Real-Time Ethiopian Currency Recognition for Visually Disabled Peoples Using Convolutional Neural Network

Mekonnen Legess Meharu (✉ mekonnen.legess@aastu.edu.et)

Addis Ababa Science and Technology University   https://orcid.org/0000-0001-6249-6038

Hussien Seid Worku

Addis Ababa Science and Technology University

Research

Abstract

A survey report made by the Ethiopian Ministry of Health along with several non-governmental organizations in 2006 G.C, there were about 5.3% of the Ethiopian population lives with blindness and low vision problems. This research work aims to develop a Convolutional Neural Network-based model by using pre-trained models to enable vision-impaired peoples to recognize Ethiopian currency banknotes in real-time scenarios. The models attempt to accurately recognize Ethiopian currency banknotes even if the input images come up with partially or highly distorted and folded Birr notes. 8500 (1700 for each class) banknotes data are collected within real-life situations by using 9 blind persons. The models were evaluated with 500 real-time videos of different conditions. The whole training, classification, and detection tasks have been demonstrated by adopting Tensorflow Object Detection API and the pre-trained Faster R-CNN Inception, and SSD MobileNet models. All the codes are implemented using Python. The model tested using numerous Ethiopian currencies at different banknotes status and light conditions. In the case of Faster R-CNN Inception model an average accuracy, precision, recall, and F1-score of 91.8%, 91.8%, 92.8%, and 91.8% are obtained respectively and in the case of SSD MobileNet model an average accuracy, precision, recall, and F1-score of 79.4%, 79.4%, 93.6%, and 84.4% are obtained respectively within a real-time video. Therefore as the first research work, the model has shown good performance in both models but Faster R-CNN provides a promising result with an average accuracy of 91.8%.

1. Introduction

Eye vision problems are considered as one of the major public health issues in the Africa continent. The WHO report indicates that 90% of the blind peoples were living in developing countries [1]. According to the national survey of the Ethiopian Ministry of Health along with several non-governmental organizations in 2006 G.C, there were about 5.3% of the total population of the country lives with blindness and low vision problems (1.6% blind and 3.7% low vision) [2].

In the current world, there is a trend to forget that peoples who are challenged to lead a normal life live among us. A person lives with some sort of vision problem faced numerous difficulties to perform a day to day activity, which seems simple tasks for us such as information access (printed media and mail), mobility, shopping, cooking, recognizing objects, and many other independent living skills [3] and also they suffer numerous serious challenges because of the consequences which come by it such as three times more likely to be unemployed, three times more likely to be involved in a motor vehicle accident, three times more likely to suffer from depression and anxiety disorders, three times near for sexual and another arrestment, and two times more likely to have a fall while walking compare with a person without vision problem [4].

As mentioned before one of the major problems visually disabled people faced in their day-to-day routine is recognizing things like currency especially paper currency. Currency is used almost everywhere [5]. Even if the electronic transaction like mobile banking and other electronic forms of payment is growing up, still hand to hand cash transaction widely used for day to day routines in Ethiopia. Both paper and coin

currencies exist and the name of the Ethiopian currency is "Birr". There are five paper currencies those are One Birr, Five Birr, Ten Birr, Fifty Birr, and One hundred Birr each of them stored a specific size, color, and other features, which makes identification task simple/easy. From a miner observation almost all visually impaired peoples identify the coin currencies simply by touching the specific tactile markings which mounted in each coin, but to identify the paper currencies they faced various challenges, the major one is they will be dependent on others to know the currency banknote value by asking a well-known question "how much is this?" for other individuals because there are no tangible patterns or other forms of marks on Ethiopian currency banknotes which enables a blind or visually impaired person to identify its value. Thus, to minimize this dependence, the blind peoples have come up with numerous ways of handling mechanisms such as measuring the currencies by putting between their fingers, store different denominations into a different pocket, organize the money in ascending order using the size of the currency, and measure the size of the banknote using readymade paper. The existing approaches are provide an amazing support for vision impaired community but the techniques didn't avoid the challenges at all because it is easy to a mix-up when they receive new banknotes, forget the specific pocket, and when the paper currency status especially when the status shift to worn. So, there is a gap which makes the community feeling dependent on others and makes them discomfort for their life. Thus, in order to fulfil this gap and minimize the discomfort and negative feeling challenges can best be solved through emerging technologies; we propose to develop a model using convolutional neural network to provide real-time Ethiopian currency recognition for the visually impaired person.

Convolutional Neural Network (CNN) is biologically inspired by Hubel and Wiesel's early work [6], which are designed to imitate the behaviour of a visual cortex using monkey. The task of keeping input images with the 2D structure inside CNNs is done by the two main layers such as Convolutional Layer and Pooling Layer. So, the neurons in a layer will only be connected to a small region of the layer before it which is similar to the biological visual cortex [7]. In the meantime even if the CNNs are showed a great performance on simple tasks like character recognition they fell out because of the growing of the problem complexity as well as the computing resource limitation until the second birthday of CNNs released by Krizhevsky, Sutskever, and Hinton by presenting the greatest image classification accuracy improvement on ImageNet [7].

There are several state-of-the-art approaches are existed for the area of object detection but these two techniques are nominated because of their well-known capability with regard to speed and accuracy. Thus, this study describes the only the two approaches such as Faster R-CNN and Single Shot Multi-Box Detector.

Faster R-CNN come up with the main aim of using shared convolutional layers for detection and for the generation of region proposal. Different researchers discovered that the feature maps produced by the networks of object detection networks also can be implemented to region proposals generation. Region proposal network is one portion of the Faster R-CNN network which is the fully convolution part that produces the feature proposals. In this research work the custom dataset which created by this study is

trained and evaluated by the pre-trained model which created by the integration of Faster R-CNN and Inception-V3 with the aim of detection and feature extraction respectively.

There is an option of a train a Faster R-CNN network either for the detection or for the generation of the region of interest which means simply for the case of feature extraction. The most common training procedure and description of this network is, first and foremost training of two distinct networks are performed then combination and fine-tuning techniques implemented on the two networks. In the case of fine-tuning, some layers are preserved fixed and some layers are trained one after the other [8]. The feature maps are generated by the shared fully convolutional layers from a single image that is received by the trained network as input. Thus, the region proposal network produces its region proposals as output after receiving the feature maps which generates previously. Finally, the feature maps together with the region proposals being an input for the last detection layers which include a region of interest pooling layer then classification [9]. The computational cost of region proposals is very low in the case of shared convolutional layers. In Convolutional Neural Network work out the region proposals with a small computation cost is to provide extra benefit. In the case of detection windows with a variety of sizes and shapes special anchor boxes are implemented instead of the pyramid of different filter sizes. The anchor is the essential idea of the sliding window [8].

Single Shot MultiBox Detector (SSD) [10] was presented as a method to detect objects from image or sequence of images (video) by using a single shot. SSD is one stage detection approach means in the case of region-based object detection approaches like Faster R-CNN the required tasks are done by using performing critical stages those are the region of interest generation and classify the generated regions in a distinct step, however in the case of SSD prediction of the region (bounding box) and classification tasks are done in a single shot. Performing the two critical operations in a single shot makes SSD a good nominee for real-time detections because of the speed which earned from its nature. This model has the capability to save computational time because the generation of the regional proposal is not used and there is no resampling of image segments. SSD has less accuracy when it compares with Faster R-CNN.

SSD handles objects of different sizes by using features maps from different convolutional layers as input to the classifier. This network produces a large number of regions (bounding boxes) with the scores of an object class in those boxes. Non-maximum suppression is used to eliminate boxes below a certain threshold so that only the boxes with higher confidence values proceed for classification. SSD architecture allows end-to-end training and improving the speed of the detector. This architecture does everything in one shot, thus, it is faster than other architectures but it lags the detection accuracy. SSD model is easy to train and simple to integrate with any systems which require an object detection module because as described earlier the SSD model encapsulates all the computation in a single network by removing the proposal generation and feature resampling stages. The authors evaluates the accuracy of this model on the PASCAL VOC, MS COCO, and ILSVRC datasets and obtained a comparable accuracy and much faster than others. The SSD object detection is composed of feature maps extraction, and convolution filters to detect objects parts.

The TensorFlow object detection API is nothing but a CNN based framework built on top of Tensorflow for object detection. The easiness to build, train, and setup capacity make it is popular among the researcher community. There are numerous pre-trained models are exist on it for object detection, those models are trained on the Common Objects in Context (COCO), KITTI, and Open Image datasets. These models can be used either for inference if anyone is interested in categories only in this dataset or for initializing our models when training on the custom dataset. Pre-trained models which trained on the COCO dataset with their speed of execution, accuracy and the type of output are listed in the TensorFlow model zoo.

2. Related Work

The Recognition system for Pakistani paper currency was proposed by Ahmed Ali and Mansoor [11]. The authors attempted to provide an accurate and intelligent recognition solution for Pakistani paper currency, which has different denominations and properties like size, color, and pattern variations by using image processing techniques. The outcome of the study was claimed to avoid the purchase of expensive recognition hardware and minimizing human effort. The overall construction of the proposed system was based on personal computer, scanner, and classifiers. In the research work, pre-processing techniques like noise removal, RGB to gray conversion, and gray to binary conversion were performed after the currency banknote image was captured from the scanner. One of the instance-based learning algorithms called k-nearest neighbors (KNN) was selected as an algorithm for the study and Euler number, height, width, aspect ratio, and area are identified as features or characteristics to decide the classification by the authors. These features are extracted from the training images and stored in the database as a MAT-file format. The KNN classifier works by staring at the classification of unknown instances which will be done by relating the unknown to the glorious in keeping with some distance/similarity perform. Generally, the proposed approach comprises four different procedures those are the acquisition of images, preprocessing, feature extraction, and classification. They have acquired a total of 100 images which means 20 from each currency notes (10, 20, 50, 100, 500, and 1000) with the help of a scanner.

A Fast-Mobile Money Reader was proposed to enable blind peoples to exchange United States currency banknote with no fear by using their smartphones [12]. Scale Invariant Feature Transform (SIFT) algorithm as a faster approach for feature classification is selected by the authors for the sake of the feature extraction process. In this work instead of picking different features or characteristics found on the currency bills using hand, they propose to follow a robust machine learning approach to train the data. Only the four United States currency banknotes ($1, $5, $10, and $20) are stored for testing and training. They scaled each of the images to 300 pixels by 300 pixels with a 200-pixel white border is added around the image. For the sake of experimentation, they create artificial training images by rotating the existing images through 90, 180, and 270 degrees, and scaling by 0.5 and 1.5 its original size. The proposed system taking continuous snapshots until at least 60% of the currency banknote exposed in front of the phone's camera.

Jegnaw Fentahun [13] was proposed to design and develop automatic recognition of Ethiopian paper currency by holding three major aims such as identify Ethiopian paper currency, identifies counterfeit currency banknote from genuine and categorize them in their denominations by using the main color, distribution of color, hue value and SURF as a discriminative feature of the currency banknotes. The proposed system consists of two major components such as currency denomination component which accept scanned image as input then processed through its sub-components pre-processing, feature extraction, and currency categorization to classifying the input currency banknote into one of the five denomination of Ethiopian currency (1 Birr, 5 Birr, 10 Birr, 50 Birr, and 100 Birr) and currency verification component which receives either the output of currency denomination component or an image which captured by the camera as an input to perform its responsibility which is to verify the specific paper currency whether it is genuine or counterfeit. To classify the banknotes the correlation coefficient-based template matching was implemented and to verify the originality of the banknote segmenting the thin golden vertical strip which is on the paper denomination of 50 and 100 Birr was done.

In another research work investigating the case of feature extraction SIFT, GLCM, color momentum, CNN and combination of SIFT, GLCM and color momentum techniques, and Feed-Forward ANN as a classifier for the design of Ethiopian paper currency recognition system [14] was proposed. The major image processing phases such as image acquisition by using a scanner, pre-processing which is responsible to remove noise, convert RGB to grayscale and normalized the size of the input banknote image to reduce the influence of the noise to the recognizer, feature extraction which is responsible to extract the descriptive features from the given banknotes by using Convolutional Neural Network (CNN) model and classification which was performed by using feed-forward artificial neural network classifier was followed by this research work. In this research work to train and test the proposed model a total of 2400 banknotes image was collected through the scanner and 70%, 15%, and 15% was used for training, validating and testing respectively

Almost all of the reviewed researches in this section was/will play a significant role in the problem area of Ethiopian currency recognition, however, they have their limitations/gaps such as many of the research works are trained and tested by using a small number of datasets, some of them require the static environment to perform recognition process, some of them are expensive or needs/requires special knowledge to use it. In addition, most of the technological solutions require blind or vision impaired individual take a picture of the full currency banknotes by presenting the banknotes in front of the camera which does not seem ideally and also practically easy for an individual who lives with a vision disability, some of them require a static background and distance, position or environment but this is also not practically applicable for blind or vision impaired peoples, some others didn't assume different protrusion (e.g. fingers) occurs in between the camera and currency banknote, and others didn't consider the input picture which takes by blind or visually impaired individuals may come up with different issues like folding, lighting conditions, only take the small piece of the banknote and so on. Thus developing a real-time currency recognition model using CNN classifier will have the potential to full fill the mentioned problems, by giving a better assist or uniform consultation to the blind or vision impaired community.

3. Objective Of The Paper

The general objective of this study is to develop a CNN based model that enables vision-impaired or blind peoples to recognize Ethiopian currency banknotes in a real-time context. To achieve the general objective of the study, the following specific objectives are formulated:

- In-depth literature study on background knowledge such as neural network algorithms, vision impaired and blindness, Ethiopian currency banknotes, different countries' approaches to recognize banknotes, and other related concepts.
- Datasets for training and testing the model will be collected.
- Develop the recognition model using a pre-trained CNN model.
- Evaluate the performance of the developed model
- Draw conclusions and forward recommendations.

4. Methods

The major aim of this study is to create a model which haves an ability to recognize Ethiopian currency banknotes thus to achieve this aim an in-depth literature review was done on background and other important knowledge's which includes visually disability, currency recognition, previous related works, and Convolutional Neural Network (CNN). Performing review activity on numerous previous researches is an important track cleaning activity because which can enable to share their idea to know what has been done and needs to be done in that particular problem area [15]. Related literature in the problem domain area was reviewed from various sources including books, journal articles, conference papers, reports, and the internet.

This section of the study presented the fundamental methods carried out in this research. The complete methodology of this study is consisting of five major phases such as data collection, data preparation as well as preprocessing, justification of framework as well as pre-trained model selection; re-train the pre-trained models, and evaluation of the re-trained model.

4.1. Data Collection

In the data collection phase the five different values of Ethiopian banknotes from new up to worn was gathered from different banks and then recorded a lot of custom video dataset which shows the real-life scenario of the vision-impaired community. All the data is collected by putting Samsung Galaxy A10 mobile phone beside vision-impaired individual ears for 10 seconds. To obtain the presence of the real-world challenge, all the videos were attempted to be recorded within uncontrolled environments. In the case of recording the videos, the instances of different protrusion including fingers and shadows and also the various lighting conditions including moonlight, daylight, and artificial light were attempted to be included. All the videos were recorded by considering the moonlight, daylight, artificial light, and the transition of one lighting condition. To reduce the time-consuming and tedious routine task which is

manual labeling, only 1700 images of each banknote are selected so that a total of 8,500 images are selected as a dataset. Ideally, the data should look as close as possible for the real-world situation, that is why the custom video was recorded which was collected at the real-world situation of the vision-impaired community. The images presented in the dataset have similar width and height which is 256 × 256 × 3 (Width x Height x Channel).

4.2. Data Preprocessing and Data Preparation

The images presented in the dataset have similar width and height which is 1080 × 1080 then resizes the images to the target size which is 256 × 256 based on the work described in [7] was performed.

In the data preparation task image labeling using bounding box and data separation was performed. Image labeling is an essential task for the supervised machine learning techniques because the output result of the model is determined by the labels we feed the model in the training stage. An open-source graphical labeling tool "LabelImg" and the ground truth bounding box technique were used to label the dataset which holds a collection of the five Ethiopian banknotes as shown in Fig. 1. For each image, the labeled information was saved as an eXtensible Markup Language (XML) file in PASCAL VOC format, the format compatible for CNN pre-trained models. The XML file stores important information such as image name, folder name, size of the image as (width, height, and depth), each bounding box coordinates as (xmin, ymin, xmax, ymax), class name, and others. We attempt to be rational in the case of include or discard banknotes from labeling. Banknotes that were fully or partially visible and recognizable were included, whereas banknotes that were unrecognizable because of size or position were excluded.

After the completion of image labeling task randomly splitting the dataset into training and test dataset which is used for train the model and is used for evaluating the trained model respectively was performed before the actual training begins. The training set contains images with their corresponding XML file generated by the image labeling task and similarly, the test directory contains images with their corresponding XML files which are used to evaluate the trained model. The data splitting task was done by adopting the ratio of 9:1 (90% training and 10% testing). This means that 90% from the dataset which is 7650 images are used to train the model, and 10% from the dataset which is 850 images are used to test the trained model.

4.3. Pre-Trained Model Selection

To train the Ethiopian currency banknote dataset two pre-trained models are selected those are SSD with MobileNet v1 and Faster R-CNN with Inception v2. These pre-trained models are nominated by using the literature review described in the Introduction part of this study. To select pre-trained models from the existing one's speed, accuracy, detection approaches, and problem domains are mainly takes into consideration.

In the case of SSD with MobileNet v1, the extraction of features is performed using the Mobile Network (MobileNet) and the detection task performed by SSD. The reason behind selecting this pre-trained model

is by considering the problem domain means blind peoples need to know the required information as quickly as possible, speed, and its lightweight nature capacity to perform object detection on a device with low computational power such as a smartphone or Raspberry Pi. Mobile network (MobileNet) is a lightweight deep neural network that is efficient for mobile and embedded devices. The principle behind this architecture is the division of the standard convolutional filter into depth-wise convolution and point-wise convolution filter [16].

In the case of Faster R-CNN with Inception v2, the feature extraction task is performed by the Inception algorithm and R-CNN set of rules are applied for the detection task. The reason behind selecting this pre-trained model is by considering the importance of accuracy and detection nature of the model.

4.4. Environment Setup

Train CNNs from zero (scratch) requires a lot of data and high-performance computing powered hardware. The training and evaluation of the model are implemented using a Tensorflow object detection API which is configured in Windows 10 environment.

All the training, preprocessing, and experimental tasks were done on HP ProBook 450 G3 laptop with Intel(R) Core(TM) i7-6500U CPU @ 2.50GHZ processor and 16 GB RAM having Windows 10 operating system and HP Pavilion power 15 laptop with Intel(R) Core(TM) i5 7th generation CPU @ 2.1 GHZ processor, 8 GB RAM, and 2 GB AMD Radeon graphics having Windows 10 operating system. It used either the front-facing webcam of the laptop or external webcam to demonstrate real-time recognition. All the codes presented in this study are written by using Python programming languages. The idea behind using python as programming languages is its capabilities of easy to learn and the availability of matured resources to perform computer vision and real-time techniques [17].

Tensorflow is a completely open-source framework developed by Google in 2015 by holding the ambition of being a playing place for machine learning. This framework is written in C++, Python, and Cuda. There are so many reasons behind the implementation of this framework some of them are speed for computation which makes Tensorflow appropriate for the practical industry and academic research, expressive architecture, matured online support, and resource availability [18].

Open Source Computer Vision Library (OpenCV) is a completely free and open-source software library for the area of machine learning specifically computer vision. OpenCV has the capability to support numerous programming languages such as Python, C++, Java, etc. and also platform-independent means working on a variety of platforms including Windows, Linux, Android, and iOS [19]. Thus, it can be easily accessed and used as a tool. Based on the above reasons, the real-time Ethiopian currency banknote recognition model is to use OpenCV by integrating with other machine learning libraries and tools.

4.5. Train the Pre-Trained Model

Transfer learning is applied by using the pre-trained models SSD MobileNet and Faster R-CNN Inception to train the prepared Ethiopian currency custom dataset. As mentioned in the framework selection section, Tensorflow Object Detection API was used to train the dataset. Thus before beginning the actual training process numerous must-to-do steps must be done such as XML to CSV conversion, label map creation, Tensorflow record generation, and training pipeline configuration.

4.5.1. XML to CSV Conversion

As mentioned before, training and testing datasets hold XML files which are generated by data labeling task with the name of its corresponding image (.jpeg) file. Each XML file contains important values such as image file name, width, height, category/class name, the four corners points of the bounding box (xmin, xmax, ymin, ymax), and others. Thus, the XML files were converted into two (test_labels.csv and test_labels.csv) CSV files which hold essential information for all images in the train and test dataset by editing the *xml_to_csv.py* file which comes together with the API. These two test_ labels.csv and train_ labels.csv files provide tables of 858 and 7655 rows respectively since some of the images contain more than one class.

4.5.2. Creating a Label Maps

Map every label into an integer value (ID) because of the training and detection processes. Thus, the label map file by the name *label_map.pbtxt* was created with the five classes and their integer representation as shown in Fig. 2.

4.5.3. Tensorflow Record (TFRecord) Generation

Object Detection API requires all the labeled training data to be in the TFRecord file format. Thus, the CSV file and the training images were converted to a TFRecord file by adopting and *generate_tfrecord.py* file which comes together with the API and modifying the row labels to by One Birr, Five Birr, Ten Birr, Fifty Birr, and One Hundred Birr as shown in Fig. 3.

4.5.4. Pipeline Configurations and Fine-Tuning

Transfer learning is selected rather than creating the model from scratch. Before triggering the actual training process configuring the required pieces of information for the object detection training pipeline comes first. Thus, for both pre-trained models by using the provided configuration files (*ssd_mobilenet_v2_coco.config and faster_rcnn_inception_v2_pets.config*) as the basis and then some modifications have been made to the default configuration file. For both cases, the number of classes in our cases we have five classes, the locations of checkpoint file which delivered by Tensorflow, train and test TFrecord file those created for the training and test datasets respectively, the label map file which holds the target classes/categories, and other important parameters are required by the configuration file. Accordingly, all the required modifications and the definition of which model and what parameters to be

used for training were done. There are also various important adopted as well as customized parameters that exist for two of the selected pre-trained models as shown in Table 1.

Table 1
Important parameters for Faster R-CNN with Inception v2 pre-trained model

Name of Parameter	Value(s) Faster R-CNN	Value(s) SDD
Number of Layers	6	-
Learning Rate	0.0002 is the initial learning rate by considering shift to 0.00002 after 10,000 steps, and then to 0.000002 after 20,000 steps	0.004 is the initial rate by way of considering 0.95 as a decay factor after 800,720 iterations
Aspect Ratios	0.5, 1.0, and 2.0	1.0, 2.0, 0.5, 3.0, and 0.3333
Batch Size	1	5
Score Converter	SIGMOID	SOFTMAX

4.5.5. Train the Models

After assuring the above important steps are done correctly, the training can be launched. The step count and loss value in each step can be display on the screen, as shown in Fig. 4 and Fig. 5. The sum of classification and localization loss is displayed as a total loss value. From Fig. 4 and Fig. 5 it can be noted that the classification loss starts at a high value and gradually decreases as the algorithm learns as the iterations progress. This process was monitored and determined whether the model is ready to be deployed or it needs more training time or any other changes by using Tensorboard. The learning curve of the model was visualized by observing the "Total/Loss" graph. The training takes 48 hours and manually stooped at 40,000 iterations for Faster R-CNN with Inception v2 as well 5 days and manually stooped at 20,075 for SSD with MobileNet because of time constraint and mainly the value of total loss decreased and being near for 0. Localization loss dealing with how worthy properly classified objects are localized and classification loss dealing with how worthy objects are classified In general, lesser loss implies better training of the model.

5. Evaluation And Result

In this research work, various performance evaluation matrixes are used in order to evaluate the trained model, including simple accuracy, precision, recall, F1-score, and confusion matrix. All the performance matrixes are based on **True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN)** values.

The classification accuracy is determined by the total number of correctly classified currencies divided by the total number of validation as shown in Eq. (4 – 1). In the case of Faster R-CNN Inception, the model obtained an average of 91.8% accuracy for detection as shown in Table 3, and in the case of SSD MobileNet the accuracy obtained an average of 79.4% accuracy for detection as shown in Table 4

$$\text{Accuracy} = \frac{\text{Correctly Classified currency}}{\text{Total number of validation}} \times 100 \qquad (5\text{-}1)$$

Table 2
Accuracy of Ethiopian currencies using Faster R-CNN Inception v2 model

Value	Total Test	Correctly Classification	accuracy
100	100	92	92.0%
50	100	96	96.0%
10	100	90	90.7%
5	100	96	97.0%
1	100	84	84.0%
Average			91.8%

Table 3
Accuracy of Ethiopian currencies using SSD MobileNet model

Value	Total Test	Correctly Classification	accuracy
100	100	94	94.0%
50	100	81	81.0%
10	100	76	76.0%
5	100	92	92.0%
1	100	54	54.0%
Average			79.4%

5.1. Confusion Matrix

A more detailed wrong and correct classification for each Ethiopian currency class are showed by using the confusion matrix because it has the capability to illustrate the correct and incorrect classification percentage for each individual class in a tabular way.

Table 4
Confusion matrix for Faster R-CNN Inception Model

Class	One Birr	Five Birr	Ten Birr	Fifty Birr	Hundred Birr
One Birr	84	11	3	0	2
Five Birr	0	97	0	0	3
Ten Birr	2	6	90	1	1
Fifty Birr	0	4	0	96	0
One Hundred Birr	2	6	0	0	92

Table 5
Confusion matrix for SSD MobileNet model

Class	One Birr	Five Birr	Ten Birr	Fifty Birr	Hundred Birr
One Birr	54	15	0	0	31
Five Birr	0	92	0	0	8
Ten Birr	0	6	76	0	18
Fifty Birr	0	5	0	81	14
One Hundred Birr	0	6	0	0	94

The total number of FN's for a certain class means the sum of values in the corresponding row except for the TP value; FP's for a certain class means the sum of values in the corresponding columns excluding the TP value, and TN's for individual class means the sum of all columns and rows except that class column and row. The overall extracted values and the calculated value of precision, recall, and F1-Score for each individual class and each model are summarized as well as shown in Table 7.

Table 6
Evaluation summary for Faster R-CNN Inception model

Class	TP	TN	FP	FN	Precision	Recall	F1-Score
One Birr	84	396	16	4	0.84	0.95	0.892
Five Birr	97	373	3	27	0.97	0.78	0.865
Ten Birr	90	397	10	3	0.9	0.96	0.929
Fifty Birr	96	399	4	1	0.96	0.99	0.975
One Hundred Birr	92	394	8	6	0.92	0.94	0.93
Average/Total					0.918	0.924	0.918

Table 7
Table 7: Evaluation summary for SSD MobileNet model

Class	TP	TN	FP	FN	Precision	Recall	F1-Score
One Birr	54	400	46	0	0.54	1.0	0.701
Five Birr	92	368	8	32	0.92	0.74	0.820
Ten Birr	76	400	24	0	0.76	1.0	0.864
Fifty Birr	81	400	19	0	0.81	1.0	0.895
One Hundred Birr	94	329	6	6	0.94	0.94	0.94
Average/Total					0.794	0.936	0.844

5.2. Discussion

With regards to the SSD MobileNet model, the summary evaluation values are presented in *Table 8.* Thus from observing the table the F1-score value of One Birr class is relatively low 70.1% means there is a middling capability in order to classify One Birr input into its correct class/category. However the F1-score value of the rest four classes Five Birr, Ten Birr, Fifty Birr, and One Hundred Birr are good 82.0%, 86.4%, 89.5%, and 94.0% means that the values for the false positives and the false negatives are very low, which indicates that most of the time the model classifying the input into its correct category. Correspondingly, the accuracy of the models has been calculated and obtained 54.0%, 92.0%, 76.0%, 81.0%, and 94.0% respectively for One Birr, Five Birr, Ten Birr, Fifty Birr, and One Hundred Birr classes. In addition to that, the average classification accuracy is also determined from Table 4 − 2 which is 79.4%. One Hundred Birr class obtained the highest accuracy 94%, precision 94%, and F1 score 94% for Ethiopian currency classification.

On the other hand, the Faster R-CNN Inception model evaluation values summary is presented in Table 7. This model can obtain relatively higher average accuracy, precision, and F1-score values when compared to the SDD MobileNet model as shown in Table 7. Additionally, the model obtained F1-Score values of 89.2%, 86.5%, 92.9%, 97.5%, and 93% for One Birr, Five Birr, Ten Birr, Fifty Birr, and One Hundred Birr classes respectively means simply most of the time the model classifying the input values into its correct category. The founded accuracy values for the five classes 92.0%, 96.0%, 90.7%, 97.0%, and 84.0% also indicate that the classification task performs relatively with good accuracy. In this model, the false positive and false negative values are relatively low.

6. Conclusions

In this study, the researchers attempted to develop and tested models that provide a capability to detect Ethiopian currency banknotes in a real-time scenario by using transfer learning. The models are able to

classify Ethiopian currencies into their respective categories. In-depth reviews of currency recognition studies are performed. Pre-trained Faster R-CNN Inception and SSD MobileNet models are used and also both the models are trained by using a custom dataset and evaluated in the real-time scenario. In this research work, both single-stage and two-stage detection approaches are applied. The detection process takes a frame from a live video as an input and attempts to classify it as One Birr, Five Birr, Ten Birr, Fifty Birr, or One Hundred Birr.

Even if there are a few pieces of research conducted in the area of Ethiopian currency recognition but their domain is very far from the domain of this research work thus, as being the first Ethiopian currency recognition research in the domain of blindness and vision impairment, the evaluation result of the model can be considered as one that has a good performance. Both the models are evaluated by using numerous status means from new up to worn Ethiopian currencies. The classification accuracy of the models is evaluated by using 500 currencies provided by using real-time video. In this research work, the Faster R-CNN Inception model obtained an average accuracy of 91.8%, average precision of 91.8%, an average recall of 92.8%, and an average F1-score of 91.8% with a real-time video validation. Likewise, the SSD MobileNet model obtained an average accuracy of 79.4%, average precision of 79.4%, an average recall of 93.6%, and an average F1-score of 84.4% with a real-time video validation.

This study has grasped the general objective over the development of the model using CNNs, transfer learning, and Object Detection API framework of Tensorflow with the preferred outcomes. It is mandatory to mention about pre-trained models are initially trained with dissimilar datasets when compared with the dataset employed for training in this research work. The default pipeline configuration and Hyper-parameters are adopted then work successfully with slight modification after the implementation of miner investigation on it. Even if there are some problematic issues such as the wrong classification and failure to detect observed in the model evaluation phase the numerous strategies have worked as expected.

Abbreviations

- **ANN =** Artificial Neural Network
- **CNN** = Convolutional Neural Network
- **OpenCV** = Open Computer Vision
- **PASCAL VOC =** Pascal Visual Object Classes
- **SIFT =** Scale – Invariant Feature Transform
- **SSD =** Single Shot MultiBox Detector
- **SURF =** Speed Up Robust Features
- **TFRecord =** Tensorflow Record

Declarations

Availability of Data

The dataset is available online https://data.mendeley.com/datasets/r76dwc7nnw/1

Competing Interests

The authors confirm that there are no competing interests. The article is not under any review process and not in the subject of any other submission.

Consent for publication

The authors consent for publication.

Acknowledgements

Authors' contributions

The main contributions of this research work are listed as follows:

- Real-life situation Ethiopian currency banknote dataset is our main contribution.
- The study exhibited transfer learning techniques and approaches in order to develop a real-time Ethiopian paper currency, recognition model.
- The study shows the basic challenges to develop real-time Ethiopian paper currency recognition model for the research domain area and the possible methodologies to solve those challenges.

Funding

References

1. World Health Organization, "GLOBAL DATA ON VISUAL IMPAIRMENTS 2010," Geneva, Switzerland, 2010.
2. Yemane Berhane and et al. , "Prevalence and causes of blindness and Low Vision in Ethiopia," Ethiopian Journal of Health Development, vol. 21, no. 3, pp. 204-210, April 2007.

3. Arvind Sharma, "COMPUTER VISION GUIDED NAVIGATION SYSTEM FOR VISUALLY IMPAIRED," ResearchGate, p. 13, June 2016.

4. World Health Organization. (2019, October) World Health Organization. [Online]. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

5. Andrii Kyrychok , "THE OVERVIEW OF INVESTIGATION IN THE FIELD OF BANKNOTE DESIGN FOR VISUALLY IMPAIRED PEOPLE," National Technical University of Ukraine, Kyiv, Reports on research 3, 2018.

6. D. H. Hubel and T. N. Wiesel , "Receptive fields and functional architecture of monkey striate cortex," The Journal of Physiology, vol. 195, no. 1, pp. 215 - 243, 1968.

7. A. Krizhevsky , G. E. Hinton, and I. Sutskever , "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105, 2012.

8. Shaoqing Ren , Kaiming He , Ross Girshick , and Jian Sun , "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," vol. V3, no. arxiv: 1506.01497, January 2016.

9. Sebastian Raschka , Joshua Patterson , and Corey Nolet , "Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence," MDPI information, vol. 11, pp. 1-48, April 2020.

10. Wei Liu , Dragomir Anguelov , Dumitru Erhan , and Christian Szegedy , SSD: Single Shot MultiBox Detector.

11. Ahmed Ali and Manzoor, "Recognition System for Pakistani Paper Currency," World Applied Sciences Journal, vol. 28, pp. 2069-2075, 2013.

12. Nektarios Paisios, Alexander Rubinsteyn, and Lakshminarayanan Subramanian, Exchanging Cash with no Fear: A Fast Mobile Money Reader for the Blind.

13. Jegnaw Fantahun Zeggeye and Yaregal Assabie , "Automatic Recognition and Counterfeit Detection of Ethiopian Paper Currency," International Journal of Image, Graphics and Signal Processing(IJIGSP), vol. 8, pp. 28-36, 2016.

14. Asfaw Sheferaw and Million Meshesha , "Ethiopian Paper Currency Recognition System: An Optimal Feature Extraction," IEEE-SEM, vol. 7, no. 8, pp. 130 - 137, August 2019.

15. Bruce W. Tuckman, Conducting Educational Research, Fifth ed., Angela Williams Urquhart, Ed. Ohio State, USA: Earl McPeek.

16. Heikki Huttunen , CAMERA BASED OBJECT DETECTION FOR INDOOR SCENES, 2018.

17. Javier Andreu Perez, Fani Deligianni, Daniele Ravi, and Guang-Zhong Yang, Artificial Intelligence and Robotics. UK: UK-RAS Network, 2016.

18. Aditya Sharma. (2020, June) https://www.datacamp.com. [Online]. https://www.datacamp.com/community/tutorials/cnn-tensorflow-python

19. Alexander Mordvintsev k and Abid, OpenCV-Python Tutorials Documentation., 2017.

20. Douwe Osinga , Deep Learning Cookbook, 1st ed., Rachel Roumeliotis and Jeff Bleiel , Eds. CA, USA: O'Reilly Media, Inc., 2018.

Figures

Figure 1

Using the LabelImg and ground truth bounding box to label the dataset

Figure 1

Using the LabelImg and ground truth bounding box to label the dataset

```
1    item {
2      id: 1
3      name: 'One Birr'
4    }
5
6    item {
7      id: 2
8      name: 'Five Birr'
9    }
10
11   item {
12     id: 3
13     name: 'Ten Birr'
14   }
15
16   item {
17     id: 4
18     name: 'Fifty Birr'
19   }
20
21   item {
22     id: 5
23     name: 'One Hundred Birr'
24   }
25
```

Figure 2

label map file for Ethiopian currency dataset

```
1   item {
2     id: 1
3     name: 'One Birr'
4   }
5
6   item {
7     id: 2
8     name: 'Five Birr'
9   }
10
11  item {
12    id: 3
13    name: 'Ten Birr'
14  }
15
16  item {
17    id: 4
18    name: 'Fifty Birr'
19  }
20
21  item {
22    id: 5
23    name: 'One Hundred Birr'
24  }
25
```

Figure 2

label map file for Ethiopian currency dataset

```
29
30  # TO-DO replace this with label map
31  def class_text_to_int(row_label):
32      if row_label == 'One Birr':
33          return 1
34      elif row_label == 'Five Birr':
35          return 2
36      elif row_label == 'Ten Birr':
37          return 3
38      elif row_label == 'Fifty Birr':
39          return 4
40      elif row_label == 'One Hundred Birr':
41          return 5
42      else:
43          None
44
45
```

Figure 3

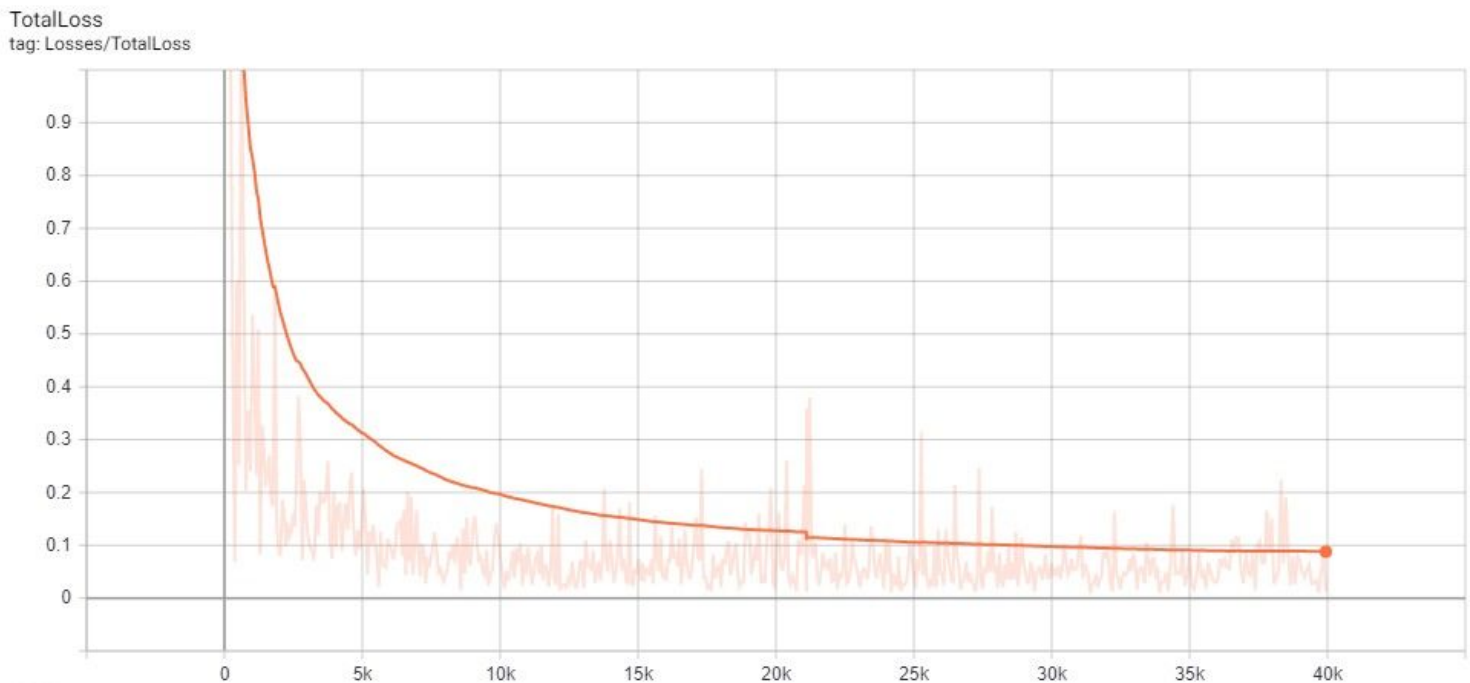Modified row labels with its corresponding integer value

```
29
30    # TO-DO replace this with label map
31    def class_text_to_int(row_label):
32        if row_label == 'One Birr':
33            return 1
34        elif row_label == 'Five Birr':
35            return 2
36        elif row_label == 'Ten Birr':
37            return 3
38        elif row_label == 'Fifty Birr':
39            return 4
40        elif row_label == 'One Hundred Birr':
41            return 5
42        else:
43            None
44
45
```

Figure 3

Modified row labels with its corresponding integer value

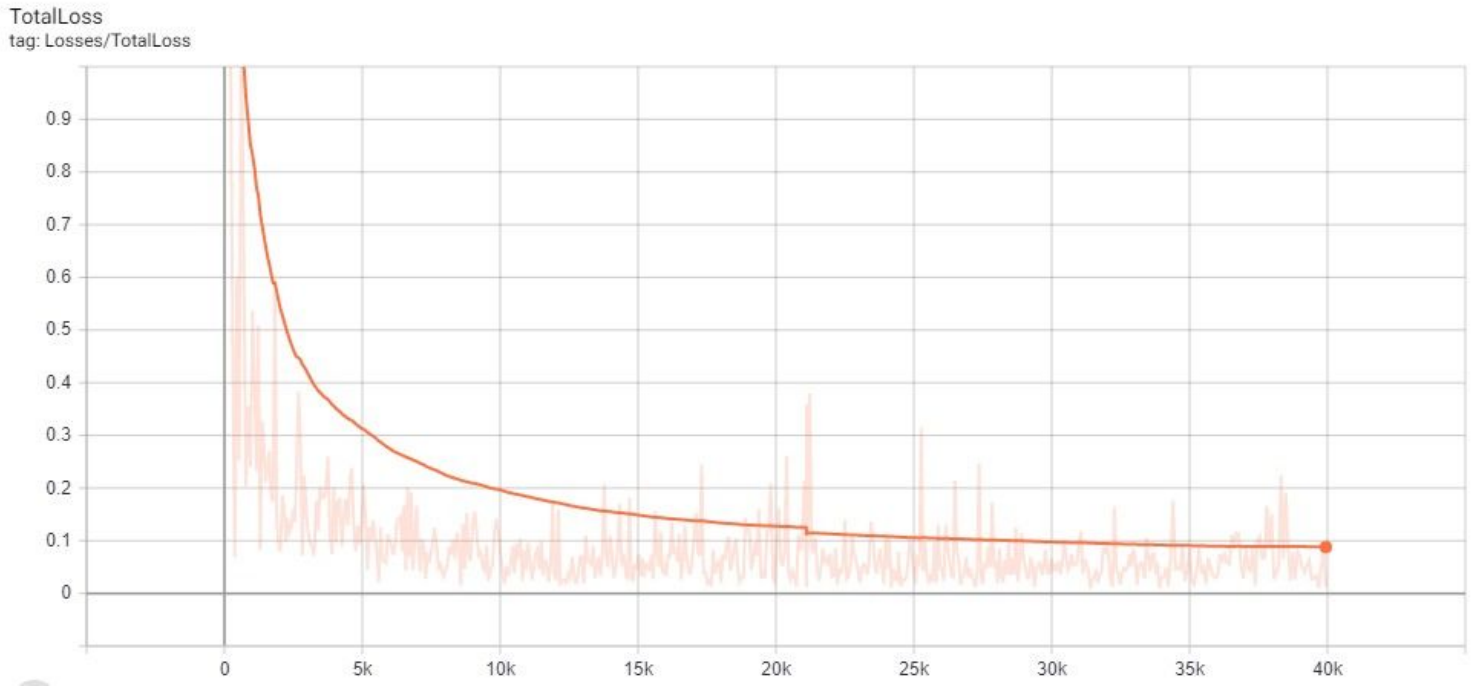**Figure 4**

Total Loss graph for Faster R-CNN model

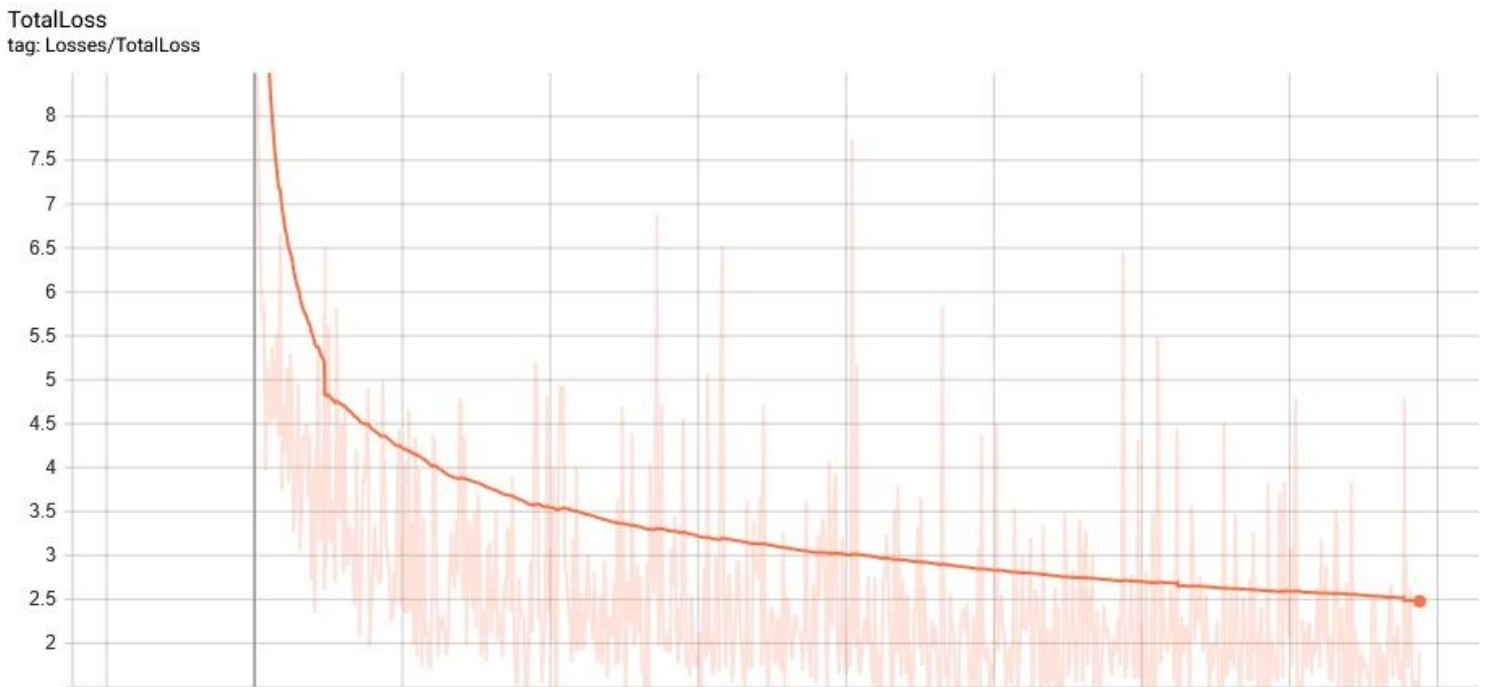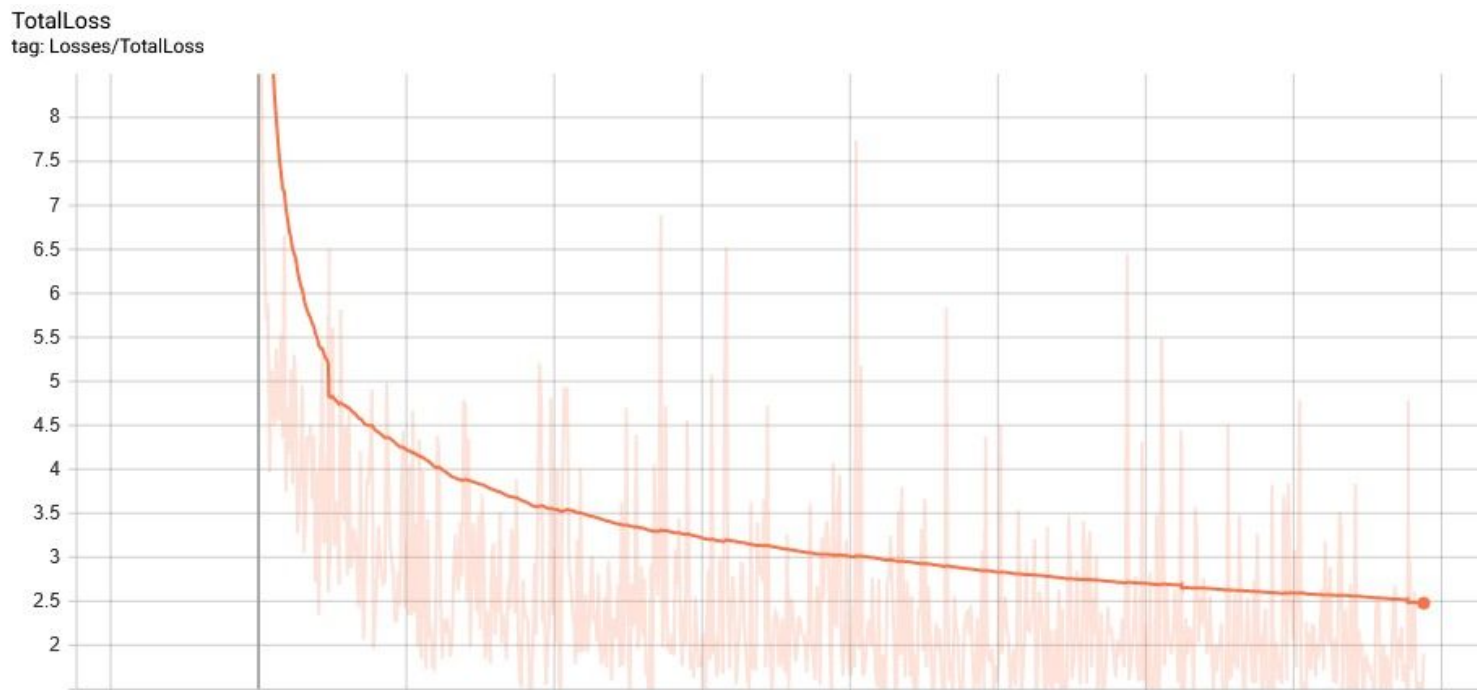**Figure 4**

Total Loss graph for Faster R-CNN model

Figure 5

Total Loss graph for SSD with MobileNet model

TotalLoss
tag: Losses/TotalLoss

Figure 5

Total Loss graph for SSD with MobileNet model