Algorithm

YOLO (You Only Look Once) is a real-time object detection algorithm that frames object detection as a single regression problem. It divides the input image into an $S \times S$ grid, where each grid cell is responsible for predicting bounding boxes and corresponding class probabilities for objects whose centers fall within that cell. (Terven, Córdova-Esparza, & Romero-González, 2023)

YOLOv8, one of the newer version in the YOLO series, introduces several improvements over its predecessors.

Unlike previous anchor-based versions, YOLOv8 uses an anchor-free approach, which reduces model complexity and improves detection robustness. It also incorporates a decoupled head design, where classification and regression are handled by separate branches, leading to better specialization and performance. YOLOv8 is designed through neural architecture search (NAS), enabling optimal balancing of speed and accuracy for different deployment settings. (Terven et al., 2023)

The core output of the model is defined by:

$$\mathbf{Y} = f_\theta(\mathbf{I}) = \left\{ \left(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i, \hat{C}_i, \hat{p}_{i1}, \ldots, \hat{p}_{iC} \right) \right\}_{i=1}^{S^2 \cdot B} \tag{1}$$

Where f_θ is the YOLO model with learnable parameters θ, $\mathbf{I}$ is the input image, and $\mathbf{Y}$ is the output set of predictions. Each prediction includes:

- $(\hat{x}_i, \hat{y}_i)$: Center coordinates of the bounding box (normalized)
- $(\hat{w}_i, \hat{h}_i)$: Width and height of the box
- $\hat{C}_i$: Confidence score (objectness)
- $\hat{p}_{ij}$: Class probabilities for each of the C classes

During inference, YOLOv8 processes the entire image in a single forward pass, producing predictions in real time. Predictions with confidence scores below a certain threshold are discarded. To handle overlapping detections, the algorithm applies Non-Maximum Suppression (NMS) using an Intersection-over-Union (IoU) threshold to retain only the most confident bounding boxes.

The training process optimizes a composite loss function that integrates localization, objectness, and classification losses:

$$\mathcal{L}_{\text{YOLOv8}} = \lambda_{\text{CIoU}} \cdot \mathcal{L}_{\text{CIoU}} + \lambda_{\text{obj}} \cdot \mathcal{L}_{\text{BCE-obj}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{BCE-cls}} \tag{2}$$

Where:

- $\mathcal{L}_{\text{CIoU}}$: Complete IoU loss for bounding box regression
- $\mathcal{L}_{\text{BCE-obj}}$: Binary cross-entropy loss for objectness score
- $\mathcal{L}_{\text{BCE-cls}}$: Binary cross-entropy loss for classification

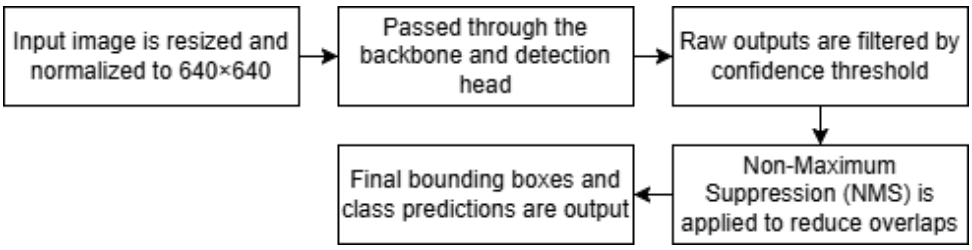

Figure 3: YOLOv8 inference pipeline

Figure 3 illustrates the inference pipeline of the YOLOv8 model used for currency recognition. The process starts with resizing and normalizing the input image to 640×640 pixels to match the model's expected input

requirement. The preprocessed image is then passed through the model's backbone and detection head, where feature extraction and object prediction are performed. The raw outputs, which include bounding boxes and class scores are filtered based on a confidence threshold to eliminate low-probability detections. To further refine the results, Non-Maximum Suppression (NMS) is applied, removing redundant overlapping boxes and retaining only the most confident predictions. The final output consists of the remaining bounding boxes along with their corresponding class labels and confidence scores, representing the recognized currency denominations.

7 Demo.

The demonstrations in Figures 1, 2, and 3 show the result of the model. It illustrates that the model is well trained and capable of accurately detecting different currency denominations.

Figure 4: NZ $10 note

Figure 5: NZ $20 note

Figure 6: NZ $50 note

8 Evaluation and Metrics

Model	**Precision**	**Recall**	**mAP@0.5**
YOLOv8	0.9897	0.9971	0.9950
YOLOv5 (Achar et al., 2021)	0.9027	0.9357	0.9027
EfficientDet (Bandara, 2025)	0.9983	N/A	0.9877
Faster R-CNN (Hanif et al., 2024)	0.763	0.821	0.869
Vision Transformer (Islam et al., 2024)	0.9994	0.9993	N/A

Table 1: Comparison of object detection models on precision, recall, and mean Average Precision (mAP@0.5).

As table 1 presents a comparative analysis of five object detection models which is YOLOv8, YOLOv5, EfficientDet, Faster R-CNN, and Vision Transformer based on their precision, recall, and mAP@0.5 scores. YOLOv8 achieves the highest balance across all three metrics, with a precision of 0.9897, recall of 0.9971, and mAP@0.5 of 0.9950. YOLOv5 also performs quite well, with precision and recall scores above 0.90. EfficientDet have precision of 0.9983 and mAP of 0.9877, however recall was not available. Vision Transformer achieves precision and recall of 0.9994 and 0.9993, but mAP is not available. In contrast, Faster R-CNN performs the worst with a precision of 0.763 and recall of 0.821. Overall, YOLOv8 demonstrates the most consistent and robust performance among the models evaluated.

9 Conclusion.

The project has successfully demonstrated the deployment of the YOLOv8 model for currency denomination recognition using a deep learning platform. Through comparative analysis, YOLOv8 appear to be the most suitable model for currency identification. The combination of a robust object detection algorithm with a properly prepared dataset enabled the model to run at high velocity and accuracy. Through visual demonstrations, the system proved capable of differentiating between $10, $20, and $50 New Zealand currency correctly, confirming its potential use in real-world applications in the future, especially in helping the visually impaired or augmenting computerized financial systems.

Future work will involve expanding the dataset size to include more currency denominations, capturing lighting variations, angles, wear-and-tear, and note conditions to further improve model generalization. Additional labels such as security features or serial number can be included to enable secondary tasks such as authenticity verification.

A second key area of emphasis is to deploy the trained model on mobile devices to enable real-time offline recognition. This will further advance the solution towards being more practical and usable for everyday use. Ultimately, these developments are aimed toward building an end-to-end, user-friendly currency recognition system for release in assistive and fiscal applications.

References

Abass, E. S., Mohamed, A. E. F., Amer, A., Hafez, M., Solyman, A., & Fawzy, M. (2023). Currency recognition using east for text detection and tesseract ocr for text recognition [Conference Proceedings]. In *2023 2nd international engineering conference on electrical, energy, and artificial intelligence (eiceeai)* (p. 1-9). Retrieved from `https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=10590444&ref=` doi: https://doi.org/10.1109/EICEEAI60672.2023.10590444

Achar, S. D., Shankar Singh, C., Sumanth Rao, C., Pavana Narayana, K., & Dasare, A. (2021). Indian currency recognition system using cnn and comparison with yolov5. In *2021 ieee international conference on mobile networks and wireless communications (icmnwc)* (p. 1-6). doi: 10.1109/ICMNWC52512.2021.9688513

Bandara, A. (2025). *Money recognition for the visually impaired: A case study on sri lankan banknotes.* Retrieved from `https://arxiv.org/abs/2502.14267`

Chatterjee, K., Obaidat, M. S., Samanta, D., Islam, S. H., & Joseph, N. P. (2023). Machine learning-based currency information retrieval for aiding the visually impaired people [Conference Proceedings]. In *2021 international conference on computer, information and telecommunication systems (cits)* (p. 1-5). doi: https://doi.org/10.1109/CITS52676.2021.9618567

Das, D., Kundu, D., Sazzad, S., & Rahman, A. (2023). Utilizing deep learning with cnn model for precise identification of bangladeshi currency notes to aid visually impaired people [Conference Proceedings]. In *2023 5th international conference on sustainable technologies for industry 5.0 (sti)* (p. 1-6). doi: https://doi.org/10.1109/STI59863.2023.10464626

Hanif, M. Z., Saputra, W. A., Choo, Y. H., & Yunus, A. P. (2024, Aug). Rupiah banknotes detection comparison of the faster r-cnn algorithm and yolov5. *JURNAL INFOTEL*, *16*(3). doi: 10.20895/infotel.v16i3.1189

Islam, M. F., Hasan, J., & Rahman, M. S. (2024). Bangladeshi paper currency recognition with improved dataset using vision transformer [Conference Proceedings]. In *2024 6th international conference on electrical engineering and information & communication technology (iceeict)* (p. 226-229). Retrieved from `https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=10534465&ref=` doi: https://doi.org/10.1109/ICEEICT62016.2024.10534465

Shoumik, T. M., Chowdhury, S. J., Mostafa, T., Amit, A. M., Chowdhury, S. A. H., Aadi, O. A., ... Rasel, A. A. (2022). Bangladeshi paper currency recognition using lightweight cnn architectures [Conference Proceedings]. In *2022 ieee international conference on artificial intelligence in engineering and technology (iicaiet)* (p. 1-6). doi: https://doi.org/10.1109/IICAIET55139.2022.9936749

Tasnim, R., Pritha, S. T., Das, A., & Dey, A. (2021). Bangladeshi banknote recognition in real-time using convolutional neural network for visually impaired people [Conference Proceedings]. In *2021 2nd international conference on robotics, electrical and signal processing techniques (icrest)* (p. 388-393). doi: https://doi.org/10.1109/ICREST51555.2021.9331182

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, *5*(4), 1680–1716. Retrieved from `https://www.mdpi.com/2504-4990/5/4/83` doi: 10.3390/make5040083

Thomas, M., & Meehan, K. (2021). Banknote object detection for the visually impaired using a cnn [Conference Proceedings]. In *2021 32nd irish signals and systems conference (issc)* (p. 1-6). doi:

https://doi.org/10.1109/ISSC52156.2021.9467850

Tong, D., & Yan, W. Q. (2023). Visual watermark identification from the transparent window of currency by using deep learning [Book Section]. In B. Mohamed (Ed.), *Applications of encryption and watermarking for information security* (p. 59-77). Hershey, PA, USA: IGI Global. Retrieved from https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-6684-4945-5.ch003 doi: https://doi.org/10.4018/978-1-6684-4945-5.ch003