Research Article/Review Article

Journal of Computing, Science &Technology.Vol.1, Issue 2,xx, 2024

Journal of Computing, Science &Technology

https://focjournal.unidel.edu.ng/

editorjcst@unidel.edu.ng

ISSN:......Print

ISSN: .... Online

Convolutional Neural Network Framework for Banknote Recognition

Damilola Olusoga[1], Olumide Sunday Adewale[2] & Kehinde Agbele[3]

[1,3,]Department of Mathematics & Computer Science, Elizade University, Ilara-Mokin, Nigeria,

[2]Department of Computer Science, Federal University of Tech. Akure, Nigeria

damilolaolusog@gmail.com[1], adewale@futa.edu.ng[2], kehindeagbele@eu.edu.ng[3]

Corresponding Author's Email: damilolaolusog@gmail.com

Article Info
Date Received: 28-03-2024
Date Accepted: 05-05-2024

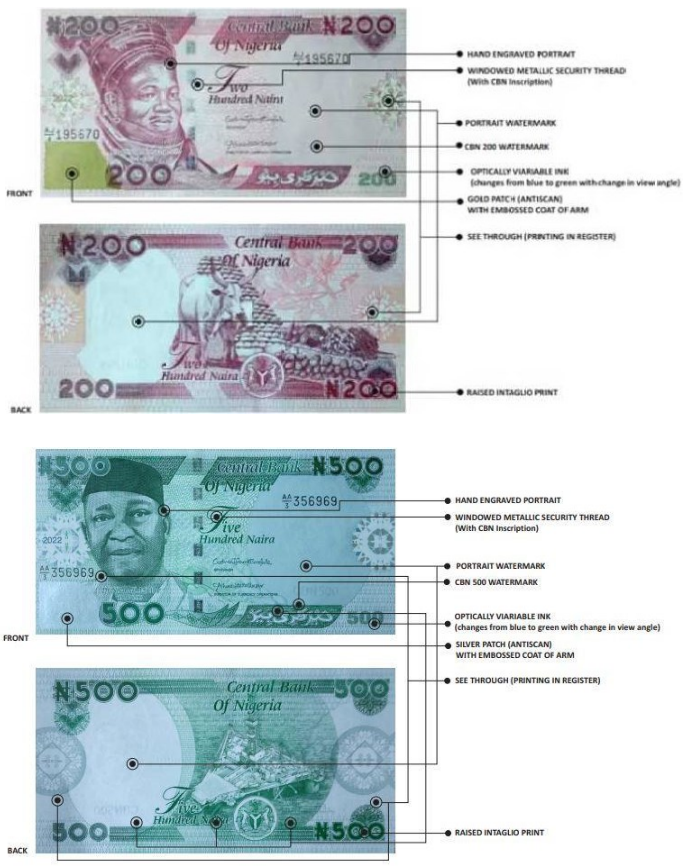
Keywords:
banknote recognition,
convolution neural network,
banknote security features

ABSTRACT

A human visual system can identify and validate currency notes. But because human vision is limited, it can occasionally be challenging for individuals to tell the difference between genuine money and counterfeit without the use of technology. For many applications, deep learning approaches have been shown to be more successful. This research presents a deep learning convolutional neural network system aimed at enhancing the identification accuracy of large denomination Nigerian banknotes.

1.0 INTRODUCTION

The Naira banknotes are protected by numerous security procedures, which make it easy to recognize genuine notes. The distinguishing features that are easily recognized by touch and visibility are the security thread, raised print, and watermark. The denominational numerals on the reverse and obverse, the lettering, and the portrait are among the additional areas of the notes that are embossed. Raised printing on the notes adds a tactile aspect, and tiny letters "CBN" are printed on both sides of the security thread, which, when held up to light, looks broken but Is not. The Naira notes are also protected from photocopying. Being in the presence of ultra-violent light also makes some features visible. For example, the serial number on every banknote is black in the dark and turns green when it comes into contact with UV light. The denominations of banknotes that Nigeria issues include ₦5, ₦10, ₦20, ₦50, ₦100, ₦200, ₦500, and ₦1000. The larger denominations measure 151 X 78 mm and are printed on paper substrate, while the smaller denominations are printed on polymer substrate and are 130 X 72 mm. Certain components in the paper and polymer substrates are unique to banknotes. The materials and printing process provide the currency banknotes the distinct qualities they require to be in circulation for a lengthy period of time. These distinctive qualities also give the banknotes a unique feel and appearance, which discourages counterfeiting. The features of the Nigerian security banknotes include see-through, windowed metallic thread, kinegram, iridescent band, intaglio raised printing, portrait watermark, optically changeable ink, and colors for different values [1] (banknote samples are displayed in (Figure 1a, 1b & 1c). Recalling these symbols and the security measures on each banknote is necessary for banknote recognition. The most powerful instruments and image data processing may be used to classify banknotes using innovative technology called banknote recognition. In this paper, a deep convolutional neural network model-based banknote recognition model that can identify and classify recently issued Nigerian banknotes is presented.



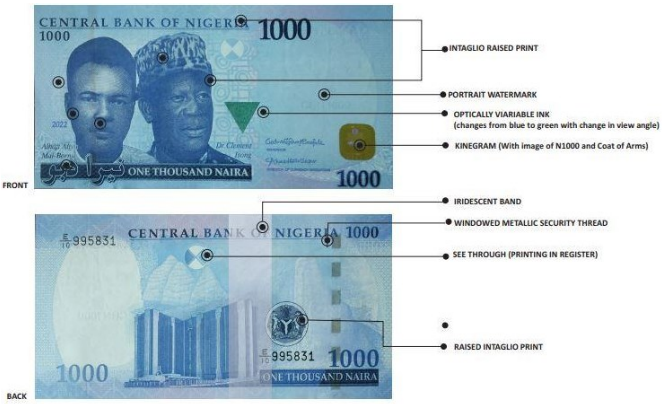


Figure 1: Banknotes & security features (a) Two hundred naira note (b) Five hundred naira note and (c) One thousand naira note

2.0 RELATED WORK

Several studies have been reported on banknote recognition technique several studies have been reported on banknote recognition techniques. A feature-based machine learning approach for effective banknote recognition was presented in [2]. The goal of the work was to recognize banknotes accurately and efficiently. Using feature-based approaches and machine learning algorithms, the suggested strategy offered a scientific and practical approach to banknote recognition, achieving accurate and efficient banknote recognition [3]. Because of “wear and tear”, feature-based machine learning techniques would not be able to handle variations in banknote look and require considerable feature engineering.

[4] demonstrated a method for recognizing banknotes that combines local and texture features. In this work, a novel method for banknote recognition was investigated. To increase the accuracy of banknote recognition, the authors devised a fusion technique that combines texture and local features. to improve recognition accuracy and resilience. This approach offered a promising trend in banknote recognition. The fusion strategy, however, may lead to an increase in computing complexity, which would make it less appropriate for real-time applications on devices with limited resources.

A real-time banknote recognition system utilizing a single shot multibox detector and mobilenet was proposed in [5]. The proposed system is a recent effort that combines the single shot multibox detector framework for object detection with a lightweight convolutional neural network architecture optimized for mobile devices in order to create a banknote recognition system with real-time processing capabilities. The detector was a step forward in the creation of effective and useful banknote recognition systems appropriate for practical uses in the real world where quick and precise processing is needed. Though this system operates in real-time, it might have trouble correctly recognizing banknotes against busy backdrops or when they overlap [6].

A deep learning-based banknote recognition system for the blind was presented by [7]. The study concentrated on creating a mechanism to help the visually handicapped recognize banknotes. The system used deep learning techniques to solve the problem of banknote identification for visually challenged people. The system attempted to precisely classify and recognize various currency denominations by utilizing deep learning models, such as Convolutional Neural Networks or comparable architectures [8]. The suggested technique marked a substantial advancement in improving the self-sufficiency and reach of people with visual impairments when managing daily tasks requiring banknotes and financial transactions. However, if the banknotes are substantially damaged or partially torn, the system's performance might suffer, which would decrease the recognition accuracy [9].

Convolutional neural networks were used in the banknote recognition method described [10]. The application of convolutional neural networks (CNNs) to banknote recognition was the main emphasis of this study. CNNs are appropriate for banknote recognition since they have demonstrated impressive performance in image recognition applications. The study tackled issues related different illumination, folds, wrinkles, and occlusions that are frequently present on banknotes. The system sought to achieve high accuracy in recognizing and categorizing various currency denominations by utilizing CNNs. The study shed light on a recent effort that uses deep learning methods—in particular, CNNs—to address practical issues related to banknote recognition tasks, thereby advancing banknote recognition systems. While CNNs work well for image recognition tasks, addressing different lighting conditions, banknote creases, folds, and occlusions may provide difficulties for this method [11].

Each of the approaches of banknote recognition in the reviewed literature had advantages and disadvantages. The accuracy and effectiveness of banknote recognition systems are constantly being improved by developments in deep learning and computer vision techniques. Further research and studies is still needed to address issues such as occlusion, distortion, and fluctuations in lighting.

3.0 METHODOLOGY

The proposed deep convolutional neural network model for banknote classification comprises fourteen convolutional layers, two fully connected layers, two output layers, along with additional components such as Max-pooling, Sigmoid activation, batch normalization, dropout, ReLU activation, and flatten layers as shown in Figure 2.

Banknotes Image: 224 x 224 x 3

Pre-processing

Image normalisation

Figure 2: Banknote recognition architecture

The model will be trained using the fully connected layer and the whole convolutional basis of the network. The picture is downsized to (224*224) grayscale and used as network input, meaning that (224, 224) shaped the matrix. The model used kernels that covered the whole picture of the banknote, measuring 3×3 and had a stride of 2. Max pooling is examined with a 2x2 pixel window that has two slices. Convolution layers in the suggested model move across the input image using a receptive and ReLU activation function. The deep convolutional neural network model is used to pre-process the banknote image in order to improve the detection outcome for the banknote categorization. Based on the bounding boxes detection, the banknote portions will be cropped from the original input image. The model's three 3x3 convolutional layers each feature two strides and the same size of spatial padding. Following every set of three consecutive convolutional layers, the model will employ max-pooling with a 2×2 window size and a stride of 1. Additionally, the threshold operation implemented by the suggested model ensures that any negative input values within the feature map are set to zero. ReLU activation function: Boost generalization, expedite computation, and expedite model training. A deep convolutional neural network with constrained hardware resources, short deployment time, and low processing power is implemented. The suggested model outlines the process and procedures for doing iterative tests to find the optimal CNN model.

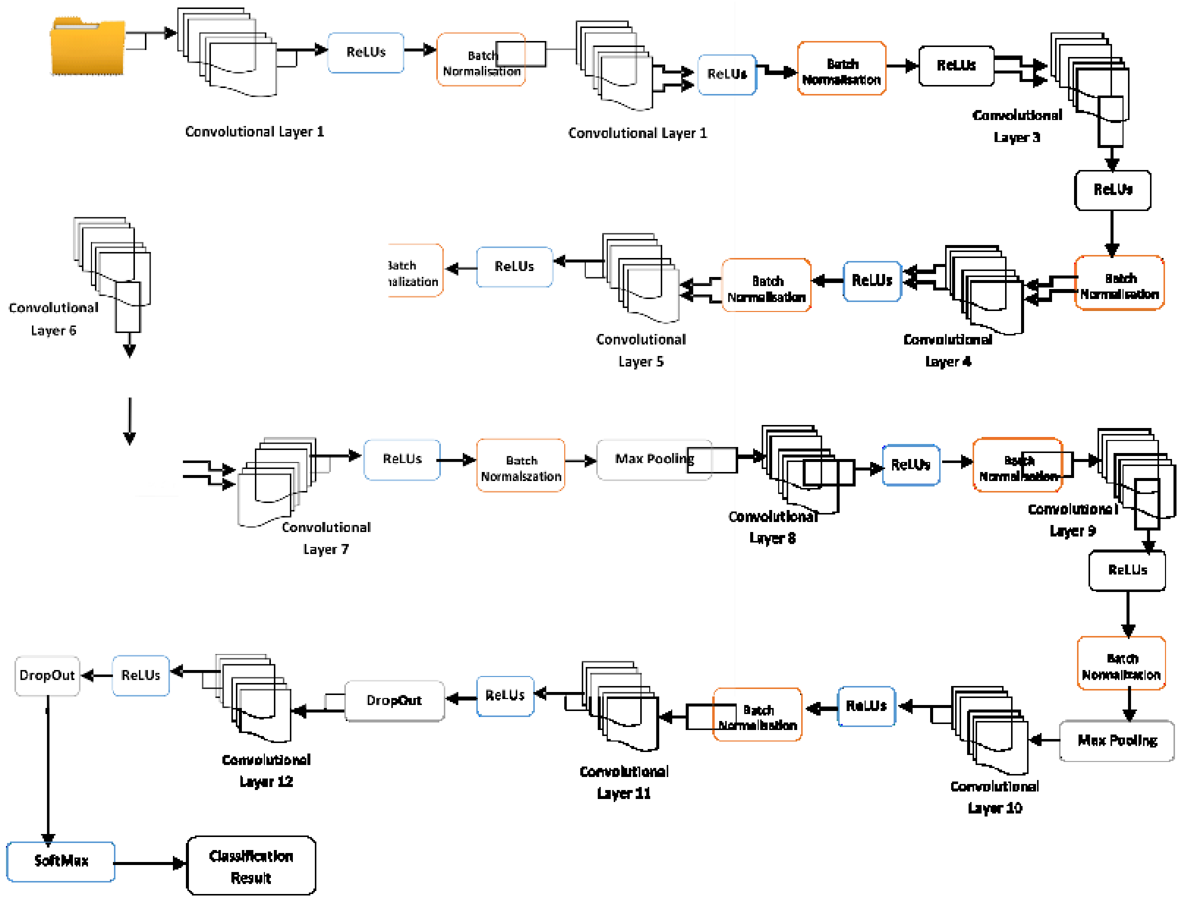


Figure 3: The Convolutional Neural Network Components

The various interactions of the components make up the suggested model as shown in Figure 3. The input layer comes first. Grayscale banknote images of 224 x 224 pixels are accepted by the input layer, which then forwards the data to the first convolution layer without doing any calculations. Since there are no learnable features in this layer, there are zero parameters. The convolutional layer is the second. Two completely linked layers and fourteen convolutional layers are present. Moreover, it is the output layer. The 224x224x1 input picture is filtered by the first layer of the model using 32 kernels with a filter size of (3x3) and a stride of 1 pixel. The completely connected layer comes in third. The output layer is one of two dense layers where the model operates at the fully linked layer. The model comprises two neurons in the output layer (binary classification) and 64 neurons in each of the first two dense levels. The final output of the fourteenth convolutional layer enters the flattened layer and converts the 2D input into a vector value that has been flattened. The output of the preceding layer is then accepted and flattened by the first dense layer, which then computes the class score and the number of neurons in the layer. Two neurons with sigmoid activation functions are present in the output layer since the model's goal is to classify binary classification.

When digital image processing is used, digital images are conveyed by vectors and are represented similarly to a matrix in computer vision. An NxN matrix, for instance, displays a N*N image. Equation 1 illustrates how the then-dimensional vector becomes a one-dimensional vector by sequentially aligning the image's rows of pixels.

$$imgvec = \{x_1, x_2, \dots, x_{N^2}\} \quad (1)$$

where the first N elements, starting at x_1 and going up to x_{N2} correspond to the image's first row, the next n elements to the image's second row, and so on. Thus, given 20 sample photos, each represented by an image vector, the huge image matrix describing the class similarities between those images may be expressed. By condensing the image space to a subspace of a lesser dimension while keeping as much relevant information from the source images as feasible, the dimensionality issues are avoided. To effectively solve the issues, principal component analysis (PCA) is the appropriate method. A useful technique for reducing the dimensionality of a multidimensional dataset for analysis or visualization in statistics is principal component analysis (PCA).

The Scale Invariant Feature Transformation is used to extract pertinent security components from the selected banknote (SIFT). The SIFT characteristic provides a sufficient description of the image's invariance to rotation, scaling, and small changes in viewing direction. The capabilities of the SIFT feature extraction approach are essential for correctly classifying banknotes. Giving a detailed description of the nodes, orientation assignment, feature localization, and scale-space extrema detection are important in this case. Scale-space extrema detection is one technique for identifying the whole scale and image points to classify the banknote. The image is blurred in various sizes throughout this procedure in order to find scale-invariant characteristics that call for a difference of Gaussian (DoG) analysis using the input picture data (blurred pictures). A constant multiplicative factor k determines the various scale on which the DoG must handle the separate scale-space extrema. The mathematical model is described as follows:

Using the convolution operator (*) from a Gaussian kernel variable G(x, y, σ) and the input image I(x, y), let the scale-space be defined as L(x, y, σ).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)$$

The parameter σ represents the scale of the key point, and the standard deviation of the Gaussian function is as follows:

$$G(x, y, \sigma) = \frac{1}{\partial \prod \sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3)$$

After a factor k is applied to the DoG of two scales, the DoG Function D(x, y, σ) is obtained as follows:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)$$

When analysing and comparing descriptions of genuine and counterfeit banknote images, the local extremum values are pivotal in distinguishing between the local maxima and minima of D (x, y, σ) of key points, serving as a reference point. SIFT feature extraction techniques, employing varied thresholds, are utilized to select suitable or likely candidates. The gradient and magnitude of the key point are important in identifying the image region because they provide a histogram of gradient orientations q(x, y) that is weighted by the gradient magnitudes m(x, y) generated from nearby features. The identified key point features are differentiated using the SIFT methodology for local descriptors, while the local image gradient scale within the input image is evaluated.

3.1 Data Pre-Processing

The objectives of pre-processing for banknote recognition images include refining banknote attributes essential for the banknote classification model and enhancing image quality through the mitigation of undesired distortions. Banknote image processing aids in classifying an image into predefined categories during the recognition phase by furnishing a refined representation of the image's primary information. From the image's raw pixels, the deep learning system extracts specific attributes from the image. By starting from scratch and creating convolutional neural network models, it explicitly pre-processes the dataset. Size normalization, color scale conversion (from RGB to grey), and histogram equalization are done on the data before processing. Normalizing image size comes first. The initial banknote images of varying denominations are resized to 224x224 pixels for a variety of banknote image sizes while reducing the size of the photos in a 2:1 ratio to save computing time. These will improve precision and address the over-fitting problem. Second, convert RGB to grayscale. Grayscale images are created by converting the RGB banknote images. A grayscale image is important since it reduces the model's computational requirements. The grayscale image created by the luminosity approach yields better results for the model based on the pre-processing of the image. Equation 5 illustrates the method's usage of the luminosity method, which assigns weight to each three-color channel based on its wavelength, and the average of the RGB channel to produce the grayscale image.

$$greyimg = 0.17R + 0.62G + 0.21B \quad (5)$$

Light refraction is the measure of image intensity. Refracted intensity is image "I" if image I0 has the same intensity (Saravana *et al*., 2022). Equation 6. This suggests that achieving a more satisfactory image involves minimizing the color blending, notably by decreasing the red color component and the green component situated between red and blue. Furthermore, it is essential to improve the detecting precision or accuracy of the ratio of these three hues.

$$I = I_0(1 - \eta)e^{-\alpha d} \quad (6)$$

The refraction of light is image intensity. If image I0 has an intensity of I, then image "I" has a refracted intensity

(Saravana et al., 2022). A more acceptable image is obtained by repeating the color combination in order to minimize the red color contribution and the green between the red and blue colors, which is crucial for improving the detection precision or accuracy of the ratio of these three colors, as indicated by equation 6.

Third, improving the quality of the image. The method of enhancing image quality while preserving the informative content and integrity of the original banknote data. The majority of banknotes exhibit inaccurate currency usage and service duration parameters due to color, texture, and integrity loss. To contrast the visual data, histogram equalization techniques are taken into consideration. Enhancing the image quality involves enhancing the banknote image's contrast to improve its appearance.

3.2 Feature extraction using deep convolutional neural network

To produce patterns, feature extraction requires dynamic computing effort and a range of situations (Figure 4). Before banknote classification, the fourteen convolutional layers are used to categorize the banknotes. The convolutional neural network model automatically extracts many properties from the input picture during training. Batch normalization, the ReLU activation function, five max-pooling layers, and fourteen continuous convolutional layers—which are composed of a sequence of filters or learnable kernels—are the methods used in convolutional neural network-based feature extraction to extract local features from the input picture. The three-by-three filters are part of the convolution layer. Figure 4 illustrates that pattern formation for feature extraction necessitates a dynamic computational effort. The banknotes are categorized using fourteen convolutional layers prior to banknote classification. Throughout training, the convolutional neural network model automatically picks up on a variety of traits extracted from the input picture. In feature extraction based on convolutional neural networks, local features are derived from the input image through a continuous sequence of fourteen convolutional layers, which include five max-pooling layers, batch normalization, and ReLU activation functions. These convolutional layers consist of 3x3 filters, allowing them to learn kernels or series of filters throughout the process.

The features map contains patterns from the provided banknote picture. ReLU activation functions are used to route each feature map value. By transferring the recovered feature map to the Max pooling layer, the convolutional neural network's computational cost is minimized and its feature map resolution is decreased. The convolution layer, batch normalization, max pooling layers, and ReLU activation function are the phases in the feature extraction technique that use the banknote as an input.

The convolution layer uses a feature detector on the picture to analyze the input data and computes the total products at each point. Convolution on the input image yields feature maps of its size and filter number, which are then used to calculate the number of strides and padding. Equations 7 and 8 demonstrate how the batch normalisation technique is used to standardise the inputs to a layer for every mini-batch, train deep networks with fewer training epochs, and stabilise the learning process.

$$x_m = \frac{x_m - \mu}{\sqrt{\sigma^2 + \beta + \varepsilon}} \quad (7)$$

$$y_m = \gamma x_m + \beta \quad (8)$$

Where the input is denoted by x_m, m=1, 2, …, M, and the variance and mean of the input values are represented by σ and ε, respectively, by a tiny positive number. The two parameters that are acquired during the training process are Gamma (γ) and Beta (β).

In order to introduce nonlinearity into a feature map, we employ the ReLU (rectified Linear Unit), a reflection syndrome on the convolution process, which is calculated as

$$ReLU(x) = \max(0, x)$$

$$ReLU(x) = \begin{cases} x \ if \ x > 0 \\ \\ 0 \ Otherwise \end{cases}$$

Sampling, also called pooling, is the process of lowering the size of corrected feature maps created by the ReLU step in order to choose the best value. It does this by using spatial pooling in repeated rounds of max pooling procedures. The two-dimensional matrix is made into a dimension by flattening it for best compatibility and inserting neural networks for classification judgments that follow the whole linked layer to learn the feature from the previous stages of picture classification.

In the suggested model, there are two fully connected layers, including the output layer. Through the first complete connected layer, the output of the convolution layer, pooling layers, and flattening layers are accepted. The convolution layer and pooling layer's 2D dimensional output was transformed into 1D by the flattened layer before being sent into the fully linked layer. Vector value is used to represent the greyscale picture characteristics class. Figure 6 shows how detection is handled by the fully connected layer of the network. The features that were taken out of the earlier layers of the proposed deep convolutional neural network will be used to classify the banknote. Additionally, the proposed model will use backpropagation and a transformation method to classify the input banknotes into a probability distribution across classes. The cross-entropy function and the softmax function function together in this case. The difference between the intended distribution, d(x), and the probability distribution of the softmax function's output, f(x), is computed using this function. The loss (L) of class number (C) is seen in Equation. 9.

$$L_i = \sum_{c=1}^{C} \left[d(x^i)\right]_c log\left[f(x^i)\right]_c \quad (9)$$

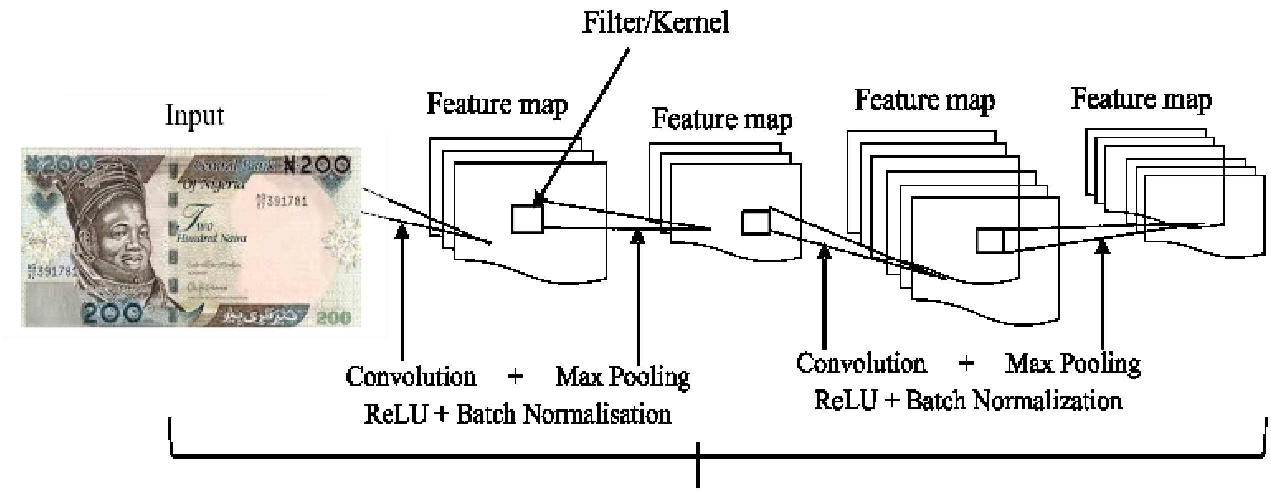


Figure 4: Feature Extraction and Feature Learning

4.0 PERFORMANCE EVALUATION METRICS

The suggested model's efficacy will be assessed using a number of indicators. These metrics will be used as assessment tools to ascertain the competency of the model. Recall, accuracy, precision, F1-score, and confusion matrices are some of the most often used metrics for evaluating the performance of our model.

4.1 Classification Accuracy

One measure used to assess classification models is accuracy. It shows the percentage, usually given as a number, of accurate predictions our model made. The degree to which anticipated values in test data closely match actual values is reflected in accuracy. The confusion matrix that the model creates can also be used to determine a classification's accuracy. Based on the confusion matrix, the accuracy was determined using the following mathematical expression as shown in Equation 10:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots \quad (10)$$

where TP, TN, FP, FN represent true positive, true negative, false positive and false negative respectively.

4.2 Precision

The classifier's precision in predicting positive occurrences can be measured by calculating the ratio of true positives to all positives predicted by the model. The precision of the model will decrease as the number of false positive predictions increases. Equation 11 is used to calculate precision:

$$Precision = \frac{TP}{TP+FP} \quad \ldots \quad (11)$$

Recall

This metric, which can be represented mathematically in Equation 12, counts the proportion of accurate positive predictions produced out of all possible positive predictions:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

F1 score

It is the mathematically represented weighted average of Precision and Recall, as shown in Equation 13:

$$F1score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (13)$$

4.3 Confusion Matrix

A performance evaluation metric called the confusion matrix is used to evaluate a classification model. It entails calculating the proportion of test records that the model predicts accurately and inaccurately. The confusion matrix shows how the model gets confused or makes mistakes while classifying examples, providing important insights into the model's performance. It offers a thorough understanding of the model's performance, including the precision of forecasts for various classes, the kinds of accurate and inaccurate predictions made, and the model's overall effectiveness. The confusion matrix provides a thorough explanation of the model's performance by presenting its findings in a matrix format.

5.0 FUTURE WORK

In the future, histogram equalization will be used to improve model extraction. It is a technique for altering pixel values in an image to more uniformly distribute intensities, therefore boosting contrast. Detection and the golden strip used to differentiate real banknotes from fakes were among the elements of the watermark. Data augmentation is the fourth. The training procedures based on random rotation and shifting make use of data augmentation techniques. Training on massive amounts of data will result in a high accuracy rate as a result. Through data augmentation, new data will be generated from the original input data. We enhanced the data collection by applying augmentation techniques like zoom, horizontal flip, and Gaussian blur. Subsequently, the various datasets are tested and the outcomes are compared using the trained convolutional neural network models.

6.0 CONCLUSION

Convolutional neural networks are a strong contender for image categorization in deep learning. As a result, a convolutional neural network for banknote categorization was reported in this paper. The proposed banknote recognition framework model will be implemented in the future. The proposed research framework will enhance the identification accuracy of large denomination Nigerian banknotes.

REFERENCES

[1] Krueger, R. (2012). On a Blank Slate: Cash and Cash Requirements for Future Currency Unions– Background Paper.

[2] Park, C., Cho, S. W., Baek, N. R., Choi, J., & Park, K. R. (2020). Deep feature-based three-stage detection of banknotes and coins for assisting visually impaired people. IEEE Access, 8, 184598-184613.

[3] Nasayreh, A., Jaradat, A. S., Gharaibeh, H., Dawaghreh, W., Al Mamlook, R. M., Alqudah, Y., & Abualigah, L. (2024). Jordanian banknote data recognition: A CNN-based approach with attention mechanism. Journal of King Saud University-Computer and Information Sciences, 102038.

[4] Zeggeye, J. F., & Assabie, Y. (2016). Automatic recognition and counterfeit detection of Ethiopian paper currency. International Journal of Image, Graphics and Signal Processing, 8(2), 28.

[5] Meharu, M. L., & Worku, H. S. (2020). Real-time Ethiopian currency recognition for visually disabled peoples using convolutional neural network.

[6] Lasecki, W. S., Song, Y. C., Kautz, H., & Bigham, J. P. (2013, February). Real-time labeling for deployable activity recognition. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 1203-1212).

[7] Awad, S. R., Sharef, B. T., Salih, A. M., & Malallah, F. L. (2022). Deep learning-based Iraqi banknotes classification system for blind people. Eastern-European Journal of Enterprise Technologies, 1(2), 115.

[8] Zhang, Q., Yan, W. Q., & Kankanhalli, M. (2019). Overview of currency recognition using deep learning. Journal of Banking and Financial Technology, 3, 59-69.

[9] Joshi, S., & Khanna, N. (2017). Single classifier-based passive system for source printer classification using local texture features. IEEE Transactions on Information Forensics and Security, 13(7), 1603-1614.

[10] Jang, U., Suh, K. H., & Lee, E. C. (2020). Low-quality banknote serial number recognition based on deep neural network. Journal of Information Processing Systems, 16(1), 224-237.

[11] Pham, T. D., Lee, Y. W., Park, C., & Park, K. R. (2022). Deep learning-based detection of fake multinational banknotes in a cross-dataset environment utilizing smartphone cameras for assisting visually impaired individuals. Mathematics, 10(9), 1616.