

# Vision Detect AI: Live Object and Currency Recognition Using Real-Time Video Streams

<sup>1</sup>Jhansirani Atluri, <sup>2</sup>Dr. K LittleFlower

<sup>1</sup>M.Tech Scholar, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, [jhansi.atluri@gmail.com](mailto:jhansi.atluri@gmail.com).

<sup>2</sup>Associate Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth (Deemed to be University), Maisammaguda, Hyderabad, Telangana 500100, India, [littleflower.kunchala@mrvv.edu.in](mailto:littleflower.kunchala@mrvv.edu.in)

## Abstract

The rapid development of computer vision and artificial intelligence in the last several years has led to an explosion in the need for automated systems capable of real-time currency and object recognition. Fast and accurate identification systems may significantly improve operational efficiency and decrease human error for businesses in the retail, banking, security, and automated finance industries. Especially in crowded places, manually identifying things and money is a pain and a certain recipe for mistakes. Consequently, modern robots rely heavily on recognition systems driven by AI. This project demonstrates a comprehensive approach to real-time item and currency recognition using the powerful YOLO deep learning algorithm. Its speed and accuracy in object identification tasks have made YOLO, which stands for "You Only Look Once," famous. When processing photos using YOLO, unlike with traditional multi-stage detection methods, just one forward pass is needed. This unique architecture enables fast detection, which is suitable for real-time applications.

**Keywords:** Currency recognition; Currency value detection; Image processing; Image acquisition; Accuracy; Financial and Retail applications; Image augmentation

## Introduction

In this digital era, automation has been a boon to many industries, such as retail, finance, security, and smart surveillance systems. Due to the growing complexity of business operations and the volume of transactions, there is a significant need for intelligent systems that can perform duties quickly and accurately. Two examples of such tasks that are fundamental to regular operations are item recognition and currency identification. The human error rate is high and the amount of time required to manually count, verify, or identify objects is big. Public service centers, supermarkets, and banks are examples of high-traffic places where even little identification errors may have a major influence on operational efficiency and financial losses. A system that can automatically identify items and their denominations in real-time is, therefore, very important.

The capabilities of computer vision systems have been substantially improved by recent advancements in artificial intelligence and deep learning. Thanks to Convolutional Neural Networks (CNNs), machines are now able to effectively understand and evaluate visual input. Yolo is one of the greatest models for real-time object identification that you can find. "You Only Look Once," or YOLO for short, is a competing method to traditional region-based detection algorithms that rely on a single forward neural network to identify and categorize items. Detection speed is much enhanced without accuracy being compromised by this approach. A system capable of real-time item and currency recognition is the goal of the proposed project, which intends to use the YOLO deep learning framework. In the setting of a single image or live video feed, the technology can differentiate between several item kinds and different monetary denominations. Photos captured under different lighting conditions, with different backgrounds, orientations, and partial occlusions make up the big and structured dataset required to train the model. In real-world scenarios, this kind ensures the system's reliability. It is not often

simple to recognize different denominations of cash since they share characteristics such as size, color, or patterning. Training the model on such fine-grained visual differences as layout structures, security features, printed numbers and symbols, etc. is therefore necessary.

We can increase performance while cutting computational cost by using transfer learning methodologies. In order to begin training, we will use the pre-trained weights from earlier editions of YOLO. This approach improves detection accuracy and accelerates model convergence even with a smaller custom dataset. Included with the system is an intuitive graphical user interface (GUI). With a camera, users may shoot video in real-time or submit photos for later editing. The found objects and monetary amounts are highlighted with boundary boxes and confidence ratings to offer immediate and clear feedback. Even though objects in these scenarios could be moving fast or appear against busy backgrounds, the system is designed to perform well in these conditions. One application where the system's real-time processing capabilities might be useful is in automated checkout counters and banking kiosks, which operate at rapid speeds. This project aims to merge deep learning with practical user interface design to create an intelligent, scalable, and efficient recognition system that can meet the needs of modern automation.

## Literature Survey

Rapid advancements in computer vision and artificial intelligence have caused a sea change in the development of automated recognition systems in recent years. The importance of intelligent systems is growing across several sectors, such as retail, healthcare, transportation, finance, and security, due to the need for efficient identification and verification procedures. Now that they are so integral to daily life, object detection and cash recognition are at the very top of this list of crucial tasks. Operations are sped up, human error is reduced, and effort is saved by using automated object recognition. Currency recognition systems are very useful for banks, automated teller machines, vending machines, and self-checkout lines. With more and more people opting for cashless and contactless solutions, the necessity for trustworthy currency detection technology has only grown.

The foundation of older recognition algorithms were methods for manually extracting features. These methods were based on histogram analysis, template matching, edge detection, and color segmentation. When used to real-life images with complex backgrounds or variable lighting, these algorithms often yielded erroneous findings, despite their efficacy in controlled laboratory environments. Other factors that impacted them were occlusions, rotations, and size alterations. Because real-world systems need to be adaptable and robust, researchers began to investigate machine learning techniques. The introduction of deep learning has a profound impact on computer vision. A Convolutional Neural Network (CNN) emerged victorious in a photo recognition and classification task. One characteristic of CNN-based systems is their ability to automatically learn hierarchical feature representations from raw images. This eliminates the need for human feature engineering while also dramatically improving accuracy. Compared to other object detection frameworks, YOLO stands out due to its real-time detection capability. A novel unified detection architecture called YOLO can predict bounding boxes and class probabilities at the same time. When analyzing an image, YOLO takes a holistic view rather than the area-based approach used by region-based detectors.

The effectiveness of the YOLO principle makes it a good fit for dynamic, high-pressure situations. Automatic banking kiosks, retail billing systems, and surveillance monitoring are just a few examples of the many applications that need systems for real-time detection. Accuracy is of the utmost importance when dealing with several currencies, as even a little mistake in classification could cause inconsistencies in the management of monies. The proposed method is based on YOLO's architecture and can distinguish between various forms of currency and other objects in a single image. It is essential to use transfer learning to increase a system's efficiency. Instead of starting from scratch when training a deep learning model, using pre-trained weights may improve generalization and accelerate convergence. The system is trained to identify objects and monetary denominations across various environments and lighting situations by means of a diverse dataset. Scaling, rotating, and brightness fluctuation are data augmentation strategies that boost robustness. Additionally, the system's UI is designed to be user-friendly for added convenience. Detection might make use of a live camera feed or user-uploaded images. The user may quickly

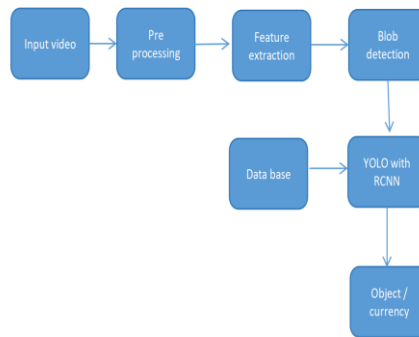
see their confidence level in the interface's bounding boxes. These intelligent solutions, once implemented, boost operational efficiency and pave the way for future intelligent automation.

## Methodology

The proposed technique allows for the detection of several item categories and monetary denominations within a single frame. The model is trained using images captured under a range of lighting, background, and orientation conditions. The dataset is rich in variety and well-structured. Supplementing the dataset with several variations enhances the system's robustness. While being trained, the model is instructed to identify various objects and currency notes based on their distinct attributes. Because these tokens are all around the same size and color, distinguishing between various amounts of money requires excellent visual sight. In order to differentiate between the many little patterns seen on banknotes—such as the numbers, symbols, and designs—an algorithm is developed. In order to reduce training time and boost efficiency, the project employs transfer learning methodologies. Famous YOLO models' pre-trained weights form the basis of the first training layer. The system may put traits that have been learned from large datasets to use using this approach. Transfer learning improves the accuracy of both convergence and detection. By fine-tuning the model, it becomes adaptable to the project's specific item and currency dataset. The architecture of the system strikes a balance between speed and precision. An easy-to-use interface facilitates both navigation and communication inside the system. Using this interface, you may capture images or use the live camera feeds for detection. Items and denominations of currency that have been discovered are shown with bounding boxes and confidence ratings. Immediate feedback enhances the practical applications' applicability. Thanks to its low latency frame processing, the system continues to run smoothly even in very dynamic environments.

The experimental evaluation findings reveal that the proposed model achieves excellent recognition accuracy across all test conditions. The pace of detection remains constant even when many things appear simultaneously. The dependability of the system is unaffected by slight variations in light and partial obstructions. A model's performance may be measured by its accuracy, recall, and mean average precision. Each frame's processing duration is optimal for real-time deployment.

Some real-world uses for YOLO-based recognition systems include automated checkout at retail establishments, currency counting devices, banking kiosks, and security surveillance. Automatic currency recognition allows the visually impaired to quickly identify denominations. Integrating the technology into monetary transaction-facilitating mobile applications is another potential use case. The new development also provides the groundwork for augmented reality applications that use object detection to trigger digital overlays. Finally, the proposed real-time item and currency recognition system demonstrates that deep learning is capable of handling practical identification challenges. The system is able to accomplish both speed and accuracy because it employs transfer learning techniques and YOLO's efficient architecture. Use is further enhanced by the inclusion of an intuitive user interface. Advancements in this research might have a positive impact on smart automation systems, digital banking, intelligent retail, and automated recognition, among other fields.



proposed system

The proposed system for object and currency recognition using deep learning is shown in the accompanying flow diagram. The first phase, called "input video," involves importing video that has been stored or is currently being captured by a camera. This input video provides the bulk of the visual data used by the system. Every every frame of the captured video goes through a series of processing steps. All of the frames include visual data, such as objects and financial notes. The system extracts frames from the live video stream at a predetermined frame rate. The frames will be sent to the preprocessing phase after this.

As part of the preprocessing phase, the raw frames undergo a number of enhancement procedures. Using noise reduction technologies, the annoying picture distortions may be eliminated. Image resizing is done to ensure that the input dimensions meet the requirements of the detection model. It is possible to do color space conversion in order to highlight characteristics. By adjusting the contrast and normalization, you may enhance the image quality. Once the frames are adjusted to ensure precise detection, the procedure is finished. When the system is preprocessed, it becomes more resilient to changes in lighting. It reduces computational complexity by standardizing input data as well.

After the preprocessing phase, comes the feature extraction phase. Important visual details from each frame are extracted at this point. Feature extraction may help you find patterns in your images, such as edges, textures, shapes, and color distributions. We are able to distinguish between different goods and monetary amounts based on these features. Using data and deep learning techniques, hierarchical features may be automatically learnt. High-level features are those that are distinctive to objects, while low-level features include things like edges and corners. Importance of feature extraction lies in the fact that it provides context to pixel data that would otherwise be useless. The accuracy of the detection is directly related to the quality of the characteristics that are extracted. next feature extraction, the next stage is blob detection. The purpose of blob detection is to separate regions of an image that vary significantly from their background in terms of brightness or color from the rest of the image. These regions could stand for different things or different amounts of money. Finding possible research areas is made easier using blob identification. Reduced processing overhead is achieved by selectively highlighting relevant areas of the image. This is the last stage before the data is formally classified. By detecting blobs, the method zeroes in on certain areas. Things become quicker and more efficient as a result of this. After YOLO processes the data, it feeds it into the RCNN module. The system's basic detection and classification capabilities are now on show. By simultaneously calculating the bounding boxes and class probabilities, the YOLO approach can detect objects in real-time. After partitioning the image into grids, it processes it using a single forward pass. This ensures detection happens quickly enough for use in real-time scenarios. Applying RCNN techniques improves localization accuracy. The model fixes the spots it discovers to make it more precise. In this hybrid approach, speed and accuracy are preserved. The database is a crucial component of the system's framework. The labeled datasets are saved throughout the training process. The database contains images of various goods and various denominations of money. While detecting, features are extracted and compared to representations in the training model. The database may also keep

track of performance metrics and detection logs. It enables continual system updates, which are an improvement. With data management, storing and retrieving information is a breeze.

Finally, the recognized item or currency's result is generated in the output stage. The system will automatically create bounding boundaries around any detected objects or cash. Each detected item's confidence score and class name are shown. The outcomes may be seen live on the screen. In real-time, users get visual feedback. Quick decisions are made in retail or financial contexts because of this. The system may pick up on many items simultaneously. The accuracy is maintained even while the scene is in motion. All things considered, the image depicts a logical procedure for identifying monetary and non-monetary items. Every step is crucial for ensuring reliable operation. The input video may provide you with data in real-time. The preprocessing of images improves their quality. Feature extraction is a method for capturing meaningful patterns. Target regions may be located using blob detection. Accurate localization and classification are accomplished using RCNN YOLO. The database may make training and data management easier. The end product comprises results that are both easy to understand and put into practice. When put together, these components provide a robust automated recognition system. A problem-free flow of information from input to output is ensured by the procedure. Thanks to its modular design, future enhancements are within reach. Thanks to the system's scalability, new item categories may be readily added. This saves time and cuts down on human error. The architecture is designed to facilitate deployment in real-time. When it comes to artificial intelligence identification systems, deep learning is clearly a potent instrument.

## Implementation

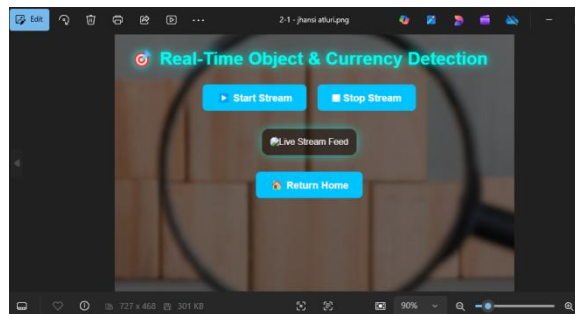
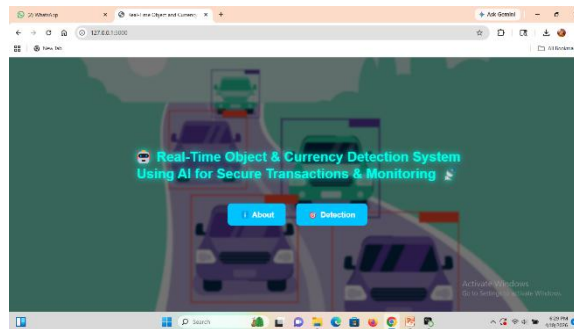
The proposed system for identifying items and currencies is divided into several functional sections to ensure ordered development and effective performance. Every module works together to complete the procedure. The modular architecture improves scalability, debugging ease, and maintainability. The first module is the one responsible for input acquisition. This part is responsible for processing the camera's live feed or user-uploaded photos. To process the video, it first decomposes it into its individual frames. With the input module's help, you can be sure that each frame will be captured at the sweet spot where speed and accuracy meet. Number two is the preprocessing module. The detection model is given into this module just before it enhances the input frames. Adjusting sizes, normalizing, reducing noise, and converting colors are all within its capabilities. Make sure it meets the input size criterion of YOLO when you resize. Scaling pixel values with normalization increases model convergence. Reducing noise allows for clearer pictures with fewer false positives. One important role this module plays is stabilizing system performance under different lighting conditions. The final component is the training and dataset module. This module takes care of importing datasets, managing annotations, and training models. It employs labeled images of various objects including currency. Data enrichment techniques, such as flipping, rotating, and altering brightness, are used to improve generalization. This module uses transfer learning to improve the pre-trained YOLO weights. The training module iteratively adjusts the model's parameters and finds loss values. Performance is also evaluated using metrics like as accuracy and recall.

The fourth part is the detection and classification module. This module uses the YOLO model for real-time inference. It predicts class probabilities and bounding boxes for each processed frame. Non-Maximum Suppression is used to eliminate superfluous overlapping boxes. Verifying the accurate localization of objects and currency notes is the responsibility of this module. Multiple things may be identified simultaneously in a single image. Confidence ratings are used to generate detection results. The sixth part is the interface module for the user. Users and systems may communicate with the help of this component. We can see the processed frames' bounding boxes here. The user interface allows for the start and stop of detection. This function also allows you to submit photos and see the results of live streams. The last unit covers database administration. It records information about training data references, model checkpoints, and detection logs. The modular architecture ensures flexibility and can readily accept future extensions of item categories or other functionality.

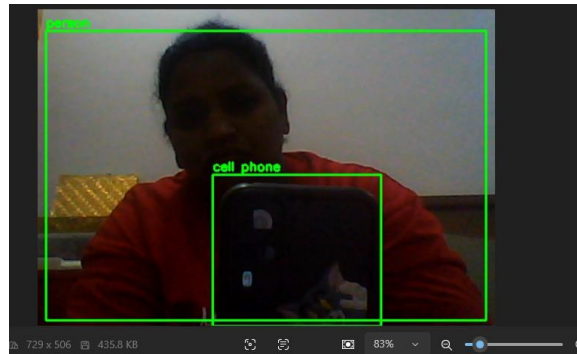
## Algorithms

The YOLO object detection technique is the backbone of this system. The object identification challenge is reduced to a single regression issue using YOLO. It creates a grid from the supplied picture. Class probabilities and bounding boxes are predicted for each grid cell. Only one forward run through the neural network is required for the algorithm to process the picture. With this layout, you can be certain that real-time applications will benefit from the fast detection. From the input picture, YOLO's backbone network retrieves hierarchical characteristics. These characteristics stand in for various forms, textures, and designs. The feature maps at various scales are aggregated by the neck component. This enhances the precision of object identification, regardless of size. The final class predictions and bounding box coordinates are generated by the head component. To find out how likely it is that an item is there, the algorithm takes confidence ratings into account. To eliminate duplicate detections, Non-Maximum Suppression is used. This method eliminates overlapping boxes by selecting the one with the greatest confidence score. A combination of localization, confidence, and classification losses make up YOLO's loss function. The localization loss is the difference between the expected and observed coordinates of the bounding boxes. The loss in classification is the sum of all the expected and actual class labels. Loss of confidence measures how well a thing is really present.

## Results



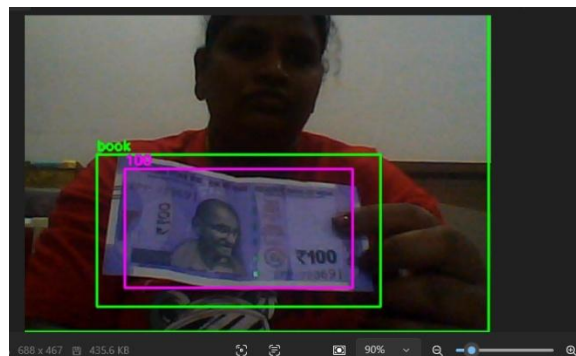
Predictions



prediction



Object detection



Currency detection

## Conclusion

The proposed real-time item and currency recognition system demonstrates, finally, that practical identification problems may be successfully addressed by deep learning. With the growing need for automated identification systems in sectors such as retail, banking, and security, intelligent solutions such as these are assuming more significance. In fast-paced environments, manual identification methods could be slow, inaccurate, and inefficient. The use of a deep learning-based system significantly loosens these constraints. Quick detection and categorization with just one forward pass is the goal of this project's YOLO implementation. This unified architecture allows for the simultaneous localization and recognition of many objects and denominations of currency. The system is trained on a diverse and expansive dataset to ensure its resilience in various contexts. Employing data augmentation techniques improves the model's adaptability to different lighting, orientation, and backdrop scenarios. Transfer learning reduces training time without compromising accuracy by making use of pre-trained weights, thereby improving efficiency. Incorporating preprocessing procedures improves the detection reliability and the picture quality. The modular implementation architecture ensures both scalability and maintainability.

Results from experiments show that the system achieves good recall and accuracy. The results of the real-time performance testing show that the detection process is flawless and devoid of any discernible latency. Recognizing many objects and currency notes within a single frame demonstrates the system's resiliency. A more accessible and user-friendly user interface is the goal of its design. Bounding boxes and confidence ratings simplify the data for easier understanding and manipulation. In order to reduce false positives, the system employs effective Non-Maximum Suppression techniques.

Although there are limitations to traditional recognition methods, this endeavor successfully circumvents them. It eliminates the need for expensive hardware setups by using camera-based detection. Businesses of all sizes may take advantage of the system's affordable deployment opportunities. One example of a real-time application that it enables is automated checkout systems and financial kiosks. Integrating state-of-the-art deep learning frameworks ensures scalability and reliability. Smart automated recognition systems may be well-established with the help of the proposed strategy. It improves operating efficiency and decreases the requirement for human verification processes. The system maintains its stability throughout a range of environmental conditions. Its future interoperability with other technologies will be facilitated by its modular design. The results show that YOLO-based detection can successfully identify items and money in real-time. Thanks to this endeavor, smart automation and computer vision applications have improved.

## REFERENCES

- [1] Ghanem AM, Youness HA, Wahba M, et al. Recognizing Egyptian currency for people with visual impairment using deep learning models. *Sci Rep.* 2025.
- [2] Fanzury N, Hwang M. Real-time identification of mixed and partly covered foreign currency using YOLOv11 object detection. *AI.* 2025.
- [3] Rajeswari K, Kaviyarasan S, Purusothaman R, et al. A deep learning approach to currency recognition for the visually impaired people. *ShodhKosh J Visual Perform Arts.* 2024.
- [4] Carletti V, Saggese A, Spagnolo P. Object detection and recognition based on deep learning. *Sensors.* 2026.
- [5] Islam MN, et al. NSTU-BDTAKA: An open dataset for Bangladeshi paper currency detection and recognition. *Data Brief.* 2024.
- [6] Majadi N, et al. Bangladeshi currency detection dataset and benchmark study. *Data Brief.* 2024.

- [7] Reddy PS, Sathvik B, Anirudh NS. A novel deep learning-based object detection framework for visually impaired assistance. *J Namibian Stud.* 2023.
- [8] Al-Edreesi E, Al-Gaphari G. Real-time Yemeni currency detection using deep learning. *arXiv.* 2024.
- [9] Zhang Q, et al. Deep learning techniques for automated currency recognition systems. *J Bank Financ Technol.* 2021.
- [10] Redmon J, et al. YOLOv5/YOLOv7 advancements for real-time object detection. *arXiv.* 2021–2022.
- [11] Wang CY, et al. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *CVPR.* 2022.
- [12] Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. *arXiv.* 2020.
- [13] Jocher G, et al. Ultralytics YOLOv8: Next-generation object detection framework. 2023.
- [14] Liu W, et al. SSD: Single Shot MultiBox Detector revisited with improvements. *IEEE TPAMI.* 2021.
- [15] Carion N, et al. End-to-End object detection with transformers (DETR). *ECCV.* 2020.
- [16] Zhu X, et al. Deformable DETR: Efficient transformer-based object detection. *ICLR.* 2021.
- [17] Lin TY, et al. Focal loss for dense object detection (updated applications). *IEEE TPAMI.* 2020.
- [18] He K, et al. Mask R-CNN advancements for real-time applications. *IEEE TPAMI.* 2021.
- [19] Liu L, et al. Development and challenges of object detection: A survey. *Neurocomputing.* 2024.
- [20] Tan M, Le Q. EfficientDet: Scalable and efficient object detection. *CVPR.* 2020.