JAPANESE COINS AND BANKNOTES RECOGNITION FOR VISUALLY IMPAIRED PEOPLE

JAPANESE COINS AND BANKNOTES RECOGNITION FOR VISUALLY IMPAIRED PEOPLE

Bui Thi Thanh Huyen
Applied Informatics major, Graduate School of Science and Engineering,
Hosei University,
Supervisor: Professor Jinjia Zhou

***Abstract*- Recent deep learning techniques are successfully integrated into devices to assist visually impaired people in their daily lives, particularly detecting coins/banknotes. Previous works have focused on well-captured devices and examined high-quality images. In this work, we design a framework to recognize Japanese Coin/Banknote (JCB) for low-quality images under various criteria. Discriminate features usually disappear in low-quality images. Consequently, using the depth image in addition to RGB image in processing can be enhanced the accuracy of our system. In this work, we first leverage depth information by using a Monocular Depth Prediction network. Additionally, a pre-trained Deep Convolutional Neural Network process RGB and Depth images, respectively. At last, we combine two networks by an ensemble method to produce more accurate detections. By processing depth images in addition to RGB images, the detection results are thus accurate. As a result, our work achieves 74.1% mean Average Precision (mAP).**

Keywords: Visually Impaired People, Japanese Currency Recognition, Depth Estimation, Object Detection, Deep Learning

I. INTRODUCTION

Visual impairment is a decreased ability to see, it means that a person's eyesight cannot see objects as clearly as usual. According to the report of the World Health Organization (WHO), there were an estimated 2.2 billion people have a near or distance vision impairment around the world in 2020. Of these, 237 million people are thought to have moderate or severe distance vision impairment. Reduced or absent eyesight affects all aspects of daily living, interacting with the community, and the ability to work. Identifying objects visually is a simple task for normal humans, but not really for individuals with blindness and impaired vision. One of the most important problems facing visually impaired people is currency recognition. Therefore, developing assistive technology to support people with visual impairment is thus necessary and very important. In this work, we design a Coin/banknote recognition framework to assist visually impaired people.

Various algorithms have been proposed to recognize Coin/Banknote, from traditional approaches like SIFT matching[1], [2] and CircularHough Transform[3] to Deep Learning-based approaches like Faster R-CNN[4], SSD[5], and YOLO[6]. These methods implement on the dataset with a simple background and non overlapped coin/ banknote. Few researchers have addressed the problem of detecting Coin/Banknote on smartphone cameras [2], [4], [7] that processing on high-quality images. Although these approaches are impressive, there are not suitable for detecting coin/banknote in the real-world scenario.

Recent developments in object detection have achieved a great result on RGB images. However, these methods raise some weaknesses for detecting the real-world object; the RGB image can not represent the depth information in the 3D world led to a lack of information. To overcome this drawback, using the depth image addition to the RGB image can be considered. A depth image shows more description of the image, which enriches the representations of each target object. Many approaches for RGB-D images[8], [9] have been studied to show the usefulness of leveraging depth images. Moreover, these methods still need a depth sensor that can capture depth information and RGB images.

In this study, we present a Japanese coin/banknote detection framework that can detect low-quality images. Our system contains an object detection network well-trained on low-quality images collected by our low-cost video-capture glasses under various criterions such as occlusion, glare, noise, and so on. Firstly, we introduce a novel dataset – Japanese Coins and Banknotes dataset (JCB). To effectively leverage the depth information, we implement the Monocular Depth Prediction[10] to create depth images. After that, our system process deep CNN networks on both RGB image and depth images, respectively. Finally, an ensemble method combines two networks to produce more accurate detections.

II. METHODS

A. Our proposal

Our work mainly focuses on processing low-quality images. The main contribution of our work consists of 4 aspects: (1) Firstly, we introduce a novel dataset - Japanese Coin and Banknote Dataset (JCB dataset) collected by our low-cost video-capture glasses under various criteria. The selection of object categories based on Japanese currency. It contains nine categories (1 Yen, 5 Yen, 10 Yen, 50 Yen, 100 Yen, 500 Yen, 1,000 Yen, 5,000 Yen, 10,000 Yen). (2) Low-quality image easily gives false-positive detection due to noises, lack of information. To overcome this drawback, we implement the Monocular Depth Prediction to generate the depth images from RGB images. (3) Furthermore, our system pre-train a YOLOv4 network on both depth images and RGB images. (4) Finally, an ensemble learning model combines two networks to produce more accurate detections. Figure 1 shown our proposed network.

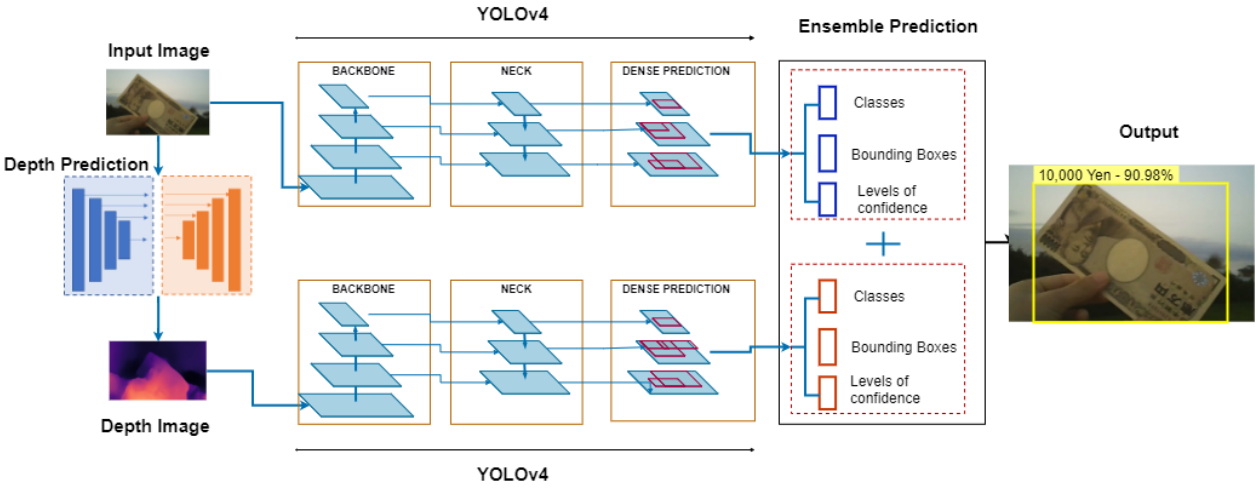

Figure 1: Overview of our framework. We use the YAOAWE Glasses to generate the RGB image. By using the Monocular depth estimation network, the Depth images are generated. Two YOLOv4 models independently process RGB and Depth images. An ensemble method processes the output of two models and compute the result for the target object.

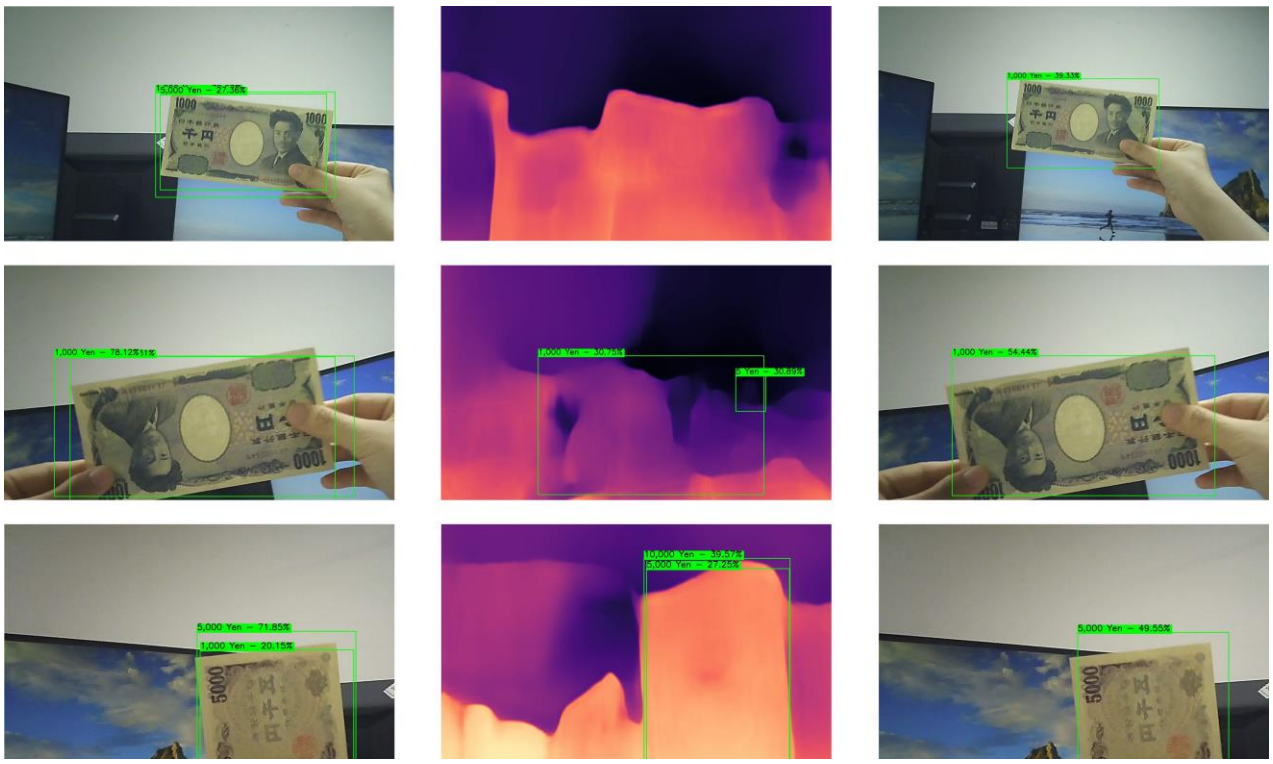

Figure 2: Our qualitative results on the YOLOv4 model. The first two columns are the results of RGB-YOLOv4 and Depth-YOLOv4 models. The last columns show our result on ensemble learning method.

B. Depth images generation

The depth information has various advantages, providing geometry contour from the color/depth, depth disparity. The trained model is usually distracted by complicated background; plus, low-quality image easily gives false-positive detection due to noises, lack of information. Consequently, using the depth image in processing can be enhanced the accuracy of our system. Recently, numerous works have been proposed to leverage depth information. However, these approaches have examined depth images from the RGB-D sensor. To address this problem, we implement the Monocular Depth Prediction estimating depth disparity to our system to create depth images.

C. Training YOLOv4 for RGB images and Depth Images

In this work, we apply one of state-of-the-art, real-time object detection network YOLOv4 for Japanese

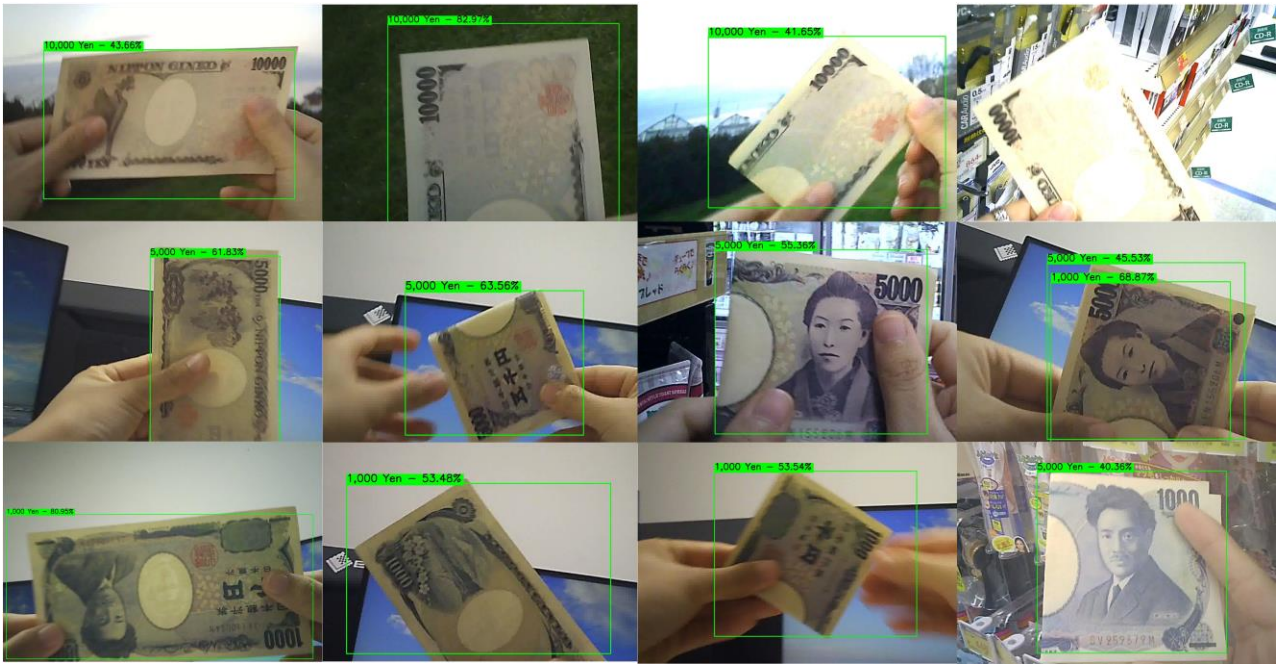

Figure 3: Our quantitative results with the ensemble method. The last column shows our fail cases.

Coin and Banknote dataset. Both RGB and Depth images are trained independently.

D. Combine RGB and Depth detections

The ensemble method is a machine learning technique that combines several models in order to produce more accurate detection. Some works combine the output of detections by applying Non-Maximum Suppression[11], Soft-NMS[12], fusion[13] to reduce misdetection. However, these approaches only process the bounding boxes to remove the false detection.

In our work, we consider both the bounding box and the class of detected objects. Our work uses a voting strategy – Unanimous for the JCB dataset. In this strategy, a box is considered if all of the models generate the same object in a region. Furthermore, the box is considered if it is generated by RGB model with high probability.

E. Data Preparation

We present a low-quality dataset of Japanese Coin and Banknote (JCB), which consisting of 7097 images. The data were obtained using our video-capture YAOAWE glasses, which has a low-cost camera integrated to collect images and videos. Based on the Japanese currency, nine categories were collected: 6 of coins and 3 of banknotes. We manually annotate the bounding boxes of each image. The resulted dataset has 7097 images of nine types of Japanese currency.

III. EXPERIMENTAL RESULTS

In this section, we conduct YOLOv4 to predict nine types of Japanese Coin/Banknote. Our study uses the pre-trained model for the training. The experiments were conducted on a GPU GeForce RTX 208 Ti.

We evaluate two YOLOv4 networks for RGB and depth images and compare the results with the ensemble method. These results are given in Table 1. We use a confidence score of 0.25 to evaluate mAP for each model. Using 788 annotated images, our framework achieved an mAP of 0.741. As shown in this table, the ensemble method improves the average precision of each class.

Table 1: Our performance (Average precision %) on each coin/banknote with IOU thresh=0.25

Coin/ Banknote	Depth-Yolov4(%)	RGB-Yolov4(%)	Ensemble Method(%)
1 Yen	33.12	82.8	88.39
5 Yen	32.83	66.15	88.67
10 Yen	18.21	17.65	42.59
50 Yen	24.03	95.14	94.87
100 Yen	22.65	91.47	95.52
500 Yen	15.21	46.37	57.27
1,000 Yen	7.93	30.12	49.36
5,000 Yen	3.76	61.41	71.2
10,000 Yen	16.63	70	79.05
mAP	19.37	62.35	74.1

Observe the quantitative results, some false detections occur in our model. As shown in **Error! Reference source not found.**, the last column shows our fail cases. The first image is unable to detect due to brightness; the next two images, these images have the wrong bounding boxes or multi-bounding boxes in an object.

As a result, our work achieves 74.1% mean Average Precision (mAP) with the ensemble method. Subjectively, we show our qualitative results on many aspects such as foggy, folded banknotes, complicated background, lack of brightness, as shown in Figure 2. In several cases, our model still failed on recognition (e.g., wrong class, unable to detect).

IV. CONCLUSION

Our study provides a framework for Japanese Coins and

Banknotes recognition. In conclusion, we have obtained the novel dataset of Japanese Coins and Banknotes under various criterions such as occlusion, glare, noise, and so on. Furthermore, we use a Monocular Depth Prediction to generate the depth image dataset from the RGB dataset. Our work implements the two YOLOv4 models for RGB images and depth images. By using the ensemble method, our work achieves better results than single model RGB-YOLOv4 and Depth-YOLOv4.

REFERENCES

[1] T. Yingthawornsuk, N. Chumuang, and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," in *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, Dec. 2017, pp. 418–423. doi: 10.1109/SITIS.2017.75.

[2] I. Abu Doush and S. AL-Btoush, "Currency recognition using a smartphone: Comparison between color SIFT and gray scale SIFT algorithms," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 484–492, Oct. 2017, doi: 10.1016/j.jksuci.2016.06.003.

[3] R. S. Hassoubah, A. F. Aljebry, and L. A. Elrefaei, "Saudi riyal coin detection and recognition," in *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, Dec. 2013, pp. 62–66. doi: 10.1109/ICIIP.2013.6707556.

[4] C. Park, S. W. Cho, N. R. Baek, J. Choi, and K. R. Park, "Deep Feature-Based Three-Stage Detection of Banknotes and Coins for Assisting Visually Impaired People," *IEEE Access*, vol. 8, pp. 184598–184613, 2020, doi: 10.1109/ACCESS.2020.3029526.

[5] Q. Zhang and W. Q. Yan, "Currency Detection and Recognition Based on Deep Learning," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Nov. 2018, pp. 1–6. doi: 10.1109/AVSS.2018.8639124.

[6] R. C. Joshi, S. Yadav, and M. K. Dutta, "YOLO-v3 Based Currency Detection and Recognition System for Visually Impaired Persons," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, Feb. 2020, pp. 280–285. doi: 10.1109/IC3A48958.2020.233314.

[7] X. Liu, "A camera phone based currency reader for the visually impaired," Jan. 2008, pp. 305–306. doi: 10.1145/1414471.1414551.

[8] T. Ophoff, K. Van Beeck, and T. Goedemé, "Exploring RGB+Depth Fusion for Real-Time Object Detection," *Sensors*, vol. 19, no. 4, Art. no. 4, Jan. 2019, doi: 10.3390/s19040866.

[9] K. Zhou, A. Paiement, and M. Mirmehdi, "Detecting humans in RGB-D data with CNNs," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, May 2017, pp. 306–309. doi: 10.23919/MVA.2017.7986862.

[10] C. Godard, O. Mac Aodha, M. Firman, and G. Brostow, "Digging Into Self-Supervised Monocular Depth Estimation," *arXiv:1806.01260 [cs, stat]*, Aug. 2019, Accessed: Dec. 09, 2020. [Online]. Available: http://arxiv.org/abs/1806.01260

[11] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," *arXiv:1705.02950 [cs]*, May 2017, Accessed: Jul. 05, 2021. [Online]. Available: http://arxiv.org/abs/1705.02950

[12] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS -- Improving Object Detection With One Line of Code," *arXiv:1704.04503 [cs]*, Aug. 2017, Accessed: Jul. 05, 2021. [Online]. Available: http://arxiv.org/abs/1704.04503

[13] P. Wei, J. E. Ball, and D. T. Anderson, "Fusion of an Ensemble of Augmented Image Detectors for Robust Object Detection," *arXiv:1803.06554 [cs, eess]*, Mar. 2018, Accessed: Jul. 05, 2021. [Online]. Available: http://arxiv.org/abs/1803.06554