B.Sc. in Computer Science and Engineering Thesis

Developing a Bangla Currency Recognizer for Visually Impaired People: a Deep-Learning Based Approach

Submitted by

Hasan Murad
201305040

Supervised by

Dr. Mohammed Eunus Ali

Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

October 2018

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, "Developing a Bangla Currency Recognizer for Visually Impaired People: a Deep-Learning Based Approach", is the outcome of the investigation and research carried out by us under the supervision of Dr. Mohammed Eunus Ali.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Hasan Murad
201305040

CERTIFICATION

This thesis titled, **"Developing a Bangla Currency Recognizer for Visually Impaired People: a Deep-Learning Based Approach"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in October 2018.

Group Members:

Hasan Murad

Supervisor:

Dr. Mohammed Eunus Ali
Professor
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

I would like to declare the greatness of Almighty Allah for giving me the greatest opportunity of my life to study in the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology. I am extremely thankful to my beautiful parents for their constant support and love for me. I would like to express my heartiest gratitude to my respectful thesis supervisor Dr. Mohammed Eunus Ali sir for his sincere and valuable guidance, continuous support and encouragement. I am thankful to Nafis Irtiza Tripto(Lecturer, Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology) for helping me in publishing a short paper in ICTDX conference, 2019.

Finally, I would like to thank all my teachers, friends, seniors, and juniors in Bangladesh University of Engineering and Technology.

Dhaka
October 2018

Hasan Murad

Contents

List of Figures

List of Tables

ABSTRACT

With the recent development of artificial intelligence and the integration of deep learning in computer vision, different associative technologies for the visually impaired and blind people have gained increasing attention from various research communities. In this study, we have developed a camera-based computer vision technology to automatically recognize Bangladeshi bank notes to assist visually impaired people in Bangladesh. We have utilized MobileNet as our desired deep learning architecture for classification due to its scalability and performance. We have evaluated the performance of our model using a novel dataset consisting of nearly 8000 images of Bangladeshi bank notes. In order to incorporate our system as a mobile application, we have conducted an extensive feasibility study and requirement analysis with blind people to enhance the usability of the application. Our proposed system is robust and highly effective with heterogeneous backgrounds. Finally, to verify the effectiveness and efficacy of the proposed solution, we have evaluated and validated our application with the users from blind community, which achieved much appreciation from them.

Chapter 1

Introduction

The introduction of human computer interaction based technology brings new dimensions and alleviates various day to day problems faced by disabled persons. Vision being an essential key sensory modality in humans, the loss of it certainly affects the performance of almost all activities of daily living and thereby hampering an individuals' quality of life, general lifestyle, interactions with environment. Advances in Artificial Intelligence (AI) are beginning to make the lives of visually impaired people much richer by allowing them to access to the physical world through computer vision technologies [5–10]. These technologies that facilitate accessibility, safety, and an improved quality of life for visually impaired people have a very relevant social impact [11]. In this study, we aim to solve currency recognition problem as it is a major difficulty faced by visually impaired people in Bangladesh.

1.1 Motivation

The latest studies of the World Health Organization[1] report that 253 million people are visually impaired globally, in which 36 million are blind and 217 million have moderate to severe vision impairment. A rapidly aging population means that the impact of vision loss is expected to rise. Likewise, a considerable percentage of the population in Bangladesh is affected by visual impairments. From current estimation, Bangladesh has almost 800,000 blind people[2]. While interacting with the environment, visually impaired people face a number of challenges in their daily life because so much information is encoded visually. One particular problem that blind people encounter in their daily life is to know the value of the currency that they are holding. Currency recognition is therefore one of the most challenging ubiquitous problem faced by visually impaired people due to increased amount of financial transactions occurring face to face.

[1] http://www.who.int/blindness/en/
[2] https://goo.gl/dHo44K

According to the American Foundation for the Blind[3], one way that a blind person can identify paper currency is to fold each denomination in different ways. Although the idea of folding the currency seems promising, it needs others' help to organize bills. There are also various traditional techniques available to recognize Bangla currency for the visually impaired persons. People with 100% visual impairment require notes with different sizes and tactile marks that they can feel easily by touching the bank note. Individuals with low vision need contrast colours and large fonts to identify currency. However, these techniques have been proven ineffective to distinguish among different bank notes. The new bank notes introduced frequently have further multiplied the challenge. Moreover, in a developing country like Bangladesh, visually impaired people often fall a victim to deception while transacting money. Therefore, in order to ensure safety and confidence during financial transactions conducted by visually impaired people, they need assisting technology to recognize Bangla currency.

Although there is a number of studies on camera-based currency recognition that have been published in literature [12–17], many of these studies are restricted to specific and standard environment or background. For example, the entire paper note must be visible without occlusion, wrinkles, etc. However, implying these restrictions on using camera based currency recognition for visually impaired people is not feasible. Therefore, an automatic banknote recognition system for blind people should have the capability to recognize banknotes in a wide variety of real world scenarios, such as varied illumination, occlusions, cluttered background, different viewpoints, and worn or wrinkled notes, etc. Deep learning has absolutely dominated the computer vision are over the last few years. It has become a widely used approach for image classification and object recognition in an unconstrained natural environment. Therefore, we concentrate on a deep learning based approach to develop a mobile based Bangla Currency Recognizer for the visually impaired people.

Currently, there are several smartphone applications available for recognizing currencies of different countries; example as LookTel Money Reader[4], IDEAL Currency Identifier[5]. However, they do not provide an interface for Bangla currency recognition so far. Moreover, these apps are not totally satisfactory considering background effect, simplicity, and low end phone configuration. For example, IDEAL Currency Identifier fails to recognize paper currency with a small degree of background effect and LookTel Money Reader requires a high quality camera. There is still no promising smartphone application to recognize Bangla currency.

[3] http://www.afb.org/

[4] http://www.looktel.com/

[5] https://play.google.com/store/apps/details?id=org.ideal.currencyid&hl=en

1.2 Research Problem

We are going to develop a mobile based application for Bangla currency recognition for visually impaired people. We need to carefully investigate the possible difficulties faced by the visually impaired people while operating with various smart phone applications. Moreover, application requirements for visually impaired people are significantly different from sighted people. We have to focus on the technology engagement of the visually impaired people and identify unique requirements for our application from visually impaired people. As there is no existing Bangla currency dataset, we need to create a challenging dataset with almost all possible wild natural background. Finally, we have to investigate to find a suitable deep learning based model which is applicable a mobile based currency recognizing application.

1.3 Solution Overview

Development of a successful application for visually impaired people mainly depends on the simplicity and effectiveness of the design. To ensure comprehensibility of our application, we develop a real time currency recognizer. The users only need to open the application and then place it in front of a note. In Figure 1.1, we see a simple currency recognition test.

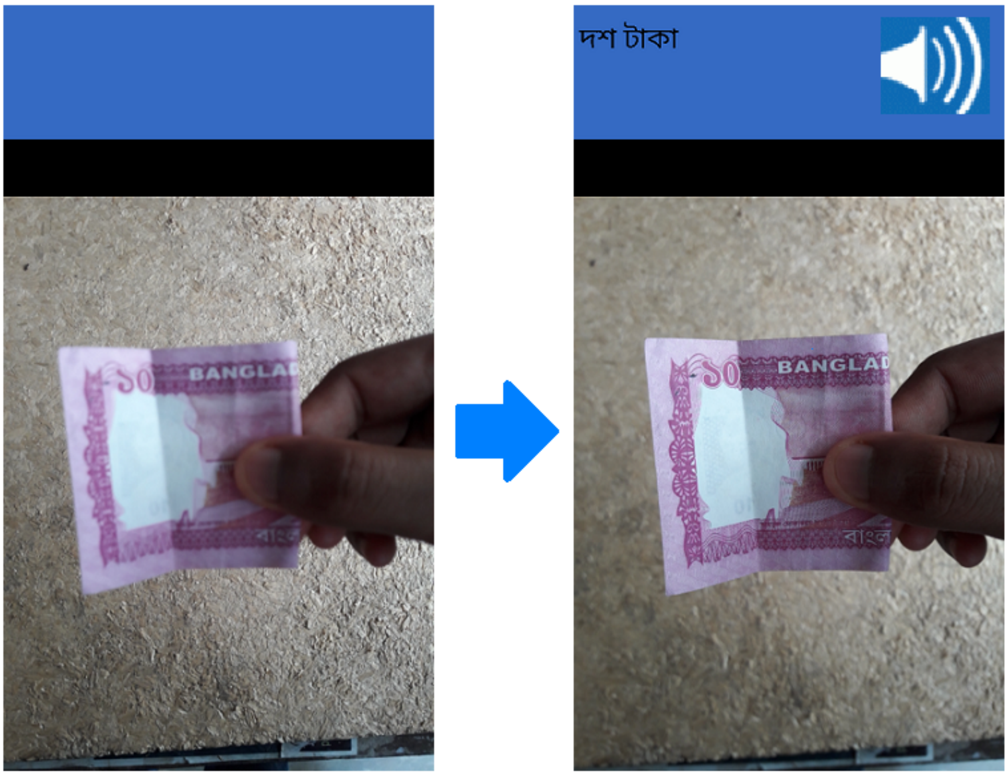

Figure 1.1: A simple currency recognition test

The salient goal of our research is to find out a deep learning based model which can recognize different bank notes with high accuracy and high reliability in a unconstrained environment. Figure 1.2 provides a high level overview of the architecture of our proposed model.

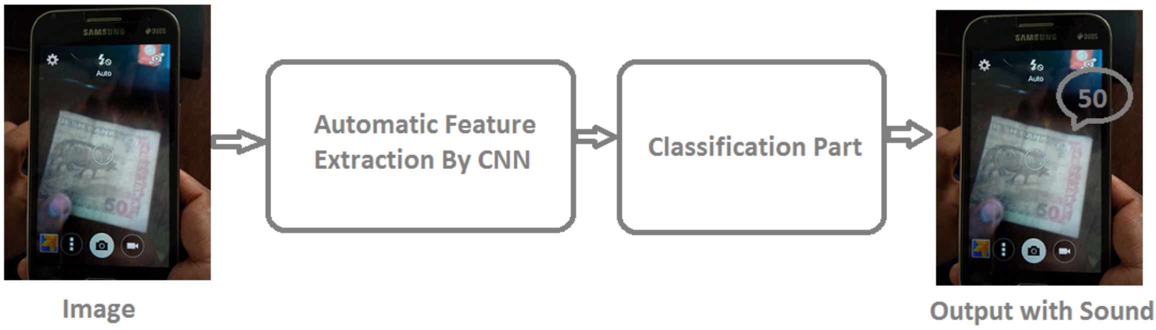

Figure 1.2: Architecture of our model

To ensure high accuracy, high reliability, low latency, and small storage of the application, we use a deep learning based model called MobileNet [18] as our model. We train, validate, and test our model on a novel Bangla paper currency dataset which contains images in diverse lighting conditions, heterogeneous backgrounds, currency note condition, and orientation.

As blind people can comfortably use android phone with the help of assisting technologies, we develop our application in android platform. The application recognizes currency in video mode. From the video, we collect image frame. After some prepossessing, we send it into our pretrained model. Finally, the model recognizes the corresponding class. After identifying the bank note, we notify the user with a clear sound in multiple languages. Since, our application is an offline based currency recognizer, it does not require any internet service. In the application development process, we implement all the basic requirements found in the feasibility study.

1.4 Contribution

The primary contribution of our research work is the develop a mobile based assistive technology for the visually impaired and blind people to recognize Bangladeshi currency. We summarize all the main contributions of our work bellow.

Since there was no publicly available dataset on Bangledeshi bank notes, we have created a novel Bangladeshi currency dataset consisting of more than 8000 images. As Bangladeshi currency dataset is now a publicly available dataset, projects and reseach works are going on based on it.

We have also developed an Android application in order to aid visually impaired people in practical scenario. We distribute the application and provide sufficient training to the real targeted audience through a local blind community named Bangladesh Visually Impaired People's Society (BVIPS).

For viability of our system, we have conducted a feasibility analysis in blind society of Bangladesh and evaluated the performance of the application in real test scenario. Our proposed system has achieved high accuracy and great performance measure while evaluating in real scenario with blind people.

Finally, our research work will provide a general direction to develop any mobile based assistive technology for the visually impaired and blind people. We create a public git repository in which all the necessary files related to data, training, validation, testing, and instructions are uploaded[6].

[6] https://github.com/HasanMuradBuet/Bangla-Currency-Recognizer.git

Chapter 2

Background

In this chapter, we concentrate on what visual impairment is, its classification, and possible causes of visual impairment. Moreover, we discuss the popular assistive technology systems available for visually impaired people. Finally, we provide a brief introduction to machine learning, deep learning, and deep learning based image classification techniques.

2.1 Visually Impairment

Visually impairment describes the functional limitation of the vision system. Visual impairment cannot be corrected with standard glasses or contact lenses. Moreover,, it will reduce a person's ability to function at vision related tasks.

2.1.1 Classification of Visually Impairment

According to World Health Organization[1] there exists different distinct categories of visually impairment in terms of Snellen chart based measurement. People with normal vision are able to read the 20 ft line at 20 ft is called 20/20 vision. Classification of visually impairment according to best possible glasses correction is given in Table 2.1.

2.1.2 The Causes of Visual Impairment

The causes of visual impairment are different for different categories. The major global causes of moderate to severe vision impairment are given below.

[1] https://www.teachingvisuallyimpaired.com/vision-classifications.html

Table 2.1: Classification of Visually Impairment

Name of the category	Best possible glasses correction limit
Near-normal vision	20/30 to 20/60
Moderate low vision	20/70 to 20/160
Severe low vision	20/200 to 20/400
Profound low vision	20/500 to 20/1,000
Near total blindness	Less than 20/1,000
Total blindness	No light perception

Glaucoma

Glaucoma is a condition caused due to the rise of normal fluid pressure inside the eyes. If the increased pressure within the eye remains untreated, it can lead to optic nerve damage. Untreated Glaucoma will result in progressive, permanent vision loss.

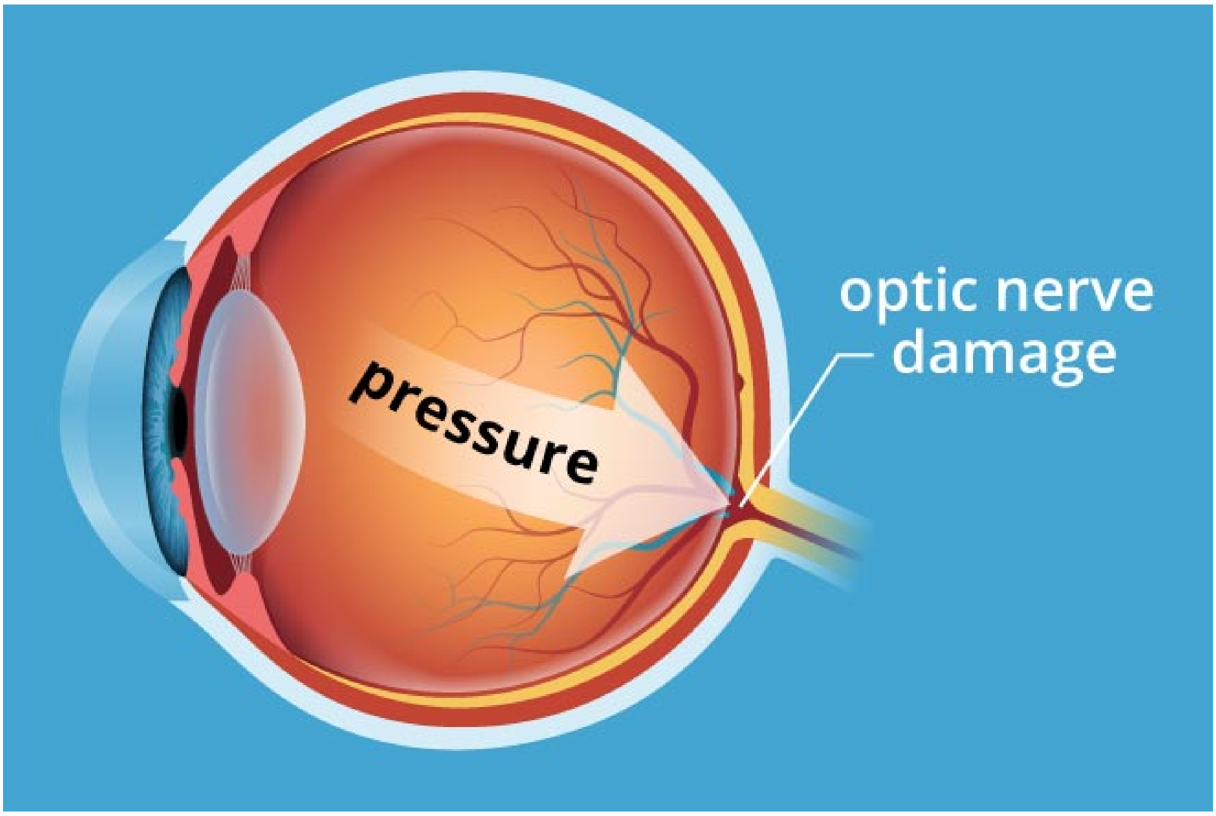

Figure 2.1: Glaucoma [1]

Age-Related Macular Degeneration

Macular degeneration is the most common case for the aged people over 60. It is caused due to the progressive deterioration of macula. Macula is one of the most critical region of retina. people affected by macular degeneration face many difficulties related to vision system like reading, driving, and sewing.

Cataract

A cataract is the condition of eye in which there is a clouding of the normally clear lens of eye. Cataracts are now becoming very common among the aged people. People affected by cataracts face difficulties while reading, driving a car especially at night and lower lighting condition.

Diabetic Retinopathy

Diabetes is a common disease that leads to damage our vision system. It damages the smaller arteries and blood vessels at the back of the eyes over the retina. If the diabetes patients remains unchecked, it results in progressive, permanent vision loss.

Near Sightedness

Near sightedness is a condition in which people can clearly see nearby objects but distant objects appear blurry. It is also known as short sightedness or myopia. High degree of myopia may result in vision impairment.

2.2 Assistive Technology for Visually Impaired People

Invention and development of assistive technology for the visually impaired and blind people has become increasingly popular among the researchers nowadays. Advances in computer vision, medical interventions, multisensory research, and wearable technology have facilitated the development of a significant number of assistive technology solutions for visually impaired and blind people. Research on assistive technology for the visually impaired and blind people has traditionally focused on object recognition, mobility, and navigation. Some of the assistive technologies are given below.

2.2.1 Screen-reading Software

Screen-reading is a kind of software which can convert all the contest appeared on the screen into speech. Those softwares are generally compatible with most other programs. Most of the cases add-on features are available for screen-reading in Windows and Linux. But it is a build-in function in Mac operating system.

Figure 2.2: OCR Technology [2]

2.2.2 Magnification Software

Magnification software are mainly popular among people with low vision problem. It acts like a high-powered magnifying glass which assists to magnify an specific location on the screen pointed by the mouse cursor.Those softwares are generally compatible with most other programs installed in our operating system. It is a build-in function in Mac operating system while in Windows and Linux, it is available as an add-on features. ZoomText and Magic are two examples of screen-magnification programs.

2.2.3 Refreshable Braille Displays

Refreshable Braille Display is a special kind of display that can create braille characters by lowering and raising combination pins. Visually impaired and blind people can read and write files using braille characters and also save those files quietly. Refreshable Braille Display is also portable and can easily able to interface with a computer device.

2.2.4 Optical Character Recognition (OCR) Systems

OCR can convert a scan document into an editable text document. It is one of the most useful software for visually impaired and blind people. They can go through the hard-copy of textbook by just scan and convert it into soft document and then create sound by a screen reader.

2.2.5 Portable Magnifiers

Portable Magnifiers is a video magnifier which is attached with a hand held cameras. Visually impaired and blind use it to read road signs or levels. By installing the GPS devices, it can provide voice instructions while walking.

2.3 Machine Learning

Machine learning is a core sub-field of artificial intelligence in computer science. It adds a new dimension by giving computers the ability to learn without being explicitly programmed. Machine Learning introduces numerous algorithms based on computational learning theory. Those leaning algorithm can learn for data, identify patterns and make decisions. Such algorithms make it possible to get rid of strictly static program instructions. Machine Learning is now employed in almost all the other fields of learning such as computer vision, robotics, search engine, etc.

2.3.1 Machine Learning Approach

There have been traditionally two fundamentally different types of Learning Approach in machine learning: supervised learning and unsupervised learning.

Supervised Learning

Supervised learning is a machine leaning task which is done based on ground truth. To apply supervised learning approach, we must have sufficient amount of labeled training data. Examples of supervised learning are Support Vector Machines, Linear Regression Logistic Regression, Neural Networks, etc. Figure 2.3 shows a supervised learning approach.

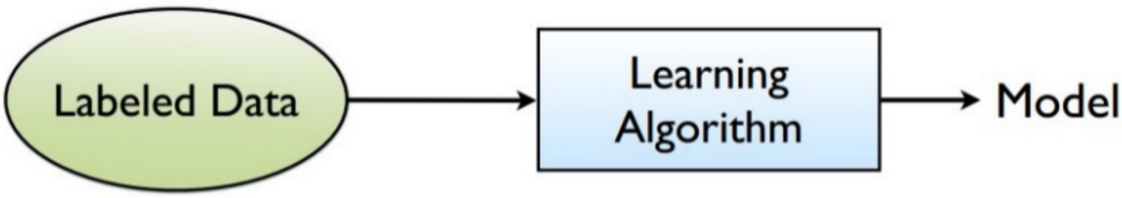

Figure 2.3: Supervised learning

Unsupervised Learning

Unsupervised learning is the most possible approach for learning when we do not have sufficient amount of labeled training data. Sometimes, it is very difficult to find label data. Moreover, collected dataset may contains noisy and false labeling. Examples of unsupervised learning are Clustering, Representation Learning, and Density Estimation, etc. Figure 2.4 shows a unsupervised learning approach.

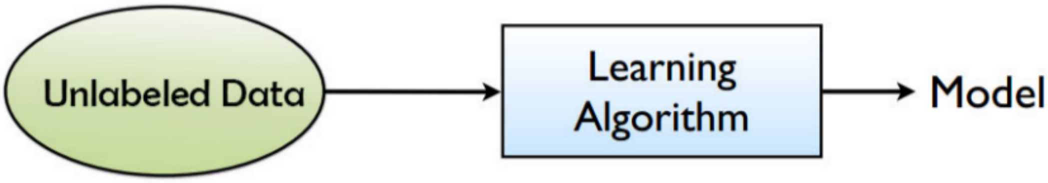

Figure 2.4: Unsupervised learning

2.4 Artificial Neural Network

Artificial Neural Network is a computational machine learning model inspired by the biological neural networks. It is a simple network which consists of an input layer at first, then multiple hidden layers and finally an output layer. The following Figure 2.5 shows a small Artificial Neural Network.

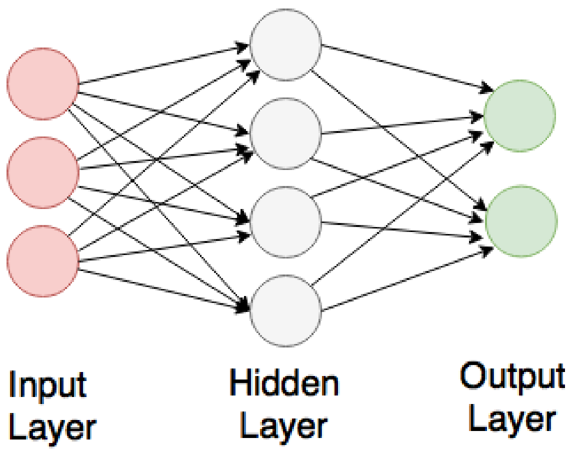

Figure 2.5: Artificial Neural Network

2.5 Deep Learning

Deep Learning Networks are nothing but neural networks with multiple layers. Therefore by using deep neural network we can easily extract complex high level features. We concentrate on deep convolution neural network which is widely used for extracting features from image data.

2.5.1 Convolutional Neural Network

Convolution neural network(CNN) is an effective tool for image classification and image recognition. In general, CNN has four main operations: Convolution step, Non linearity(ReLU), Sub-sampling or Pooling layer. Figure 2.6 shows the core building block of CNN.

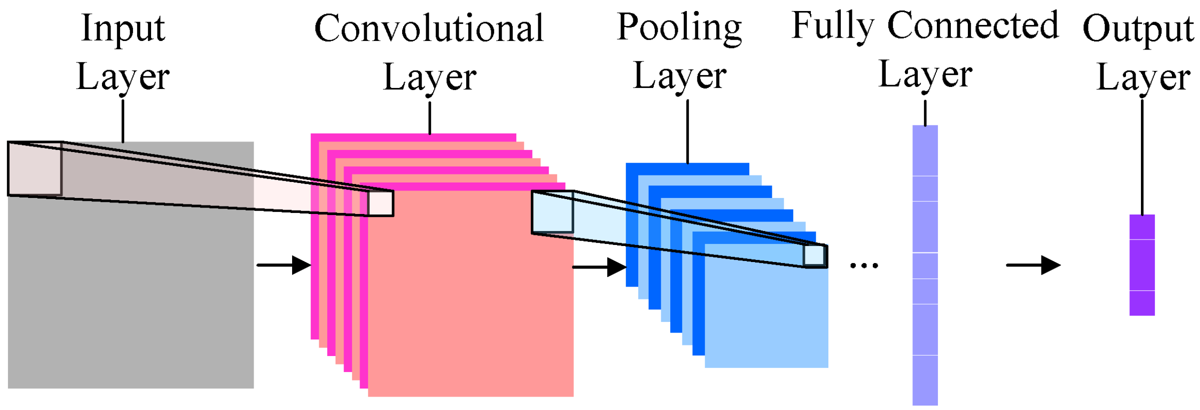

Figure 2.6: CNN [3]

Convolution step

It is the most important layer of CNN. In this step, we preserve the spatial relation by learning features from image with a filter. We slide the filter over the image, compute the dot product and generate a feature map matrix as output.

Non linearity(ReLU)

Rectified Linear Unit(ReLU) introduces non-linearity after the convolution step. It is a element-wise operation which replace all negative number with zero.

Sub-sampling or Pooling layer

The pooling layer mainly reduces the dimensionality of a image matrix but do not remove important features. There exists different types of pooling layers: Max pooling, average pooling, sum pooling etc.

Classification layer

The classification layer consists of some fully connected layers and a final softmax layer. The fully connected layer is nothing but a simple multi-layer perceptron. The convolutional layer and pooling layer provide a representation of high level features and fully connected layer and the last softmax layer use these features to classify the input image.

2.6 Computer Vision

Computer vision has now become one of the most exciting sub-field of Artificial Intelligence and Computer Science. It aims at giving computers a visual understanding of the world. Computer vision includes all the methods that are necessary for acquiring, processing, analyzing, and understanding images. Some of the most sophisticated sub-area of computer vision are discussed bellow.

2.6.1 Image Classification

Image classification means assigning a label to an input image from a set of possible categories. It has now become a core research problem computer vision that has numerous practical applications. But image classification is a difficult task as there are a variety of challenges such as background clutter, image occlusion, scale variation, image deformation, illumination conditions, intra-class variation, viewpoint variation, etc.

2.6.2 Object Detection

Object detection refers to the ability to locate and identify instances of real-world objects in an digital image. There are a wide range of practical applications of object detection. Some exciting examples of object detection are face recognition, people counting, vehicle detection, video object co-segmentation, etc.

2.6.3 Object Tracking

Object tracking is the process of tracking a path followed by a specific object in a given video scene. There are many real world application of object tracking. Autonomous driving systems such as self driving vehicles must track object to to avoid any kind of accident.

2.6.4 Semantic Segmentation

Semantic segmentation of an input image is very important for semantically understanding of an digital image. It divides the whole image into different group of pixels and label each group properly. The main applications of semantic segmentation are autonomous driving, GeoSensing, facial segmentation, precision agriculture, etc.

2.7 Application of Deep Learning in Computer Vision

Deep leaning is a recent addition to machine learning which gradually replaced traditional machine learning algorithms. As deep learning has a great advantage in the automatic feature extraction, It is now widely used in the field of computer vision. Deep learning is now applied to hundreds of problems in the field of computer vision which achieve state-of-art performance.

2.7.1 Image Captioning

Image captioning is the process of providing simple description of an image in textual representation. It is now becoming an essential feature in social media like Facebook and Twitter.

2.7.2 Face Detection

Face detection means identifying human faces in a digital image. There exists variety of applications of face detection. For example, Facebook and Google Photos can automatically tag the right person in an image.

2.7.3 Generating Photos

Generating photos is also an exciting application of deep learning. Astronomers are now using deep learning to create photos of galaxies as well as volcanoes.

2.7.4 Reading text in the Wild

The detection and recognition of characters from images captured in wild life is a most recent application of deep leaning in computer vision. Most common applications are recognizing road sign, store name, etc.

2.8 Deep Learning Based Image Classification

Deep Learning methods have seen a lot of success in the domain of image-based data. Deep Learning introduces an automated feature engineering that replaces the manually handcrafted feature engineering. Therefore, deep learning based feature extraction is the best choice for image classification.

2.8.1 State-of-the-art deep learning image classifiers

Some of popular state-of-the-art deep learning image classifiers which achieved high accuracy on ImageNet [2] dataset are given below.

MobileNets

MobileNets [18] is a small and effective model for mobile based application. It greatly minimizes the number of parameters need to be computed and provides low-latency. There exists 16 different pre-tarined models for different possible application sizes.

VGG Net

The VGG Net [19] is a simple but very deep convolution network. The VGG architecture consists of a stack of only 33 convolution layers. The two available versions are VGG-16 with 16 convolution layers and VGG-19 with 19 convolution layers.

Inception-V3

Inception-V3 [20] is a deep convolutional neural network Based on the concept of inception module. Inception module introduces the concept of using several convolution filter on input

[2] http://image-net.org/

image to reduce the number of parameters and finally concatenation outputs of different convolution filters. The Inception-V3 architecture is designed by stacking 22 layers of inception modules.

Chapter 3

Related Work

For last few decades, there have been many efforts for developing electronic aids to improve quality of life and the safety of individuals with special needs [11,21–25]. Research on assistive technology for the visually impaired and blind people has traditionally focused on mobility, navigation, and object recognition [26]. Several techniques have been developed to identify banknotes from camera captured images for helping blind or visually impaired people [13–17]. The related body includes the classical image vision based methods as well as current deep learning based techniques for mobile based currency recognition. Moreover, we provide an overview about works done on Bangladeshi currency recognition.

3.1 Research on Currency Recognition

Most of current methods for currency recognition rely on manually hand-crafted feature extraction methods before classification. The traditional procedures are edge detection, masking, threshold, segmentation, wavelet transformation [27–32]. The classification part uses simple machine learning methods like decision tree, Artificial Neural Network, or Support Vector Machines(SVM) to classify the image using extracted features.

Zhang et al. [27] applied edge detection algorithm to find the edge characteristic information [33, 34] and used it to classify Chinese banknotes. They also utilized linear transform of image gray level to remove the influence of the noise in background of an image.

Random masking with Genetic Algorithms [35, 36] is a widely applicable technique in bank currency recognition. Takeda et al. [28] used Axis Symmetric Mask to avoid multiple training patterns to recognize euro currency.

The local principal components analysis (PCA) [37, 38] is a new method for improving the classification reliability. PCA reduces the dimensionality of datasets for analyzing purposes and is used for compressing image data to be classified.

Choia et al. [29] applied edge detection, thresholding and counting of wavelet coefficients in wavelet domain for extracting features from bank notes. Wavelet transform is also applicable to object identification, and texture analysis. For segmentation and classification of banknotes Patel et al. [30] used Canny Edge Detector.

For improving the recognition ability and the transaction speed Neural Network(NN) is an widely used technique in bank currency recognition [39–41]. Hassanpour et al. [14] used Hidden Markov Models for paper currency recognition.

Fast Fourier transform(FFT) [31, 32] reduces the influence of noise and acts as pre-processor for currency recognition. In that case inputs of the neural network are the fourier coefficients or its amplitudes in frequency domain.

3.2 Bangladeshi Currency Recognition

For Bangla Currency Recognition Rahman et al. [42] applied Oriented FAST and Rotated BRIEF(ORB) to perform feature matching. The proposed system can recognize Bangladeshi paper currency notes with 89.4% accuracy on white paper background and with 78.4% accuracy tested on a complex background.

Jahangir et al. [15] utilized Neural network with axis symmetrical masks for Bangladeshi banknote recognition.The propose system can recognize currently available 8 notes with an average accuracy of 98.57% with a restriction of full scan of the note. For detecting fake Bangladeshi banknotes, Ahmed et al. [43] used image processing based Feature extraction, Speeded UP Robust Feature (SURF) and Canny Edge & Hough transformation algorithm of OpenCV.

For detecting fake Bangladeshi banknotes, Ahmed et al. [43] used mage processing based Feature extraction, Speeded UP Robust Features (SURF) and Canny Edge Hough transformation algorithm of OpenCV.

However, most of the existing currency recognition methods fail to recognize currency in heterogeneous background. Therefore, in this paper, we focus to build a deep learning based system that would be able to recognize Bangladeshi bank notes in real life scenarios.

3.3 Deep Learning Based Image Classifier for Mobile Based Application

The extraction of sufficient, stable, and distinctive features is essential for accuracy and robustness of any recognition system. MobileNet is a small and efficient deep learning based model for mobile application development. The major problems with other deep leaning based models like VGG [19], Inception [20], ResNet [44] are high latency and large size of models. However, for mobile and embedded system development, we require a smaller and faster model. To define smaller and more efficient MobileNet model, it introduces two simple global hyper parameters called width multiplier and resolution multiplier. The width multiplier can uniformly thin each layer of the network while the resolution multiplier can control the input size and the internal representation of each layer.

The core building block of MobileNet is the depthwise separable convolution. It effectively reduces the computational cost and size of the model. Traditional convolution network has convolution layer, batch normalization, and ReLU while MobileNet has depthwise separable convolutions, batch normalization, and ReLU.

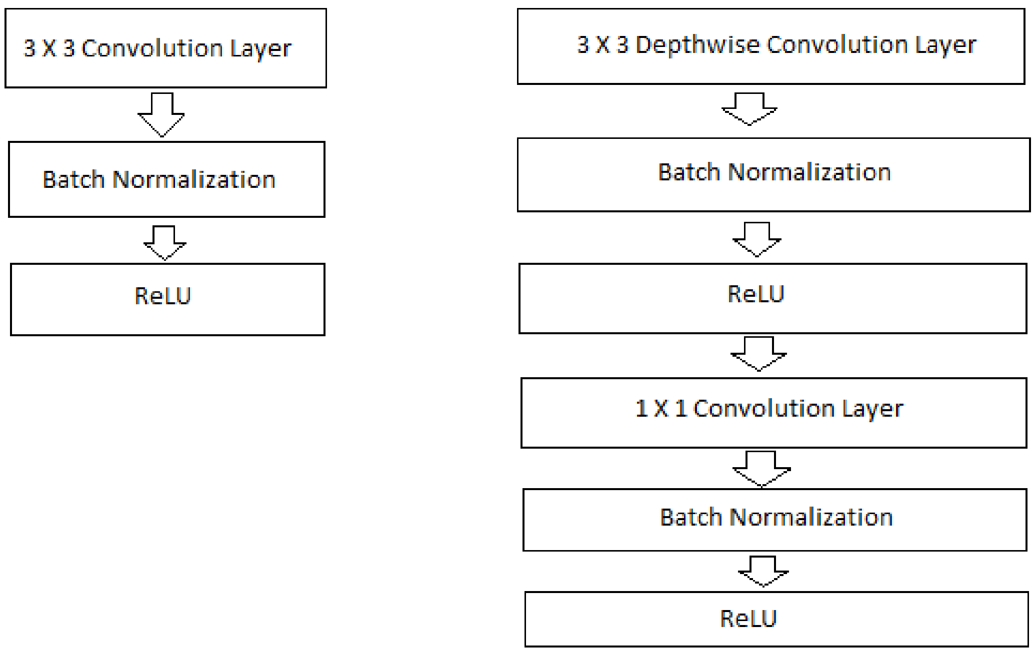

Figure 3.1: The core building block of MobileNet

Model Checkpoint	Million MACs	Million Parameters	Top-1 Accuracy	Top-5 Accuracy
MobileNet_v1_1.0_224	569	4.24	70.7	89.5
MobileNet_v1_1.0_192	418	4.24	69.3	88.9
MobileNet_v1_1.0_160	291	4.24	67.2	87.5
MobileNet_v1_1.0_128	186	4.24	64.1	85.3
MobileNet_v1_0.75_224	317	2.59	68.4	88.2
MobileNet_v1_0.75_192	233	2.59	67.4	87.3
MobileNet_v1_0.75_160	162	2.59	65.2	86.1
MobileNet_v1_0.75_128	104	2.59	61.8	83.6
MobileNet_v1_0.50_224	150	1.34	64.0	85.4
MobileNet_v1_0.50_192	110	1.34	62.1	84.0
MobileNet_v1_0.50_160	77	1.34	59.9	82.5
MobileNet_v1_0.50_128	49	1.34	56.2	79.6
MobileNet_v1_0.25_224	41	0.47	50.6	75.0
MobileNet_v1_0.25_192	34	0.47	49.0	73.6
MobileNet_v1_0.25_160	21	0.47	46.0	70.7
MobileNet_v1_0.25_128	14	0.47	41.3	66.2

Figure 3.2: Publicly available versions of MobileNet [4]

Currently, there are 16 different MobileNet models publicly available and they differ in input size and the width multiplier. Therefore, we can easily utilize a suitable model for our application based on latency and size.

Chapter 4

Proposed Solution

We provide a basic overview of the proposed solution to our problem in this Chapter. Initially, we describe the creation of a novel Bangladeshi currency dataset. Then we present the architecture used for our model. Later we discuss data prepossessing and augmentation techniques to expand our dataset before training process. Finally, we concentrate on the selection of an appropriate training method to build our classifier.

4.1 Bangladeshi Currency Dataset

Since there is currenlty no publicly available dataset of Bangladeshi bank note images, we have created a Bangladeshi currency dataset that contains more than 8000 images. There are currently eight different categories of Bangladeshi paper currency (2, 5, 10, 20, 50, 100, 500 & 1000 Taka). Our dataset has equal number of images for each category. Figure 4.1 shows some sample images from the dataset.

In order to create a standard dataset, we have followed several procedures. We have mainly focused on capturing images in real life scenarios and so there were not any restrictions with the background of images. This helps our solution to become more robust, realistic, and user oriented. Since Smartphones have a variety range of camera configuration nowadays, the collectors have captured images with various types of phone cameras. As a result, the dataset is full of images with different resolutions and backgrounds. As paper currency gets older with repeated use, we have collected lots of images for both old and new notes. Illumination is also an important issue in recognizing currency. Therefore, we have captured images in different lighting conditions.

Figure 4.1: Bangladeshi currency dataset. For each bank note, there are multiple images from different background, illumination and condition

4.2 The Architecture of Our Model

There are two distinct parts of our proposed architecture. The first one is the automatic feature extraction process and the later one is the classification task. Figure 4.2 represents a high level overview of our proposed architecture.

For high level automatic feature extraction, we consider different versions of MobileNet architecture as mentioned previously. We pass the input image through the feature extractor. It provides a vector of meaningful features that are relevant to classification.

Our next challenge is to classify the input image by using the feature vector. The classification

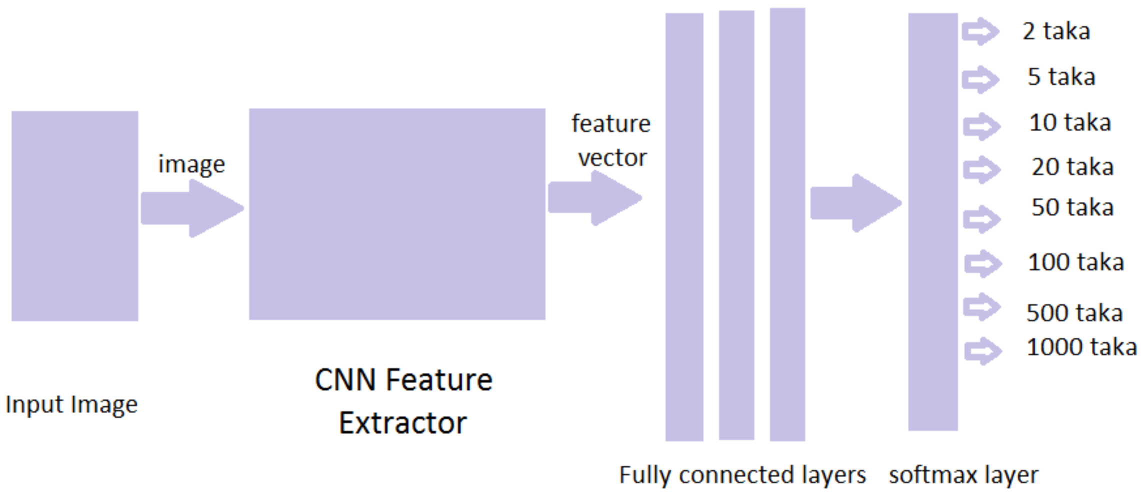

Figure 4.2: Architecture of our model

part consists of variable number of fully connected dense layers. The final layer is a softmax layer which generates a probability between 0 to 1 for each class. From this probability vector, we can determine the corresponding class of input image.

4.3 Prepossessing and Data Augmentation

Before training the desired model, we need to perform several image prepossessing steps. The MobileNet architecture only accepts input image of size 224×224. We strictly need to follow the size restriction in order to achieve a higher accuracy. The first step of the prepossessing tasks is to crop the image (if necessary) to get square sized image. It ensures uniform aspect ratio which is mandatory for MobileNet architecture. Next we need to resize input image to 224×224 pixels. Normalization is also very important for faster convergence during training process. Therefore, we need to normalize each pixel value of image between 0 to 1. This normalized image is the actual input to the MobileNet model.

Deep learning based architecture requires a larger dataset in order to build an effective model. Data augmentation is an essential method for increasing the size of any dataset. As we are using MobileNet as our feature extractor, data augmentation is compulsory for ensuring high accuracy and avoiding overfitting. Most of the machine learning frameworks provide built-in library for data augmentation. Although there are different methods for image augmentation, we select online data augmentation because of its effectiveness to reduce overfitting. In this method, we can apply augmentation for each batch of input in each iteration of training process. Online augmentation augments the total training dataset by the number of iterations times. Hence, we

can easily multiply the size of the training dataset by simply setting higher number of iterations.
‘

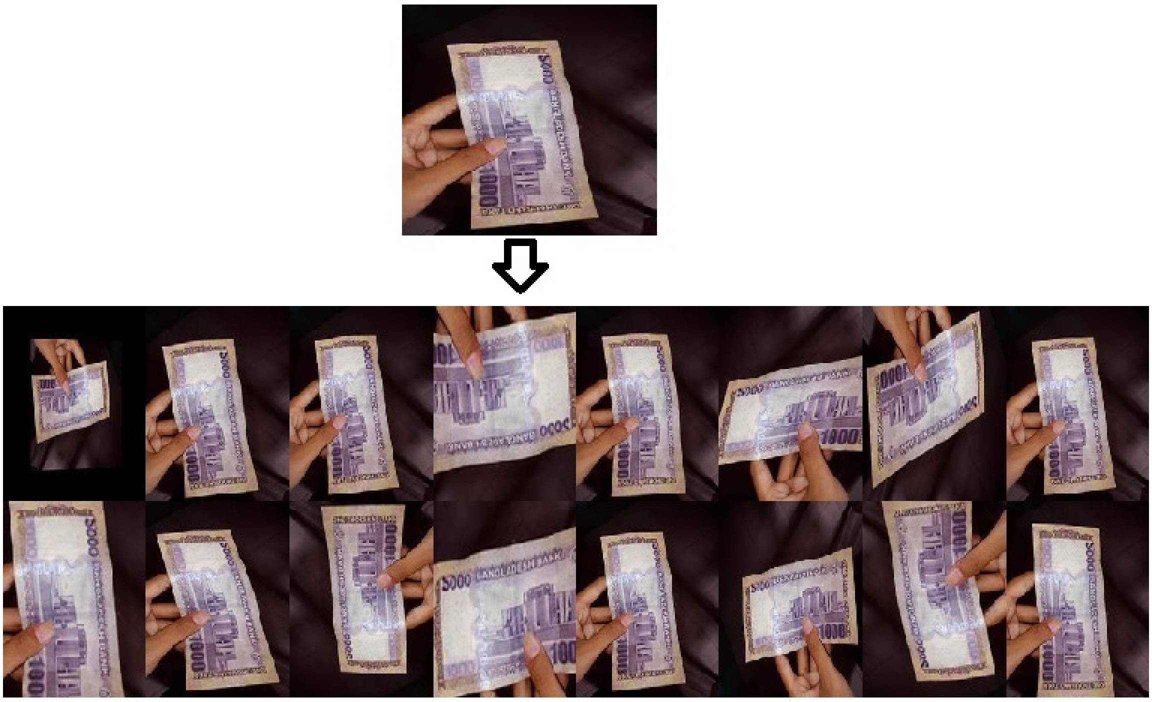

Figure 4.3: Original and augmented images

There are many different types of augmentation techniques that we can apply on the training dataset. Since user can capture image in any rotating angle, random angle rotation in both clockwise and anti-clockwise is a powerful augmenting approach. The distance between camera and note is an important factor for proper image capturing. Hence, zooming is also a high priority augmentation technique. Besides, we also apply shifting and shearing for image augmentation. All these data augmentation techniques ensure the generalization of our created model.

4.4 Training Method Selection

In our training process, we investigate the performance of different versions of MobileNet. We train our model by varying only the width multiplier of MobileNet model. There are different methods of training a deep learning based model like transfer learning, finetune, training from scratch.

In transfer learning, we only need to train the output layers after automatic feature extraction. We initialize the model with pretrained weights. However, transfer learning is not suitable for our system because the accuracy is not promising for application development. On the other hand, training from scratch is an inefficient way as it takes a lot of time to converge.

Finetune is the most significant method to train the model. In this approach, we initialize our model with pretrained weights and finetune the entire model. This procedure provides higher accuracy than the previous methods.

Chapter 5

Application Development

In order to utilize our proposed system for the aid of visually impaired people, we need to build a mobile application based on this model. In this Chapter, we concentrate on requirement analysis of the application. We also present the application design and implementation process.

5.1 Requirement Analysis

Researches in very best practise of technology design for visually impaired community [45,46] focus on both simplicity and accessibility of the application. Sierra & Togores [46] introduced the concept of "Low Vision Mobile App Portal" which provides a way to access mobile apps specifically designed for visually impaired users. Therefore, application requirements for visually impaired people are significantly different from sighted people.

In order to conduct feasibility study and requirement analysis of our desired application, we visit Bangladesh Visually Impaired People's Society (BVIPS). It is an autonomous organization dedicated for the interest of visual impaired community. For the well-being of the blind people, the organization works together with social welfare ministry. They arrange training programs and workshops to introduce blind people with the latest technologies.

We have conducted rigorous face to face interview with visually impaired community in BVIPS. There are about 200 active members in this organization. We have selected three male and two female members randomly for our face to face interview. They are aged between 25 to 50. All of our selected interviewees are total blind. We have made a set of questionnaire for our interview. The questions basically focus on the technology engagement of the visually impaired people. We study whether they can use any digital device such as smartphone, tablet, computer, etc. We have found that they are capable of using both mobile and computer based applications. They utilize assisting technologies like voice control, gesture based control or both while using these devices. The questionnaire also includes the basic requirements of the app like preferred size, mode of operation (online/offline), preferred language. Overall, the participants appreciated the motive of our research. They indicated currency recognition as their one of the major problems. They also pointed out some features for this application. Therefore, we need to consider both general and unique requirements while developing application for visually impaired people.

5.1.1 Unique Requirements for Visually Impaired People

Unique requirements refer to the special requirements for blind people due to their disability. In the interview, the participants mentioned several special requirements.

Capturing image is a difficult task for blind people without proper guidance. They cannot easily position their phone camera to focus an object. Therefore, run time recognition of currency is essential for them. They prefer a video mode based application where the value of the currency can be identified from the video streams of the camera. Using backward camera is more comfortable for them.

Blind people navigate any application by using TalkBack application. It generates sound from the labeled setting options of any application. Blind people can easily recognize setting options by hearing the sound. Moreover, there are several built-in accessibility features in recent versions of smartphones. Therefore, We need to label all the setting options so that TalkBack or built-in accessibility features can operate with our application properly. Moreover, we need to avoid navigation options with color only.

The setting options must be as simple as possible. Nested setting options are difficult to navigate for blind people. The number of clicks to access any feature should be at minimal level. Therefore, We need to implement all setting options in single depth.

After recognizing a note, if we only provide the result in graphical interface, it will be pointless for blind people. Therefore, the application should notify the result with a clear and non-robotic sound. They prefer both English and Bangla language. We need to implement all these requirements to ensure usability and functionality of our application.

5.1.2 General Requirements

Besides including the unique requirements for visually impaired people, the application also need to fulfill the general requirements of a mobile application. Recognition time is an important issue for users satisfaction. Since high latency is inappropriate in our application, we must ensure faster currency recognition. We also need to develop our application within the standard application size. However, machine learning based application requires larger space. Therefore, a light weighted model is suitable for mobile version.

As high speed internet service is still not available in the remote area in Bangladesh, we have to develop an application that can recognize currency without any internet requirement. Lighting condition is a crucial factor in image capturing. Users might operate this application in low illuminating environment. Hence, utilizing mobile flash for proper illumination is more convenient.

5.2 Application Design and Implementation

Application design for visually impaired people is more difficult task then for general people. Development of a successful application for visually impaired people mainly depends on the simplicity and effectiveness of the design. Simplicity and effectiveness ensure the usability of the application as visually impaired people are the targeted audience. To ensure comprehensibility of our application, we develop a real time currency recognizer. Moreover, we have incorporated all the unique and general requirements for visually impaired people. To use The Bangla currency recognizer, the users only need to open the application and then place it in front of a note. As blind people can comfortably use android phone with the help of assisting technologies, we develop our application in android platform. In figure 5.1, we see a simple currency recognition test with our application.

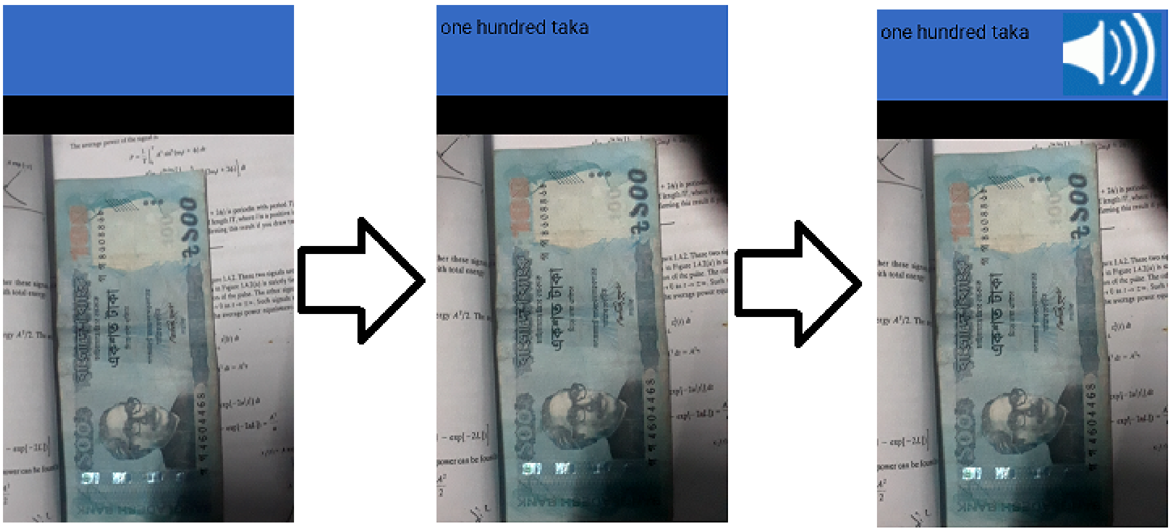

Figure 5.1: A simple currency recognition test

The application recognizes currency in video mode. We use the built-in android hardware library called *Camera2*[1] to handle video mode camera control. The *Camera2* package provides an interface to individual camera devices connected to an Android device. By using this package-

[1] https://developer.android.com/reference/android/hardware/camera2/package-summary

age, we handle all the functions related to capture or stream images from a camera device. Moreover, we control mobile flash by using this package.

From the video mode, we collect stream of image frame. After some prepossessing according to Section 4.3, we send it into our pretrained model. Finally, the model recognizes the corresponding class.

After identifying the bank note, we notify the user with a clear sound. A built-in text to speech library named *TTS*[2] is available in android to create sound in multiple languages. We use this text to speech conversion tool for generating sound in both Bangla and English.

Since, our application is an offline based currency recognizer, it does not require any internet service. We ensure low latency and small storage of the application by using MobileNet as our model. Therefore, in application development process, we try to implement all the basic requirements mentioned in the feasibility study.

[2] https://developer.android.com/reference/android/speech/tts/TextToSpeech

Chapter 6

Experiments and Results

In this Chapter, we analyze the performance of our proposed architecture with Bangladeshi currency dataset. Initially, we divide the dataset and set the parameters for training. Finally, we present experimental results of our training process.

6.1 Division of Dataset

Before starting our training process, we divide Bangla currency dataset into three subset: training, validation, and testing. We use 6400 images and apply data augmentation to increase the total number of images for training purpose. Validation dataset has 800 images which is essential for tuning hyper parameters. Finally, We use rest 800 images to test our model after validation. In each subset, there is an equal number of images for each class.

6.2 Parameter Setting

We use Python Keras[1] framework with Tensorflow[2] as a background to implement all methods for training, tuning, and testing. We need to set some parameters before training our model. For optimization, we use Adam optimizer as it is one of the most effective optimization techniques. To evaluate the performance of our model after each iteration, categorical accuracy is used as desired performance metric. We use categorical cross entropy as loss function. We attach dropout layer after each dense layer to ensure regularization and the dropout rate is set to 0.3. The model which gives the best validation accuracy will be saved after each iteration. We choose a batch size of 32.

[1] https://keras.io/
[2] https://www.tensorflow.org/

6.3 Results

We train our model by using different versions of MobileNet. Width multiplier and resolution multiplier are the two hyper parameters that define each model. The highest value of resolution multiplier is 224. We fix the resolution to its highest value to get better accuracy and only vary the width multiplier from 0.25 to 1.0. We set the number of epoch to 100. Hence, size of augmented train dataset is 100×6400 or 640000. We can see the results in the following Table 6.1.

Table 6.1: Training summary

Model Name	**Training**	**Validation**	**Testing**
MobileNet_v1_1.0_224	99.91%	99.81%	99.80%
MobileNet_v1_0.75_224	98.20%	97.01%	97.02%
MobileNet_v1_0.50_224	97.90%	94.00%	93.80%
MobileNet_v1_0.25_224	93.10%	90.01%	90.20%

For the previously mentioned parameter setting, we get convergence of our model after 30 epoch on average. In Table 6.1 we see that MobileNet_v1_1.0_224 gives us that highest accuracy of 99.81% in validation and 99.80% in testing. Moreover, we also provide the training progress curve for MobileNet_v1_1.0_224 in Figure 6.1. We observe that the accuracy becomes high after 20 iterations.

We can easily identify any false classification by observing the confusion matrix in Figure 6.2. It is evident that our model is almost accurate in recognizing any class while testing.

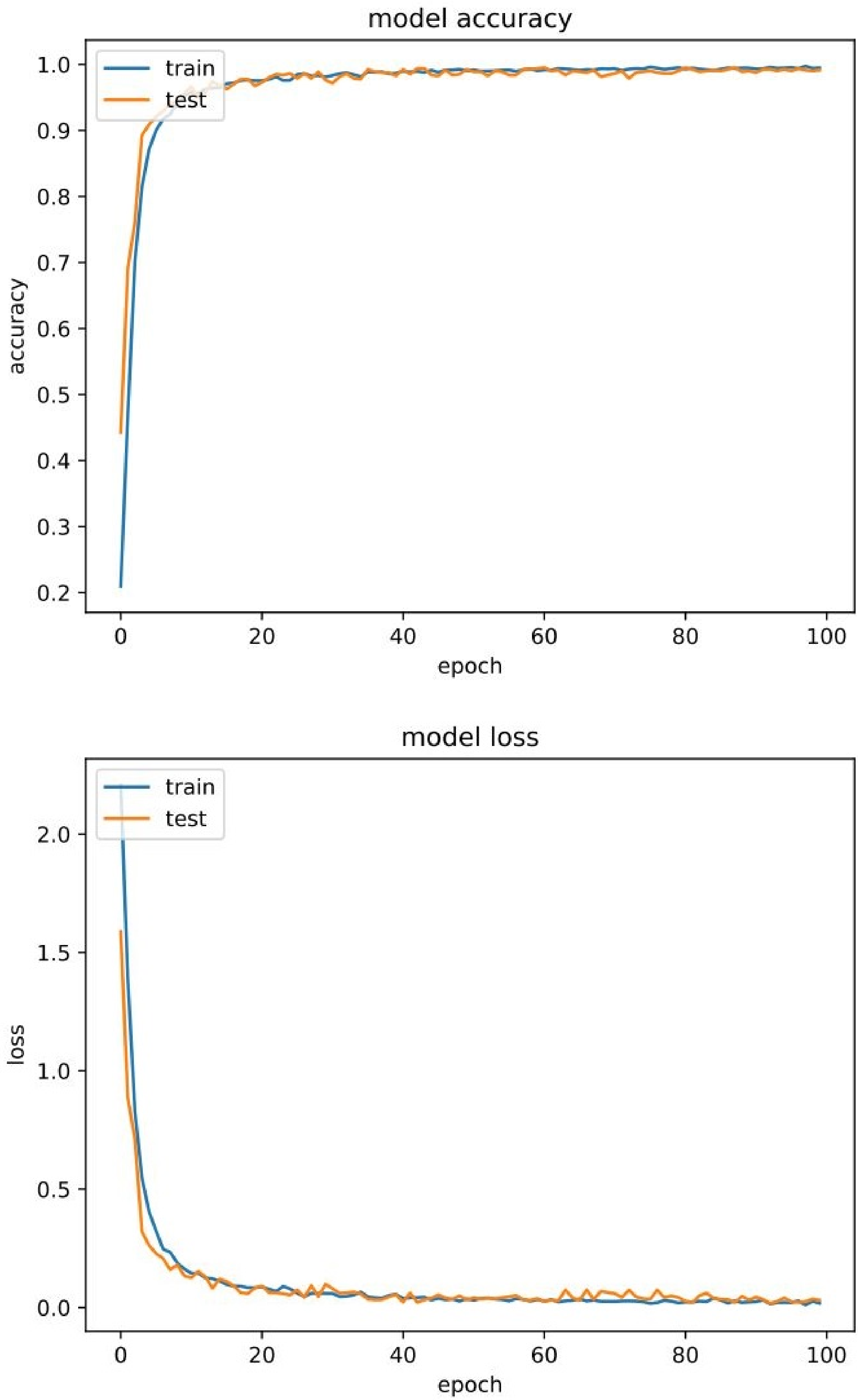

Figure 6.1: Training progress curve

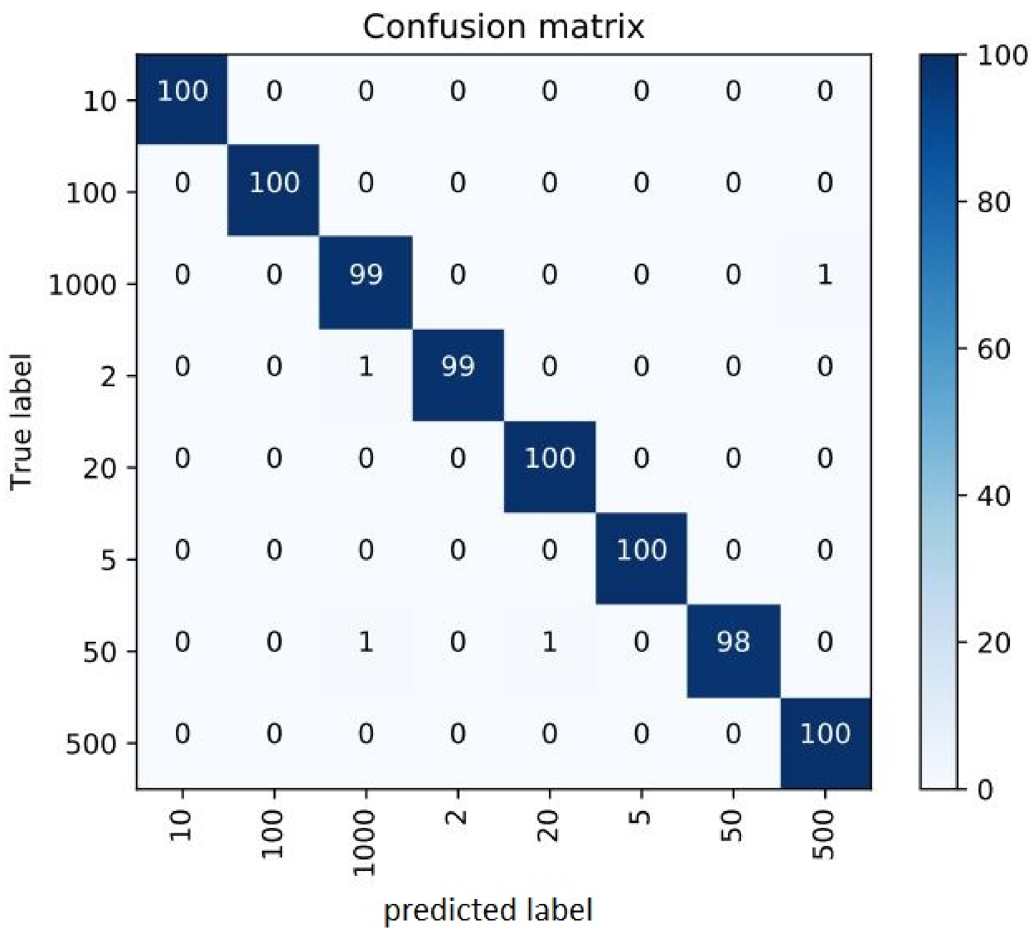

Figure 6.2: Confusion matrix

Chapter 7

Evaluation

Since we have utilized our built model to develop a mobile application, we need to measure the performance of the system. We evaluate our application by using two different approaches. First approach is to evaluate with individual tester to find any limitation of our application. In the next step, we distribute our application to blind users and get their feedback.

7.1 Evaluation with Testers

We perform an alpha testing on our application to find out any defeat or false recognition before releasing it to blind people. The performance is measured in terms of true recognition of each currency. We randomly select a specific group consisting ten people. We provide them an early version of our application. As proper lighting plays an important role while capturing an image, they test the application in different lighting conditions. Testers also inspect notes of different qualities. They use different phone cameras to conduct the test. we summarize the result in Table 7.1.

Table 7.1: Testing summary for individual user

Tester ID	**Number of Test**	**Right**	**Wrong**
1	50	49	1
2	50	47	3
3	50	48	2
4	50	49	1
5	50	49	1
6	50	48	2
7	50	47	3
8	50	50	0
9	50	48	2
10	50	50	0

The testers investigate our application in practical scenarios and get a high accuracy in most of the cases. The only fault with our application that has been found so far is recognizing background as a valid note occasionally. Therefore, some testers report these errors as false classifications. Though it is a rare event, we try to find the explanation behind it. Though Bangla currency dataset contains more than 8000 images, it cannot cover all the variations of background. If we extend the size of our dataset, we can minimize the error. We currently optimize the error by setting a probability threshold while recognizing. Our model recognizes the note when it gets more than 90% probability in a particular class. We provide some background images classified as notes in Figure 7.1.

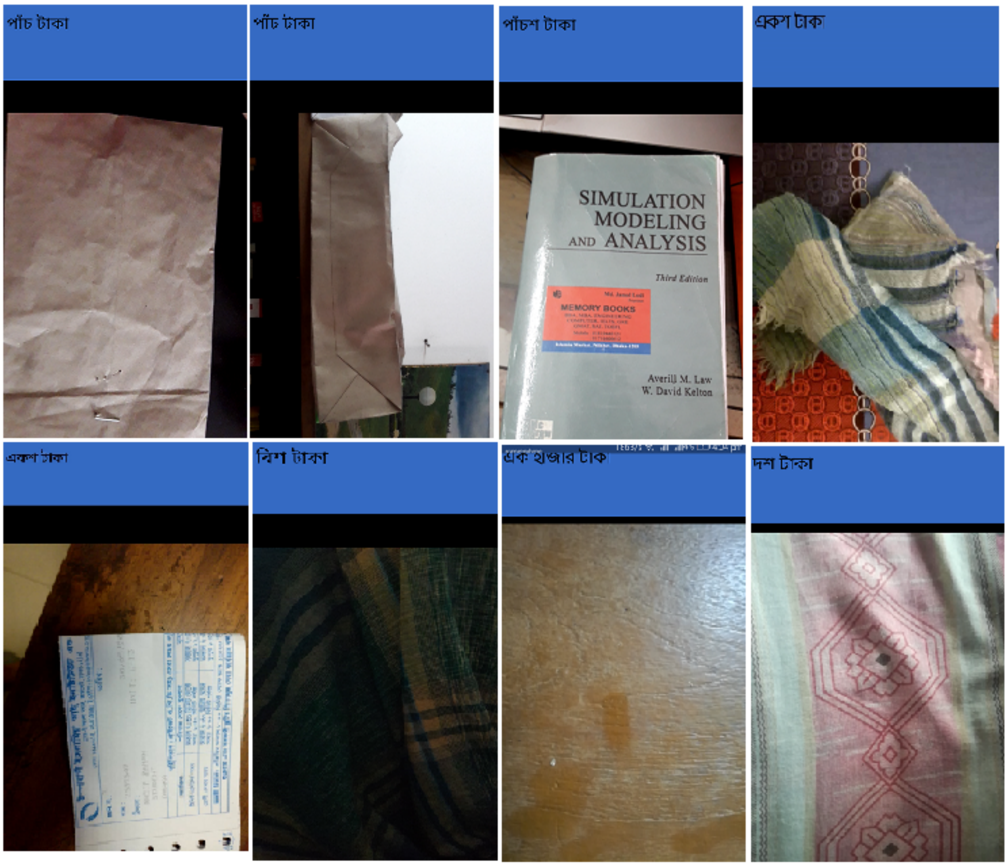

Figure 7.1: Background images classified as bank notes

7.2 Evaluation with Visually Impaired People

We have provided the first version of Bangla currency recognizing application to Bangladesh Visually Impaired People's Society (BVIPS). To introduce our application, we deliver them a general demonstration and training. In that training program, they learn how to use our application to recognize currency. They do not usually use their phone camera. As a result, they cannot focus the bank note properly by using their phone camera at the beginning. Most of time, the paper note remains out of the camera range. After several trials, they manage to focus appropriately. Distance between camera and the note is a crucial factor for proper capturing of image. Therefore, in our demonstration, we also suggest them to maintain a minimum distance.

Table 7.2: Feedback summary from participants

Real User ID	**Usefulness**	**Simplicity**	**Accuracy**
1	5	4.5	5
2	5	5	4.5
3	5	4.5	5
4	5	5	5
5	5	5	4.5

BVIPS takes the responsibility to distribute Bangla currency recognizer to the blind people. They are very excited to share the application to the visually impaired community. In Figure 7.2, we see that they have shared in social media to contact with them to collect the application[1]. Moreover, They arrange a training program every week and demonstrate the application. They have also agreed to provide regular feedback to us about the performance of the application.

We have also conducted an interview with visually impaired people in BIVPS to investigate the usefulness, simplicity and accuracy of the application. Usefulness refers to the benefit get from our application. Simplicity means the user friendliness and simple design of this application. Accuracy is the measurement of correct recognition capability. They have provided a grade between 0 to 5 to each of the criterion. We provide a summary in Table 7.2. The application helps them to instantly recognize paper notes during financial transaction. They appreciate the simplicity of our application design and are satisfied with high recognition accuracy.

[1] https://goo.gl/pukZoT

Figure 7.2: Screenshot taken from a Facebook post by BVIPS. They appreciate the app in their post and ask their users to install it

Chapter 8

Conclusion and Future Work

In this chapter, we will summarize our research work on developing the Bangla currency recognizer. Finally, we discuss possible future extension of our research work.

8.1 Conclusion

In this work, we have successfully developed a mobile based Bangla currency recognizer application for visually impaired people. For this purpose, we have created a novel Bangladeshi currency dataset which consists of image of notes captured in diverse lighting conditions, heterogeneous backgrounds, currency note condition, and orientation. We have utilized a recent deep learning based model called MobileNet due to its scalability and high performance and achieved 99.8% accuracy in testing. In order to deploy the model in a camera based mobile application, we have rigorously conducted a feasibility study and requirement analysis in blind community. In the development process of the application, we strictly focus on implementing all the unique requirements of visually impaired people so that it will enhance the usability and performance of the application. Finally, we distribute the Bangla currency recognizer among the visually impaired people through a local blind community named Bangladesh Visually Impaired People's Society (BVIPS). Our mobile application has also been evaluated by visually impaired people and received much appreciation. We have also established an effective communication channel with the blind community for regular evaluation, feedback, improvement, training workshop, and proper distribution of the application.

8.2 Future Work

Our proposed architecture is a general architecture for image classification and recognition problem. Therefore, we will incorporate the architecture to develop a large scale currency recognizer which includes foreign currencies like Indian rupee, US dollar, euro, etc. Moreover, for the scalability and high accuracy of our proposed architecture, it will be a standard image recognition architecture in mobile application development industry.

By working with the blind community, we have found that visually impaired and blind people are using very limited number of assistive technologies and there is a great opportunity to work for them.We deeply study and identify all the basic and unique requirements for visually impaired people, it will help us in the development more assistive applications for visually impaired community.

Figure 8.1: Attention based text extraction approach to recognize currency by reading the corresponding value on the note

We have investigated a more general deep learning approach called attention-based text extraction [47]. It introduces an novel technique in the field of real-world text extraction. In this approach, we can recognize currency by reading the corresponding value written on corner of the note. It will be a generalized method as model trained on Bangladeshi currency dataset can recognize currency of other countries. Figure 8.2 shows the attention based approach to currency recognition.

Figure 8.2: Bangla text extraction in wild environment

Finally, we will study the possibility to develop an assistive technology that can read the street name, store front, poster,etc for visually impaired people by applying attention-based text extraction method in wild environment.

References

[1] https://www.allaboutvision.com/conditions/glaucoma.htm.

[2] https://bit.ly/2CWBEiv.

[3] https://bit.ly/2PyQfYf.

[4] https://ai.googleblog.com/2017/06/mobilenets-open-source-models-for.html.

[5] J. Coughlan and R. Manduchi, "Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 03, pp. 379–397, 2009.

[6] J. A. Hesch and S. I. Roumeliotis, "Design and analysis of a portable indoor localization aid for the visually impaired," *The International Journal of Robotics Research*, vol. 29, no. 11, pp. 1400–1415, 2010.

[7] H. Hile, A. L. Liu, G. Borriello, R. Grzeszczuk, R. Vedantham, and J. Kosecka, "Visual navigation for mobile devices.," *IEEE MultiMedia*, vol. 17, no. 2, pp. 16–25, 2010.

[8] V. Ivanchenko, J. Coughlan, and H. Shen, "Crosswatch: A camera phone system for orienting visually impaired pedestrians at traffic intersections," in *International Conference on Computers for Handicapped Persons*, pp. 1122–1128, Springer, 2008.

[9] V. Ivanchenko, J. Coughlan, and H. Shen, "Real-time walk light detection with a mobile phone," in *International Conference on Computers for Handicapped Persons*, pp. 229–234, Springer, 2010.

[10] J. J. Liu, C. Phillips, and K. Daniilidis, "Video-based localization without 3d mapping for the visually impaired," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 23–30, IEEE, 2010.

[11] R. Manduchi and J. Coughlan, "(computer) vision without sight," *Communications of the ACM*, vol. 55, no. 1, pp. 96–104, 2012.

[12] S.-H. Chae, J. K. Kim, and S. B. Pan, "A study on the korean banknote recognition using rgb and uv information," in *Communication and Networking*, pp. 477–484, Springer, 2009.

[13] F. M. Hasanuzzaman, X. Yang, and Y. Tian, "Robust and effective component-based banknote recognition by surf features," in *Wireless and Optical Communications Conference (WOCC), 2011 20th Annual*, pp. 1–6, IEEE, 2011.

[14] P. F. H. Hassanpour, "Using hidden markov models for paper currency recognition," in *Expert Systems with Applications*, vol. vol. 5, pp. pp. 10105–10111, 1992.

[15] N. Jahangir and A. R. Chowdhury, "Bangladeshi banknote recognition by neural network with axis symmetrical masks," in *2007 10th international conference on computer and information technology*, pp. 1–5, Dec 2007.

[16] T. Kosaka and S. Omatu, "Bill money classification by competitive learning," in *Soft Computing Methods in Industrial Applications, 1999. SMCia/99. Proceedings of the 1999 IEEE Midnight-Sun Workshop on*, pp. 5–9, IEEE, 1999.

[17] F. Takeda and S. Omatu, "High speed paper currency recognition by neural networks," *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 73–77, 1995.

[18] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.

[21] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 25–35, 2010.

[22] N. A. Giudice and G. E. Legge, "Blind navigation and the role of technology," *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, pp. 479–500, 2008.

[23] R. Manduchi, J. Coughlan, and V. Ivanchenko, "Search strategies of visually impaired persons using a camera phone wayfinding system," in *International Conference on Computers for Handicapped Persons*, pp. 1135–1140, Springer, 2008.

[24] H. Shen, K.-Y. Chan, J. Coughlan, and J. Brabyn, “A mobile phone system to find crosswalks for visually impaired pedestrians,” *Technology and disability*, vol. 20, no. 3, pp. 217–224, 2008.

[25] E. Tekin, J. M. Coughlan, and H. Shen, “Real-time detection and reading of led/lcd displays for visually impaired persons,” in *Proceedings/IEEE Workshop on Applications of Computer Vision. IEEE Workshop on Applications of Computer Vision*, p. 491, NIH Public Access, 2011.

[26] A. Bhowmick and S. M. Hazarika, “An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends,” *Journal on Multimodal User Interfaces*, vol. 11, no. 2, pp. 149–172, 2017.

[27] E.-H. Zhang, B. Jiang, J.-H. Duan, and Z.-Z. Bian, “Research on paper currency recognition by neural networks,” in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, vol. 4, pp. 2193–2197 Vol.4, Nov 2003.

[28] F. Takeda and T. Nishikage, “Multiple kinds of paper currency recognition using neural network and application for euro currency,” in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 2, pp. 143–147 vol.2, 2000.

[29] J. L. E. Choia and J. Yooni., “Feature extraction for banknote classification using wavelet transforms,” in *In Proc. 18th International Conference on Pattern Recognition*, vol. 2, pp. pp 934–937, 2006.

[30] V. N. Patel, U. K. Jaliya, and K. N. Brahmbhatt, “Indian currency recognition using neural network pattern recognition tool,” in *ICRISET2017. International Conference on Research and Innovations in Science, Engineering and Technology. Selected Papers in Computing* (R. Buyya, R. Ranjan, S. D. Roy, M. Raval, M. Zaveri, H. Patel, A. Ganatra, D. G. Thakore, T. A. Desai, Z. H. Shah, N. M. Patel, M. E. Shimpi, R. B. Gandhi, J. M. Rathod, B. C. Goradiya, M. S. Holia, and D. K. Patel, eds.), vol. 2 of *Kalpa Publications in Computing*, pp. 67–72, EasyChair, 2017.

[31] T. I. S. O. K. K. F. Takeda, S. Omatu, “High speed conveyed bill money recognition with neural network,” in *Trans. Inst. Elect. Eng. Japan*, vol. vol. 112-C, pp. pp. 101–110, 1992.

[32] T. I. S. O. F. Takeda, S. Omatu, “Bill money recognition using neural network with fft as pre-processor,” in *Trans. Inst. Syst. Control and Info. Eng.*, vol. vol. 5, pp. pp. 265–2 73, 1992.

[33] G. A. H. Hassanpour, A. Yaseri, "Feature extraction for paper currency recognition," in *In International symposium on signal processing and its applications (ISSPA), Sharjah, UAE*, 2007.

[34] J. D. Z. B. E.H. Zhang, B. Jiang, "Research on paper currency recognition by neural networks," in *In Proceeding of the second international conference on machine learning and cybernetics*, pp. pp. 2193–2197, 2003.

[35] F. Takeda and S. Omatu, "A neuro-paper currency recognition method using optimized masks by genetic algorithm," in *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, vol. 5, pp. 4367–4371 vol.5, Oct 1995.

[36] F. Takeda and S. Omatu, "Image processing and pattern recognition," in *Academic Press*, pp. 133–160, 1998.

[37] S. O. A. Ahmadi and T. Kosaka, "A study on evaluating and improving the reliability of bank note neuro-classifiers," in *In Proc. SICE Annual Conference, Japan*, vol. 2, pp. pp 2550–2554, 2003.

[38] S. O. A. Ahmadi and T. Kosaka, "Improvement of the reliability of bank note classifier machines," vol. 2, pp. pp 1313–1316, 2004.

[39] F. Takeda and T. Nishikage, "Multiple kinds of paper currency recognition using neural network and application for euro currency," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 2, pp. 143–147 vol.2, 2000.

[40] M. Fukumi and N. Akamatsu, "A method to design a neural pattern recognition system by using a genetic algorithm with partial fitness and a deterministic mutation," in *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, vol. 3, pp. 1989–1993 vol.3, Oct 1996.

[41] F. Takeda and S. Omatu, "High speed paper currency recognition by neural networks," *IEEE Transactions on Neural Networks*, vol. 6, pp. 73–77, Jan 1995.

[42] M. M. Rahman, B. Poon, M. A. Amin, and H. Yan, "Recognizing bangladeshi currency for visually impaired," in *Machine Learning and Cybernetics* (X. Wang, W. Pedrycz, P. Chan, and Q. He, eds.), (Berlin, Heidelberg), pp. 129–135, Springer Berlin Heidelberg, 2014.

[43] Z. Ahmed, S. Yasmin, M. N. Islam, and R. U. Ahmed, "Image processing based feature extraction of bangladeshi banknotes," in *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, pp. 1–8, Dec 2014.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[45] Á. Csapó, G. Wersényi, H. Nagy, and T. Stockman, "A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research," *Journal on Multimodal User Interfaces*, vol. 9, pp. 275–286, Dec 2015.

[46] J. S. Sierra and J. Togores, "Designing mobile apps for visually impaired and blind users," in *The Fifth International Conference on Advances in Computer-Human Interactions*, pp. 47–52, Citeseer, 2012.

[47] Z. Wojna, A. Gorban, D.-S. Lee, K. Murphy, Q. Yu, Y. Li, and J. Ibarz, "Attention-based extraction of structured information from street view imagery," *arXiv preprint arXiv:1704.03549*, 2017.

Index

Appendix A

Codes

A.1 Training Code

```
from keras import applications
from keras.preprocessing.image
import ImageDataGenerator
from keras import optimizers
from keras.models import Sequential, Model
from keras.layers import Dropout, Flatten,
Dense, GlobalAveragePooling2D
from keras import backend as k
from keras.callbacks import ModelCheckpoint,
LearningRateScheduler, TensorBoard, EarlyStopping
from keras.models import load_model
from keras.callbacks import Callback
from keras.optimizers import Adam as Adam

import os
import sys
import zipfile

import math
from keras import backend as K
from keras.utils.generic_utils import CustomObjectScope
import shutil


```

```

def training_data(train_data_dir,img_height,
                        img_width,batch_size):
    train_datagen = ImageDataGenerator(rescale=1. / 255,
                                rotation_range=180,
                                width_shift_range=0.1,
                                height_shift_range=0.1,
                                shear_range=0.1,
                                zoom_range=[0.8, 1.5],
                                fill_mode='nearest')
    training_set = train_datagen.flow_from_directory(
                    train_data_dir,
                    target_size=(img_height, img_width),
                    batch_size=batch_size,
                    class_mode='categorical')
    listing = os.listdir(train_data_dir)
    number_of_train_images = 0
    for folder in listing:
        inner = os.listdir(train_data_dir + '/' + folder)
        number_of_train_images += len(inner)
    return training_set,number_of_train_images




def validate_data(validation_data_dir,img_height,
                        img_width,batch_size):
    validation_datagen = ImageDataGenerator(rescale=1. / 255)

    validation_set = validation_datagen.flow_from_directory(
                        validation_data_dir,
                        target_size=(img_height, img_width),
                        batch_size=batch_size,
                        shuffle=False,
                        class_mode='categorical')
    listing = os.listdir(validation_data_dir)
    number_of_test_images = 0
    for folder in listing:
        inner = os.listdir(validation_data_dir + '/' + folder)
```

```
            number_of_test_images += len(inner)

    return validation_set,number_of_test_images




def model_creator(alpha_1, img_width, img_height,
                img_channels,nb_classes):
    model = applications.mobilenet.MobileNet(
    alpha=alpha_1, weights='imagenet', include_top=False,
    input_shape=(img_width, img_height, img_channels))



    x = model.output
    x = Flatten()(x)
    x = Dense(1024, activation="relu")(x)
    x = Dropout(0.3)(x)
    x = Dense(512, activation="relu")(x)
    x = Dropout(0.3)(x)
    x = Dense(512, activation="relu")(x)
    x = Dropout(0.3)(x)
    x = Dense(256, activation="relu")(x)
    predictions = Dense(output_dim=nb_classes,
                    activation="softmax")(x)
    model_final = Model(input=model.input, output=predictions)

    adam = Adam(lr=1e-4)
    model_final.compile(optimizer=adam,
        loss='categorical_crossentropy',metrics=['accuracy'])
    return model_final


if __name__ == "__main__":

    zip_ref = zipfile.ZipFile(sys.argv[1], 'r')
    zip_ref.extractall("./test/train")
    zip_ref.close()
```

```

    str = sys.argv[2]
    file = open(str[2:],'r')

    i = 0
    img_width = 0
    for val in file:
        if i == 0:
            val.split()
            alpha_1 = float(val[0])
            i = i+1
        else:
            img_width = int(val)
    img_height = img_width
    train_data_dir = "./test/train"
    model_path = sys.argv[3] #model saving path

    img_channels = 3 #RGB
    nb_classes = 8 #final categories
    batch_size = 8

    nb_epoch = 5 #iterations


    print (alpha_1)
    print (img_width)
    print (sys.argv[1])
    print (model_path)

    training_set,number_of_train_images = training_data(
    train_data_dir,
    img_height, img_width,batch_size)

    steps_per_epoch = math.ceil(
    number_of_train_images/batch_size)
```

```
    model_final = model_creator(alpha_1, img_width,
     img_height, img_channels,nb_classes)


    history = model_final.fit_generator(training_set,
                    steps_per_epoch = steps_per_epoch,
                    epochs=nb_epoch,
                    initial_epoch=0)

    model_final.save(model_path)

    shutil.rmtree("./test")
```