Pincode detection using Deep CNN for Postal Automation

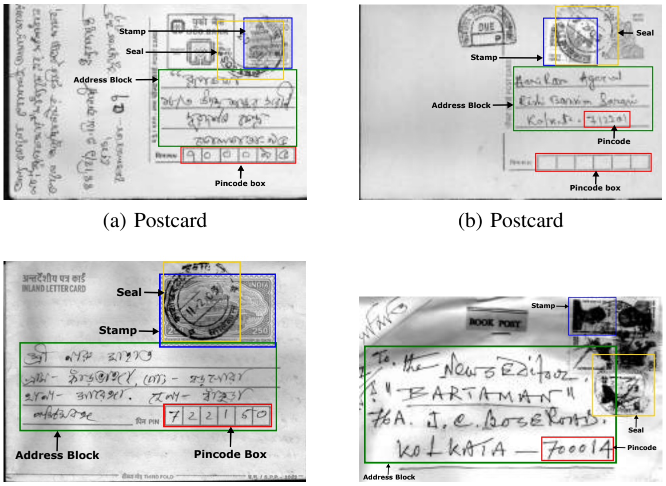
Nabin Sharma*, Abira Sengupta†, Rabi Sharma†, Umapada Pal† and Michael Blumenstein*
*University of Technology Sydney, Broadway, Ultimo, NSW 2007, Australia.
Email: {Nabin.Sharma, Michael.Blumenstein}@uts.edu.au
†CVPR Unit, Indian Statistical Institute, Kolkata 700108, India. Email: umapada@isical.ac.in

***Abstract*—Postal automation has been a topic of research over a decade. The challenges and complexity involved in developing a postal automation system for a multi-lingual and multi-script country like India are many-fold. The characteristics of Indian postal documents include: multi-lingual behaviour, unconstrained handwritten addresses, structured/unstructured envelopes and postcards, being among the most challenging aspects. This paper examines the state-of-the-art Deep CNN architectures for detecting pin-code in both structured and unstructured postal envelopes and documents. Region-based Convolutional Neural Networks (RCNN) are used for detecting the various significant regions namely, Pin-code blocks/regions, destination address block, seal and stamp in a postal document. Three network architectures namely Zeiler and Fergus (ZF), Visual Geometry Group (VGG16), and VGG_M were considered for analysis and identifying their potential. A dataset consisting of 2300 multi-lingual Indian postal documents of three different categories was developed and used for experiments. The VGG_M architecture with faster-RCNN performed better than others and promising results were obtained.**

I. Introduction

Indian postal automation has been a topic of research for a decade and many published articles are available [1], [9], [10], [5], [6], [7], [8]. Many systems are available for the postal automation in USA, UK, France, Australia and Cannada. No systems exists for India postal automation, due to numerous complexities and challenges involved. India being a multi-lingual and multi-script country, the addresses are either written in a regional language, or in English, or both. The addresses are usually handwritten on a simple or a structured envelope, in an unconstrained environment. Multi-lingual and unconstrained nature of Indian postal documents/envelopes are the major bottle neck in the development of Indian postal automation system.

One of the essential task in postal automation is to detect/localize the destination address block (DAB) and the Pin-code. Postal codes in India are six digit numbers which identifies a postal zone uniquely. The major problems in identifying the DAB are the presence of other meaningful regions namely, postal stamp, post-office seal, return address and other graphics. Detection of pin-codes in such an unconstrained environment is a challenging task. Indian postal documents can be broadly categorized as structured (e.g. postcard, inland letters, etc.) and unstructured (e.g. simple and printed envelopes, business letters etc.). It was also found that, even though a pre-printed Pin-code box is present in structured documents/envelopes, individuals tend to write the pin-code outside the box. Examples of some postal documents with different regions marked are shown in the figure 1. Figure 1(a, b) are examples of Postcards and the areas of interest such as Pin-code box (in red), destination address block(in green), stamp (in blue) and seal (in yellow) are marked. Additionally, Figure 1(b) shows that the Pin-code was written outside the pre-printed Pin-code box, making its detection a challenging task. Figure 1(d) is an example of Pin-code written on an envelope without any pre-printed Pin-code box. Figure 1(d) is a very common senario in the Indian postal system where Pin-code is written in an unconstrained manner.



(a) Postcard (b) Postcard

(c) Inland Letter (d) Envelope

Fig. 1: Sample of different types of Indian postal documents and the regions of interest used for postal automation.

Among the recent works on India postal automation are [1], [9], [10], [5], [6], [7], [8]. Roy et. al. [1], [9] present a system for Indian postal automation. Roy et. al. [1] used Run Length Smoothing Algorithm (RLSA) along with positional information to detect the destination address block (DAB), seal, stamp and Pin-code box. The Pin-code recognition was done using MLP neural network. In [9], NSHP-HMM was used for the recognition of city names. On the contrary, most of the other works focused on Pin-code string recognition [8], city name recognition [7], [6], [5] and handwritten script identification [9], which forms the various steps in processing multi-lingual Indian postal documents. Although many published article exists in the literature, detection of Pin-code box and unconstrained handwritten pin-codes is still a challenging task, and has not been explored much.

In this paper, we proposed to use the state-of-the-art Deep Convolutional Neural Networks for detection of various regions namely, DAB, Pin-code box, handwritten Pin-code, seal and stamp in India postal documents. Specifically, we analyze the potential of Faster Region-based Convolutional Neural Networks (RCNN) [14] for the detection of the areas-of-interest and adapt it to the current problem. Three different network architectures namely Zeiler and Fergus(ZF)[15], Visual Geometry Group(VGG16)[16] and VGG_CNN_M_1024[17] were used in the study. The primary intension of the present study is to model region segmentation task in postal documents, as a standard object detection problem. The study explores object detection methods which can detect Pin-code box/region at real-time, in a single pipeline and can eventually be used for Pin-code based sorting of Indian postal documents.

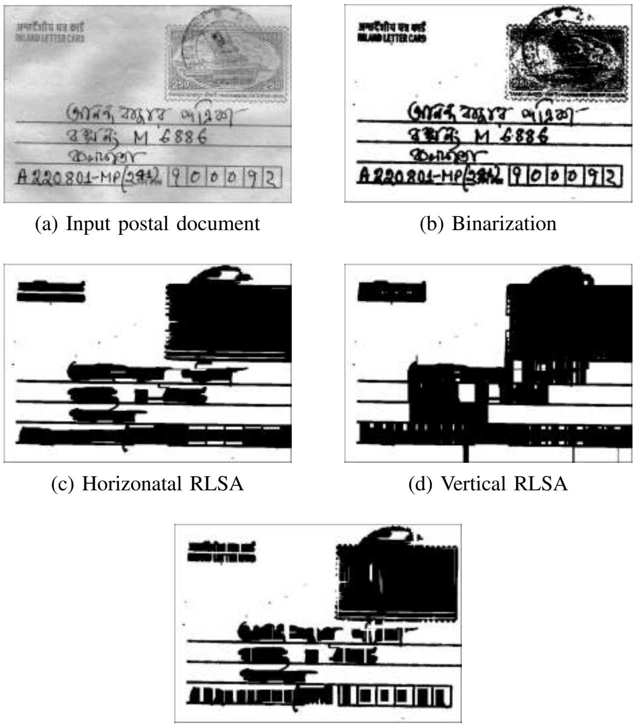
The paper is organized as follows. The related works on pin-code detection and Deep CNN based object detection were discussed in section 2. The proposed method in presented in section 3. In section 4, experimental analysis and result are discussed. Finally, the paper is concluded in section 5.

II. Related Works

In this section, the current state-of-the-art on the Indian postal automation and object detection using Deep CNNs are disscused. Related works on pincode detection from Indian postal documents are discussed in section 2A. In section 2B, the recent Deep CNN-based object detection methods are reviewed.

A. Pincode detection methods

Among the relevant works on Indian postal automation, Roy et. al. [1], [9] presented a system for Indian postal automation. Traditional image processing and machine learning techniques were used for the detection of Pincodes from postal document in [1], [9]. This techniques involved pre-processing the postal documents to convert them into binarized image or two-tone image, followed by a smoothing operation to remove isolated and spurious pixels. The binarized image was used decompose to the postal document into text and non-text (e.g. seal, stamp, etc.) blocks using the Run-Length Smoothing Algorithm (RLSA). RLSA was applied to the binarized image in vertical and horizontal directions. The two result images were combined using a logical AND operation. An example of the results obtained after each operations are shown in the figure 2.



(a) Input postal document

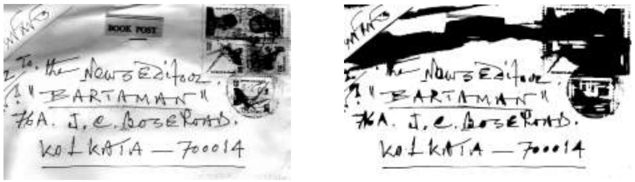
(b) Binarization

(c) Horizonatal RLSA

(d) Vertical RLSA

(e) Logical AND of (c) and (d)

Fig. 2: Processing structured Indian postal documents to detect pincode box



(a) Input postal document

(b) Logical AND of vertical and horizontal RLSA

Fig. 3: Processing unstructured Indian postal document to detect pincode region

Component analysis is applied on the result obtained after logical AND operation, to obtain blocks. Each block is then checked for postal seal or stamp based on the black pixel density. It can be seen from the image in figure 2e that the stamp/seal block have a very high black pixel density. The identified non-text part are then deleted and the remaining components are analysed for DAB and pincode box detection. Positional information such as, DAB is generally located in the lower-right portion of the document is used to segment DAB. Pincode boxes general have its length around six times the width, as a pincode in India has 6 digits. This clue was used to form rules to detect the pincode box.

This techniques are heuristic in nature and involves a number of thresholds. This reduces the its generalization capability on unconstrained situations. Moreover, this method will fail to detect pincode when the pincode box is absent or while processing unstructured envelope. In case of unstructured envelopes, no specific location is provided for writting the pincode, which makes it a very challenging task. An example of unstructured envelope and result of logical AND operation after vertical and horizontal RSLA is shown in figure 3. The result of logical AND operation shows that it is very challenging to detect the pincode regions, as the other text component also satisfies the six digit pincode rule. Moreover, most of the earlier works focused on optical character recognition of the handwritten pincodes [8] and city names [7], [6], [5]. Whereas, detection of pincodes regions, especially in unstructured postal documents is still a challenging problem.

Hence, in this paper we examine the potential of Deep CNNs for detecting Pincode in both structured and unstructured documents, using a common pipeline of operations, by modelling it as a standard object detection problem.

B. *CNN-based object detection methods*

In this section, the current state-of-the-art methods for object detection using Deep Convolutional Neural Networks (CNN) are discussed. In particular, a brief overview of R-CNN [13], Fast R-CNN [12], and Faster R-CNN [14] is presented.

Recent advances in object detection techniques presented the community with Region-Based Convolutional Neural Network (R-CNN) and its successors (Fast and Faster R-CNN). R-CNN [13] uses Selective Search (SS) to compute (2k) object proposals of different scales and positions. For each of these proposals, image regions are warped to fixed size $(227X227)$ pixels. The warped image regions are then fed to the CNN for detections. The proposed network architecture uses classification head for classifying region into one of the classes. The SS does not necessarily provide perfect proposals. Therefore, to make up for the slightly wrong object proposals, regression head uses linear regression to map predicted bounding boxes to the ground-truth bounding boxes. R-CNN is very slow at test time where every individual object proposals are passed through CNN. The feature extracted are cached to the disk. Finally, a classifier such as SVM is trained in an offline manner. Therefore, the weights of the CNN did not have the chance to update itself in response to these offline part of the network. Moreover, the training pipeline of the R-CNN is complex.

In Fast R-CNN [12] the order of the extracting region of proposals and running the CNN is exchanged. In this architecture whole image is passed once through the CNN and the regions are now extracted from convolutional feature map using ROI pooling. This change in architecture reduces the computation time by sharing the computation of convolutional feature map between region proposals. The region proposal are projected to the corresponding spatial part of convolutional feature volume. Finally, fully connected layer expect the fixed size feature vector and therefore the projected region is divided into grid and Spatial Pyramid Pooling (SPP) is performed to get fixed size vector. SPP deals with the variable window size of pooling operation and thus end-to-end training of the network is very hard. The generation of the region proposals is the bottle neck at the test time. In above mentioned approaches, CNN was used only for regression and classification. The idea was further extended to use CNN also for region proposals. The latest offspring from the R-CNN family, the Faster R-CNN [14] proposed the idea of small CNN network called Region Proposal Network (RPN), build on top of the convolutional feature map. A sliding window is placed over feature map in reference to the original image. The notion of anchor box is used to capture object at multiple scales. The center of the anchor box having different aspect ratio and size coincide with the center of sliding window. RPN generates region proposals of different sizes and aspect ratios at various spatial locations. RPN is a two layered network which does not add to the computation of overall network. Finally, regression provides finer localization with the reference to the sliding window position.

Although Faster R-CNN and its predecessors perform well with high accuracy, they are computationally very expensive and time consuming, make them undesirable for real-time applications. Faster-RCNN works at a rate of 7 frames per second, while maintaining high accuracy.

Based on the brief investigation of the state-of-the-art, Faster R-CNN was considered in this study for experiments on detecting Pincodes and other regions of interest from postal documents. Different CNN architectures were used with Faster R-CNN for analysis.

III. Proposed Method

A. *Dataset preparation*

The postal documents for the current work has been collected from an Indian post-office. A flatbed scanner was use for document digitization. Images are in gray tone with a resolution of 300 dpi and are saved in TIF format. For training Faster RCNN, ground truth/annotation were created for all documents in PASCAL Voc XML format. Specifically, Pin-code box, Pin-code region (unconstrained handwritten pin-code), AddressRegion (destination address block), Stamp and Seal were considered for annotation. A sample postal document with respective regions marked and corresponding XML annotation is shown in Figure 5. Pin-code is box marked in red, Address block in green, seal and stamp in yellow and blue respectively.

B. *Methodology*

Unlike the state-of-the-art methods, pre-processing techniques such binarization, smoothing, etc. were not applied on the postal documents. Raw images were considered as input to the system. Faster RCNN [14] with Caffe [19] deep learning library was considered for our experiments. The Caffe-based pre-trained models are publically available for most of the object detectors. There are less number of images in our dataset for a deep learning system to train from scratch. Hence, to take full advantage of network architectures, a transfer learning technique was used. Pre-trained model from ImageNet [18] was used for fine-tuning and adapting to our problem domain. In case of transfer learning, the CNN layers were initialized with the weights from the pre-trained Imagenet models, without the fully connected (FC) layers, as they are more dataset centric. Moreover, as the postal document dataset is quite different from the Imagenet dataset, initializing the

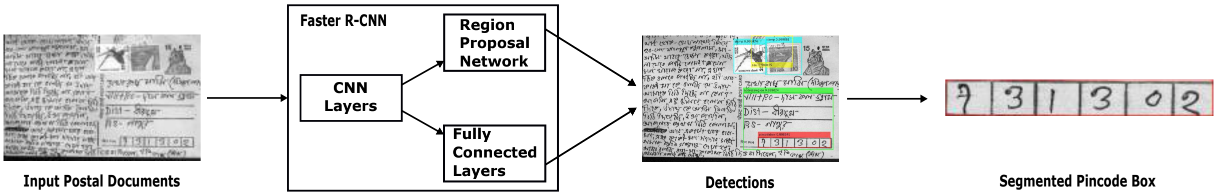
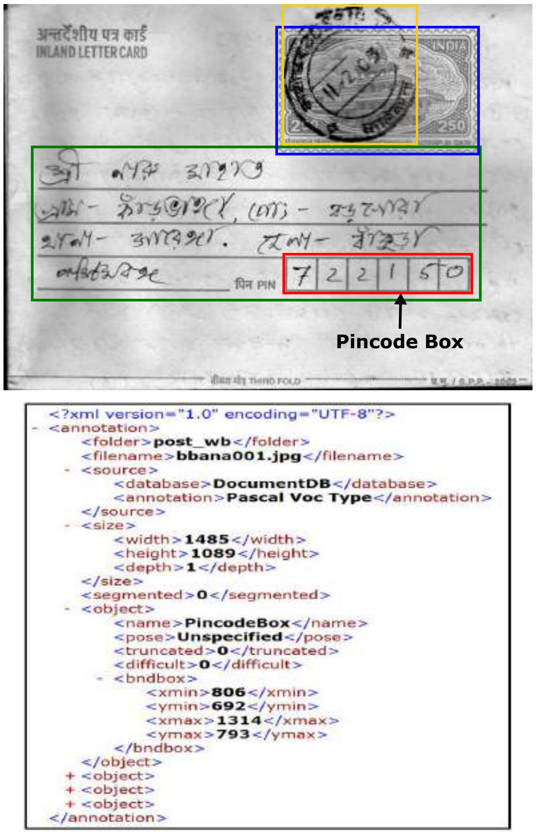


Fig. 4: Flow diagram for Pincode detection using Deep CNN



```
  <?xml version="1.0" encoding="UTF-8"?>
- <annotation>
      <folder>post_wb</folder>
      <filename>bbana001.jpg</filename>
    - <source>
          <database>DocumentDB</database>
          <annotation>Pascal Voc Type</annotation>
      </source>
    - <size>
          <width>1485</width>
          <height>1089</height>
          <depth>1</depth>
      </size>
      <segmented>0</segmented>
    - <object>
          <name>PincodeBox</name>
          <pose>Unspecified</pose>
          <truncated>0</truncated>
          <difficult>0</difficult>
        - <bndbox>
              <xmin>806</xmin>
              <ymin>692</ymin>
              <xmax>1314</xmax>
              <ymax>793</ymax>
          </bndbox>
      </object>
    + <object>
    + <object>
    + <object>
  </annotation>
```

Fig. 5: Sample postal document with annotation.

CNN layers with the pre-trained weights helps to re-train/fine-tune the network to converge faster. The FC layers uses the features from the CNN layers and classifies the data as applicable to the postal document problem. A flow diagram of the pipeline used in pincode detection is given in figure 4.

We have used various state-of-the-art network architectures such as ZF [15], VGG16 [16], and VGG_CNN_M_1024 [17] to train the system and evaluate the performance on the dataset. ZF is a 8 layered architecture containing 5 convolutional layers and 3 fully-connected layers. Whereas, VGG16 is a much deeper layered architecture with 16 layers, comprising 13 convolutional layers and 3 fully connected layers. On the contrary, VGG_CNN_M_1024 is a wider network architecture.

IV. Experimental Results and Discussion

The dataset used for experiments comprises of 2300 postal documents. Three different type of postal documents were considered, namely, postcards, inland letters and unstructured handwritten envelopes. The datset was divided in three subsets, for training, validation, and testing, with ramdom sampling. The train set consist of 50% for total dataset, whereas validation and test set consist of 10% and 40% of the total samples, respectively. Three different network architectures namely, ZF, VGG_CNN_M_1024 (VGG_M in short), and VGG16 were used for experiments. Implementations details and the detection results are discussed in the subsections given below.

A. Implementation Details

We trained our models with Nvidia Quadro P6000 GPU, 24GB, on a Ubuntu server (Core i7 processor, 64 GB RAM) with a learning rate of 0.001 and batch size of 64. The RPN batch size is kept constant at 128 for region based proposal networks (RPN). Regions proposal networks were trained end-to-end using backpropagation and stochastic gradient desecnt (SGD). In order to reduce redundancies arising from RPN proposals, non-maximum supression (NMS) was applied to the proposals based on the class scores. Performance of each network architecture at a different iterations was also analysed. In the training phase, the snapshot of trained models were saved at an interval of $10k$ iterations. Detections with overlap greater than the 50% Intersection Over Union (IOU) threshold with the corresponding ground-truth bounding box are considered as true positive and all other detections as false positive as shown in Eq.1 [20].

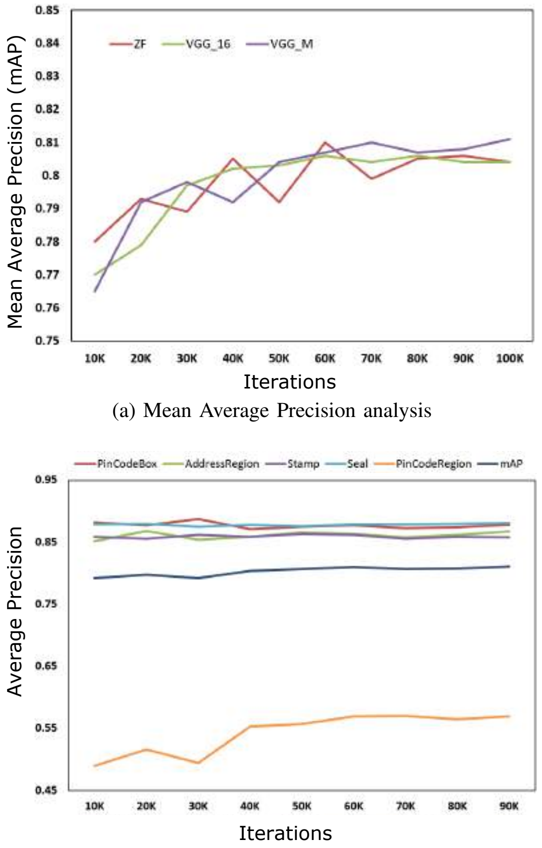
$$IOU = \frac{area\left(BBox_{pred} \cap BBox_{gt}\right)}{area\left(BBox_{pred} \cup BBox_{gt}\right)} \tag{1}$$

where, $BBox_{pred}$ and $BBox_{gt}$ denotes predicted bounding box and ground truth bounding box respectively. The ground truth box with no matching detection are considered false negative detection. To evaluate the detection performance, we use Average Precision (AP) calculated from the area under the Precision-Recall (PR) curve [20]. While, mean Average Precision (mAP) is used for a set of detections and is the mean over classes, of the interpolated AP for each class.

B. Detection Results

The detection results of different regions on postal documents are detailed in Table I. The results obtained considering the different architectures are given in the respective columns of the table. Among all the iterations, best results obtained for each network architectures are reported in Table I. The Table I shows that VGG_M performed better than ZF and VGG16.

Mean average precision of 0.810 (60K iterations), 0.806 (60K iterations) and 0.811 (100K iterations) was obtained for ZF, VGG16 and VGG_M, respectively. Average time taken for processing each image for detection was 0.044 seconds, 0.130 seconds and 0.048 seconds, for ZF, VGG16 and VGG_M, respectively. For Pin-code region detection, VGG_M performed better with an average precison of 0.569. Lower precision for the detection of Pin-code regions reflects the complexity involved in its detection due the unconstrained nature. On the contrary, performance of Pin-code box detection was better than other regions, due to its specific structure and location on postal documents.



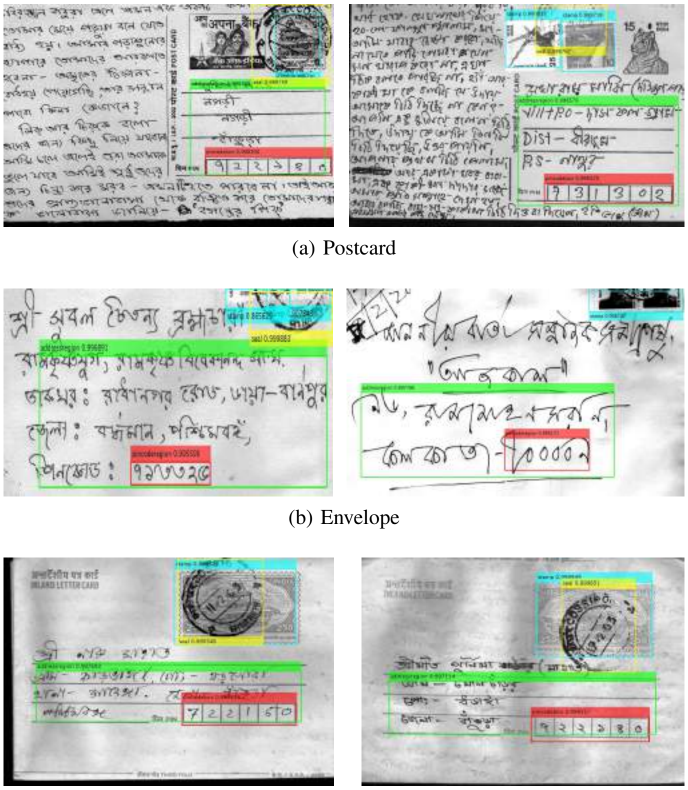
(a) Mean Average Precision analysis

(b) AP analysis of each class using VGG_M

Fig. 6: mAP and AP analysis at different iterations

TABLE I: Performance of various network architectures on test dataset.

Class	ZF	VGG16	VGG_M
PinCodeBox	**0.881**	0.873	0.878
PincodeRegion	0.554	0.544	**0.569**
AddressRegion	0.869	**0.870**	0.867
Stamp	**0.866**	0.860	0.858
Seal	0.879	0.881	**0.881**
mAP	0.810	0.806	**0.811**
Iterations	60K	60K	100K
Time (per image)	0.044 sec	0.130 sec	0.048 sec



(a) Postcard

(b) Envelope

(c) Inland Letter

Fig. 7: Sample detection results using VGG_M trained model. Pin-code boxes and region are marked in red, DAB in green, seal in yellow and stamp in cyan.

An analysis of mAP obtained at an interval of $10K$ iterations for each CNN architectures is given in the graph shown in Figure 6(a). Figure 6(b) shows the average precision obtained for each class using VGG_M model at 100K iterations. Sample detection results obtained from VGG_M trained model are shown in Figure 7. Samples of detection results on Postcards, unconstrained handwritten envelopes and Inland letters are shown in the Figure.

Although, Roy et. al. [1] reported results on detection of DAB, stamp/seal and Pin-code box, a comparative study cannot be performed as the experiments were not conducted on the same dataset, due its unavailability. Moreover, in their method [1], stamp and seal were considered as a single category and only Pin-code boxes were considered for detection. Explicit analysis on the detection of Pin-code region was not done/reported in the work. On the contrary, the current study explicitly considered, stamp, seal, Pin-code box, Pin-code region and DAB as different categories and detects them in a single pipline.

An errors analysis was also done on the incorrectly detected documents. It was found that multiple Pin-code regions were detected when there were other numeral strings such as phone number etc. present in the documents. Additionally, unclear, cluttered or too much of text on envelopes also resulted in false positives. An OCR based post processing will be investigated to resolve the false positive issue with Pin-code region detection.

V. Conclusion

In this study, detection of various regions on an Indian postal document is modelled as a standard object detection problem. Analysis of state-of-the-art object detection technique is explored in this paper, inorder to understand its potential in for Indian Postal Automation. Given the complexity involved in sorting Indian postal documents at real-time, use of Deep CNN looks very promising. Use of a single pipline for detection of different regions on a postal document is also explored for the first time, to the best of our knowledge. The results obtained from the experiments are very encouraging. The system was also able to detect unconstained handwritten Pin-codes as well as Pin-code box, successfully. To the best of our knowledge, this is the first study which considers Deep CNNs towards the develepment of a real-time Indian Postal automation system. Outcome of the present study is the basis for our future research. OCR guided post-processing of the detected Pin-code region candidates will enhance the performace of detection. Modification of the CNN architecture and multi-lingual OCR of the Pin-codes will be also considered. A larger dataset in under preparation for further experiments and will surely improve the performance of the Deep CNN networks.

References

[1] K. Roy, S. Vajda, U. Pal and B. B. Chaudhuri: A system towards Indian postal automation. Ninth International Workshop on Frontiers in Handwriting Recognition, pp. 580-585. 2004

[2] U. Mahadevan, and S. N. Srihari: Parsing and Recognition of City, State, and ZIP Codes in Handwritten Addresses. In Proc. of Fifth ICDAR, pp. 325-328, 1999

[3] X. Wang, and T. Tsutsumida: A New Method of Character Line Extraction from Mixed-unformatted Document Image for Japanese Mail Address Recognition. In Proc. of Fifth ICDAR, pp. 769-772, 1999

[4] S. N. Srihari, and E.J. Keubert: Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System. In Proc. of Forth ICDAR, pp. 892-896. 1997

[5] U. Pal, R. K. Roy and F. Kimura: Multi-lingual City Name Recognition for Indian Postal Automation. 2012 International Conference on Frontiers in Handwriting Recognition, Bari, pp. 169-173, 2012

[6] U. Pal, R. K. Roy and F. Kimura: Handwritten Street Name Recognition for Indian Postal Automation. 2011 International Conference on Document Analysis and Recognition, Beijing, pp. 483-487, 2011

[7] U. Pal, R. K. Roy and F. Kimura: Bangla and English City Name Recognition for Indian Postal Automation. 2010 20th International Conference on Pattern Recognition, Istanbul, pp. 1985-1988, 2010

[8] U. Pal, R. K. Roy, K. Roy and F. Kimura: Indian Multi-Script Full Pincode String Recognition for Postal Automation. 2009 10th International Conference on Document Analysis and Recognition, Barcelona, pp. 456-460, 2009

[9] K. Roy, S. Vajda, U. Pal, B. B. Chaudhuri and A. Belaid: A system for Indian postal automation. Eighth International Conference on Document Analysis and Recognition (ICDAR'05), pp. 1060-1064 Vol. 2, 2005

[10] K. Roy, U. Pal and B. B. Chaudhuri: Neural network based wordwise handwritten script identification system for Indian postal automation. Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, pp. 240-245, 2005

[11] P. S. R. Prasanna, S. Balaji, T. H. Khezhie, C. Vasanthanayaki and S. Annadurai: Destination address interpretation for automating the sorting process of Indian Postal System. TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, pp. 858-862 Vol.2., 2003

[12] R. Girshick: Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, pages 14401448, 2015

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik: Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580587, 2014

[14] S. Ren, K. He, R. Girshick, and J. Sun: Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 9199, 2015

[15] M. D. Zeiler and R. Fergus: Visualizing and understanding convolutional networks. In European conference on computer vision, pages 818833. Springer, 2014

[16] K. Simonyan and A. Zisserman: Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2014

[17] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman: Return of the Devil in the Details: Delving Deep into Convolutional Nets. British Machine Vision Conference (BMVC), 2014

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei: Imagenet A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. IEEE Conference on, pages 248255. IEEE, 2009

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell: Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, pages 675678. ACM, 2014

[20] Everingham, Mark and Eslami, SM Ali and Van Gool, Luc and Williams, Christopher KI and Winn, John and Zisserman, Andrew: The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, pages 98-136. 2015