Currency Detection for Visually Impaired Iraqi Banknote as a Study Case

Raghad Raied Mahmood [1] **Dr. Majid Dherar Younus** [2], **Dr. Emad Atiya Khalaf** [3]

Department of Computer And Information Engineering, College of Electronics Engineering, Nineveh University

_____

Abstract

It is relatively simple for a normal human to interpret and understand every banknote, but one of the major problems for visually impaired people are money recognition, especially for paper currency.

Since money plays such an important role in our everyday lives and is required for every business transaction, real-time detection and recognition of banknotes become a necessity for blind or visually impaired people For that purpose, we propose a real-time object detection system to help visually impaired people in their daily business transactions.

Dataset Images of the Iraqi banknote category are collected in different conditions initially and then, these images are augmented with different geometric transformations, to make the system strong. These augmented images are then annotated manually using the "LabelImg" program, from which training sets and validation image sets are prepared.

We will use YOLOv3 real-time Object Detection algorithm trained on custom Iraqi banknote dataset for detection and recognition of banknotes. Then the label of the banknotes is identified and then converted into audio by using Google Text to Speech (gTTS), which will be the expected output.

The performance of the trained model is evaluated on a test dataset and real-time live video. The test results demonstrate that the proposed method can detect and recognize Iraqi paper money with high mAP reaches 97.405% and a short time.

Keywords – Visually impaired, Object detection, Currency Recognition, Iraqi banknote, YOLOv3, Text to speech.

_____

1. Introduction

According to the study conducted worldwide by the World Health Organization (WHO), about 285 million people suffer visually impaired, of whom 39 million were blind, 246 million had low vision. The number of visually impaired people is exploding with the growth of the newborn population, eye diseases, accidents, aging, and so on, and every year, this number grows by up to 2 million worldwide [1][2].

The abilities of the visually impaired to interpret and understand every banknote are limited or influenced. For that reason, many visually impaired people will bring a sighted friend or family member to help them in their daily business transactions [3].

Previous research has suggested many strategies to overcome the issues of visually impaired people (VIPs) to interpret and understand every banknote, but the proposed ideas of these strategies are generally high in complexity, and not cost-effective etc.[4].

Currency plays a significant task as a medium of payments for products and services. Every country has its own money, which comes in a variety of colors, sizes, shapes, and patterns.

It is very complicated for anyone who is visually impaired to recognize and count various types of money. Owing to repeated use, tactile markings on the banknote's surface disappear or fade away, making it difficult for visually impaired people to detect and distinguish banknotes correctly by touch.

We suggest a system based on image processing and machine learning breakthroughs. The proposed Iraqi banknote detection, as well as recognition system to help people to identify banknotes in a real-time scenario.

The challenge is a self-built Iraqi banknote dataset, and after augmentation and manual annotation, transfer learning is performed on the YOLOv3 model, whereby the system includes a camera module and an audio jack. The camera will capture the banknote's image that is in front of the person. Thereafter, it gets processed using deep learning methods and, in turn, the output, which is the name of the banknote, will be converted into audio for the user through the audio jack. A system is proposed to aid Iraqi visually impaired people in dealing with day-to-day business activities.

2. Related Work

To support the visually impaired, many technologies have been developed. Some relevant works connected to this segment are described below:

In 2016, N. A. Semary. et al. [5], Proposed system based on simple image processing utilities. The basic techniques utilized in the proposed system include image foreground segmentation, histogram enhancement, region of interest (ROI) extraction and finally, template matching based on the cross-correlation between the captured image and our data set. The experimental results demonstrate that the proposed method can recognize Egyptian paper money with high quality reaches 89% and short time.

In 2020, K. Singh [6], Proposed system to detect Indian paper currencies including six kinds of currency paper. In the proposed work, first, take the input of the given image and pre-processed the given image and convert the RGB image into the grayscale image. After pre-processing, apply a Sobel algorithm for the extraction of the inner as well as the outer edges of the image. Clustering will be done using the YOLO V3 algorithm. In which it forms the clustering of features one by one. After that, they recognized the input image as a 200, 500, or 2000 and compared the features of the image and classified it as 200, 500, 2000, or not with the help of the YOLO V3 algorithm.

In 2021, N. A. Semary. et al. [7], proposed framework that mainly focuses on the identification of various currency notes and also whether the note is fake or original concerning the security features of a particular currency note. For faster identification of currency notes, the YOLO V3 architecture has been utilized to extract the features of the new Indian currency note.

3. Methodology

To implement the real-time Iraqi banknote detection and recognition system using YOLOv3 deep learning algorithm, Images are acquired and then, pre-processing, augmentation and annotation are done to train the YOLOv3 network model.

3.1 Image Augmentation and Annotation

For image acquisition, the Canon 750D camera with 24.2MP resolution is used in different scenarios like occlusion, illumination (lightning at the front, side and scattered), etc. Around 700 images of 7 denominations of Iraqi banknote were acquired from the camera. Image augmentation is done further to make a large image dataset that prevents the training model from overfitting and retains the correct details of dataset images. Then, these 700 images were increased to 7,700 images through different image augmentation techniques, which yield the dataset for all category Iraqi banknote. The methods of image augmentation include a rotation with 6 angles (90,-90,-45, 180, 135,-135,), half crop, and dark (-1.0,-0.5, 0.5,) as shown in Figure-1.
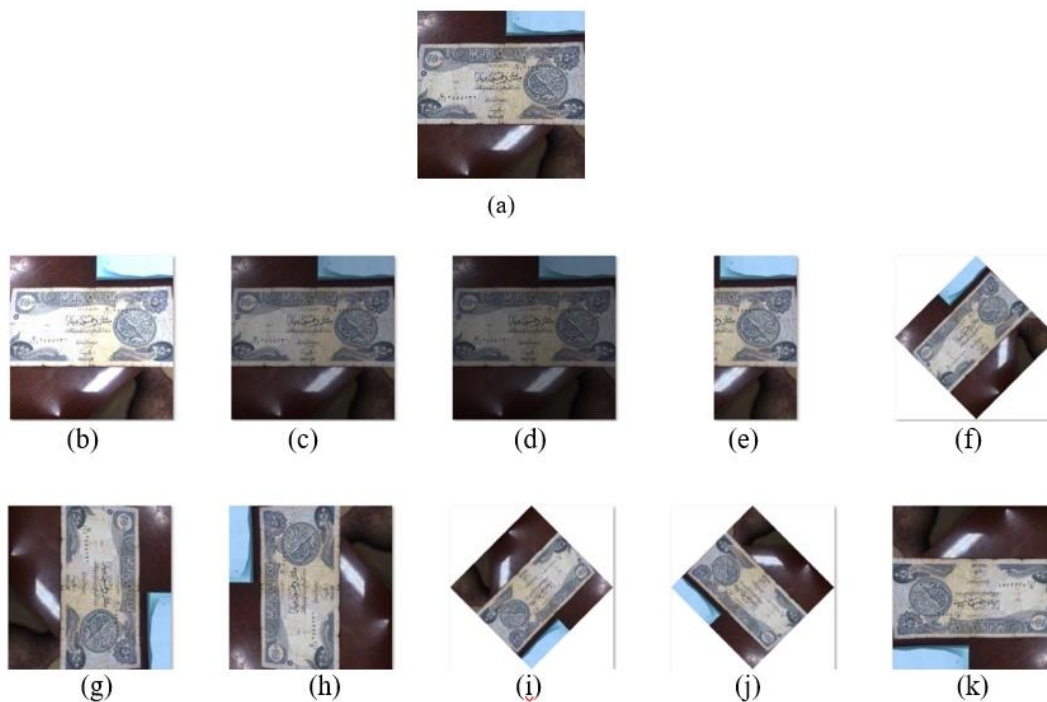

Figure -1. Different image augmentation techniques on acquired images (a) Original Image, (b) 0.5 dark, (c) -0.5 dark, (d) -1.0 dark , (e) half crop, (f) -45 angle rotate, (g) 90 angle rotate, (h) -90 angle rotate, (i) 135 angle rotate, (j) -135 angle rotate, (j) 180 angle rotate

_____

After having augmentation, this dataset of different Iraqi currency is divided into training and testing sets.70% of images were selected randomly for the training set in each denomination set of Iraqi banknote and the rest for testing sets.

To specify where the custom objects are located on the specific image, the images of different Iraqi banknote are manual annotation with the tool named "LabelImg", where the corresponding annotation files are saved in .xml format for all the dataset images. To do so, bounding boxes around the Iraqi banknote are drawn on every image. Then use script in OIDv6 (Open Images Detection Dataset version 6)Tool Kit to convert .xml labels to the appropriate YOLOv3 training format.

Samples of some Iraqi banknote in the different denominations used for training are given in Figure-2.

Figure -2. Samples of Iraqi currency Images in Dataset (a) 250 Dinar, (b) 500 Dinar, (c) 1000 Dinar, (d) 5000 Dinar, (e) 10000 Dinar, (f) 25000 Dinar, (g) 50000 Dinar

3.2 Transfer Learning and Model Training

YOLO network converts the problem of detection to a problem of regression. It develops a coordinated of bounding box and probability of each class straight by regression.

YOLO is an algorithm that uses convolutional neural networks CNN for object detection and not only predicts class labels but often detects target location. Therefore, it will not only classify the image into a category, but also detect multiple objects within an image.

YOLO v3 deeper architecture of feature extractor called Darknet-53. which contains of 53 convolutional layers, each followed by batch normalization layer and Leaky ReLU activation[8]. Figure-3 below shows the darknet-53 architecture.

_____

Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3 / 2	128 × 128
1× Convolutional	32	1 × 1	
Convolutional	64	3 × 3	
Residual			128 × 128
Convolutional	128	3 × 3 / 2	64 × 64
2× Convolutional	64	1 × 1	
Convolutional	128	3 × 3	
Residual			64 × 64
Convolutional	256	3 × 3 / 2	32 × 32
8× Convolutional	128	1 × 1	
Convolutional	256	3 × 3	
Residual			32 × 32
Convolutional	512	3 × 3 / 2	16 × 16
8× Convolutional	256	1 × 1	
Convolutional	512	3 × 3	
Residual			16 × 16
Convolutional	1024	3 × 3 / 2	8 × 8
4× Convolutional	512	1 × 1	
Convolutional	1024	3 × 3	
Residual			8 × 8
Avgpool		Global	
Connected		1000	
Softmax			

Figure - 3 Darknet-53 Architecture

The flow diagram for the training model from the collection of Iraqi banknote images and augmentation is shown in Figure -5

Images Acquisition → Images Augmentation → Images Annotation

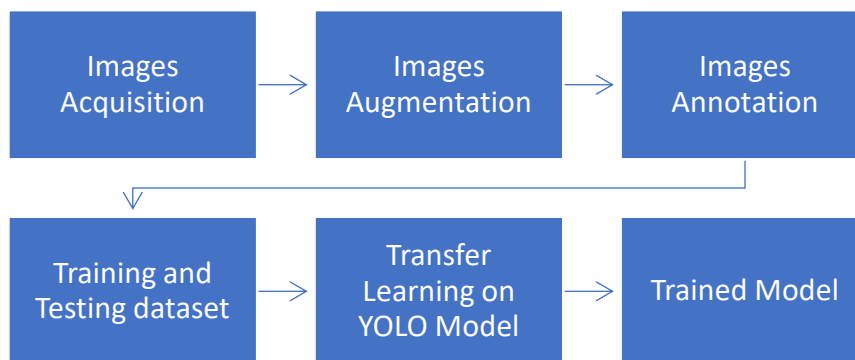Training and Testing dataset → Transfer Learning on YOLO Model → Trained Model

Figure -5 Training Model Flow Diagram

The model is trained with the datasets as mentioned above. This Algorithm applies a single Neural network to the Full Image. It means that this network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities[9].

YOLO predicts multiple bounding boxes per grid cell. For this, we select only a few boxes based on:

- **Score-thresholding:** Boxes that have detected a class with a value less than the threshold should be discarded.
- **Non-max suppression:** Calculate the intersection over the union to avoid selecting boxes that are overlapping.

During training of the model the objective is to optimize the loss function which is measure useing the sum squared error between the predictions and the ground truth.

The loss function [8] of YOLO v3 can be summarized as follows:

- **Confidence loss**: determine whether there are objects in the prediction frame.
- **Box regression loss:** computed only when the prediction box contains objects.
- **Classification loss:** determine the class of the object in the prediction frame.

Figure 4 depicts the YOLOv3 loss function.

$$
\begin{aligned}
\text{Regression loss} \quad & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
\text{Confidence loss} \quad & + \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
& + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
\text{Classification loss} \quad & + \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}
$$

Figure -4 YOLOv3 Loss function

The flow diagram of our user side system is depicted in the figure below.

Figure -6 Flow Diagram of User Side System

Imag

camera

(Input)

(YOLOv3 Model)

gTTS

Headset

(Audio Output)
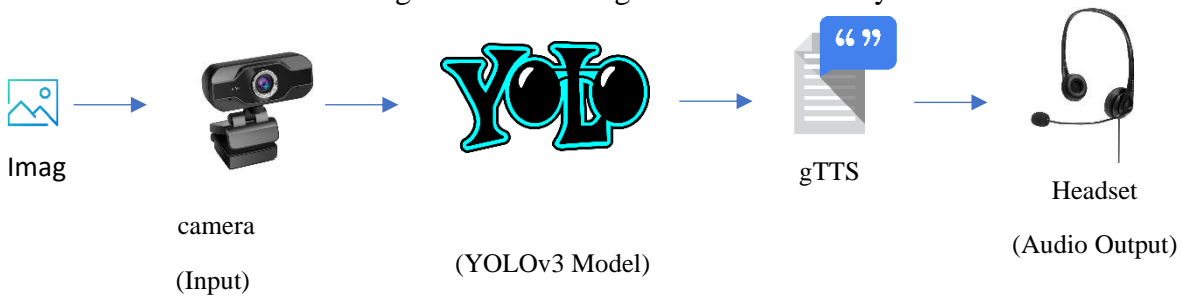
Our system's process begins with an image acquisition requirement, which is met by the camera. As an image is fed into a model trained on a custom dataset, it generates output for recognized banknotes as well as a mark and bounding box for each one. It can detect a multiple number of banknotes in a single image, and the proposed system has no form limitations. The label, which is a text output, is then converted into speech for different labels or banknote

recordings, which can then be played. As a result, the visually impaired person receives the audio output for each detected and recognized banknote using a headset.

4. Experiments And Analysis

The hardware requirements of the device to perform transfer learning and model training is an NVIDIA GPU card with CUDA Compute Capability. In this thesis, the GPU used is the GeForce GTX 1650 with these properties: (Compute Capability: 7.5, core Clock: 1.56GHz, core Count: 16, device Memory Size: 4.00 GiB, device Memory Bandwidth: 119.24 GiB/s).

The performance of the proposed Iraqi banknote recognition method has been evaluated on the datasets, which were collected manually and also evaluated on real-time live video.

In this section we will focus on the evaluation of the system by using the dataset collected manually that includes images of both the front side and backside for each category of Iraqi banknote (250 Dinar, 500 Dinar, 1000 Dinar, 5000 Dinar, 10000 Dinar, 25000 Dinar, 50000 Dinar). The dataset has no limitation to the amount of banknotes it can have any number of banknote images.

The system has been evaluated using a mAP (mean Average Precision) metric [10] which is the average of AP, mAP measuring to evaluate the performance of both classifications and localization of using bounding boxes in the image.

We use the concept of Intersection over Union (IoU). IoU measures the overlap between 2 boundaries which is used to measure how much our predicted boundary overlaps with the ground truth (the real object boundary):
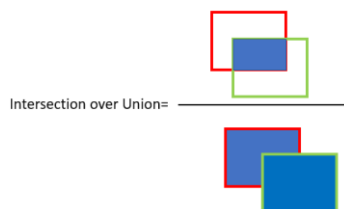
- Red - ground truth bounding box;
- Green - predicted bounding box.

To calculate mAP. First, we set a threshold value for the IoU to 0.5 to determine if the object detection is valid or not, in our case:

- If IoU ≥0.5, classify the object detection as True Positive(TP).
- If Iou <0.5, then it is a wrong detection and classifies it as False Positive(FP).
- When ground truth is present in the image and the model failed to detect the object, we classify it as False Negative(FN).
- True Negative (TN): TN is every part of the image where we did not predict an object. This metrics is not useful for object detection, hence we ignore TN.

To get mAP we calculate precision and recall for all the objects presented in the images. Precision and Recall are calculated using true positives (TP), false positives (FP) and false negatives (FN):

_____

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

We also consider the confidence score for each object detected by the model in the image. Bounding boxes above the threshold value are considered as positive boxes and all predicted bounding boxes below the threshold value are considered as negative. So, the higher the confidence threshold is, the lower the mAP will be, but we'll be more confident with accuracy.

Table 1 shows the evaluation results of the proposed system performance. The YOLOv3 based Iraqi banknote detection and recognition system achieves 97.405% mAP on different images.

No	Denomination of the banknote	No. of test images	Average Precision AP(%)
1	250 Dinar	250	93.544%
2	500 Dinar	250	94.209%
3	1000 Dinar	250	99.995%
4	5000 Dinar	250	98.876%
5	10000 Dinar	250	99.454%
6	25000 Dinar	250	97.384%
7	50000 Dinar	250	98.374%
mAP50		97.405%	

Table 1. Evaluation Performance of The System

The testing results have proved the power of training, variation in dataset and YOLOv3 model to detect and recognize Iraqi banknote. The output results are shown in Figure-7.

Figure -7 Output Results for Different Denominations of Iraqi Banknotes

_____

5. Conclusion

The proposed system is designed for currency detection and recognition that recognizes an Iraqi banknote to help blind people in their daily lives. The system is trained on a variety of images with various backgrounds and augmentation techniques, resulting in a high accuracy system for Iraqi banknote detection and recognition. After installing and programming the Yolov3 model in small digital signal processing devices available on the market, the visually impaired person can effectively use it in real-world situations involving paper currency detection, we can use any imaging device, mobile phone camera, or DSP processor-based device, with real-time processing speed. In addition, the dataset has no limitation to the amount of banknotes. It can have any number of banknote images, and it is possible to add foreign languages that can be used worldwide.

References

[1]   "WHO | World Health Organization." https://www.who.int/en (accessed Oct. 08, 2020).

[2]   J. CHEN, "Research on Image Processing for Assisting the Visually Impaired to Access Visual Information," no. September, 2015.

[3]   R. Rajwani, D. Purswani, and P. Kalinani, "Proposed System on Object Detection for Visually Impaired People," *Int. J. Inf. Technol.*, vol. 4, no. 1, pp. 1–6, 2018.

[4]   F. Rahman, I. J. Ritun, and N. Farhin, "Assisting the visually impaired people using image processing," no. July, 2018.

[5]   N. A. Semary, S. M. Fadl, M. S. Essa, and A. F. Gad, "Currency recognition system for visually impaired: Egyptian banknote as a study case," *2015 5th Int. Conf. Inf. Commun. Technol. Access. ICTA 2015*, no. December, 2016, doi: 10.1109/ICTA.2015.7426896.

[6]   K. Singh, "CURRENCY DETECTION FOR VISUALLY," vol. 7, no. 5, pp. 999–1002, 2020.

[7]   W. Rarani, V. Rode, C. Mahatme, D. Chavhan, and P. K. Gholap, "Indian Currency Note Recognition System using YOLO v3 Methodology," pp. 1349–1356, 2021.

[8]   J. Redmon and A. Farhadi, "YOLO v.3," *Tech Rep.*, pp. 1–6, 2018, [Online]. Available: https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[9]   A. F. Joseph Redmon∗, Santosh Divvala∗†, Ross Girshick¶, "You Only Look Once: Unified, Real-Time Object Detection Joseph," *J. Chem. Eng. Data*, vol. 27, no. 3, pp. 779–788, 2016, doi: 10.1021/je00029a022.

[10]   R. Padilla, S. L. Netto, and E. A. B. Da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," *Int. Conf. Syst. Signals, Image Process.*, vol. 2020-July, no. July, pp. 237–242, 2020, doi: 10.1109/IWSSIP48289.2020.9145130.