UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Ingeniería

https://revistas.udistrital.edu.co/index.php/reving/issue/view/1221

DOI: https://doi.org/10.14483/23448393.21408

Research

Mobile Application for Recognizing Colombian Currency with Audio Feedback for Visually Impaired People

Aplicación móvil para el reconocimiento de moneda colombiana con retroalimentación de audio para personas con discapacidad visual

Camila Bolaños-Fernández[1] and Eval Bladimir Bacca-Cortes[1]*

[1]Universidad del Valle (Cali, Colombia)

Abstract

Context: According to the census conducted by the National Department of Statistics (DANE) in 2018, 7.1 % of the Colombian population has a visual disability. These people face conditions with limited autonomy, such as the handling of money. In this context, there is a need to create tools to enable the inclusion of visually impaired people in the financial sector, allowing them to make payments and withdrawals in a safe and reliable manner.
Method: This work describes the development of a mobile application called *CopReader*. This application enables the recognition of coins and banknotes of Colombian currency without an Internet connection, by means of convolutional neural network models. CopReader was developed to be used by visually impaired people. It takes a video or photographs, analyzes the input data, estimates the currency value, and uses audio feedback to communicate the result.
Results: To validate the functionality of CopReader, integration tests were performed. In addition, precision and recall tests were conducted, considering the YoloV5 and MobileNet architectures, obtaining 95 and 93 % for the former model and 99 % for the latter. Then, field tests were performed with visually impaired people, obtaining accuracy values of 96 %. 90 % of the users were satisfied with the application's functionality.
Conclusions: CopReader is a useful tool for recognizing Colombian currency, helping visually impaired people gain to autonomy in handling money.

Keywords: mobile application, convolutional neural network, visually impaired people, Colombian currency recognition

Article history

Received:
12th /Oct/2023

Modified:
10th /Mar/2024

Accepted:
23th /Apr/2024

Ing, vol. 29, no. 2, 2024, e21408

* **Correspondence:** bladimir.bacca@correounivalle.edu.co

Resumen

Contexto: Según el censo realizado por el Departamento Nacional de Estadística (DANE) en 2018, el 7.1 % de la población colombiana tiene una discapacidad visual. Estas personas enfrentan condiciones con autonomía limitada, como lo es el manejo de dinero. En este contexto, es necesario crear herramientas que permitan la inclusión de las personas con discapacidad visual en el sector financiero, permitiéndoles realizar pagos y retiros de manera segura y confiable.
Método: Este trabajo describe el desarrollo de una aplicación móvil llamada *CopReader*. Esta aplicación permite el reconocimiento de monedas y billetes de la moneda colombiana sin conexión a Internet, mediante modelos de redes neuronales convolucionales. CopReader fue desarrollada para ser utilizada por personas con discapacidad visual: toma un video o fotografías, analiza los datos de entrada, estima el valor de la moneda y utiliza retroalimentación auditiva para comunicar el resultado.
Resultados: Para validar la funcionalidad de CopReader, se realizaron pruebas de integración. Además, se llevaron a cabo pruebas de precisión y *recall*, considerando las arquitecturas YoloV5 y MobileNet, donde se obtuvo 95 y 93 % para el primer modelo y 99 % para el segundo. Luego, se realizaron pruebas de campo con personas visualmente discapacitadas, obteniendo valores de exactitud del 96 %. El 90 % de los usuarios quedaron satisfechos con la funcionalidad de la aplicación.
Conclusiones: CopReader es una herramienta útil para el reconocimiento de la moneda colombiana, ayudando a las personas con discapacidad visual a ganar autonomía en el manejo del dinero.

Palabras clave: aplicación móvil, red neuronal convolucional, personas con discapacidad visual, reconocimiento de moneda Colombiana

Table of contents

1. Introduction

The human visual system is in charge of up to 90 % of the interaction with the environment (1). This system allows gathering, processing, and obtaining images from the surroundings. The brain is part of this system, as it is where images are processed. Another important part includes the eyes and the optic nerve, which gather all the necessary information from the environment and transmit it to the brain. Therefore, a total or partial reduction of visual capabilities is regarded as an impairment, which could be categorized into two groups: blindness and low vision (2). Visually impaired people have many difficulties when it comes to identifying low- and head-level obstacles and currency, and they lack awareness of their surroundings. These difficulties affect their emotional state, leading to depression, frustration, and anxiety (3).

Nowadays, 2200 million people have a visual impairment, and 11.9 million of them are blind or have severe visual issues (1). In Colombia, 62.17 % of the total population with disabilities have a visual impairment (4), which is why Law 1346 of 2009 adopted conventions regarding people with disabilities. For blind people, thanks to this law, braille language is required in elevator buttons, public signaling, and bus stops, among others. However, one of the most significant issues of visually impaired people is unemployment; 62 % of them have no formal job (5), since most companies do not possess a sufficient infrastructure to ensure inclusiveness. Another difficulty faced by visually impaired people is access to the financial system. In Colombia, the Bank of the Republic implemented two inclusive methods to help visually impaired people to identify Colombian currency: the Braille system (2010) and tactile marks (2016) (6). However, both measures suffer from the wear generated by constant use, making these marks lose their relief.

Additional solutions to Braille systems or printing marks on currency have focused on technological options, such as Orcam (7), a visual-based system with multiple recognition functionalities and audio feedback that is lightweight and compact but very expensive. Another option is using smartphones, which are currently popular. In this context, mobile apps such as CashReader (8) have been developed, and there are many others. These focus on one specific currency, but not the Colombian one.

Therefore, in this work, the development of a mobile application called CopReader is described. CopReader enables the recognition of coins and banknotes of Colombian currency without an Internet connection, by means of convolutional neural network (CNN) models. CopReader was developed to be used by visually impaired people. It is simple to use: it takes a video or pictures, analyzes the input data, estimates the currency value, and uses audio feedback to communicate the result.

This paper is organized as follows: section 2 outlines some related works; section 3 describes the development of CopReader and the dataset used; section 4 delves into the development of CopReader; section 5 presents and discusses the results; and section 6 provides our conclusions.

2. Related works

The development of inclusive technologies based on computer vision and machine learning is an interesting research field. Table I summarizes the most representative works related to this research. These works were compared while considering the recognition goal, the platform used, the input data, the algorithm employed, the size of the dataset, the type of feedback, and the accuracy obtained. These criteria are important technical aspects to develop this kind of inclusive technology.

Table I. Related works

Ref.	Rec. Goal	Platform	Input Data	Algorithm	Dataset	Feedback	Accuracy
(9)	Objects	Smartphone	Video	YoloV3	N/A	Audio	N/A
(12)	Banknote	Emb. Syst.	Photo	Perceptron M.	13 200 img.	Glove	96 %
(10)	Objects	Smartphone	Video	YoloV3, RCNN	80 000 img, 80 classes	Audio	85.5 %
(13)	Objects	Glasses	Photo	YoloV3	330 000 img.	Audio	96.3 %
(14)	Addresses	Smartphone	Photo	Perceptron M., MobileNet, CNN	600 000 img.	Video	94.6 %
(15)	Car license plates	Emb. Sys.	Photo	YoloV2	500 img.	Video	94.9 %
(16)	Coins	Smartphone	Photo	Perceptron M.	48 img.	Audio	82 %
(17)	Coins	Smartphone	Photo	AlexNet	8 320 img.	Audio	72.2 %
(18)	Coins	Smartphone	Photo	SIFT Matching	700 img.	Audio	65.7 %
(11)	Objects and banknote	Smartphone	Photo	CNNdroid	N/A	Audio	66.6 %
(19)	Banknote	Smartphone	Photo	MobileNet	12 160 img.	Audio	96.6 %
(20)	Coins	Emb. Sys.	Photo	AlexNet	1 600 img.	Video	N/A
(21)	Banknote	Emb. Sys.	Photo	SVM, AlexNet	504 img.	Prot.	99.7 %
(22)	Banknote	Emb. Sys.	Photo	kNN, CNN	80 img.	Prot.	95.6 %
(23)	Banknote	Emb. Sys.	Video, Photo	CNN	70 000 img.	Prot.	94.3 %

The works presented in Table I show that it is important to recognize banknotes, coins, and objects in general, as this provides visually impaired people with more autonomy in daily activities. Works such as (9, 10), and (11) focus on low and head-level objects, which are potentially dangerous for the aforementioned population. The studies on banknote and coin recognition shown in Table I are especially designed for some currencies, and they do not support others.

Another important aspect is the platform used. Most of the reviewed works use smartphones to implement the proposed solution, but employing embedded systems is also popular. (13) propose the use of smart glasses, albeit with no onboard processing capabilities. These glasses also need remote backend server support, which means a permanent Internet connection. The works that use smartphones run their solutions offline and take advantage of the current capabilities of these devices. Using embedded systems such as the Raspberry Pi (12, 15, 20) is a popular option, as it is very flexible. However, short response times are needed, not 2 s or even 3.6 s, as reported in these works. These values are not enough, as per the needs stated by surveyed test subjects, who need to obtain a solution just by placing the currency in front of the camera.

Visually impaired people usually walk around their environment carefully, but their motions are performed at normal speed. Therefore, a high data acquisition rate is a significant requirement. Most of the works reported in the literature (as shown in Table I) employ a shot of the scene to be processed, but some of them (9, 10, 23) use video to recognize objects.

Object recognition (*i.e.*, objects in general, coins, or banknotes) is a challenging task, where changes in the background, lighting conditions, occlusions, and/or partial information hinder the obtainment of high accuracy measures. Thus, it is common to use deep learning algorithms, such as Yolo, RCNN, MobileNet, AlexNet CNNdroid, or the Multilayer perceptron. Other machine learning algorithms have been used, *e.g.*, SVM (21) or classical feature matching based on SIFT (Scale-Invariant Feature Transform) (18). Normally, visually impaired people need portable and reliable solutions, which technically means the use of an algorithm with high confidence but low parametrization. Therefore, the works reviewed used Yolo (9, 10, 13, 15), MobileNet (14, 19), AlexNet (17, 20, 21), or R-CNN (10, 11, 14, 22, 23).

Using deep learning or machine learning algorithms implies using a dataset that has a high number of samples and is well organized. Therefore, the dataset used and the accuracy achieved in a given work are related. Proposals such as (9, 11, 15, 21), and (22) could be over-trained, and their real-time performance (which is not reported by the authors) could be very different. However, in general, the works reviewed in Table I use large datasets to ensure diversity.

User feedback is a very important aspect of human-machine interaction. Some cases, such as (14,15), and (20) are solutions oriented to slight visual impairments. Nevertheless, for people with high levels of visual impairment or blindness, audio feedback is preferable – or, alternatively, the use of gloves (12) to transmit the result of recognition tasks via an established touch code. Some other works (21–23) present prototypes that are tested on embedded systems but are not fully developed for actual interaction with visually impaired people.

The accuracy of any proposed solution is important, as it indicates whether the approach is reliable. Works such as (11, 16, 17), and (18) report low accuracy values, since the authors use classical matching techniques (18), which are not enough to deal with the complexity of the problem complexity. In other cases, the authors use datasets with a very small number of images (11, 16). Successful currency recognition is a complex problem if one only uses partial information, if there is occlusion, if blurry images are captured, if the banknotes are dirty or damaged, or if they coins are worn. Furthermore, in the case of (17), hyperparameter optimization showed improved accuracy. It is worth noting that the works that used Yolo, MobileNet, R-CNN, and AlexNet reported accuracy values ranging from 85.5 to 96.6 %.

The works shown in Table I mention useful properties to develop our proposal, the CopReader mobile application: 1) using a smartphone is preferable, since visually impaired people are familiar with using these devices in conjunction with mobile applications such as Google Talkback (24); 2) the input data should include photos and videos; 3) the deep learning algorithm should be reliable and have a small number of hyperparameters; 4) user feedback should include audio, an inclusive alternative for most visually impaired people.

3. CopReader development process and dataset building

Fig. 1 depicts the development of CopReader, which consists of two stages. The first stage is building the dataset and developing, training, and validating the neural network model. This process is iterative, meaning that the model is continuously validated, the results are analyzed, and the model is re-trained if it does not meet the expectations. The second stage involves using CopReader with the trained neural network model.

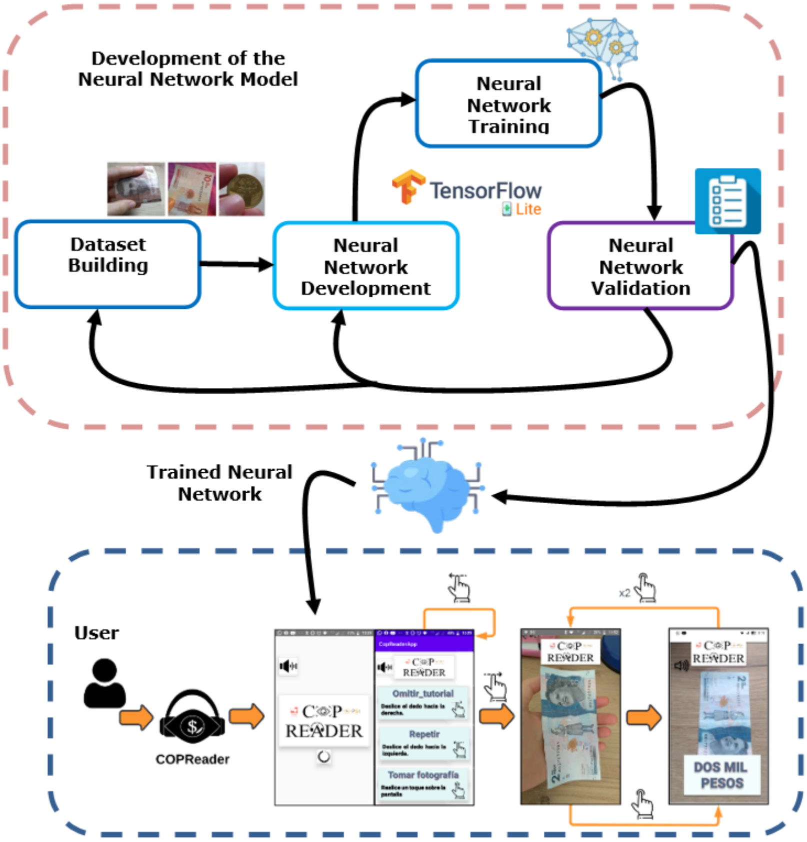

Figure 1. CopReader development process

Building the dataset is crucial for applications based on neural network models. In this specific case, to recognize Colombian currency, the dataset must be created from scratch. To this effect, it is important to clarify the classes of coins and banknotes that CopReader will recognize. Fig. 2 shows the coins and banknotes currently used by Colombia's Bank of the Republic. The following classes were defined: \$50 pesos coin, \$100 pesos coin, \$200 pesos coin, \$500 pesos coin, \$1000 pesos coin, \$2000 pesos banknote, \$5000 pesos banknote, \$10 000 pesos banknote, \$20 000 pesos banknote, \$50 000 pesos banknote, and \$100 000 pesos banknote. Fig. 2 shows the one face of the banknotes and coins, but the dataset included both faces.

Image acquisition was performed using a 13 Mpixel camera and considering different conditions, such as lighting, occlusion, the state of the currency, and the point of view. Figs. 3a to 3f show examples

(a)

(b)

Figure 2. a) Coins and b) banknotes used in Colombia

of these conditions, *i.e.*, crumpled banknotes, dirty coins, written banknotes, blended banknotes, incomplete banknotes, and occluded coins. There are examples for each class of coins and banknotes.

One common cause of recognition errors is incorrectly captured input images. To account for this issue, another class was added, *i.e.*, 'without currency'. This new class is composed of images from (Figs. 3g to 3i) that are blurry, moved, or out of focus. It is important to remember that CopReader uses a vision-based data input to recognize coins and banknotes of Colombian currency; since it has no other type of data input, it cannot detect counterfeit bills – this typically requires the use of UV lamps or texture information.

The initial dataset had 400 3120 x 460 pixel images in each of the 11 classes, for a total of 4400. These images were uploaded to the RoboFlow software tool (25) for proper labeling, as well as to apply data augmentation techniques regarding brightness (±10 %), darkening (±10 %), rotation (90°), out-of-focusup 2 pixels, and horizontal flip. The dataset included 8640 images, with 720 images per class. 80 % of the images were used for training, 10 % for validation, and 10 % for testing.

The related works presented in Table I show that the most commonly used neural network models that exhibit high accuracy values and have suitable configuration hyperparameters are Yolo (9,10,13,15) and MobileNet (14, 19). Thus, in this work, we selected the Yolo V5 nano and MobileNet V2 models for the development, training, and validation phase. The models' pipeline was implemented in Google Colaboratory (26). More details in this regard are provided in section 4.2. Testing was conducted with visually impaired and non-impaired individuals.

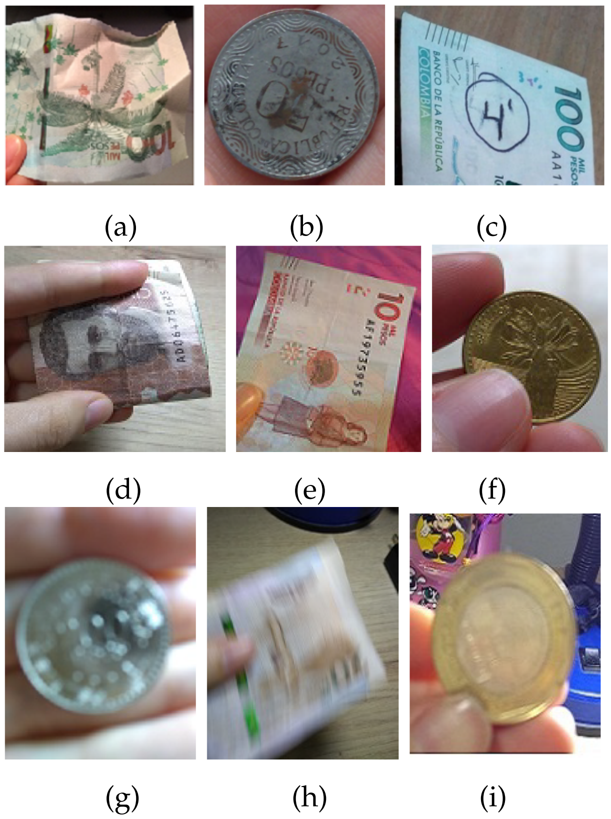

Figure 3. Images included in the dataset: a) crumpled, b) dirty, c) written, d) blended, e) incomplete, f) occluded, g) blurry, h) moved, and i) out-of-focus coins and banknotes

4. Design and implementation of CopReader

CopReader was developed while considering the analysis presented in section 2 and by following the RUP (Rational Unified Process) methodology (27) to document the software engineering process. The RUP methodology includes the following deliverables: functional and non-functional requirements, a conceptual diagram, real use cases, and sequence, relational, class diagrams. However, due to space limitations, real use cases and sequence diagrams are not included in this manuscript.

4.1. Design of CopReader

CopReader was developed while considering the following functional and non-functional requirements:

- *Functional requirements:*
  - Visually impaired users should be provided with an initial tutorial about how the application works.
  - The user should be able to omit the tutorial by sliding the screen to the right.
  - The user should be able to repeat the tutorial by sliding the screen to the left.

- The user should be able to activate the mobile phone's camera by sliding the screen to the right. The input image/video will then be analyzed by the neural network model to estimate the currency in front of the camera.
- The user should receive audio feedback regarding the coin or banknote in front of the camera.
- The user should be able to repeat the audio feedback by sliding the screen to the left.
- The user should be able to repeat the image/video acquisition process by touching the screen twice.

- *Non-functional requirements:*
  - CopReader was developed in Android Studio 4.0
  - CopReader use OpenCV 3.4 and the TensorFlow 2.0 library
  - The minimum camera resolution allowed should be 8 Mpixels.

After defining the functional requirements, the conceptual diagram of the app, shown in the inferior part of Fig. 1, was elaborated. CopReader is a mobile application that requires simple interactions, as it is aimed at visually impaired people. The app assumes that users have their headphones enabled. Then, it presents a tutorial on how to use it. This tutorial is read by a common mobile application called *Talkback*, typically installed in mobile phones as an accessibility tool. This is useful for new users, who are able to repeat the tutorial by sliding the screen to the left; otherwise, they can omit it by sliding the screen to the right. Afterwards, when users slide the tutorial screen to the right, the data acquisition process starts: the neural network model processes the input data and provides an estimation of the coin or banknote in front of the camera. This estimation is communicated to the user via audio feedback, which can be repeated by sliding the screen to the left. Users can repeat the currency estimation process by touching the screen twice.

Fig. 4 shows the class diagram of CopReader. The *MainActivity* and *Initialization* classes start the mobile application. The *Camera* class provides the data input to the *Classifier* class, which in turn processes the input data, yielding an estimation of the currency in front of the camera. Finally, the results are stored in the *Result* class, which generates the audio feedback for the user.

4.2. Implementation of CopReader

CopReader's GUI is very simple since it is aimed at visually impaired people. The bottom part of Fig. 1 shows the final implementation. This section explains how the neural network models were trained, validated, and selected to be used within the app.

The YoloV5 nano neural model was trained over 150 epochs, using a batch size of 32 and random weight initialization. Fig. 5a presents the precision-recall training results, showing values close to 1 and a mAP0.95 (*i.e.*, a mean average precision at an intersection over a union threshold of 0.95) of 0.89, indicating a high generalization capability. In addition, the box (*box_loss*) and object (*obj_loss*) loss values shown in Fig. 5a, which are related to the training and validation phases, do not show

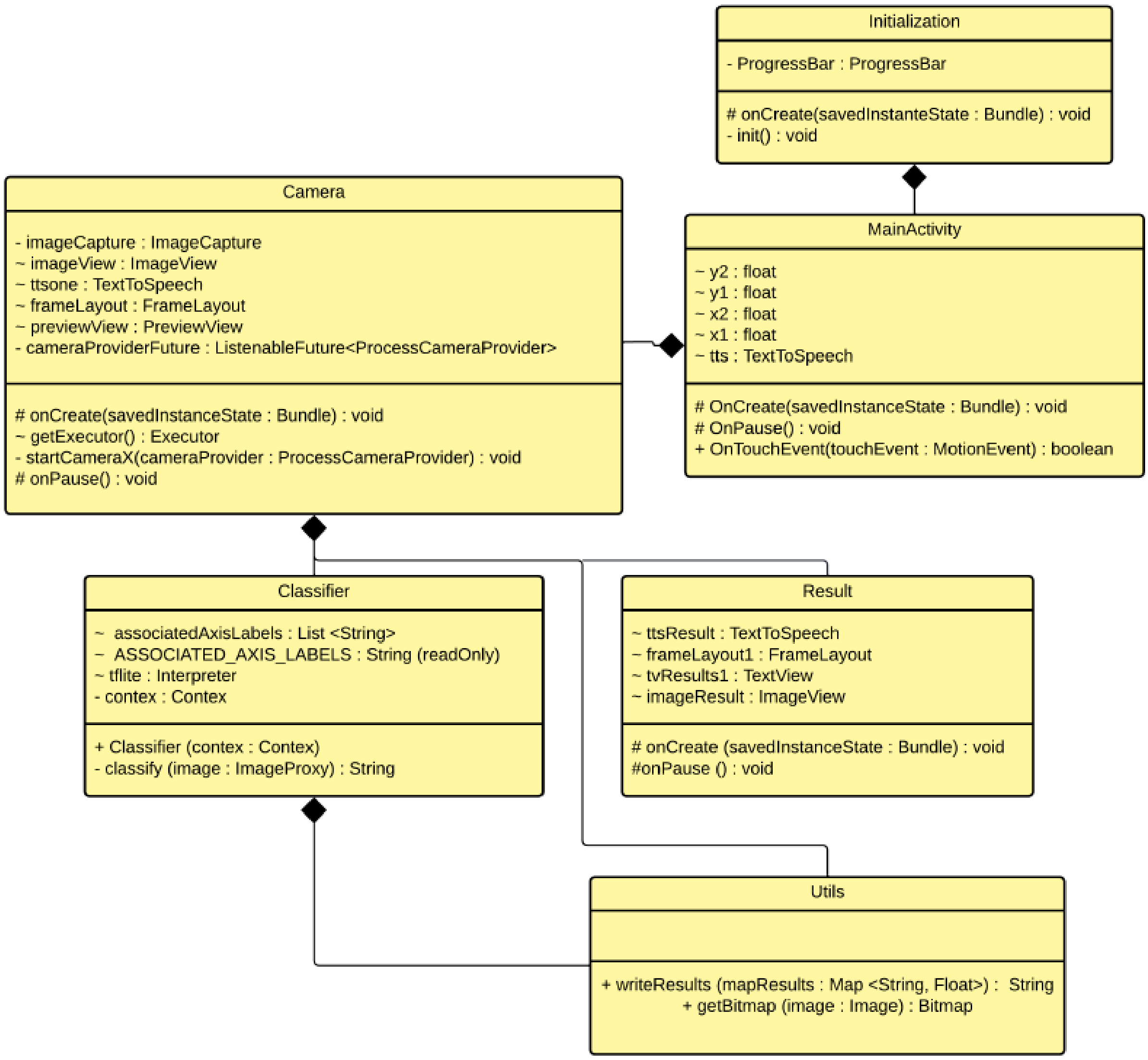

Figure 4. CopReader class diagram

evidence of overtraining. Fig. 6a shows the validation results for the confusion matrix of this neural model without data augmentation. There are no false negatives with a background, but there are false positives with a background of up to 20 %. Moreover, the \$100 000 pesos banknote exhibits a high value of false negatives (9 %).

To reduce the number of false negatives and false positives, a neural model with data augmentation was trained and validated. Fig. 5b shows the precision-recall results for this case, which are close to 1 and report a mAP0.95 of 0.85. As shown in Fig. 6b, the confusion matrix validation results show a 2 % reduction in the number of false negatives with respect to the model without data augmentation. Table II summarizes the training results for YoloV5. It can be observed that the accuracy metric is similar for both models, but the other metrics differ by up to 7 %. However, the mAP0.95 metric differs by 4 %,

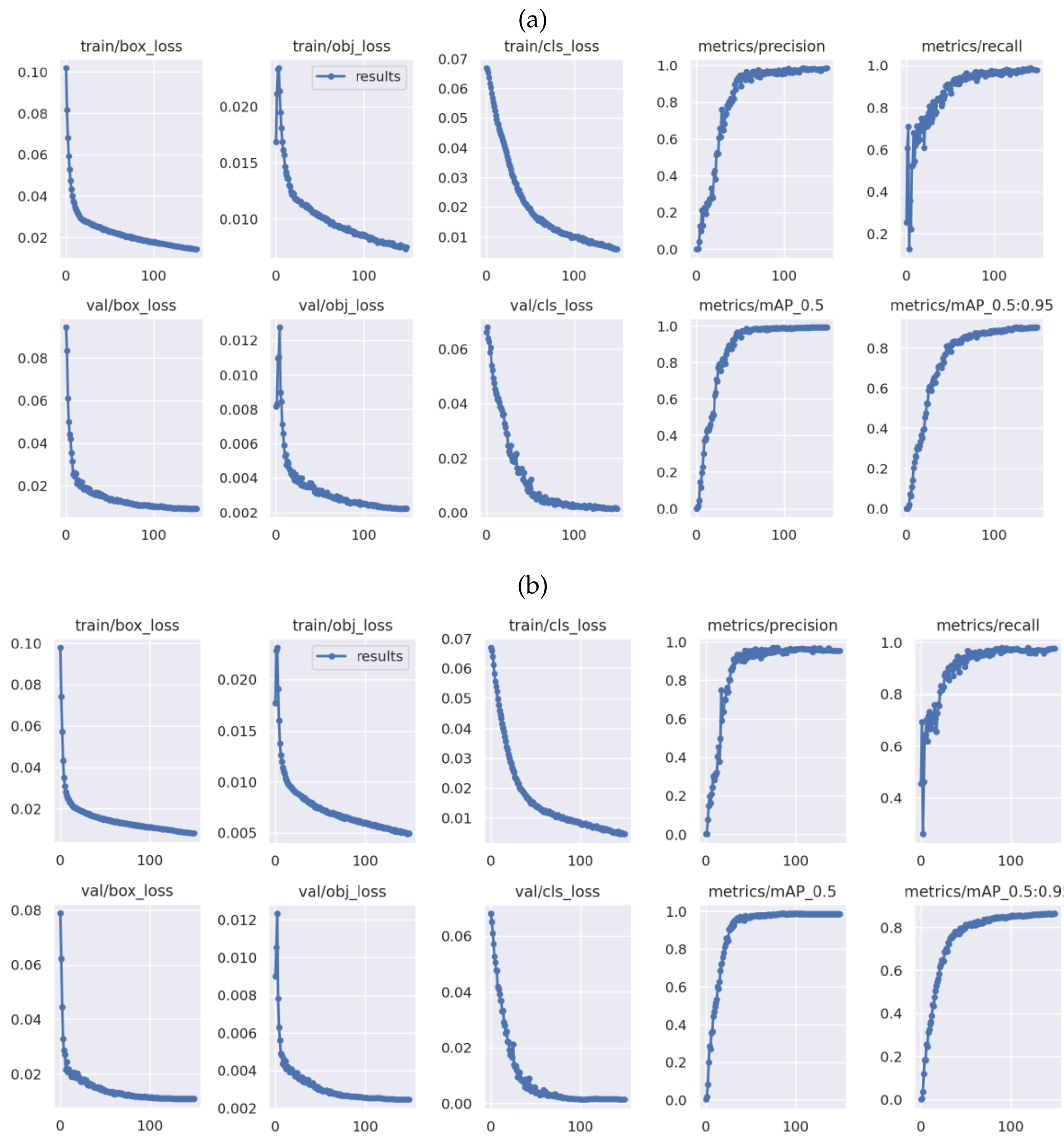

Figure 5. Training results for YoloV5 a) without and b) with data augmentation

with the second neural model (with data augmentation) exhibiting a better robustness against false negatives and false positives. Another reason for this mAP0.95 decrease could be that the Roboflow software tool does not rotate the delimiter box when rotating the sample image.

The second neural model selected was MobileNetV2. The first training attempt is presented in Fig. 7a. This attempt used 40 epochs, a batch size of 32, and random weight initialization. The obtained accuracy for validation was at least 0.95, with loss values lower than 0.2.

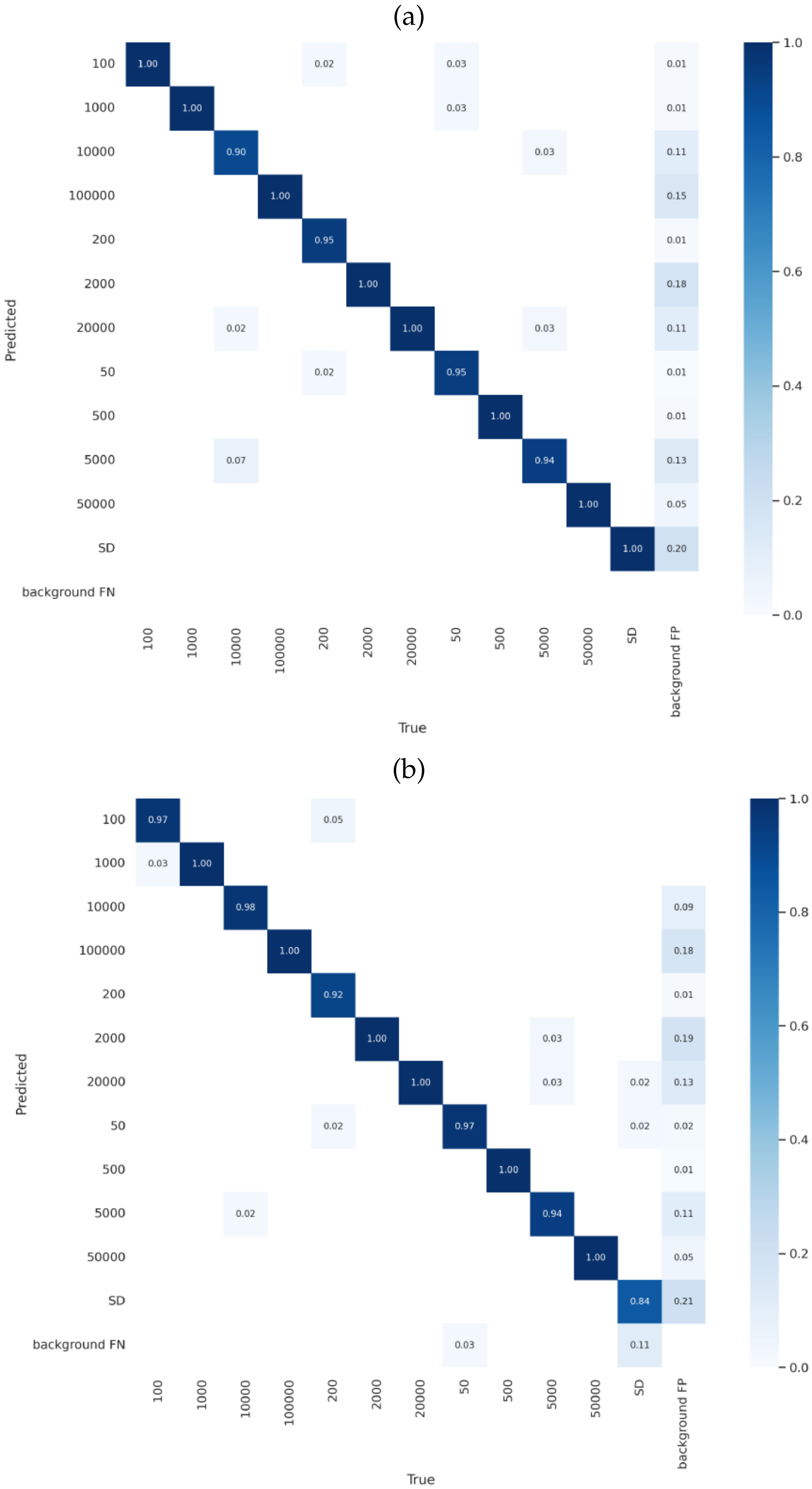

Figure 6. a) Confusion matrix for YoloV5 a) without and b) with data augmentation

Table II. Summary of training results

Metrics	YoloV5	MobileNetV2		
	Without data aug.	**With DATA AUG.**	**Without data aug.**	**With data aug.**
Accuracy	0.92	0.90	0.98	0.99
Precision	0.95	0.88	0.98	0.99
Recall	0.93	0.92	0.98	0.99
mAP	0.89	0.85	0.98	0.99

Fig. 8a shows the confusion matrix validation results. Eight out of 12 classes do not report false negatives, and two exhibit up to 5 % of false negatives. However, there are classification errors, such as those observed with the $200 (9 %) and $5000 (8 %) pesos classes. To reduce the number of false negatives and false positives, data augmentation techniques were applied, as described in section 3. Figs. 7b and 8b show the training results for this model with data augmentation, considering 30 epochs, a batch size of 32, random weight initialization, and L2 regularization, indicating accuracy values over 0.95 and loss values lower than 0.1. Fig. 8b shows the confusion matrix validation results. Note that the false negatives were reduced by 11 % with respect to MobileNetV2 without data augmentation, and the misclassification of the $200 pesos was reduced to 4 %. Table II summarizes the training results obtained for MobileNetV2. In general, the studied metrics are close to 1 (0.99), and the last training of MobileNetV2 reduced the number of false negatives by up to 6.2 %.

In light of that shown in the table II, to select the most suitable neural model, it is important to consider that there is a meaningful difference between the results obtained. In this regard, an analysis of variance (ANOVA) was performed. Two hypotheses were proposed: 1) the group means are equal (null hypothesis), and 2) they are different (alternative hypothesis). To apply this type of analysis, the following conditions must be satisfied: a normal data distribution, the homogeneity of variances, and an independent group of measures. Fig. 9 shows the residual histogram. Using these residuals, the Levene test was used to check for homogeneity, with the significance level (*alpha*) set at 0.05. The p-value computed with these residuals was Pr = 0.05015. Therefore, since Pr $>$ *alpha*, the data were homogeneous. As the experiments performed did not depend on each other, they were independent.

Afterwards, the ANOVA test was performed. Given a significance level of 0.05, the f- and p- values were computed, resulting in 41.29 and 0.03, respectively. As $f > p$ and $p < 0.05$, the null hypothesis was rejected (28). This means that there were significant differences in the group of data. Therefore, since the performance metrics of MobileNetV2 were higher, this model was selected for implementation on the smartphone. Moreover, MobileNetV2 has a lower size (3.4 Mb) which is adequate for execution in embedded systems. Before migrating this model to the device, the following steps were required: first, the neural model had to be converted to TensorFlow Lite; second, the model had to be optimized for execution on a smartphone; and third, we had to test whether this optimization could impact the performance metric. This was not the case, since the results obtained with the optimized model also achieved a 99 % accuracy.

(a)

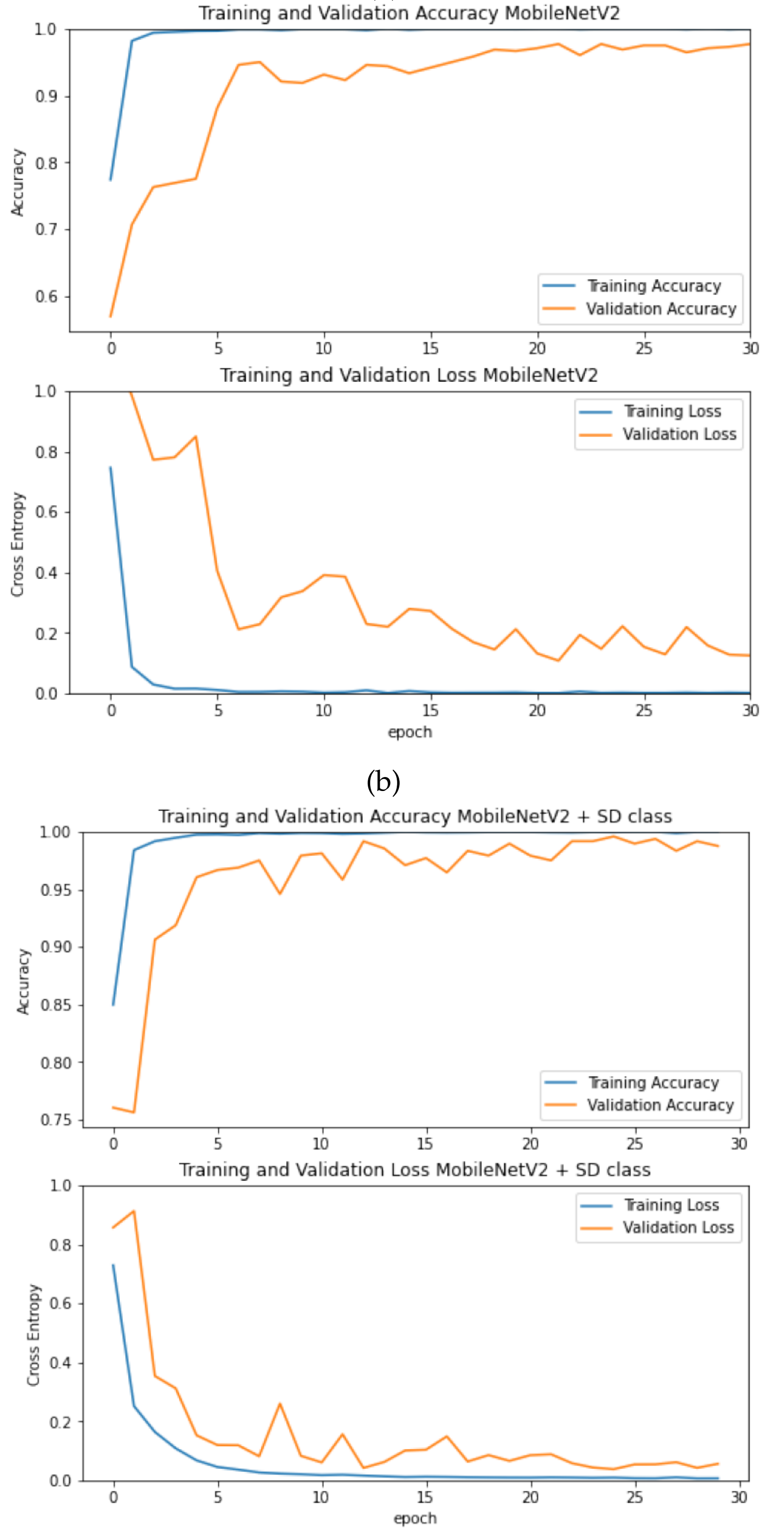

Figure 7. Training results for MobileNetV2 a) without and b) with data augmentation

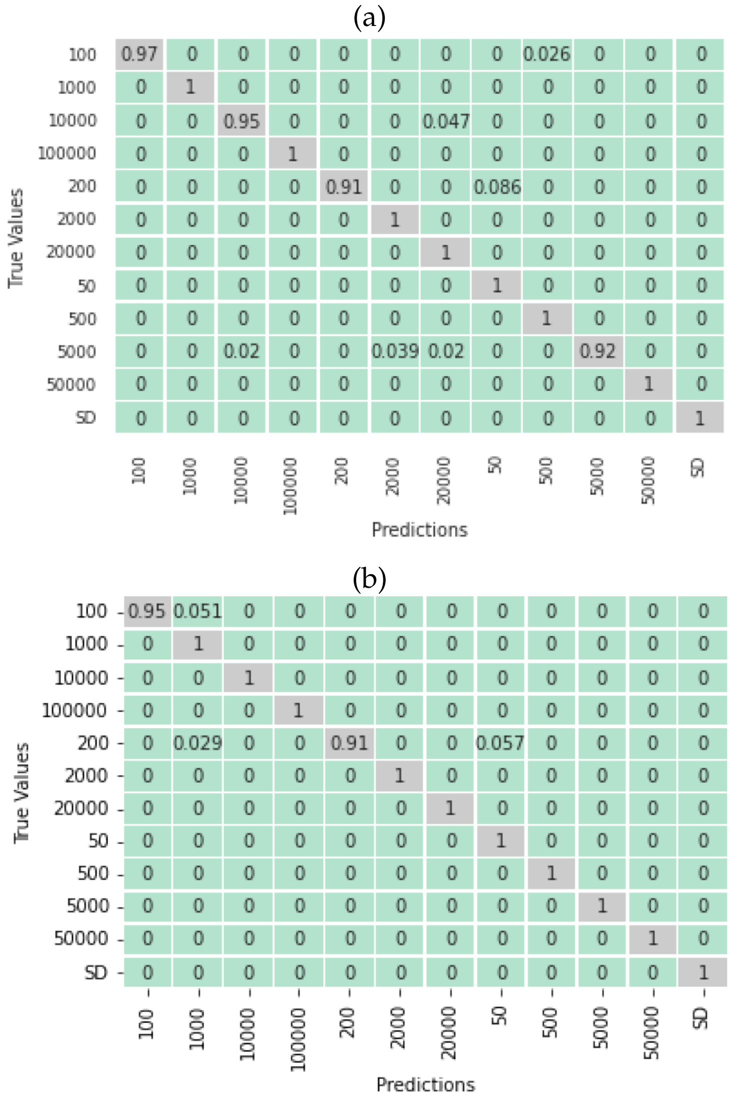

Figure 8. a) Confusion matrix for MobileNetV2 a) without and b) with data augmentation

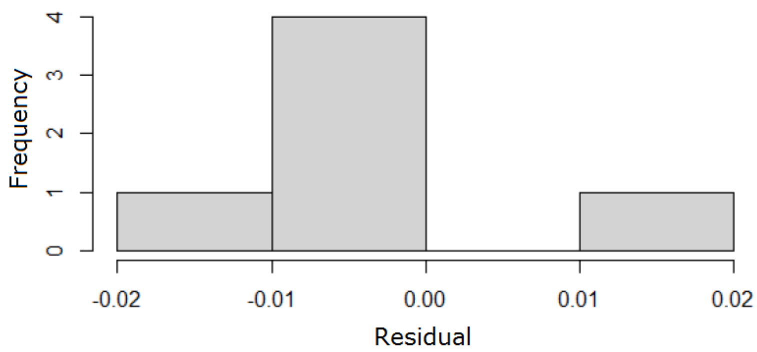

Figure 9. Residual histogram

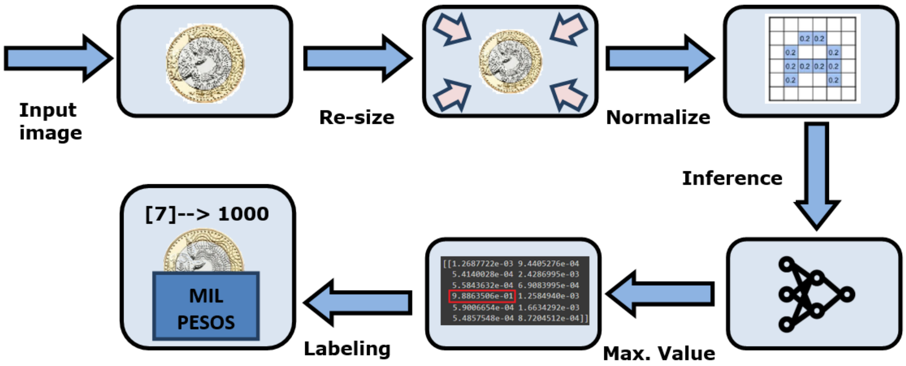

Figure 10. Smartphone process pipeline

Fig. 10 shows CopReader's smartphone process pipeline. Here, the input image is re-sized and then normalized prior to the inference process with MobileNetV2. Afterwards, the maximum value of the recognition inference is selected, associated with a label, and prepared for audio feedback.

5. Results and discussion

CopReader was validated using three quantitative tests: 1) a test with visually non-impaired users to verify its functionalities and measure performance metrics such as precision-recall; 2) a test with visually impaired users to evaluate its usefulness; 3) a test with an improved version of the app, considering the suggestions made by the visually impaired subjects, where new performance metrics were obtained.

The experimental conditions for these tests are described below:

a. The mobile devices used were a Samsung Galaxy A10s with a MediaTek Helio P22, 2 GHz processor and a Moto G5S Plus with a Snapdragon 625 octa-core, 2 GHz processor.

b. The selected number of visually non-impaired users was 10.

c. The selected number of visually impaired was 10, where nine persons were blind and one had low vision.

d. The visually non-impaired users were surveyed at Universidad del Valle, Cali, Colombia.

e. The visually impaired users were surveyed in the Hellen Keller room of the Jorge Garcés Borrero public library, with ethics committee endorsement no. 030-22 of Universidad del Valle.

In the first and second tests, 40 different coins and banknotes from each class were used by 20 different users (ten with no visual disability and ten with a visual impairment). Afterwards, all 20 users were asked to answer a survey with six simple questions. Fig. 11 shows the confusion matrix of this test, where a total precision of 0.92 and a recall of 0.9 were obtained. This figure also shows a slight increase in false negatives, as well as decreasing precision and recall values. However, these results are greater than 0.9, a good result in comparison with the related works reviewed in Table I.

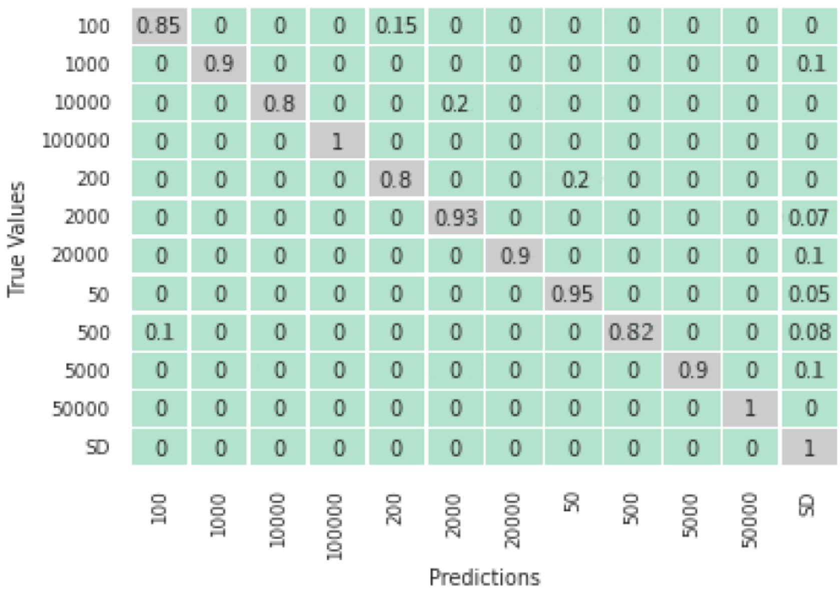

Figure 11. Confusion matrix for test 1 involving visually non-impaired users to verify the app's functionalities

All visually non-impaired users (ten in total) were asked to answer the following six questions:

1. Was the tutorial clear enough to understand how to use the mobile application?
   Answers: totally agree (5), agree (4), disagree (3), and totally disagree (2).
2. What level of satisfaction did you feel in relation to the mobile application's gestures?
   Answers: totally satisfied (5), satisfied (4), dissatisfied (3), and totally dissatisfied (2).
3. What level of satisfaction did you feel in relation to the picture acquisition process?
   Answers: totally satisfied (5), satisfied (4), dissatisfied (3), and totally dissatisfied (2).
4. What level of satisfaction did you feel in relation to the audio feedback?
   Answers: totally satisfied (5), satisfied (4), dissatisfied (3), and totally dissatisfied (2).
5. What level of satisfaction did you feel in relation to CopReader in general?
   Answers: totally satisfied (5), satisfied (4), dissatisfied (3), and totally dissatisfied (2).
6. How likely are you to recommend CopReader?
   Answers: totally likely (5), likely (4), unlikely (3), and totally unlikely (2).

These questions were inspired by the SUS usability scale (29). This scale, however, was not directly used, since it was not designed for visually impaired people. Figs. 12a to 12f show the results of the survey. It can be observed that users agree, totally agree, are satisfied, or are totally satisfied about the different properties of CopReader, such as the tutorial explanation at the start of the mobile app, the simple navigation gestures, the image acquisition process, and the operation of CopReader in general. The users are also likely to recommend this app. However, one person was not satisfied with it, stating that many gestures (more than 2) are required to identify the currency value in front of the smartphone camera.

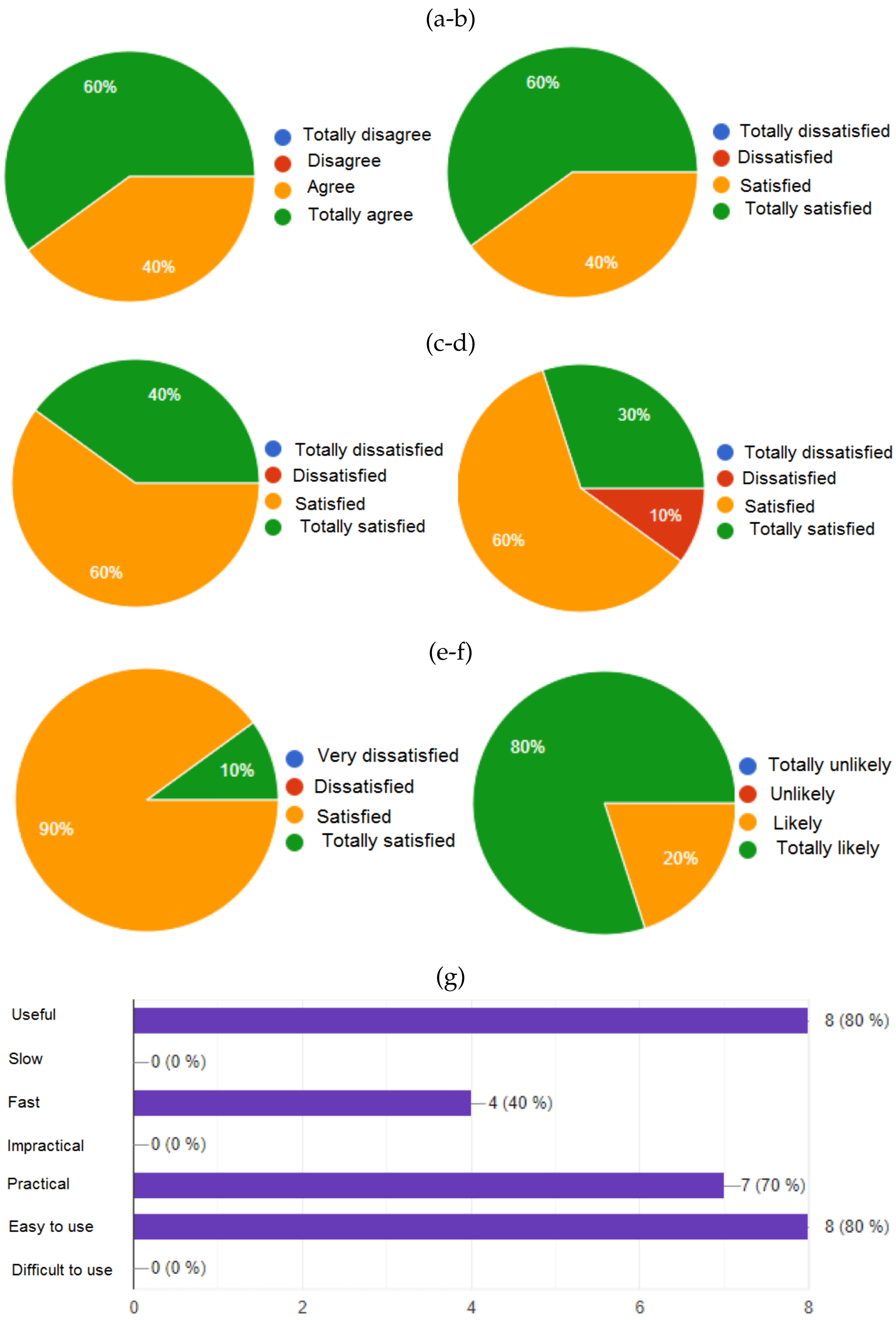

Figure 12. Survey results: a) tutorial explanation, b) use of gestures, c) image capture, d) audio feedback, e) mobile app in general, f) recommendation likelihood, g) mobile app ranking by keywords

In the last question of the survey, the users were asked to select a keyword related to CopReader. These results are shown in Fig. 12g. The users generally thought that CopReader is useful, fast, practical, and easy to use. Negative words, such as *slow, impractical*, or *difficult to use*, were not selected.

In the second test, the visually impaired subjects were divided into nine blind users and one with low vision (Fig. 13a). It is worth noting that these users used the Talkback app in conjunction with CopReader. They were surveyed with the same questions, but an additional one was introduced:

7. How likely are you to use the mobile application again?
   Answers: totally likely (5), likely (4), unlikely (3), and totally unlikely (2).

Figs. 13b to 13h show the results of the survey. The users highlighted that CopReader has a tutorial description, which is important. They felt satisfied and totally satisfied with the audio feedback and the way in which the app works. They are likely to recommend it to their acquaintances, and they will use it again. However, Figs. 12c and 12b show that 20 % and 30 % of users were dissatisfied, stating that it was difficult to locate the currency in front of the camera.

The visually impaired users were also asked for keywords regarding the CopReader mobile application. Fig. 13i shows the results, where users described CopReader as useful, practical, easy to use, and fast. Once again, negative terms such as *slow, impractical*, or *difficult to use* were not selected.

Table III. Statistical survey results

Question	Visually non-impaired users				Visually impaired users			
	Mode	**F$_{MO}$(%)**	$\bar{x}$	**Cv**	**Mode**	**F$_{MO}$(%)**	$\bar{x}$	**Cv**
1	5	60.0 %	4.6	8.7 %	5	60.0 %	4.6	8.7 %
2	5	60.0 %	4.6	8.7 %	4	50.0 %	4.1	22.0 %
3	5	60.0 %	4.6	8.7 %	4	50.0 %	3.9	28.2 %
4	4	60.0 %	4.2	19.0 %	5	90.0 %	4.8	4.2 %
5	4	90.0 %	4.1	22.0 %	4	90.0 %	4.1	22.0 %
6	5	80.0 %	4.8	4.2 %	5	60.0 %	4.6	8.7 %
7	-	-	-	-	4	80.0 %	4.2	19.0 %

All the surveys were validated using the deviation coefficient shown in Eq. (1) and the corresponding values shown in Table III.

$$Cv = \frac{|5 - \bar{x}|}{\bar{x}} \cdot 100 < 25\,\% \tag{1}$$

In this equation, $\bar{x}$ represents the arithmetic mean, where $\bar{x} > 4$ is desirable. To decide whether the results in Table III are significant, the deviations must be lower than 25 %. The results mode is also expected to correspond to the maximum value (5). Based on Table III, the deviation coefficient (Cv) is lower than 25 % in most results, the mean is greater than 4, and the mode is 5 in 54 % of the results.

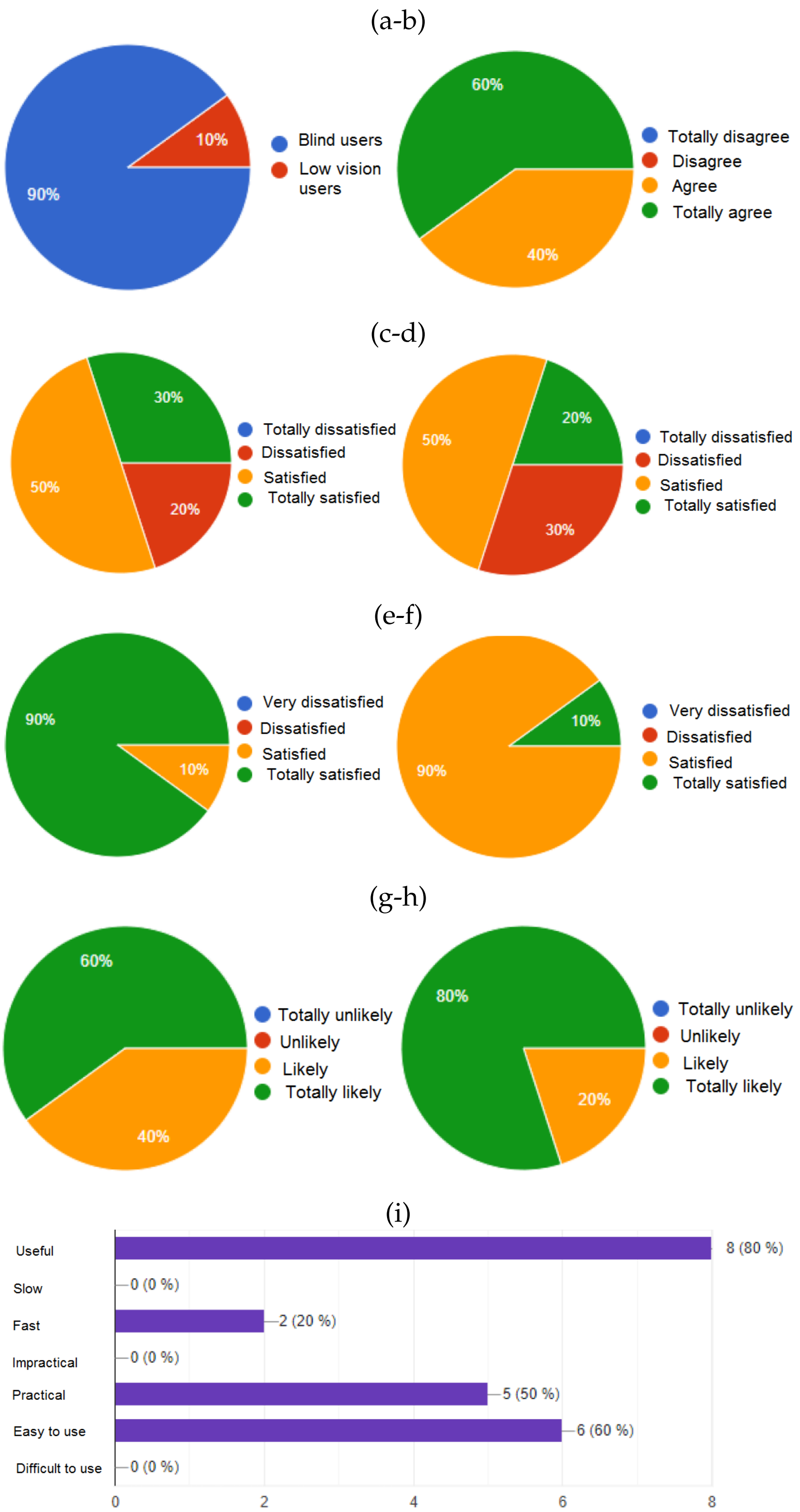

Figure 13. Results of the survey with the visually impaired subjects: a) population surveyed, b) tutorial explanation, c) use of gestures, d) image capture, e) audio feedback, f) mobile app, g) recommendation likelihood, h) likelihood of reuse, i) ranking by keywords

It is worth noting that, for the visually non-impaired users, the results for question 5 are close to the 25 % threshold, as well as questions 2 and 5 for the visually impaired users. For the latter, the deviation coefficient of question 3 is greater than 25 %. These questions are related to the CopReader gestures (question 2), the image acquisition process (question 3), and the overall functioning of the app (question 5). The users expressed that CopReader requires too many gestures (more than two) to know the value of the currency in front of the smartphone camera. Moreover, they suggested that it would be nice for the app to process videos instead of photos, as it could be difficult to locate the coin or banknote in front of the smartphone. This statistical analysis led to the third test, in which a new version of CopReader equipped with additional features was used.

Figure 14. Confusion matrix for the improved version of CopReader

For this test, considering the drawbacks observed in the previous surveys, the improved version of CopReader included video processing, a common suggestion made by the visually impaired users. This version also addressed the issue with the number of gestures needed to achieve currency recognition and obtain audio feedback while solving the difficulty of positioning the currency in front of the smartphone. This new version of CopReader employes the same pipeline shown in Fig. 1. However, instead of processing images, it processes video in YUV_420_888 format at 24 frames per second. The app was tested using 40 different coins and banknotes for each class. The resulting confusion matrix is shown in Fig. 14. In this test, the precision obtained was 0.96, and the recall value was 0.97.

6. Conclusions

This work described the development and testing of CopReader, a mobile application for visually impaired people that provides a tutorial regarding its operation, working in conjunction with Talkback to interact with the users via audio feedback, capturing images and videos and processing them to estimate the currency value of coins and banknotes placed in front of the smartphone camera.

To process the images and videos using neural network models such as YoloV5 and MobileNetV2, a dataset was built which included 8640 images, with 720 images per class. The current coins recognized by CopReader have values of $50, $100, $200, $500, and $1000 pesos, and the banknotes ae worth $2000, $5000, $10 000, $20 000, $50 000, and $100 000 pesos. YoloV5 and MobileNetV2 were trained with this dataset, obtaining precision values greater than 89 % and recall values higher than 92 %.

CopReader was validated using three different tests. First, precision and recall metrics were obtained through interaction with visually non-impaired users. This was also done with visually impaired users, and an improved version of CopReader was tested. In the first and second tests, precision and recall values of 0.92 and 0.90 were obtained, respectively. A survey was conducted with all users, according to which they agree or totally agree, or are satisfied or totally satisfied with the tutorial, the navigation gestures, and the data acquisition process. Moreover, the users are likely to recommend CopReader to others, and they found the app useful, fast, practical, and easy to use.

After these tests, the users suggested that it would be nice for the app to process video instead of photos, as it could be difficult to place coins or banknotes in front of the smartphone. These suggestions helped to improve CopReader. The number of gestures required was reduced to increase the app's acceptance. Then, a third test was performed with an improved version of CopReader, equipped with video processing. This new version processes video in YUV_420_888 format at 24 frames per second and obtains precision and recall values of 0.96 and 0.97, respectively.

The precision results obtained by related works (16–18), range from 65.7 to 82 %. In comparison, the precision and recall values of CopReader are greater than 95 %, making our proposal an interesting option for visually impaired people to increase their autonomy in daily life activities. However, since our app only uses vision-based information, it has difficulties in detecting counterfeit bills.

7. Acknowledgments

This work was conducted within the framework of the research project called *GlobalViewAid: An innovative navigation system for persons with visual impairments* of the Gelbert Foundation (Genève, Switzerland), in collaboration with the EPFL and the University of Genève, identified with the ID No. 21162 of Universidad del Valle, Cali, Colombia.

8. CRediT author statement

Camila Bolaños-Fernández: data curation, software development, and validation.
Bladimir Bacca-Cortes: conceptualization, methodology, resources, supervision, and writing.

References

[1] M. Ramamurthy and V. Lakshminarayanan, "Human vision and perception," in *Handbook of Advanced Lighting Technology*, Cham: Springer International Publishing, 2015, pp. 1-23. ↑2, 3

[2] J. C. Suárez, "Discapacidad visual y ceguera en el adulto: revisión de tema," *Med. U.P.B.*, vol. 30, no. 2, pp. 170-180, 2011. ↑3

[3] J. Alberich, D. Gómez, and A. Ferrer, *Percepción Visual*, 1st ed., Barcelona, Spain: Universidad Oberta de Catalunya, 2014. ↑3

[4] D. Parra, "Los ciegos en el censo del 2018," Instituto Nacional para Ciegos INCI, 2020. [Online]. Available: https://www.inci.gov.co/blog/los-ciegos-en-el-censo-2018 ↑3

[5] M. del R. Yepes, "La intermediación en la inclusión laboral de la población con discapacidad visual," Instituto Nacional para Ciegos INCI2. [Online]. Available: https://www.inci.gov.co/blog/la-intermediacion-en-la-inclusion-laboral-de-la-poblacion-con-discapaci\dad-visual ↑3

[6] Banco de la República, "Marca táctil en los billetes," 2016. [Online]. Available: https://www.banrep.gov.co/es/node/31529 ↑3

[7] Orcam, "OrCam MyEye," 2023. [Online]. Available: https://www.orcam.com/en-us/myeye-store ↑3

[8] M. Doudera, "Cash reader: Bill identifier," mobile app, Google Play, 2024. https://play.google.com/store/apps/details?id=com.martindoudera.cashreader&hl=en&gl=US ↑3

[9] C. Aramendiz U., D. Escorcia G., J. Romero C., K. Torres R., and C. Triana P., "Sistema basado en reconocimiento de objetos para el apoyo a personas con discapacidad visual (¿Que tengo enfrente?)," *Investig. y Desarro. en TIC*, vol. 11, no. 2, pp. 75-82, 2020. https://revistas.unisimon.edu.co/index.php/identic/issue/view/261 ↑4, 5, 7

[10] S. Vaidya, N. Shah, N. Shah, and R. Shankarmani, "Real-time object detection for visually challenged people," in *2020 4th Int. Conf. Intell. Com. Control Syst. (ICICCS)*, 2020, pp. 311-316. https://doi.org/10.1109/ICICCS48265.2020.9121085 ↑4, 5, 7

[11] M. Awad, J. El Haddad, E. Khneisser, T. Mahmoud, E. Yaacoub, and M. Malli, "Intelligent eye: A mobile application for assisting blind people," in *2018 IEEE Middle East North Africa Comm. Conf. (MENACOMM)*, 2018, pp. 1-6. https://doi.org/10.1109/MENACOMM.2018.8371005 ↑4, 5

[12] J. Tamayo, "Sistema de reconocimiento de billetes para personas con discapacidad visual mediante visión artificial," undergraduate thesis, EIA University, 2018. https://repository.eia.edu.co/entities/publication/321a4983-afec-4865-9915-4ede5b26a435 ↑4, 5

[13] J.-Y. Lin, C.-L. Chiang, M.-J. Wu, C.-C. Yao, and M.-C. Chen, "Smart glasses application system for visually impaired people based on deep learning," in *2020 Indo – Taiwan 2nd Int. Conf. Comp. Analytics Net. (Indo-Taiwan ICAN)*, 2020, pp. 202-206. https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181366 ↑4, 5, 7

[14] O. Stephen, Y. J. Jang, T. S. Yun, and M. Sain, "Depth-wise based convolutional neural network for street imagery digit number classification," in *2019 IEEE Int. Conf. Comp. Sci. Eng. (CSE) and IEEE Int. Conf. Embedded Ubiq. Comp. (EUC)*, 2019, pp. 133-137. https://doi.org/10.1109/CSE/EUC.2019.00034 ↑4, 5, 7

[15] Joshua, J. Hendryli, and D. E. Herwindiati, “Automatic License Plate Recognition for Parking System using Convolutional Neural Networks,” in *2020 Int. Conf. Inf. Manag. Technol. (ICIMTech)*, 2020, pp. 71-74. https://doi.org/10.1109/ICIMTech50083.2020.9211173 ↑4, 5, 7

[16] S. M. M. Roomi and R. B. J. Rajee, “Coin detection and recognition using neural networks,” in *2015 Int. Conf. Circuits, Power Comput. Technol. [ICCPCT-2015]*, 2015, pp. 1-6. https://doi.org/10.1109/ICCPCT.2015.7159434 ↑4, 5, 22

[17] N. Capece, U. Erra, and A. V. Ciliberto, “Implementation of a Coin Recognition System for Mobile Devices with Deep Learning,” in *2016 12th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, 2016, pp. 186-192. https://doi.org/10.1109/SITIS.2016.37 ↑4, 5, 22

[18] J. Xu, G. Yang, Y. Liu, and J. Zhong, “Coin Recognition Method Based on SIFT Algorithm,” in *2017 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, 2017, pp. 229-233. https://doi.org/10.1109/ICISCE.2017.57 ↑4, 5, 22

[19] S. Mittal and S. Mittal, “Indian Banknote Recognition using Convolutional Neural Network,” in *2018 3rd Int. Conf. Internet Things: Smart Innov. Usages (IoT-SIU)*, 2018, pp. 1-6. https://doi.org/10.1109/IoT-SIU.2018.8519888 ↑4, 5, 7

[20] A. U. Tajane, J. M. Patil, A. S. Shahane, P. A. Dhulekar, S. T. Gandhe, and G. M. Phade, “Deep Learning Based Indian Currency Coin Recognition,” in *2018 Int. Conf. Adv. Commun. Comput. Technol. (ICACCT)*, 2018, pp. 130-134. https://doi.org/10.1109/ICACCT.2018.8529467 ↑4, 5

[21] N. A. J. Sufri, N. A. Rahmad, N. F. Ghazali, N. Shahar, and M. A. As’ari, “Vision Based System for Banknote Recognition Using Different Machine Learning and Deep Learning Approach,” in *2019 IEEE 10th Control Syst. Grad. Res. Colloq. (ICSGRC)*, 2019, pp. 5-8. https://doi.org/10.1109/ICSGRC.2019.8837068 ↑4, 5

[22] U. R. Chowdhury, S. Jana, and R. Parekh, “Automated System for Indian Banknote Recognition using Image Processing and Deep Learning,” in *2020 Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, 2020, pp. 1-5. ↑4, 5

[23] R. Tasnim, S. T. Pritha, A. Das, and A. Dey, “Bangladeshi Banknote Recognition in Real-time using Convolutional Neural Network for Visually Impaired People,” in *2021 2nd Int. Conf. Robotics, Electr. Signal Process. Tech. (ICREST)*, 2021, pp. 388-393. https://doi.org/10.1109/ICREST51555.2021.9331182 ↑4, 5

[24] Google Play, “Android Accessibility Suite,” Talkback, 2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=en_US&gl=US ↑5

[25] B. Dwyer, J. Nelson, and J. Solawetz, “Software Roboflow (Version 1.0).” Roboflow Inc., 2022. [Online]. Available: https://roboflow.com/ ↑7

[26] E. Bisong, “Google colaboratory,” in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Berkeley, CA: Apress, 2019, pp. 59–64. ↑7

[27] Taryana Suryana, “Rational Unified Process (RUP),” *Ration. Unified Process*, vol. 3, no. September, pp. 1-6, 2007. ↑8

[28] D. Trenkler, "A handbook of small data sets: Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. & Ostrowski, E. (1994): Chapman & Hall, London. xvi + 458 pages, including one diskette with data files (MS-DOS), 40 British Pounds, ISBN 0-412-39920-2," *Comput. Stat. Data Anal.*, vol. 19, no. 1, p. 101, Jan. 1995. ↑13

[29] Wikipedia, "System usability scale," 2024. [Online]. Available: https://en.wikipedia.org/wiki/System_usability_scale ↑17

Camila Bolaños-Fernández

She received a degree in Electronic Engineering in 2022 from Universidad del Valle, Cali, Colombia. Currently, she serves as a software developer at SuperGIROS. Her main interests are web and mobile app development, and embedded systems applications.

Email: mariafdz51@gmail.com

Bladimir Bacca Cortes

He received a degree in Electronic Engineering in 1999 from Universidad del Valle (Cali, Valle del Cauca, Colombia); a Master's degree in Automation from Universidad del Valle in 2004; and a PhD in Technology from Universitat de Girona (Cataluña, Spain) in 2012. He is currently a full professor at Universidad del Valle, in the School of Electrical and Electronics Engineering. He is a member of the research group on Perception and Intelligent Systems. His research interests are mobile robotics, simultaneous localization and mapping systems, and computer vision.

Email: bladimir.bacca@correounivalle.edu.co