



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

Anumol RB¹, Anuja SB²

¹Student, Department of MCA, Narayanaguru College of Engineering, Anna University, Tamilnadu, India.

²Asst. Professor, Department of MCA, Narayanaguru College of Engineering, Anna University, Tamilnadu, India.

ABSTRACT

Currency recognition has always been a troublesome task for blind and visually impaired people (BVIP). Most of the current Indian legal tenders resemble in size, thus making the identification process more difficult. Automated paper currency recognition system, issues such as folded or partial views, uneven illumination, and background clutter make it non-trivial and challenging. This Paper proposes deep learning model for assisting BVIP in recognizing the Indian paper currency denomination. This Paper exploits the latent embeddings of variational auto encoders combined with IPCRNet. Two encoders are trained on currency images to generate two feature maps. The IPCRNet is based on dense connection, Multi-Dilation, and Depth-wise separable convolution layers. Contextual Block (CB) in the backend utilizing the dense connection and dilation scheme in depth-wise separable convolutional layers. The novel contextual part utilizes a depthwise separable convolution for reducing network computations. The multi-dilation scheme offers an enlarged receptive field without increasing the parameters.

Keywords: Assistive technologies, currency identification, dense network, depthwise separable convolution, visually impaired.

1 Introduction

The World Health Organization (WHO) estimated the no. of visually impaired at global level, based on the latest studies, there are about 284 million people in the world who are visually impaired, and 39 million people are blind[1]. One of the main problem faced by the people with visual disabilities is the inability to recognize the paper currency due to the similarity of paper texture and size between the different categories. One of the critical issues for Blind and Visually Impaired People (BVIP) is the recognition and authentication of currency denominations. In some cases, they need to depend on the normally sighted person (NSP) for currency identification assistance. Conventionally, the Blind and Visually Impaired People (BVIP) rely on the embedded positional patterns and differences in the size of the paper currencies for denomination recognition.

However, this approach has associated limitations as the ability to sense the engraved patterns



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

on the paper currency varies from person to person. Furthermore, the printed engravings and distinct patterns on paper currency get worn away with time. To assist BVIP in recognizing currency denominations, several automated and semi-automated currency recognition systems are proposed in the literature. An automated currency recognition system utilizes the properties such as color, size, motifs, micro-lettering, engraved patterns, and edge parameters as shown in Figure 1. However, unlike other countries' banknotes, the present paper currencies in India are almost similar in size without definite tactile attributes [2]. Observing these difficulties in manual recognition of the current currency denominations by BVIP in India, various organizations, including the Reserve Bank of India (RBI), has shown a growing interest in mobile-based solutions recently. A dedicated mobile-based solution can assist in recognizing the currency notes more effectively. BVIP relies on the voice-based features of smartphones like Talkback, Google assistant, and Siri for performing generic tasks. However, mobile-based solutions with dedicated machine/deep learning models for recognition are challenging in terms of deploy ability and adaptability. The existing models are comparatively bulkier, and integration with low-end mobile smartphones is impractical, thus causing deploy ability issues [3], [4], [5]. Additionally, the users who are legally blind or have severe low vision issues cannot handle the camera view appropriately or other required settings to get the optimum results. Therefore a stable and robust recognition system is highly needed. In deep learning-based networks have gone deeper to gain performance improvements. With more depth, a larger number of parameters is required, thus causing an exponential increase in the complexity of the networks. The bulkier models are unsuitable for resource-constrained devices such as mobile phones and edge-based devices. To minimize the dependence on servers and the internet, various lightweight networks such as we have been proposed for the tasks of classification and object detection. Schemes such as limiting the number of channels/kernel size, optimizing pooling layers, efficient coding, and representation are typically used for compressing the network sizes. GoogleNet uses a width reduction-based scheme, and MobileNet uses depth-wise separable convolution for compressing the network size. However, with such compression schemes and the simplified convolutional structure, the model tends to miss discriminative image features, affecting the overall performance.

The proposed model, IPCRNet is a lightweight neural network and utilizes MobileNet as the front-end. IPCR-Net uses a Contextual Block (CB) in the backend utilizing the dense connection and dilation scheme in depth-wise separable convolutional layers. The model has less than four million parameters, thus favoring its deployment in a resource-constrained environment. The novel contextual part utilizes a depthwise separable convolution for reducing network computations.

The multi-dilation scheme offers an enlarged receptive field without increasing the parameters, thereby increasing the accuracy via an effective integration of global and semantic features. Compared with the existing state-of-the-art backend network and approaches, our network



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

provides superior accuracy than its counterparts and is lightweight. Furthermore, to aid an effective training and evaluation of the proposed model, we have gathered a large-scale Indian paper currency dataset, IPCD. The IPCD dataset unlike existing datasets, consists of images with BVIP perspective (folded and partial note images), with varied illumination and background conditions. The proposed end-to-end Indian paper currency recognition framework (IPCRF) offers a contextual learning network, a diversified and domain perspective dataset (IPCD) to support an effective training/evaluation, and a BVIP compatible interface via android mobile application. The proposed deep learning model is trained on our dataset (IPCD)[6],[7]. Our android app utilizes the compressed trained model for real-time recognition of the underlying currency denomination.

The novelty and main contributions of this work are:

Novelty. A robust lightweight and domain specialized CNN model to capture the pervasive intra-class and inter-class dissimilarities between currency denominations classes. The proposed Contextual Block offers an effective integration of the local and global features.

Diverse Dataset. One of the largest (more than 50k images) and most diversified dataset of Indian Paper Currency images, representing real scenarios and cases.

Quantitative Analysis. A thorough quantitative analysis of the proposed network on multiple publically available datasets has been performed.

Qualitative Analysis. A quantitative analysis has been performed to investigate the transparency and intuition behind the proposed network predictions.



Figure 1: *Indian Paper Currency*



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

2 Recent Works

The currency recognition problems has been well explored. The related existing approaches/systems into three aspects: Dataset (availability of diverse datasets for training and evaluation purposes); Model (availability of lightweight and accurate recognition models); and Application (availability of a BVIP compatible interface for assisting the BVIP in currency recognition tasks). This section discusses the existing approaches and related concepts related to the mentioned individual aspects.

The advantages and drawbacks of the proposed and existing methods are summarized. The majority of the previous work uses smaller datasets. Also, very few works have used multiple datasets for evaluation which is critical for validating the generalization ability of underlying models. Standard methods and datasets dedicated to Indian currencies (INR) are limited. The detailed comparative analysis and description of available and proposed datasets [11], [13]. In this section, we primarily discuss the existing models and available systems/technologies available, along with a brief discussion of the dataset.

Lightweight yet accurate models and limited BVIP compatible real-time applications are also major concerns in the existing literature. The existing model's section discusses the models and frameworks used by currency recognition systems or approaches, categorized into two subsections: Hand-crafted feature-based Models and Deep Learning (DL) based models. The existing systems section discusses the existing assistive technologies (apps) available to the BVIP for currency identification.

[1] MobileNetV2: Inverted Residuals and Linear Bottlenecks, Mark Sandler and Andrew Howard, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. In this Paper we describe a new mobile architecture, MobileNetV2, which improves the state of the art performance of mobile models on multiple tasks and benchmarks as well as across a spectrum of different model sizes. We also describe efficient ways of applying these mobile models to object detection in a novel framework we call SSDLite. Additionally, we demonstrate how to build mobile semantic segmentation models through a reduced form of DeepLabv3 which we call Mobile Deep Labv3 is based on an inverted residual structure where the shortcut connections are between the thin bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. For object detection task, our network outperforms state-of-art realtime detectors on COCO dataset both in terms of accuracy and model complexity. Notably, our architecture combined with the SSDLite detection module is $20\times$ less computation and $10\times$ less parameters than YOLO[9],[10]. It is important to remove non-linearities [9] in the narrow layers in order to maintain representational power. We demonstrate that this improves performance and provide an intuition that led to this design. Finally, our approach allows decoupling of the input/output domains from the expressiveness of the transformation, which provides a convenient



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

framework for further analysis. We measure our performance on ImageNet classification, COCO object detection, VOC image segmentation. We evaluate the trade-offs between accuracy, and number of operations measured by multiply-adds (MAdd), as well as actual latency, and the number of parameters.

[2] Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices, Xiangyu Zhang and Xinyu Zhou, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. We introduce an extremely computation-efficient CNN architecture named ShuffleNet, which is designed specially for mobile devices with very limited computing power (e.g., 10-150 MFLOPs). The new architecture utilizes two new operations, pointwise group convolution and channel shuffle, to greatly reduce computation cost while maintaining accuracy. To evaluate the generalization ability for transfer learning, we test our ShuffleNet model on the task of MS COCO object detection. We adopt Faster-RCNN as the detection framework and use the publicly released Caffe code for training with default settings. Similar to the models are trained on the COCO trainval dataset excluding 5000 minival images and we conduct testing on the minival set. We evaluate our models on the challenging ImageNet classification and MS COCO object detection [14] tasks. A series of controlled experiments shows the effectiveness of our design principles and the better performance over other structures. Compared with the state-of-the-art architecture MobileNet. The ShuffleNet model achieves $\sim 13\times$ actual speedup (theoretical speedup is $18\times$) over AlexNet while maintaining comparable accuracy. Experiments on ImageNet classification and MS COCO object detection demonstrate the superior performance of ShuffleNet over other structures, e.g. lower top-1 error (absolute 7.8%) than recent MobileNet on ImageNet classification task, under the computation budget of 40 MFLOPs. On an ARM-based mobile device, ShuffleNet achieves $\sim 13\times$ actual speedup over AlexNet while maintaining comparable accuracy.

[3] Deep Learning-Based Fake-Banknote Detection for the Visually Impaired People Using Visible Light Images Captured by Smartphone Cameras, Tuyen Danh Pham, Chanhum Park, IEEE Access, 2020. Automatic recognition of fake banknotes is an important task in practical banknote handling. Research on this task has mostly involved methods applied to automatic sorting machines with multiple imaging sensors or that use specialized sensors for capturing banknote images in various light wavelengths. These approaches can make use of the security features on banknotes for counterfeit detection. We classify banknote images as either genuine or fake by using a CNN. AlexNet and ResNet-18 showed good performance in banknote type and fitness classification. We proposed a fake-banknote detection method that uses CNN and banknote images captured by a smartphone camera under visible light conditions. Our method was designed to classify genuine and fake banknotes regardless of the denomination and the side of the banknote exposed to the camera. We performed a four-fold cross validation on self-collected banknote image datasets of EUR, USD, KRW, and JOD banknotes. However, they require specialized devices, which are not always available for general users or visually



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

impaired people. Meanwhile, smart phones are becoming more popular and can be useful imaging devices. Moreover, the types of fake banknotes created by imaging devices such as smart phone cameras or scanners are sometimes cannot be recognized by especially the visually impaired people. Addressing these problems, we propose a method for classifying fake and genuine banknotes using visible-light images captured by smart phone cameras based on convolutional neural networks. Experimental results on a self-collected dataset of US dollar, Euro, Korean won, and Jordanian dinar banknotes showed that our method performs better in terms of fake detection than the state-of-the-art methods.

[4] Deep Feature-Based Three-Stage Detection of Banknotes and Coins for Assisting Visually Impaired People, Chanhum Park, Se Woon Cho, IEEE Access, 2020. Deep Feature-Based Three-Stage Detection of Banknotes and Coins for Assisting Visually Impaired People .Owing to the rapid advancements in smartphone technology, there is an emerging need for a technology that can detect banknotes and coins to assist visually impaired people using the cameras embedded in smartphones. Previous studies have mostly used handcrafted feature-based methods, such as scale invariant feature transform or speeded-up robust features, which cannot produce robust detection results for banknotes or coins captured in various backgrounds and environments. With the recent advancement in deep learning technology, some studies have been conducted on banknote and coin detection using a deep convolutional neural network (CNN)[11],[12]. To improve the detection performance in the VGG16-based Faster R-CNN of the first stage of detection, post processing methods were applied as the second stage of detection based on the three features. However, these studies also showed degraded performance depending on the changes in background and environment. The post processing, verification was performed as the third stage of detection by the ResNet-18-based Faster R-CNN to detect the final banknote region. Furthermore, the self-collected DKB v1 and the developed models with algorithms. To overcome these drawbacks, this paper proposes a three stage detection technology for new banknotes and coins by applying faster region-based CNN, geometric constraints, and the residual network (ResNet). In the experiment performed using the open database of Jordanian dinar (JOD) and 6,400 images of eight types of Korean won banknotes and coins obtained using our smartphones, the proposed method exhibited a better detection performance than the state-of-the-art methods based on handcrafted features and deep features.

3 Proposed Work

The proposed IPCD dataset was collected in a real-time scenario to incorporate variations in lightning conditions, backgrounds, postures, and angles. Initially, we collected around 500 image samples via a range of smartphones, ranging from low-end (with average camera quality) to high end smartphones.



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

While collecting these images, we have included conditions such as folded and full view of currency images and the indoor/outdoor environments to capture variability. Other reasons for currency images not getting included in the final dataset include very high blurriness, illuminations issues, and only a tiny portion of notes captured in the image. It is to be noted that all images in the proposed dataset have been captured through mobile phones. In total, more than 50 different mobile phone brands have been used to capture the images. For around 7% of IPCD images, the brand info was unavailable and for the remaining IPCD images, the distribution of top models used.

The front-end part constitutes a total of five blocks containing 1, 2, 2, 6, and 2 depth-wise separable convolution layers, respective. Each depth-wise separable convolution layer consists of a 3×3 depthwise (dw) and (1×1) pointwise (pw) convolution layer with batch normalization (BN) and ReLu activation layers (dw-BN-ReLu-pw-BN-ReLu). Zero padding layers have been used between blocks to preserve the resolution.

Counting depth-wise and pointwise layers separately, the front-end comprises 27 Conv layers (26 depthwise and pointwise, and one regular convolution. To reduce the resolution, stride has been used instead of the pooling operation. The final output from the front-end has a size of $7 \times 7 \times 1024$. The front-end and back-end modules of the proposed network are based on depthwise separable convolutions consisting of two parts (a) depthwise convolution and (b) pointwise convolution. Individual filters are applied to each input channel in depthwise convolution, followed by pointwise convolution (1×1) for combining the output.

The proposed CB utilizes the contextual information level feature maps through dense connection. To control the input and output channel size, $\hat{C} 1 \times 1$ is used at the beginning and end of the module. The CB has five $\hat{C} 3 \times 3$ and one $\hat{C} 1 \times 1$ layers. Instead of strictly increasing or decreasing dilation factors, CB uses a combination of both (i.e. Dilations used: 1, 2, 3, 2, 1) Figure 2. The number of kernels is set to 64 in each layer. The front-end module outputs a 7×7 feature map resolution, which goes as input to the CB. However, adding a decreasing structure allows the previous information to be contextually aggregated. For larger objects, the local information can be captured from the feature map of increasing dilation. The proposed multi-dilation scheme enables the model to efficiently capture the local and global feature maps.



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

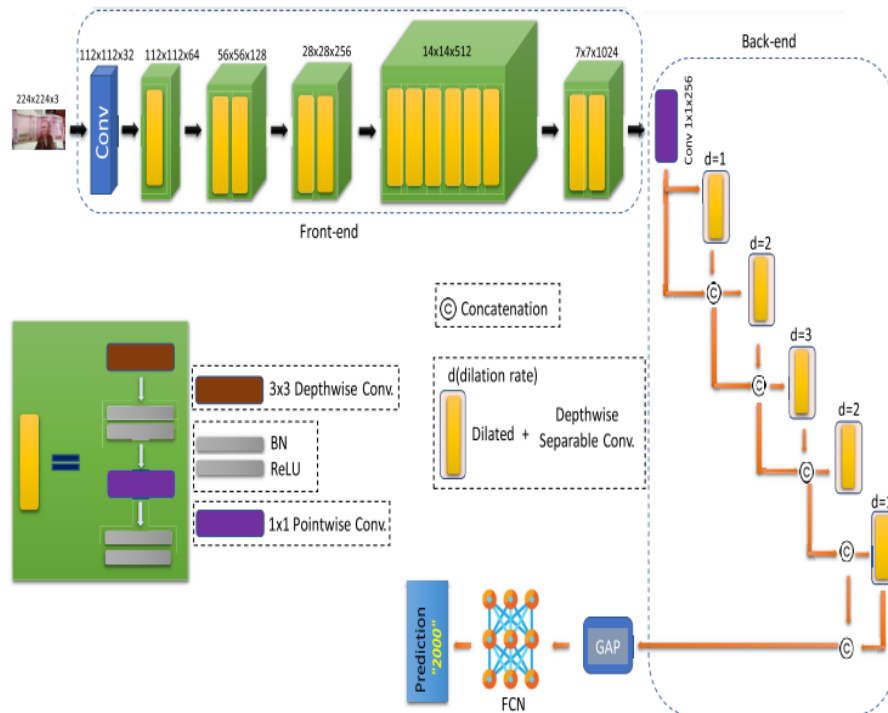


Figure 2: Architecture Diagram

3.1 IPCRNET: Proposed Indian Paper Currency Recognition Network

The proposed deep neural network architecture formulates the problem of currency denomination recognition as a multi-class classification problem. Given an image of Indian paper currency as input, our goal is to robustly recognize the denomination even with the partial or folded view. The currency images, in general, shares some similar or common patterns and thus have considerable inter-class similarities. Additionally, in the case of a partial or folded image view (much likely for BVIPs), there might be a chance that the discriminative part of the image is not visible. For accurate classification of images in such scenarios, the computed feature maps must cover multi-scale receptive field areas. The IPCRNet aims to capture high-level semantic features with an enlarged receptive field without increasing the parameter count excessively. We have utilized MobileNet as the front-end because of its lightweight and better information flow capability with lesser parameters. The front-end part consists of depthwise separable convolution (represented as a single yellow bar in Figure: 2) instead of regular convolution layers. The final computed feature map of the front-end end part is then fed to the proposed back-end part, i.e., contextual block (CB). In CB, firstly, a 1×1 convolution operation is applied to the output of the front-end module to control the excessive channel growth. CB uses controlled multi dilation schemes and dense connection schemes to capture multi-scale features. All outputs in the CB are densely connected. The layer is then fed to the Global Average Pooling (GAP) layer for flattening, followed by the fully connected network



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

for class prediction. The model comprises approx. 3.6 M parameters marginally higher (approximately 0.4M) than the base MobileNet, however, offering significant performance improvement.

A. Front-End

This section describes the front-end module of the proposed IPCRNet. The front-end part constitutes a total of five blocks containing 1, 2, 2, 6, and 2 depth-wise separable convolution layers, respectively (represented as a single yellow bar. Each depth-wise separable convolution layer consists of a 3×3 depthwise (dw) and (1×1) pointwise (pw) convolution layer with batch normalization (BN) and ReLu activation layers (dw-BN-ReLu-pw-BN-ReLu) as shown in the lower-left part. Zero padding layers have been used between blocks to preserve the resolution. Counting depth-wise and pointwise layers separately, the front-end comprises 27 Convolution layers. To reduce the resolution, stride has been used instead of the pooling operation. The final output from the front-end has a size of $7 \times 7 \times 1024$. Next, we discuss the depthwise separable convolution scheme.

1) Depthwise Separable Convolution

The front-end and back-end modules of the proposed network are based on depthwise separable convolutions consisting of two parts (a) depthwise convolution and (b) pointwise convolution. Individual filters are applied to each input channel in depthwise convolution, followed by pointwise convolution (1×1) for combining the output. This articulation results in an overall reduction in the model computations and size. Also, the 3×3 depthwise separable convolutions reduce the computation by 8 to 9 times. A sample illustration of the depthwise and pointwise convolution filters is presented. A conventional convolution filters is factorized into depthwise and 1×1 pointwise convolutions respectively, for significantly reducing the number of computational operations and parameters.

Consider an input feature map R of size $D_R \times D_R \times P$, where D_R represents the width and height of the input feature map and P is the input depth, and a convolution kernel K of size $D_K \times D_K \times P \times Q$, where D_K is the width and height of kernel and P and Q are number of input and output channels, respectively. A conventional convolutional layer using K and input feature map R will generate the output feature map S of size $D_S \times D_S \times Q$, where $D_S (=D_R)$ is the width and height of output feature map.

This conventional convolution, with stride one, can be represented as:

$$\hat{S}_{k,l,q} = \sum_{i,j,p} K_{i,j,p,q} \cdot R_{k+i-1,l+j-1,p} \quad (1)$$

The total cost associated with this conventional convolution for obtaining the complete S is: $[D_K \cdot D_K \cdot P \cdot Q \cdot D_R \cdot D_R]$. The conventional convolution operation cost involves number of input channels P , the number of output channels Q , kernel size $D_K \times D_K$ and feature map size $D_R \times D_R$. This can be reduced by considering the depthwise separable convolution in which the



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

collective filtering and combination step of conventional convolution operation is split into two steps, i.e., 3×3 depthwise and 1×1 pointwise convolutional layers.

The depthwise convolution with single filter per input channel/depth is given by:

$$\hat{S}_{k,l,p} = \sum_{i,j} K_{i,j,p} \cdot R_{k+i-1,l+j-1,p} \quad (2)$$

where $\hat{K}$ is the depthwise kernel of size $D_K \times D_K \times P$. and the associated cost is given by: $[D_K \cdot D_K \cdot P \cdot Q \cdot D_R \cdot D_R]$. Combining it with the cost of applying pointwise convolutions using the $1 \times 1 \times P$ filters, the total cost for the combined operations, or say, depthwise separable convolution can be then given as: $[D_K \cdot D_K \cdot P \cdot Q \cdot D_R \cdot D_R + P \cdot Q \cdot D_R \cdot D_R]$.

Therefore, the ratio of number of computations in depthwise separable convolution and that of conventional convolution process is given as: $[1/Q + 1/D_K^2]$. This ratio shows the reduction in computations from conventional convolution scheme to depthwise separable convolution.

B. Proposed Back-End Contextual Block

The proposed CB utilizes the contextual information level feature maps through dense connection. To control the input and output channel size, $\hat{C} 1 \times 1$ is used at the beginning and end of the module. The CB has five $\hat{C} 3 \times 3$ and one $\hat{C} 1 \times 1$ layers. Instead of strictly increasing or decreasing dilation factors, CB uses a combination of both (i.e. Dilations used: 1, 2, 3, 2, 1). This controlled increasing and then decreasing dilation scheme constrained within the available resolution improves the consistency of local feature maps. The number of kernels is set to 64 in each layer. The front-end module outputs a 7×7 feature map resolution, which goes as input to the CB. The proposed multi-dilation scheme enables the model to efficiently capture the local and global feature maps.

1) Dilated Depthwise Separable Convolution

Depthwise separable convolution filters are used in the proposed contextual module, however each filter is dilated to extract more information without increasing the model complexity and channel. The dilation process with a 3×3 filter with different dilation rates. In the receptive area is 3×3 , however with dilation = 2, the same 3×3 kernel has a receptive field as 5×5 kernel with lesser parameter. Similarly, with dilation = 3, the receptive field increases to 7×7 cross view. Locations with a circle mark denote the receptive region, and locations without the circle mark stipulate the non-receptive area.

2-D dilation can be represented as:

$$y(r,s) = \sum_{i=1}^R \sum_{j=1}^S x(r+d \times i, s+d \times j) k(i,j) \quad (3)$$

where, $y(r,s)$ and $x(r,s)$ are the input and output, and $k(i,j)$ is the kernel with height R, width S and dilation factor d. For $d = 1$, the dilation operation reduce to normal convolution. The final enlarged Receptive Field (Z) of a dilated convolution layer with filter size $k \times k$ is



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

$$Z = (d - 1) * (k - 1) + k \quad (4)$$

The stacking of these layers further enlarges the receptive field. The stacking effect on the densely connected dilated layers and the overall receptive field.

$$\hat{Z} = Z_1 + Z_2 - 1 \quad (5)$$

Dilation operation involves the expansion of the receptive field without further increasing the convolutional parameters. The enlarged receptive field favors the extraction of finer semantic details, thus increasing the model's overall accuracy. Typically, standard dilation schemes involve using the same dilation factors across layers or strictly increasing dilation factors across layers but these schemes fail to capture the local features and to extract contextual information causing aliasing in higher layers.

2) Dense Connection

We have utilized the dense connection to incorporate the multi-scale property in the proposed contextual block. Each layer's output goes to every subsequent layer. A dense connection comprises of total $\frac{L(L+1)}{2}$ connections, unlike only L connections in traditional CNN's. In traditional CNNs, the output from the layer L_i goes as input to Layer L_{i+1} . The output O_l of the l^{th} layer, where N is the non-linear transformation process at l^{th} layer within the dense block.

$$O_l = N_l(O_{l-1}) + O_{l-1} \quad (6)$$

4 Result and Discussion

The performance gain in over- all accuracy and the reliability of the proposed IPCRNet performance in more detail. To examine the accuracy and generalization ability, the performance of the different models across multiple INR currency datasets. The performance gains of the proposed model in average accuracy in comparison to the existing approaches. The proposed IPCRNet performance is comparatively better than the 2nd best performing model (D121), with a significant gain of 2.59% and 2.90% on IPCD and Coinnet datasets. It also achieves 3.61% and 1.23% on other datasets (Kaggle-U1 and Meshram). On the Kaggle-U1 dataset, the performance of top-3 approaches is the same. On larger datasets such as IPCD, the performance of different models is comparatively better, but on the smaller datasets, the performance gets lowered. However, the proposed model shows higher performance gain signifying the better generalization ability than other comparative models.

For reliability, a scheme involving the model's performance, we considered and analyzed the confidence scores of correct predictions, rather than considering and including both correct and wrong prediction cases. The model exhibiting the higher confidence on correct predictions is



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

more reliable than the model that provides correct predictions with lower confidence scores. Most models are stuck within the 97% confidence mark, but IPCRNet achieves better average confidence score of 98.38%, as shown in Table 11 and it may be also recalled that average accuracy of the proposed model is better, implying a larger number of correct predictions. Overall, the results across multiple datasets highlight that the proposed model is consistently more accurate and reliable than other models.

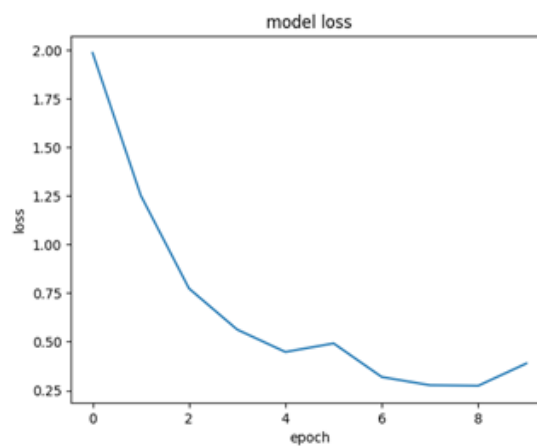


Figure 3: Model Loss

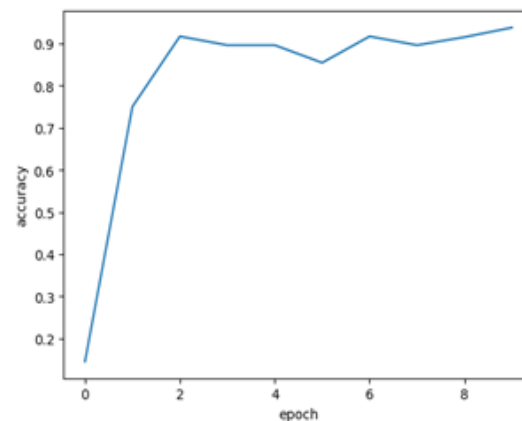


Figure 4: Model Accuracy



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

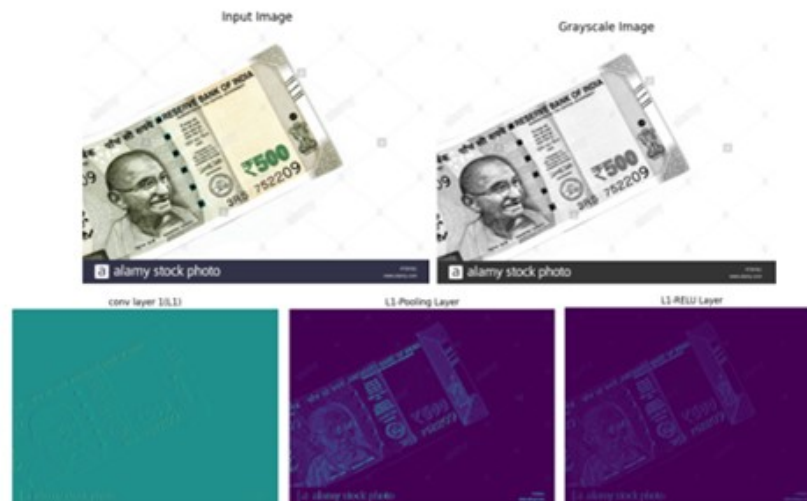


Figure 5: *Image Convolution*

5 Conclusion

This Paper focuses on the problem of Indian currency recognition for BVIP and presents an end-to-end automated solution. Here, propose an extensive large-scale Indian currency dataset (approximately 10x larger images count than the existing ones). The dataset contains images from varied backgrounds conditions and different illumination and orientations. Apart from that, images with folded and partial views are included focusing on the BVIP scenario. The proposed lightweight network (IPCRNet) uses controlled multi-dilation and depthwise separable convolution schemes with dense connection, enabling local and global contextual information aggregation. IPCRNet offers the advantage of enlarging the receptive field without a larger resolution requirement. An extensive evaluation of the proposed framework on publically available datasets has been performed for assessing the generalization and prediction capabilities. The experimental analysis demonstrates the IPCRNet competence in capturing the currency-specific features. IPCRNet is simpler and efficacious in terms of parameters (3.6M) and accuracy.

References

1. Sandler Mark; Andrew Howard; Menglong Zhu; Andrey Zhmoginov; Liang-Chieh Chen, Year: 2018, "MobileNetV2: Inverted residuals and linear bottlenecks", in Proc. IEEE/CVF Conference Visual Pattern Recognition, Vol: 46, pp. 4510–4520.



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

2. Szegedy Christian, Wei Liu; Yangqing Jia; Pierre Sermanet; Scott Reed; Dragomir Anguelov; Dumitru Erhan; Vincent Vanhoucke; Andrew Rabinovich, Year: 2015, "Going deeper with convolutions", in Proc. IEEE Conference Computer Visual Pattern Recognition (CVPR), Vol: 33, pp. 1–9.
3. Veeramsetty Venkataramana; Gaurav Singal; Tapas Badal; Year: 2020, "CoinNet: Platform independent application to recognize Indian currency notes using deep learning techniques", IEEE Multimedia Tools Application, Vol: 79, pp. 22569–22594.
4. Zhang Qian; Wei Qi Yan, Year: 2018, "Currency detection and recognition based on deep learning", in Proc. 15th IEEE International Conference Advance Video Signal Based Surveillance. (AVSS), Vol: 27, pp. 1–6.
5. Huynh Tri; Jay Pillai; Eunyoung Kim; Kristen Aw; Jack Sim; Ken Goldman; Rui Min, Year: 2019, "Bringing vision to the blind: From coarse to fine, one dollar at a time," in Proc. IEEE Winter Conference Application Computer Visual (WACV), Vol: 8, pp. 481–490.
6. Park Chanhum; Se Woon Cho; Na Rae Baek; Jiho Choi; Kang Ryoung Park, Year: 2020, "Deep feature- based three-stage detection of banknotes and coins for assisting visually impaired people," IEEE Access, Vol: 8, pp. 184598–184613.
7. Pham Tuyen Danh; Chanhum Park; Dat Tien Nguyen; Ganbayar Batchuluun; Kang Ryoung Park, Year: 2020, "Deep learning-based fake-banknote detection for the visually impaired people using visible-light images captured by smartphone cameras," IEEE Access, Vol: 8, pp. 63144–63161.
8. Joshi Rakesh Chandra; Saumya Yadav; Malay Kishore Dutta, Year: 2020, "YOLO-v3 based currency detection and recognition system for visually impaired persons," in Proc. International Conference Contemp. Computer Application (ICCA), Vol: 5, pp. 280–285.
9. Takeda, Fumiaki; Sigeru Omatu, Year: 1995, "High speed paper currency recognition by neural networks," IEEE Trans. Neural Network, Vol: 6, no. 1, pp. 73–77.
10. Frosini, Angelo; Marco Gori; Paolo Priami, Year: 1996, "A neural network-based model for paper currency recognition and verification," IEEE Trans. Neural Network, Vol: 7, no: 6, pp. 1482–1490.
11. Tanaka Masahiro; Fumiaki Takeda; Kazuya Ohkouchi; Yasuyo Michiyuki, Year: 1998, "Recognition of paper currencies by hybrid neural network," in Proc. IEEE World Congr. Computer Intell. Neural Network, Vol: 3, pp. 1748–1753.
12. Guo Junfang; Yanyun Zhao; Anni Cai, Year: 2010, "A reliable method for paper currency recognition based on LBP," in Proc. 2nd IEEE International Conference Network Infrastructure Digital Content, pp. 359–363.



Article Title: Indian Paper Currency Recognition Framework for Blind and Visually Impaired People using Deep Learning Model

13. He Kaiming; Xiangyu Zhang; Shaoqing Ren; Jian Sun, Year: 2016, “Deep residual learning for image recognition,” in Proc. IEEE Conference Computer Visual Pattern Recognit. (CVPR), pp. 770–778.
14. Deng Jia; Wei Dong; Richard Socher; Li-Jia Li; Kai Li; Li Fei-Fei, Year: 2009, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conference Computer Visual Pattern Recognition, Vol: 6 , no: 7, pp. 248–255.
15. Selvaraju Ramprasaath R; Michael Cogswell; Abhishek Das; Ramakrishna Vedantam; Devi Parikh; Dhruv Batra, Year: 2017, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE International Conference Computer Visual (ICCV), pp. 618–626.