Noname manuscript No.
(will be inserted by the editor)

Evaluation of Deep Learning models on UV ink : a Fake Money detection scheme with RPN

Anjir Ahmed Chowdhury · Argho Das · Debajyoti Karmaker

Received: date / Accepted: date

Abstract As soon as coins or money was invented, there were people trying to make counterfeits. Counterfeit money is fake money that is produced without the permission of the state or government, usually to imitate the currency and deceive the intended recipient. In Bangladesh, this is a significant problem and the problem is becoming more and more phenomenon as the days are passing by. Today's modern bank notes have several security features that makes easier to identify fake notes. One of the security features is the use of UV ink. Bank notes deliberately put random flecks of color scattered all over the surface of the money - which acts as a extra layer of protection against counterfeiters. We propose an automatic authentication model for identifying counterfeit money based on these random flecks of color which is visible under UV light. To obtain a benchmark result, existing object detection pre-trained models were used, followed by MobileNet, Inception, ResNet50, ResNet101, and Inception-ResNet architectures. After that, using the Region Proposal Network (RPN) method with Convolutional Neural Network (CNN) based classification the optimal model was proposed. The proposed model had a *96.3* percent accuracy. It is critical to reduce the circulation of counterfeit money in a country's economy to stop inflation. This study will aid in the detection of counterfeit money and, hopefully, reduce its spread.

Anjir Ahmed Chowdhury
Primeasia University
E-mail: anjir.ahmed@primeasia.edu.bd

Argho Das
American International University-Bangladesh
E-mail: arghodas555@gmail.com

Debajyoti Karmaker
American International University-Bangladesh
E-mail: d.karmaker@aiub.edu

Keywords Deep learning · Computer Vision · Object detection · RPN · CNN

1 Introduction

With all the technological improvements of printers and scanners, the threat of forgery documents has also gone up. Amongst various forgery documents, counterfeiting of banknotes has come to a serious issue, and many countries' economy is badly impacted by it [1]. Thus an authentication system is for banknotes is a prime concern in today's world. To the best of our knowledge, forensic signature or handwriting verification has been studied significantly but, unfortunately, very limited studies have been conducted for banknotes verification. Several embedded features like artwork, security thread, watermark, UV ink, etc. are used to prevent counterfeiting banknotes. The bank staffs are specially trained to detect counterfeit notes and so far they do it manually but this process is time-consuming. Moreover, this manual process is unattractive especially when it comes to a large number of documents. Thus the need for automated Computer vision and Deep Learning can play a vital role in reducing these cases.

Counterfeit money is money created without the State's or government's legal permission, typically to mimic the currency and deceive its intended recipient. Producing or using counterfeit money is considered a type of fraud or forgery, and it is punishable by law. During the early years of independence, Bangladesh inherited Pakistan's Central Bank's monetary management legacy. The initial situation was volatile, but the government quickly stabilized the position by establishing Bangladesh Bank in 1972. Following that, the availability of counterfeit money in Bangladesh has rapidly increased. Even the Bangladesh Bank detects two to three fake notes per million takas (Bangladeshi currency), and counterfeit notes are often found during the initial screening process [2]. Counterfeit money causes a decrease in the value of real money; an increase in prices, which causes inflation due to more money being circulated in the economy - an unauthorized artificial increase in the money supply; a decrease in the acceptability of paper money; and losses, when traders are not reimbursed for counterfeit money detected by banks. The country's economy is rapidly expanding, leading to the spread of fake currencies. Currently, it is estimated that 0.03 percent of the currency in circulation in the world is counterfeit [3]. In Dhaka, the Rapid Action Battalion (RAB) busted a counterfeit currency factory and confiscated more than 10 million taka in counterfeit notes [4]. The RAB also seized 40 million taka in fake notes, most of which were Tk 1,000 bills. A total of $4 million in counterfeit Indian currency was also discovered [5]. However, as the economy grows, the group's size will increase. Computer vision and Deep Learning can play a vital role in reducing these cases. All the codes are available at https://github.com/anzir29/FakeMoney_detection_scheme.

2 Related Works

Object detection is among the most fundamental and challenging problems in computer vision. It aims to identify object instances in natural images from a wide range of predefined categories [6]. Object detection techniques are analyzed [7] to detect objects in real-time on any system running the proposed model in any area. The proposed method employs multi-layer convolutional neural networks to create a multi-layer system model that can classify given objects into specified classes. This paper [8] presents a balanced feature fusion SSD (BFSSD) algorithm to improve the efficiency of SSDs. Pascal VOC2007 and VOC2012 trainval datasets are used to train their model, which is then evaluated on Pascal VOC2007 test datasets. [9], the SSD discretized the output space of bounding boxes into a collection of default boxes per function map position at various aspect ratios and scales. SSD has comparable accuracy to methods that use a different object proposal stage and is much faster, thus offering a coherent structure for both training and inference, according to experimental results on the PASCAL VOC, MS COCO, and ILSVRC datasets. The models described above were explored and used as a benchmark to assess the dataset developed in this study. The dataset used in this research is being tested by their pre-existing models. Furthermore, paper [10] presents a residual learning system for training networks far deeper than previously utilized networks. The authors tried residual nets with a depth of up to 152 layers on the ImageNet dataset, which is 8 times deeper than VGG nets but still has lower complexity. The image slice samples are passed into a Squeeze and Excitation ResNet model based on Convolutional Neural Networks in this research [11] to automatically classify brain tumors from MRI data (CNN). This paper [12] proposes a novel deep neural network architecture for microscopic image classification based on transfer learning. In this study, three separate deep CNNs were used: Inception-v3, Resnet152, and Inception-Resnet-v2. For image recognition on mobile devices and embedded devices with limited resources and ARM-based CPUs, this study [13] uses the Convolutional Neural Networks (CNN) approach with a 28 layer MobileNet architecture and works with a moderate amount of training data. The architectures described above were studied to create the proposed architecture used in this research and to examine different model complexities. In addition, a Region Proposal Network (RPN) is a complete convolutional network that simultaneously predicts entity bounds and scores at each location. The RPN has been fully trained to produce high-quality area proposals from start to finish. Using the recently common terminology of neural networks with "attention" mechanisms, RPN and Fast R-CNN were combined into a single network in this paper by sharing their convolutional features. The RPN portion tells the unified network where to look [14]. In [15], a multi-scale and multi-tasking area proposal approach is used to detect small objects in the PASCAL VOC dataset effectively. It achieves state-of-the-art object detection accuracy. This research [16] compares a novel proposal generation system called Enhanced Region Proposal Network (ERPN) to five object detection methods on the PASCAL VOC and COCO data sets. Using a new technique for anchor generation and training the network with both bounding boxes and category labels of the objects, this research [17] implemented the Class Aware Region Proposal Network (CARPN) to generate high-quality region proposals. On the other hand, Selective Search is an area proposal algorithm. It's made to be fast and has a high recall rate. It works by computing hierarchical groupings of similar regions based on color, texture, scale, and shape similarity. The power of both an exhaustive search and segmentation are combined in this paper [18] by diversifying their search and using several complementary image partitionings to deal with as many image conditions as possible. This paper [19] explores and expands an alternate approach that partitions the dataset into topic-related shards based on document similarity and scans only a few shards estimated to contain appropriate documents for the query. The proposed shard creation techniques are scalable, effective, and self-sufficient, and they produce topic-based shards with low size variance and high density of relevant documents. RPN with the selective search was used in the first stage of this study. The studies listed above aided in gathering information for the model's initial stage. In the next stage, the most popular deep learning method is the convolutional neural network (CNN), a multilayer neural network in the next stage. Since it is good at dealing with image classification and recognition problems and has improved the accuracy of many machine learning tasks, the convolution neural network (CNN) built in recent years has been widely used in image processing. It's evolved into a robust and widely used deep learning model [20]. The datasets used in this study [21] were ImageNet, CIFAR10, and CIFAR100, and the study focused on evaluating the performance of three standard networks: Alex Net, GoogLeNet, and ResNet50. According to the study, GoogLeNet and ResNet50 can recognize objects with greater precision. Alex Net. The aim of this paper [22] was to conduct classification experiments for the detected object obtained from traffic detectors using CNN and the HOG descriptor.

The performances of HOG SVM and CNN, when applied to the collected RVFTe-10 data, were excellent, with 99.9% and 99.5 percent, respectively. In the second stage of this study, the various types of CNN architectures discussed earlier were studied to develop the proposed CNN architecture.

As a result, this study aims to develop a model that can reduce the spread of counterfeit money. In this research, a custom dataset was used by scanning the taka under ultraviolet (UV) light to make the small thread visible. UV detection of counterfeit currency using thread has been around since 1976 and has proven to be very successful [23]. This thread separates the fake money from the real ones. Existing deep learning pre-build models were first put to the test to obtain a benchmark result. In this study, MobileNet, Inception, ResNet50, ResNet101, and Inception-ResNet architectures were used. The architectures generated mediocre results. After that, an optimal result was obtained using the Region Proposal Network (RPN) method with Convolutional Neural Network (CNN) based classification. The method achieved satisfactory results.

Deep neural networks for object detection are a well-established area of research. Although many object detection models have been studied over the years, single-shot and two-shot object detection are thought to be one of the best in speed vs. accuracy.

2.1 Single-shot Object detection

The model's goal in object detection tasks is to draw tight bounding boxes around desired classes in the picture, along with object labeling. Region proposal is not carried out by single-shot detection. It provides both final localization and content prediction at the same time. The famous single-shot approaches are the single-shot multibox detector (SSD) and YOLO [24].

The localization is computed by the SSD architecture in a single network pass. The SSD algorithm tiles a grid of anchors in space, size, and aspect ratio boxes onto the image. Unlike two-shot methods, the model produces a vector of predictions for each box in a subsequent network pass. A per-class confidence score, localization offset, and resizing are all stored in this vector. During preparation, SSD does not see enough small instances of each class. Zoom augmentation, which reduces or increases the size of the training videos, aids in generalization. SSD, on the other hand, is better at predicting large objects than FasterRCNN. The table 5 shows the SSD models used in this research, along with the model's speed and mean average precision (Coco mAP).

2.2 Two-shot Object detection

The two-shot detection model has two stages: region proposal, followed by region classification and position prediction refinement. For two-shot versions, faster-RCNN variants are the most common option. Faster-R-CNN detection occurs in two phases. The area proposal stage is the first phase. A small, fully connected network slides over the feature layer to predict class-agnostic box proposals with respect to a grid of anchors tiled in space, size, and aspect ratio after images are processed by a feature extractor.

These box proposals are used in the second stage to crop features from the intermediate feature map already computed in the first stage. The proposed boxes are fed into the rest of the function extractor, which is equipped with prediction and regression heads, where each proposal's class and class-specific box refinement is measured.The Faster-RCNN models used in this study are listed in the table, and their speed and mean average precision (Coco mAP).

3 Dataset

3.1 Dataset Description

The dataset used in this research consisted of 119 images. After standard augmentation, the dataset consisted of

1428 images, and the annotation of threads was 21k boxes. A security thread is a delicate ribbon or thread that runs through a banknote. When kept under ultraviolet light, the thread in the new notes glows. These protection threads make copying currency with a commercial color copier difficult. Only 1000 and 500 taka notes were used in this study because they are the largest banknotes in Bangladesh, and people counterfeit these two banknotes more frequently. Photos on both sides of the banknote were taken. Since the chances of an old banknote being counterfeit were extremely low, only pictures of new banknotes were taken. However, once the model has been trained with threads, it can operate on any kind of banknote.

(a) 1000 Taka front

(b) 1000 Taka back

(c) 500 Taka front

(d) 500 Taka back

Fig. 1: Sample images of the Dataset.

4 Proposed Method

The proposed method consists of two stages. Initially, the input image is passed through a region proposal algorithm in the first stage and then goes through a classification algorithm to check if the object is present in the foreground or the background, as shown in figure 2.

4.0.1 Region Proposal Network Stage

RPNs are designed to efficiently predict region proposals with a broad range of scales and aspect ratios. Anchor boxes are used by RPNs as references at various scales and aspect ratios. The scheme can be thought of as a pyramid of regression references that prevents enumerating several scales or aspect ratios of images or filters. For the region proposal algorithm, a selective search is selected for this research. Selective search begins by using a graph-based segmentation approach to over-segment the image based on pixel strength. It then uses these over-segments as data, adding all bounding boxes corresponding to segmented sections to the list of regional proposals and grouping adjacent segments based on similarity. More significant segments are created and added to the list of area proposals for each iteration. Intersection over Union (IoU) was used to test the accuracy of the proposed method on a specific dataset. The metric of intersection over union (IoU) is nothing more than a parameter of evaluation [25]. IoU can test any algorithm that produces expected bounding boxes as an output. The model's ground-truth bounding boxes and projected bounding boxes are needed to use Intersection over Union to test an object detector. Each image in the dataset has fewer foregrounds than backgrounds, and both samples need to be balanced. As a result, only the first 16 background images and from the below figure it can be seen that all foreground images that fall within our desired IoU are taken from the first 1000 proposed regions.

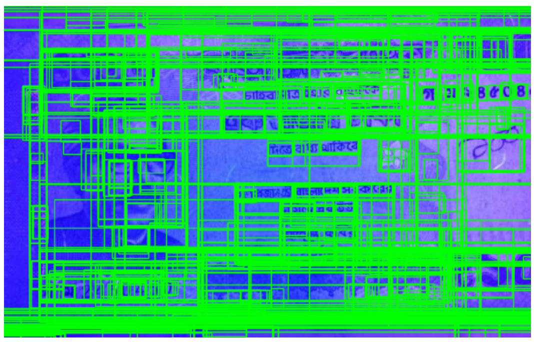


Fig. 3: Proposed regions of an input image.

By focusing on the balance of the dataset, the optimal IoU point is set for both foreground and background. As a result, various IoU points are tested to see which

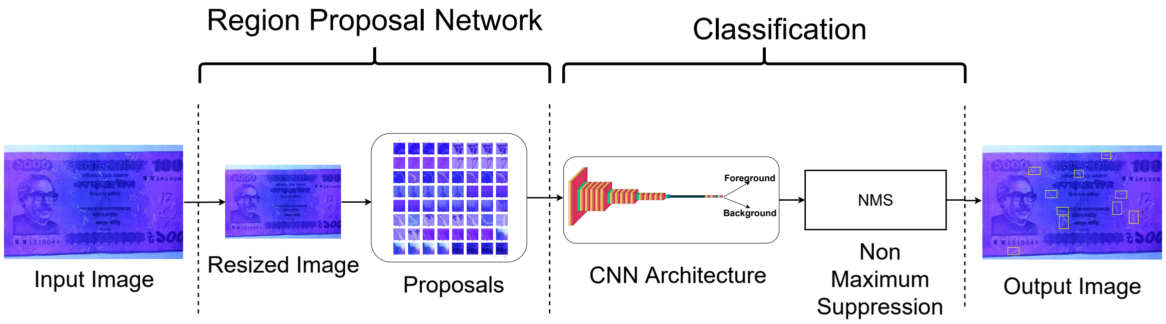


Fig. 2: Flowchart of the proposed method.

combination produces a more balanced dataset for training.

Foreground	Total	Background	Total
IoU > 0.5	22890	IoU < 0.1	22848
IoU > 0.6	17117	IoU < 0.2	22848
IoU > 0.7	12439	IoU < 0.3	22848
IoU > 0.8	6541	IoU < 0.4	22848
IoU > 0.9	1786	IoU < 0.5	22848

Table 1: Total number of Foregrounds and Backgrounds for different IoU regions.

As can be seen from the table 1, the number of backgrounds is constrained, and regardless of the IoU regions, the number of backgrounds is always greater than 16. As a result, the total number of backgrounds is still the same. However, this is not the case for Foreground. Foregrounds that come under some IoU regions are unrestricted. Any points may be checked to see if the foreground and context are nearly balanced. However, due to the small size of the objects in the foreground, the IoU for the foreground cannot be less than 0.5. Furthermore, the background should not be too close to the foreground, and the region should not be too far away from the object, which is why the IoU for the background is set at 0.3. The Foreground and Background of an image are depicted in the figure 4.

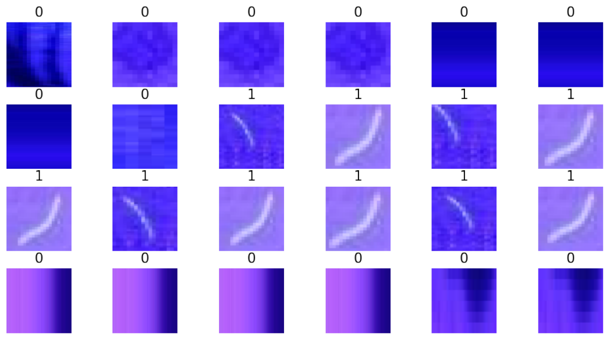


Fig. 4: Foreground and Background of an image.

4.0.2 Classification Stage

Convolutional neural networks (CNNs) are a form of an adaptive image processing system that bridges the gap between general feed-forward neural networks and adaptive filters. Multiple layers of artificial neurons make up convolutional neural networks. CNNs are a big step forward in image recognition. They're most widely used to analyze visual imagery and are often used in image classification behind the scenes. Since CNN extracts features from videos, it produces better results and removes manual feature extraction. CNN can have clearer picture results as a result of this. Region proposals from the first stage are used as input in the second stage.

From the figure 5, it can be seen that the CNN model contains 41 layers along with four maxpool layers. Initially, the input image is passed through the first layer. The data is passed drectly from the input layer to the convolutional layer. Convolutional layers are where filters are added to the original image or other feature maps. This is where the majority of the user-specified parameters are located in the network. The number of kernels and the size of the kernels are the most critical parameters. After that, the rectified linear activation function or ReLU is used. All of the values in the input volume are subjected to the ReLU layer's function

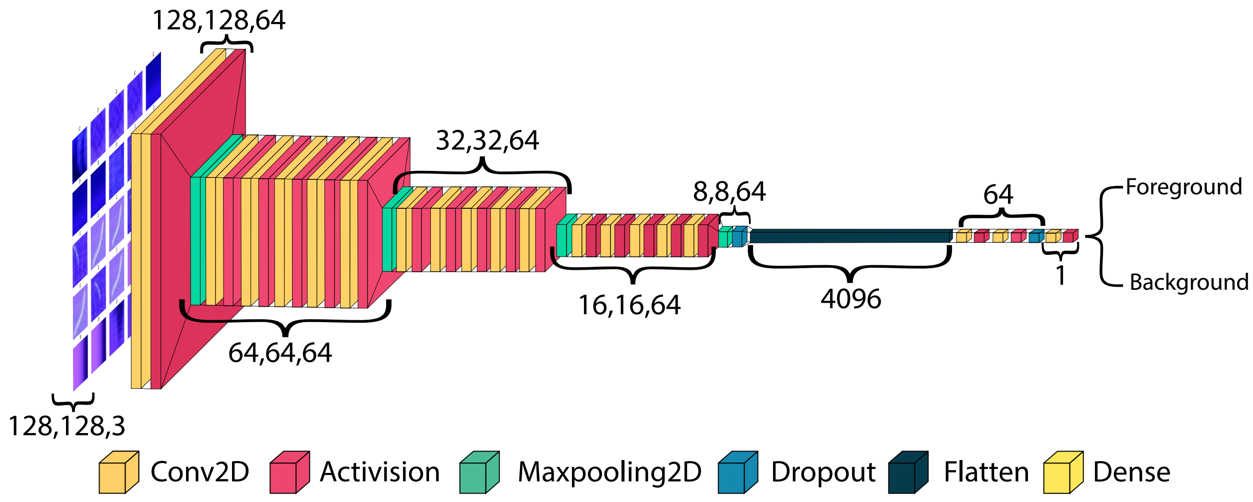


Fig. 5: Proposed CNN Architecture.

f(x) = max(0, x). This layer simply resets all negative activations to 0. Without influencing the convolutional layer's receptive fields, this layer improves the model's and overall network's nonlinear properties. After that, a pooling layer is used to reduce the spatial volume of the input signal. It's used in the middle of two convolution layers. It would be computationally costly to apply FC after the Convolutional layer without using pooling or max pooling, which was not used. Max pooling is a sample-based discretization process. The objective is to down-sample an input representation reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. As a result, the only way to reduce the spatial volume of the input image is to use maximum pooling. The first three maxpool layers are followed by five convolutional layers and ReLU activation functions. The layer after that is a maxpool layer. Next, a dropout layer is used. Dropout is a technique for preventing overfitting in a model. Dropout is a teaching strategy in which randomly selected neurons are ignored. Dropout works by setting the outgoing edges of hidden units, which are neurons that make up hidden layers, to 0. After that, the image is flattened in the next layer, which occurs in the flatten layer. Flatten is a function that takes a pooled feature map and turns it into a single column that can be transferred to a completely connected sheet. After that, the next layer is a dense layer. The dense layer gives the neural network a completely connected layer. Regularization was tested in this model, but the accuracy did not improve. Overfitting can be avoided by using regularization. Regularization is a technique for reducing model complexity. Both outputs from the previous layer are fed to all of the neurons in a Dense layer, with each neuron supplying one output to the next layer. In neural networks, this is the most fundamental layer. The ReLU activation function is then used. After that, another Dense layer is applied, followed by another ReLU function layer. Then comes a dropout layer, followed by a dense layer. The final layer uses a sigmoid activation function since it resides between the first and second layers (0 to 1). As a result, it's particularly useful for models that require the prediction of probability as an output. Since the chance of something only exists between 0 and 1, sigmoid is the best option. After going through all of the layers, the output will contain a list of proposal bounding boxes of 1. The Non Maximum Suppression technique is a computer vision technique which is used to further filter the proposed regions. It's a group of algorithms for picking one entity out of a slew of overlapping ones. After going through NMS, the output will contain a list of filtered proposals.

5 Results

The results of the pre-build models, model design of the proposed architecture, and hyperparameter selection are discussed in this section.

5.1 Architecture Selection

Cross-validation is a resampling technique used to test machine learning models on a limited set of data. The method only has one parameter, k, which determines how many groups a given data sample should be divided into. As a consequence, k-fold cross-validation is a common name for the technique. When a specific value for

Models	K_F 1	K_F 2	K_F 3	K_F 4	Test
ssd_mobilenet_v1	0.678	0.738	0.678	0.722	0.658
ssdlite_mobilenet_v2	0.712	0.725	0.767	0.767	0.703
ssd_mobilenet_v2	0.793	0.763	0.755	0.764	0.729
ssd_inception_v2	0.812	0.810	0.800	0.796	0.743
faster_rcnn_inception _v2	0.918	0.928	0.916	0.918	0.817
faster_rcnn_resnet50	0.920	0.928	0.916	0.920	0.854
rfcn_resnet101	0.914	0.914	0.910	0.911	0.846
faster_rcnn_inception_resnet_v2_atrous	0.921	0.916	0.898	0.921	0.880

Table 2: K-Folds and Test Average Precision (AP) of TensorFlow object detection Models.

k is chosen, it can be used to replace k in the model's relation, such as k=10 for 10-fold cross-validation. In applied machine learning, cross-validation is a technique for estimating a machine learning model's ability on unknown data. It's a popular approach because it's simple to understand and results in a less biased or optimistic model of prediction. It's a popular method because it's simple to understand and gives a less biased or optimistic estimate of model capacity than other methods, such as a simple train/test split.

The tensorflow models mentioned in the table below are pre-trained. It can be seen that Faster RCNN architectures achieved better results than SSD architectures. Mobilenet v2 and inception v2 have outperformed the other SSD architectures. In the Faster RCNN architectures, however, resnet101 and inception resnet v2 achieved better performance. Of all the architectures, inception resnet v2 came out on top with a score of 88%. In this study, this score is used as a benchmark. The proposed model must achieve better results than this.

The proposed regions from the first stage must be classified whether it is foreground or background. As a result, a CNN classifier is needed. Some model parameters must be checked in order to design the CNN classifier. In the proposed architecture, some model parameters have been tested such as different node sizes, number of conv2d layers, and number of dense layers. The test results are shown below.

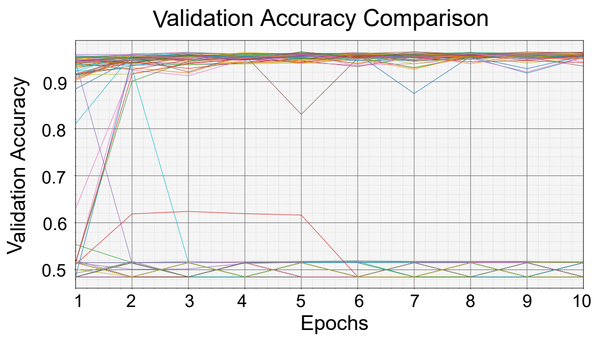


Fig. 6: Validation Accuracy of different proposed architectures.

The graph 6 illustrates that some of the configurations produced underfit results while others produced satisfactory results. In the graph 7, the top three architectures with the highest validation accuracy and loss are shown.

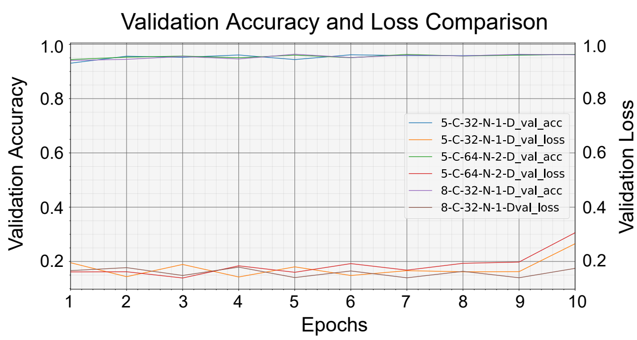


Fig. 7: Validation accuracy and loss of top three proposed architectures.

5.2 Hyperparameter selection

The best architecture was chosen based on the graph. Now a hyperparam must be selected for example, Regularization, batch size, learning rate, and dropout were the four parameters that made up the hyperparameter. Hyperparameters are all the training variables that are manually set to a predetermined value [26]. The effect of regularization accuracy on the chosen architecture was tested. The parameters in L1 have been reduced to zero. The L1 norm becomes sparse when the weights of input features are close to zero. In a Sparse solution, the majority of input features have zero weights, with just a few having non-zero weights. L1 regularization is used to select functions. L2 regularization provides a non-sparse solution by making the weights small but not zero. Since square terms exaggerate outlier error differences, L2 is not resistant to outliers, and the regularization term tries to remedy this by penalizing the weights. Among the methods Without Regularizer and L2 Regularizer had the best accuracy. However, Without Regularizer was marginally better, at 96.3 percent.

Regularizer techniques	Accuracy
Without Regularizer	0.963
L1 Regularizer	0.515
L2 Regularizer	0.953
L1 + L2 Regularizer	0.484

Table 3: Accuracy on different regularizer techniques.

The Adam optimizer was trained at logarithmic intervals with various learning rates ranging from 0.000001 to 100. The learning rates of 0.0001 and 0.00001 showed the best results, as shown in the graph 8. The highest precision was reached by 0.0001, which was 96.3 percent.

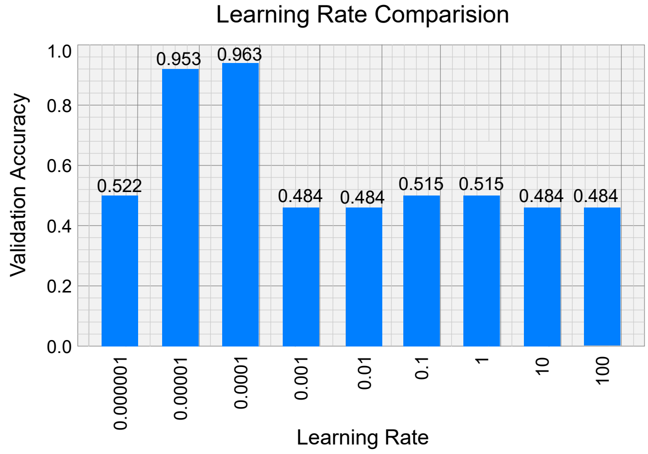


Fig. 8: Accuracy on different learning rates.

The accuracy of all the dropouts that have been tested on the proposed model is shown in the figure 9. The dropouts with the highest accuracy were 0.1, 0.2, and 0.5. However, with a 96.3 percent accuracy, 0.2 is the most accurate.

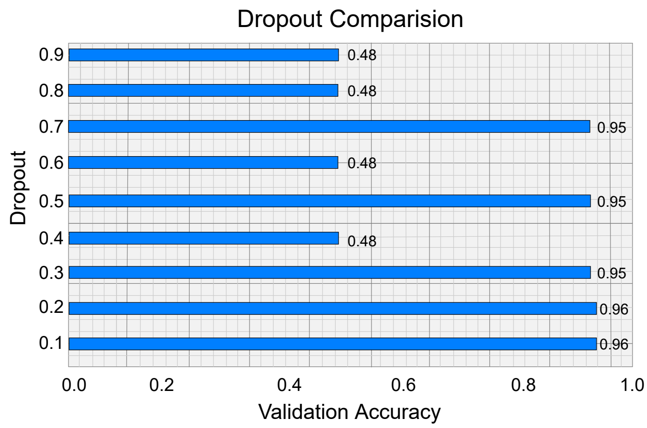


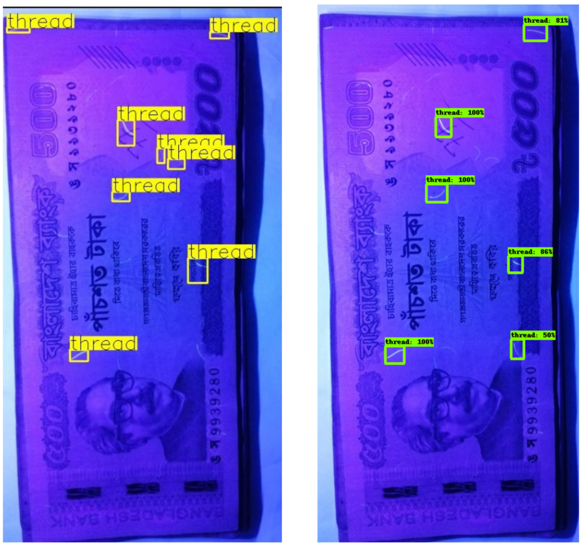
Fig. 9: Accuracy on different values of dropout.

The table 4 shows the total time spent to process one test image and also on each stage. It took 4.67 seconds to propose 1000 images in the first stage, and 36.17 seconds to predict the proposed regions and display the results in the second stage.

First Stage	Second Stage	Total
4.67s	36.17s	40.84s

Table 4: Time consume by the each test image.

After selecting all the hyperparameters and applying the NMS technique the output the proposed model and best tensorflow model is given below figure 10.



Proposed Method faster_rcnn_inception _resnet_v2

Fig. 10: Output of both proposed model and tensorflow model.

5.3 Hardware Setup

The proposed model was implemented using TensorFlow 2 as a framework and Python 3.7 as a programming language. Keras was used for the API. The supporting package included Jupyter notebook, pip3, NumPy, Matplotlib, and spicy panda. Both of these python packages were tested on a Linux framework. The hardware was built around an Intel Core i5 4th generation processor with a clock speed of 3.20 GHz. It has 16 GB of DDR3 RAM and a 1600 Mhz bus speed. The graphics processing unit NVIDIA GTX 1060 6 GB was also included.

5.4 Evaluation

Tensorflow models are evaluated on AP as a benchmark. The average precision (AP) is a method of condensing the precision-recall curve into a single number that represents the average of all precisions. The following equation is used to calculate the AP. At each threshold, the AP is the weighted number of precisions, where the weight is the increase in recall.

In order to find the best CNN architecture for the proposed method, validation accuracy for CNN classification was used. The accuracy measured on the data set is known as validation accuracy. It is not used for training but rather for validating the proposed model's generalization potential for early stopping.

6 Discussion

In this research, the hyperparameter consisted of four parameters, and the optimal value was adjusted accordingly. The batch size was set to 32, the learning rate was 0.0001, and the dropout was 0.2. On the other hand, the model chosen for this study consisted of 5 conv2d, 64 nodes, and 3 dense layers with no regularization. After training, the validation accuracy of the proposed model was 96.3%, and security thread detection per image was also high. Each test image took 40.84 seconds which can be reduced if any other search method which is faster is used instead of selective search; the total time required to complete this research can be reduced. Overall, the proposed model produced a satisfactory result.

A security thread is a security feature in many banknotes that is made up of a thin ribbon that is threaded through the note's paper to prevent counterfeiting. However, as the banknote ages, these security threads become more mutilated and ineffective. If the proposed model cannot detect the security thread on an old banknote, it could declare it counterfeit. There were several images in the dataset used in this study. Both sides of the banknote were photographed. When photographing banknotes, the light exposure must be precise for the security thread to be noticeable. The security thread will not appear in the photos if the light's exposure is incorrect. Every country includes different types of security thread in their banknotes to avoid counterfeiting [27]. Since other countries use different types of security thread for their banknotes, the model can also identify a foreign banknote as counterfeit money.

7 Conclusion

It's worth noting that in Bangladesh, the use of counterfeit money is on the rise. Two to three fake notes are detected per million takas by the Bangladesh Bank, and counterfeit notes are frequently discovered during the initial screening process. Counterfeit currency lowers the value of real money and raises prices, resulting in inflation as more money is exchanged in the economy. The country's economy is increasingly growing, which has resulted in the proliferation of counterfeit money.

Object detection technology is being used in a lot of research to detect counterfeit money. To find the proposed model, a variety of deep learning pre-build models were used, as well as a variety of architectures. The proposed model gave adequate results.

References

1. Klein RM, Gadbois S, Christie JJ. Perception and detection of counterfeit currency in Canada: note quality, training, and security features. In: Optical Security and Counterfeit Deterrence Techniques V. vol. 5310. International Society for Optics and Photonics; 2004. p. 1–12.
2. Sun TD. On obsolete and counterfeit currency; 2015. Available from: https://www.daily-sun.com/arcprint/details/95957/On-obsolete-and-counterfeit-currency/2015-12-03.
3. Mala DA. BB drafts law to check counterfeit currency; 2019. Available from: https://thefinancialexpress.com.bd/economy/bangladesh/bb-drafts-law-to-check-counterfeit-currency-1569126909.
4. RAB busts fake note factory in Dhanmondi; 2020. Available from: https://www.dhakatribune.com/bangladesh/dhaka/2020/01/10/rab-busts-fake-note-factory-in-dhanmondi.
5. Counterfeit currency ring busted in Dhaka, Tk4 cr in fake notes seized; 2020. Available from: https://www.dhakatribune.com/bangladesh/dhaka/2020/06/29/\counterfeit-currency-ring-busted-in-dhaka-tk4-cr-in-fake-notes-seized.
6. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, et al. Deep Learning for Generic Object Detection: A Survey. International Journal of Computer Vision. 2019;128(2):261–318.
7. Kumar A, Zhang ZJ, Lyu H. Object detection in real time based on improved single shot multi-box detector algorithm. EURASIP Journal on Wireless Communications and Networking. 2020;2020(1).
8. Zhao H, Li Z, Fang L, Zhang T. A Balanced Feature Fusion SSD for Object Detection. Neural Processing Letters. 2020;51(3):2789–2806.
9. Liu W, Anguelov D, Erhan D, Szegedy C, Reed SE, Fu C, et al. SSD: Single Shot MultiBox Detector. CoRR. 2015;abs/1512.02325. Available from: http://arxiv.org/abs/1512.02325.
10. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. CoRR. 2015;abs/1512.03385. Available from: http://arxiv.org/abs/1512.03385.
11. Ghosal P, Nandanwar L, Kanchan S, Bhadra A, Chakraborty J, Nandi D. Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network. In: 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP); 2019. p. 1–6.
12. Nguyen LD, Lin D, Lin Z, Cao J. Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS); 2018. p. 1–5.
13. Khasoggi B, Ermatita E, Samsuryadi S. Efficient mobilenet architecture as image recognition on mobile and embedded devices. Indonesian Journal of Electrical Engineering and Computer Science. 2019 10;16:389.
14. Ren S, He K, Girshick RB, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. CoRR. 2015;abs/1506.01497. Available from: http://arxiv.org/abs/1506.01497.
15. Ma D, Wu X, Yang H. Efficient Small Object Detection with an Improved Region Proposal Networks. IOP Conference Series: Materials Science and Engineering. 2019 05;533:012062.
16. Chen YP, Li Y, Wang G. An Enhanced Region Proposal Network for object detection using deep learning method. PLOS ONE. 2018 09;13(9):1–26. Available from: https://doi.org/10.1371/journal.pone.0203897.
17. Tao X, Gong Y, Shi W, Cheng D. Object detection with class aware region proposal network and focused attention objective. Pattern Recognition Letters. 2020;130:353–361. Image/Video Understanding and Analysis (IUVA). Available from: https://www.sciencedirect.com/science/article/pii/S0167865518306986.
18. Uijlings J, Sande K, Gevers T, Smeulders AWM. Selective Search for Object Recognition. International Journal of Computer Vision. 2013 09;104:154–171.
19. Kulkarni A, Callan J. Selective Search: Efficient and Effective Search of Large Textual Collections. ACM Trans Inf Syst. 2015 Apr;33(4). Available from: https://doi.org/10.1145/2738035.
20. Xin M, Wang Y. Research on image classification model based on deep convolution neural network. EURASIP Journal on Image and Video Processing. 2019;2019(1).
21. Sharma N, Jain V, Mishra A. An Analysis Of Convolutional Neural Networks For Image Classification. Procedia Computer Science. 2018;132:377–384. International Conference on Computational Intelligence and Data Science. Available from: https://www.sciencedirect.com/science/article/pii/S1877050918309335.
22. Jang H, Yang HJ, Jeong DS, Lee H. Object classification using CNN for video traffic detection system. In: 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV); 2015. p. 1–4.
23. Pandiscia L, Pandiscia L, the Author Leah Pandiscia received her PhD from Drexel University where she studied Biophysical Chemistry She is a Spectroscopy Applications Scientist A. Fluorescence Detection in the Detection of Counterfeit US Currency: JASCO; 2021. Available from: https://jascoinc.com/applications/using-fluorescence-detection-detection-counterfeit-us-currency/.
24. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. CoRR. 2018;abs/1804.02767. Available from: http://arxiv.org/abs/1804.02767.
25. Wajih, Rosebrock A, Anne, Ahmed W, Kiran, Jsky, et al.. Intersection over Union (IoU) for object detection; 2021. Available from: https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/.
26. Gozzoli A. Practical Guide to Hyperparameters Optimization for Deep Learning Models. FloydHub Blog; 2020. Available from: https://blog.floydhub.com/guide-to-hyperparameters-search-for-deep-learning-models/?fbclid=IwAR2WONmkJXluv1mMWEXaCrpKVjKd6M1McAxNHWw_x5NDCw1YLajJ84qr9UY.
27. Mann M, Shukla S, Gupta S. A comparative study on security features of banknotes of various countries. International Journal of Multidisciplinary research and development. 2015 06;2:83–91.
28. Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, et al. Speed/accuracy trade-offs for modern convo-

lutional object detectors. CoRR. 2016;abs/1611.10012.
Available from: http://arxiv.org/abs/1611.10012.

Supplementary Materials

Models	Speed	CocomAP
ssd_mobilenet_v1_coco	30	21
ssdlite_mobilenet_v2_coco	27	22
ssd_mobilenet_v2_coco	31	22
ssd_inception_v2_coco	42	24

Table 5: Ssd models with their speed and cocomAP [28].

Models	Speed	CocomAP
faster_rcnn_inception_v2_coco	58	28
faster_rcnn_resnet50_coco	89	30
rfcn_resnet101_coco	92	30
faster_rcnn_inception_resnet _v2_atrous_coco	620	37

Table 6: Faster-RCNN models with their speed and cocomAP [28].