2020 International Conference on Contemporary Computing and Applications (IC3A)
Dr. A.P.J. Abdul Kalam Technical University, Lucknow. Feb 05-07, 2020

YOLO-v3 Based Currency Detection and Recognition System for Visually Impaired Persons

Rakesh Chandra Joshi
Centre for Advanced Studies,
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, India
rakeshchandraindia@gmail.com

Saumya Yadav
Centre for Advanced Studies
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, India
smyaydv15@gmail.com

Malay Kishore Dutta
Centre for Advanced Studies,
Dr. A.P.J. Abdul Kalam Technical University,
Lucknow, India,
malaykishoredutta@gmail.com

2020 International Conference on Contemporary Computing and Applications (IC3A) 978-1-7281-5432-9/20/$31.00 ©2020 IEEE 10.1109/IC3A48958.2020.233314

Abstract— **It is easy for a normal human being to perceive and recognize any banknote effortlessly but it is really much difficult for any visually impaired or blind person to perform the same task. As money has an important role in daily lives for any business transaction, real-time detection and recognition of banknotes become a necessity for a person especially who is blind or visually impaired. For that purpose, the YOLO-v3 CNN model based banknote detection and recognition system is proposed which is fast and accurate. Images of different denominations and in different conditions were are collected initially and then, these images are augmented with different geometric and image transformations on images, to make the system robust. These augmented images are then annotated manually, from which training sets and validation image sets are prepared. Later, the performance of the trained model has evaluated on a real-time scene as well as a test dataset. The test result shows that the proposed YOLO-v3 model based method has detection and recognition accuracy of 95.71% and 100%, respectively. The whole system is standalone and works in real-time.**

Keywords— Machine learning; Deep learning; YOLO-v3, Real-time detection; Banknote; instance segmentation; Currency Recognition.

I. INTRODUCTION

Several frameworks and techniques had been developed for healthcare services in the last decade. The aim of these advancements is to reduce the cost of the medical diagnosis and to assist the health sector with technology where a person can self-manage the things easily as never before without having the direct supervision from the specialist. However, the people having disabilities were not the primary target of these kinds of advancements. However, there is an urgent need for technologies, which can help and assist in day-to-day lives and can better their living in a simple manner and lead a way to independence. Out of these disabilities, Visual Impairment is much significant.

Worldwide more than 2.2 billion people suffer from visual impairment which includes 1 billion people with moderate or acute distance vision debilitation or blinds, mostly over 50 years of age [1]. The main reasons behind debilitation are glaucoma, cataract, unaddressed presbyopia or refractive error. The World Health Organization also estimated that the number of visual impairment affected people would increase substantially by the year 2020 [2]. Visually impaired people normally use assistive aids, such as walking dogs or white canes. Most commonly, the white cane is chosen because of the reasons like cheap price, handy and wide acceptance in the blind community. However, these assistive has its own limitations when a variety of hurdles and situations comes in their daily life. Most often, people consider such persons as a burden and leave them all alone to take care of themselves. Thus, the visually impaired person constantly require the assistance device, which can help in their daily tasks and rehabilitation. Blindness is a major issue that is frequent and even increases with age. Older people are at higher risk of visual impairment. For visual impairments, the assistive system act a significant role in social involvement. Without having these assistive devices makes them dependent on others. In addition, the rehabilitation cost is not affordable for low-income persons.

Currencies play an important role as a medium for a transaction to have goods and services. Every country has their own currency in different denominations, which differs in color, size, shape, and pattern. It becomes very difficult for any visually impaired to recognize and count the currency in different denomination. Tactile marks at the banknote's surface vanish or faded away due to continuous use, which suffers visually impaired people to detect and identify banknotes properly by means of touch. The digital image processing is a broad area, which gives the solution to these kinds of problem, where searching and extraction of the patterns as well as identification marks is performed and then match those with original banknotes images.

The main contribution of the proposed banknotes detection, as well as recognition system, is to make easy to use the handy standalone system, which will help people to identify banknotes in the real-time scenario. The challenging self-built dataset is generated and transfer learning is performed on the YOLO-v3 model after augmentation and manual annotation.

The rest of the paper is organized as follows. First, a review of related work is described in Section II, followed by the methodology explored in Section III. Finally, the experimental results are presented, followed by the conclusion and future work in Sections IV.

II. RELATED WORK

In recent years, deep learning emerges as one of the popular methods to solve the computation and prediction problem where the dataset is trained over neural networks. These models vary in their speed and accuracy. Single Shot MultiBox Detector (SSD) as the framework is used in [3], where features of currency were extracted to recognize the currency on three denominations. Even after having the augmentation, there is no such variation in the dataset as the background of trained and tested images in the dataset was same.

CNN based research is conducted on the folded banknotes in [4] where all the banknotes are only of the single denomination. A computer vision-based technique to perform automated banknotes recognition for the assistance of visually impaired people is developed in [5], where SURF features were used for recognition purposes. This study is performed on the US banknotes where the image of the person on banknote is different, which is easier to discriminate with respect to those having the same front note features as in India.

A novel banknote image processing system based on Free-From-Deformation (FFD) model is presented in [6], which can help in the low-quality banknote processing and reduce the false rejection rate. The new architecture of a system for the recognition and verification of banknotes based on neural networks for classification and verification is described in [7]. An approach for recognition of paper currency is proposed which used sequential deep neural network and data augmentation to improve accuracy [8].

The proposed task was made to handle small data problems and unable to handle real-time processing due to high computation. A currency recognition system for Ethiopian Banknotes using a support vector machine is discussed in [9], which recognized the front part of the currency well. It was not trained on backside of the banknotes.

The banknotes classification for Turkish lira using deep convolutional neural networks that is implemented and trained on the DenseNet-121 architecture [10]. The front side and backside of Myanmar currencies (kyats) in three denominations with image processing techniques are applied in [11]. For the purpose of feature extraction, Zernike moments were used and classification is done using the k - nearest neighbor algorithm. The neural network is also used in [12] to solve these kinds of problem for visually impaired. Their study suggests that cognition frameworks and neural activities can result in more significant research to do these kinds of tasks. The portable system for blind people for Euro banknotes detection and recognition is presented in [13]. The banknote detection is based on the modified Viola-Jones algorithms [14] and Speed Up Robust Features (SURF) [15], respectively.

The YOLO-v3 network [16] is an advanced model developed from the YOLO network [17] and YOLO-v2 networks [18]. In order to reduce the detection error in YOLO-v2, the "anchor box" concept is added in Faster R-CNN [19] and appropriate priori bounding boxes are generated by k-means clustering method. As a result, to achieve identical intersection over union result, the number of required anchor boxes decreases. The network structure is improved in YOLO-v2, which uses the convolution layer at the output layer of the YOLO for replacement of the fully connected layer. Batch normalization, dimension clusters, high-resolution classifier, fine-grained features, direct-location prediction, multi-scale training are also introduced in YOLO-v2 that greatly improves detection accuracy as compared to the YOLO model.

Thus, YOLO-v3 based CNN model can be designed to make fast and accurate banknote detection and recognition system for visually impaired people and blinds, to help in their daily lives.

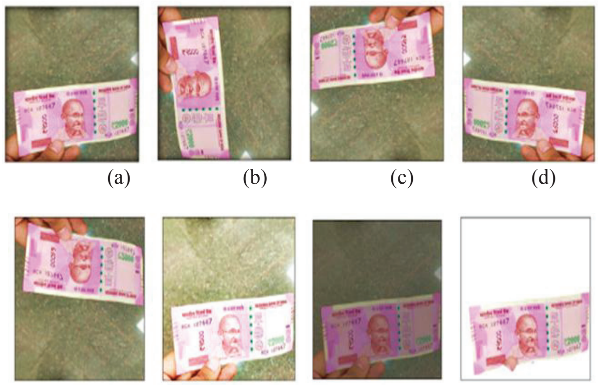
III. METHODOLOGY

To develop the real-time banknote detection and recognition system using deep learning, Images are acquired and then, pre-processing, augmentation and annotation are done to train the neural network.

A. Image Augmentation and Annotation

In this method, a camera with 1280*720 pixel resolution is used for image acquisition in different scenarios like occlusion, illumination (lightning at the front, side and scattered), etc. Around 3720 images of banknotes were acquired from the camera as well as web (Images were also in different formats like .jpg, .jpeg, .png). This dataset of different banknotes is divided into Training, validation and testing set. 65-70% of images were selected randomly for the training set in each denomination set of banknotes and rest for validation and testing sets. Image augmentation is done further to make a large image dataset that prevents the training model from overfitting and retain the correct details of dataset images. Then, these 3720 images were increased to 10,000 images through different image augmentation techniques, which yield the dataset for all category banknotes. The methods of image augmentation include a rotation, brightness, reflection, color, resizing, shear, adding noise, and background removal and translation transformations as shown in Fig. 1. To make it more complex, image data augmented also with the combination of these transformations and augmentation techniques.

After having augmentation, the images of different banknotes are arranged and numbered. Following the numbering of images, manual annotation was incorporated with the tool named as "LabelImg", where the corresponding annotation files are saved in .xml format for all the dataset images. To do so, bounding boxes around the banknotes are drawn on every image.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 1. Different image augmentation techniques on acquired images (a) Original Image, (b) 90° Horizontal rotation, (c) 180° Horizontal rotation, (d) Horizontal Flip, (e) Vertical Flip, (f) Increased Brightness, (g) Addition of noise, (h) Background Removal

The different denomination notes categorized by different labels on every image manually. Positive samples with to prevent the neural network from over-fitting, images with inadequate or ambiguous banknote pictures were not annotated. In addition to this, a banknote where an area of occlusion was more than 80% or the banknote at the corner of the image with less than 20% area, not used for annotation. The training dataset along with annotation files was used for training the YOLO-v3 detection model while remaining images were taken for validation dataset for verifying the detection and training performance of YOLO-v3 model. Samples of some banknotes in the different denominations used for training are given in Fig. 2.

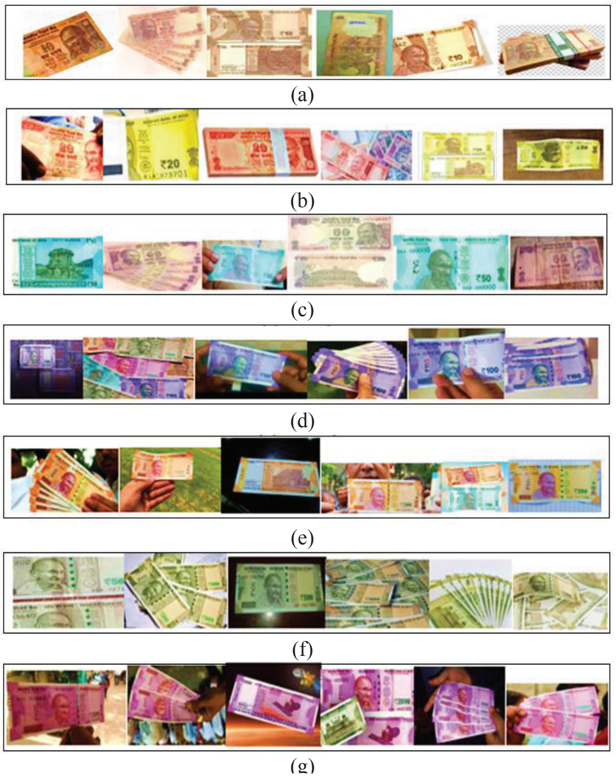


Fig. 2. Samples of Banknote Images in Dataset (a) 10 ₹, (b) 20 ₹, (c) 50 ₹, (d) 100 ₹, (e) 200 ₹, (f) 500 ₹, (g) 2000 ₹.

B. *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, instead of using the local classifier-based systems where high scoring regions are known as detection after applying the model to an image at multiple places and at different scales. This increases the speed of detection as compared to other methods like Faster R-CNN, RCNN [20] by 100 and 1000 times respectively. The accuracy and detection speed is much higher in YOLO. In place of applying the model at different places of image, the detection model is applied to complete image in YOLO. The neural network breaks down the whole image into different regions. After that, the bounding box and probability of each region are predicted.

YOLO-v2 is evolved to the new version YOLO-v3. Multi-scale prediction is used for the detection of the targeted objects, and the YOLO-v3 has a more complex network structure than YOLO-v2. The network structure of the YOLO-v3 detection model is presented in Fig. 3. Multi-scale bounding box prediction made YOLO-v3 more advantageous for small target detection compared to YOLO-v2.

The object detection process is shown in Fig. 4. The probability of an object to be contained inside a bounding box is represented by the object score. The object score should be approximately 1 for the center of ground truth box as well as for neighboring grids, whereas almost 0 for outside cells that or cells which are at the corners.

	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			

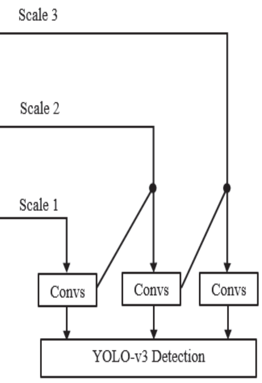


Fig. 3. Network Structure of YOLO-v3

The cell having a center of ground truth box of an object is liable for the prediction of the object. The four coordinates (**t_x, t_y, t_w, t_h**) are predicted by the network for each bounding box. If the bounding box has width and height **p_w, p_h** as anchor dimensions and the cell is offset from the top left corner of the image by (**c_x, c_y**), then, the predictions are done through the equations 1-4.

$$b_x = \sigma(t_x) + c_x \tag{1}$$
$$b_y = \sigma(t_y) + c_y \tag{2}$$
$$b_w = p_w e^{tw} \tag{3}$$
$$b_h = p_h e^{th} \tag{4}$$

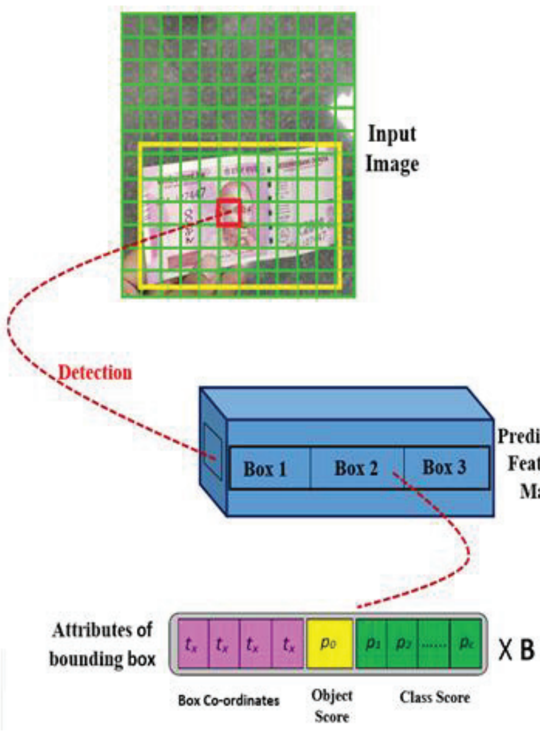


Fig. 4. Object detection

Each image of training set is divided into the grids. If the target ground truth center comes across a grid, then that grid will be responsible to detect the targeted object. 'B' number of bounding boxes with respective confidence scores and class conditional probabilities are predicted by each grid.

Flow diagram for the training model from the collection of banknote images and augmentation is shown in Fig. 5 (a) and the flow chart for the working of the proposed system as given in Fig. 5 (b). The input from the camera as an image frame is captured which is then pre-processed before giving that to the trained model. When the image is given to the model trained on the dataset, gives output for recognized banknotes and generate the label and bounding box for each of the respective banknote. It can detect the multiple number of the banknotes in a single image and there is no such type limit for the proposed system. The label, which is a text output, is then converted into the speech for different labels or recording of respective banknotes can be played subsequently. Thus, the audio output for each detected and recognized banknote is communicated to the visually impaired person through speaker or earphones.

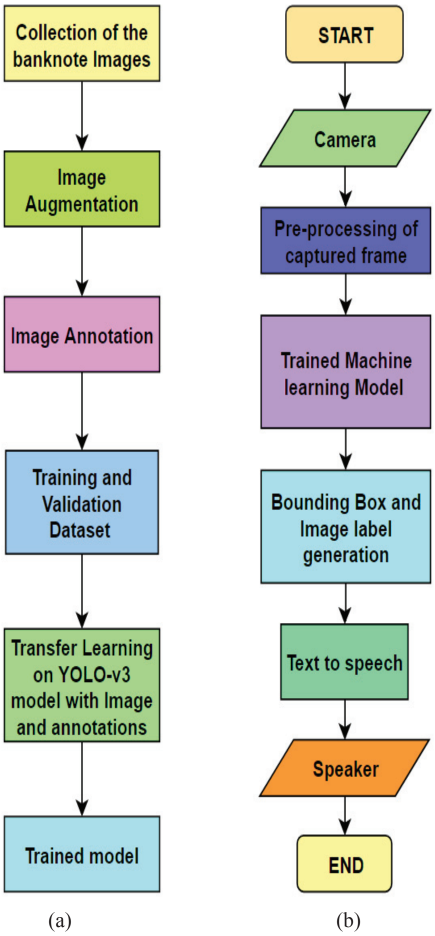


Fig. 5. Flow Diagrams (a) Model Training, (b) Proposed System

IV. EXPERIMENTS AND ANALYSIS

The device specifications to perform transfer learning and model training is Intel i-9 processor, 16 GB RAM, NVIDIA GPU with 6GB memory. The performance of the proposed paper currency note recognition method has been evaluated on the datasets, which were collected by developers manually, or through the internet. This section primarily concentrated on the evaluation of the system by using the dataset collected by the developers that include images of both front side and backside for each category of currency note (10₹, 20₹, 50₹, 100₹, 200₹, 500₹, 2000₹). Dataset has no limitation of the number of banknotes and it can have any number of banknotes an image in the dataset. These images are having a wide variety of conditions to develop real-world application. These experimental outcomes validated the effectiveness and robustness of the proposed YOLO-v3 model-based banknote recognition method. Banknotes images contain the various cases images like rotation, scaling, lighting change, partial occlusion or wrinkling, etc. In some cases, some of the conditions are combined into one condition, such as rotation and scaling. In addition to this, it also contains banknotes with different types of backgrounds. The dataset will be released for researchers in the future. With the help of this system, the user cannot only detect the banknote in a live video feed, although it will also recognize the denomination of the image. As the system is trained in various kinds of images with different backgrounds and with many augmentation techniques, it results in a highly robust system for detection and recognition of currency notes. The visually impaired user can easily use it after installation and programming the system in small digital signal processing devices available in the market that will help in the real-world cases related to paper currency recognition.

Table I shows the result after the testing of the proposed system. While the model made for a real-time scenario, the model also analyzed on the test dataset. N-fold cross-validation is done to test the performance of the system. The YOLO-v3 based banknote detection and recognition system achieves the 95.71 % average detection accuracy on different images and has 100% recognition accuracy once the banknote is detected successfully.

TABLE I. PERFORMANCE ANALYSIS OF THE PROPOSED SYSTEM

Denomination of the banknote	RESULTS				
	No. of test images	Correctly detected images	Detection Accuracy (%)	Correctly Recognized Images	Recognition Accuracy (%)
2000 ₹	150	141	94	141	100
500 ₹	150	148	98.67	148	100
200 ₹	150	143	95.33	143	100
100 ₹	150	140	93.33	140	100
50 ₹	150	144	96	144	100
20 ₹	150	140	93.33	140	100
10 ₹	150	149	99.33	149	100
Average Detection Accuracy		**95.71%**			
Average Recognition Accuracy		**100%**			

The testing results have proved the effectiveness of training, variation in dataset and YOLO-v3 model to recognize banknote. Although the proposed work is assessed in a very challenging dataset, it achieves very good detection as well as recognition results and outperforms the existing paper currency recognition algorithms, especially in the case of Indian Currency notes. The output results are shown in Fig. 6.

The proposed system is designed with the main objective for a device for currency recognition for blind users by using any imaging device or through mobile phone camera or DSP processor-based device, with real-time processing speed.

Fig. 6. Output results for different denominations of banknotes

To check the complexity handling capacity of the proposed system, both old and new type banknotes (available for 10₹, 20₹, 50₹ and 100₹ Indian rupee denomination, but not available for 200₹, 500₹ and 2000₹ Indian note due to demonetization) currency images for training and testing are included in the dataset. Even after having such complexity in the same class of banknote, it is detecting and recognizing correctly. The confusion matrix is also prepared to check the performance of the proposed system for the differentiation between the different banknote denominations. The confusion matrix for the threshold value of 0.5, is shown in Fig. 7 that clearly indicates the high recognition rate of the proposed system.

True Class \ Predicted Class	10₹	20₹	50₹	100₹	200₹	500₹	2000₹
10₹	1	0.12	0	0.073	0	0	0.060
20₹	0.157	1	0	0	0.092	0	0.05
50₹	0	0	1	0	0	0.069	0.076
100₹	0.078	0	0.05	1	0	0.042	0
200₹	0.048	0.069	0	0	1	0	0.055
500₹	0	0	0.047	0.020	0	1	0
2000₹	0	0.063	0	0.056	0.042	0	1

Fig. 7. Confusion Matrix

V. CONCLUSION AND FUTURE WORK

In this paper, the banknote or currency detection and recognition system based on the YOLO-v3 model is proposed which is standalone and working in real-time. The model is trained with multiple images of different denominations currency banknotes of different denominations as a separate class. Dataset of Banknotes is made in different conditions like cluttered background, rotation, occlusion, illumination level, scaling, etc. To overcome the overfitting of the data augmentation is done which results in the increase in the number of images having a wide variety, which ultimately helps to make detection and recognition system robust and accurate. After having the proper augmentation, annotation is done for all images that get included in the dataset. Thereafter, the dataset is split into two parts- Training and Validation set. Then, transfer learning is applied to the YOLO-v3 model and trained with the help of images and their resistive annotation files. The whole system is standalone and does not need any internet facility to perform its recognition task.

After the completion of training to the level where learning rate saturates, respective trained model files are being used for the real-time banknote detection and recognition system with a live video feed. With the help of this system, any visually impaired or blind person can recognize the banknotes, which will help in their daily lives. The proposed algorithm has achieved 95.71% average detection and 100% average recognition rate on the dataset and is robust to recognize the banknotes even in cases of partial occlusions as well as wrinkled or torn currency notes.

The future work on this system will focus to optimize the proposed system and to enlarge the training dataset by incorporating more banknotes of different countries as well. In addition, work will be done to develop an interactive interface so that it can have many functions like automatic counting of banknotes, summing of currency and a UV ray, which will also help to detect the counterfeit notes.

ACKNOWLEDGMENT

This research work is supported in parts by the Grants from the Department of Science and Technology.

REFERENCES

[1] Blindness and vision impairment, Available Online: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed on 22 Nov, 2019).

[2] Global trends in the magnitude of blindness and visual impairment, Available Online: https://www.who.int/blindness/causes/trends/en/ (accessed on 22 Nov, 2019).

[3] Q. Zhang and W. Q. Yan, "Currency Detection and Recognition Based on Deep Learning," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6.

[4] M. JIAO, J. HE and B. ZHANG, "Folding Paper Currency Recognition and Research Based on Convolution Neural Network," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 18-23.

[5] F. M. Hasanuzzaman, X. Yang and Y. Tian, "Robust and Effective Component-Based Banknote Recognition for the Blind," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1021-1030, Nov. 2012.

[6] Y. Jin, L. Song, X. Tang and M. Du, "A Hierarchical Approach for Banknote Image Processing Using Homogeneity and FFD Model," in IEEE Signal Processing Letters, vol. 15, pp. 425-428, 2008.

[7] Frosini, M. Gori and P. Priami, "A neural network-based model for paper currency recognition and verification," in IEEE Transactions on Neural Networks, vol. 7, no. 6, pp. 1482-1490, Nov. 1996.

[8] Hoang and H. Vo, "Hybrid discriminative models for banknote recognition and anti-counterfeit," 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, 2018, pp. 394-399.

[9] M. E. Ayalew Tessfaw, M. B. Ramani and M. T. Kebede Bahiru, "Ethiopian Banknote Recognition and Fake Detection Using Support Vector Machine," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1354-1359.

[10] G. Baykal, U. Demir, I. Shyti and G. Ünal, "Turkish lira banknotes classification using deep convolutional neural networks," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[11] K. N. Oo and A. K. Gopalkrishnan, "Zernike Moment Based Feature Extraction for Classification of Myanmar Paper Currencies," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 1-6.

[12] O. K. Oyedotun & A. Khashman, "Banknote recognition: investigating processing and cognition framework using competitive neural network", Cognitive Neurodynamics, Springer, vol. 11, issue 1, pp. 67-79, 2017

[13] Dunai, Larisa & Chillarón Pérez, Mónica & Peris-Fajarnés, Guillermo & Lengua, Ismael. (2017). Euro Banknote Recognition System for Blind People. Sensors. 17. 184. 10.3390/s17010184.

[14] Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001.

[15] Bay, H.; Ess, A.; Tuytelaass, T.; van Gool, L. Speed-Up Robust Features (SURF). Comput. Vis. Image Underst. J. 2008, 110, 346–359.

[16] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[17] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[18] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.

[19] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[20] Ming Liang and Xiaolin Hu, "Recurrent convolutional neural network for object recognition," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3367-3375.