Legal amount recognition in Bank cheques using Capsule Networks

Nisarg Mistry, Maneesh Darisi, Rahul Singh, Meet Shah and Aditya Malshikhare

Veermata Jijabai Technological Institute, Mumbai - 400019, India
nisargbm@gmail.com, maneeshdarisi97@gmail.com,
singhrahul1497@gmail.com, meetshah145@gmail.com,
malshikhareaditya@gmail.com

Abstract. Legal amount detection is a decade old conundrum hindering the efficiency of automatic cheque detection systems. Ever since the advent of legal amount detection as a use-case in the computer vision ecosystem, it has been hampered by the deficiency of effective machine learning models to detect the language-specific legal amount on bank cheques. Currently, convolutional neural networks are the most widely used deep learning algorithms for image classification. Yet the majority of deep learning architectures fail to capture information like shape, orientation, pose of the images due to the use of max pooling. This paper proposes a novel way to extract, process and segment legal amounts into words from Indian bank cheques written in English and recognize them. The paper uses capsule networks to recognize legal amounts from the bank cheques which enables the shape, pose and orientation detection of legal amounts by using dynamic routing and routing by agreement techniques for communication between capsules and thus improves the recognition accuracy.

Keywords: Legal amount Detection, Capsule Networks, Bank Check Recognition, Handwriting Recognition, Image Processing.

1 Introduction

Automated cheque detection systems is an active field of research. This topic has a high relevance in the finance industries where in enormous number of cheques are processed on a daily basis. To create such a system several image recognition techniques, pattern recognition techniques are used. Legal amount extraction and recognition problem is a language specific problem. People widely have divided the detection of legal amount from bank cheques into three major phases: 1. Location of legal amounts on the bank cheques and preprocessing them. 2. Segmentation of preprocessed bank cheques image into legal amount words. 3. Recognition of segmented legal amount words using diverse machine learning techniques.

Indian bank cheques have fixed layouts for various fields on the bank cheque and provide two baselines for writing the legal amount. Most basic step for detection is identifying the zones of interest. Identification of the baselines based on the layout of the bank cheque and extracting the legal amount based on the baselines is the primary task. Various techniques such as binarization[1], recursive thresholding[2], Arnold transforms[3], Least Square fitting, Hough transform[4], Gaussian filtering[5], noise elimination are used to detect legal amounts from the baselines. After the detection of the legal amount baselines, the next step is to segment the legal amount into subunits using either the holistic or analytical approach.

The next step is to recognize the segmented legal amounts. The field of handwriting recognition is a field under development with many innovative techniques coming up to tackle the problem. In recent times, CNNs or Convolutional neural networks[6] are used as the optimal and efficient deep learning neural network algorithms in most of the image classification tasks. One of the drawbacks of CNN is it uses max-pooling for down-sampling the data before it is passed to the next layers within the CNN architecture. They ignore the spatial details between objects and fail to capture spatial relationships between pixels of the image[7]. Capsule network architecture[8] was proposed to tackle the problem of invariance. It is a deep learning architecture which preserves the spatial information. The paper proposes to use capsule networks for handwriting recognition of legal amounts.

2 Literature Survey

The development of every computer vision problem lies in the variety of data sets that it has been trained or tested on. Few of the common datasets that are used for legal amount detection are CENPARMI[9], ISIHWD. ISIHWD[10] consists of 62 legal words which are written by 105 writers. It consists of 31124 handwritten words which are split into a training set of 24924 words and a test set of 6200 words to provide a writer agnostic word recognition and a common platform in writer independent word recognition. This paper also uses IDRBT[11] dataset which is a cheque image dataset consisting of 112 cheque leaves from four different banks in India with diverse texture and ink color.

A method based on baselines (guidelines) is used in [12] and [13] to extract the handwritten date, courtesy and legal amounts of Canadian bank cheques. The guideline for legal amount is found by analyzing the lengths of the lines extracted through edge detection. A searching region and a bounding region are decided for each field, and the grey-scale distributions of the handwritten strokes related to each item are extracted by tracing the connected components.

For segmentation, two of the most popular approaches are analytic approach[14] and holistic approach[15] The analytic approach involves segmenting of the entire zone of interest into multiple subparts and then recognizing the subparts. The overall accuracy of the zones of interest is found by aggregating the individual accuracy of the subparts. Thus, proper segmentation is an essential step of the analytic approach, which impacts the accuracy of the recognition of subparts. The holistic approach revolves around the detection of the legal amounts by extracting features from the legal amount without dividing them into subunits. However, both the above methods used for segmentation fall into the problem of over-segmentation and under-segmentation.

Various approaches have been put forward for legal amount recognition. [16] uses slant correction, noise detection and smoothing for preprocessing and Hidden Markov Model(HMM) model trained using Baum-Welch Algorithm and the Cross-Validation process for detection of legal amounts. Accuracies vary from 82% - 93% depending on the classification of handwritten words but suffers from the fact that the legal amounts detected are not case invariant.

[17] provides a holistic method of segmentation to avoid segmentation and stage and segmentation errors. It uses a graph numerical recognizer based input subgraphs to graphs of symbol prototypes. The approach finds similarities between the courtesy and legal amount recognition accuracies using cross validation which achieves 20% improvement in recognition. [18] involves creating a legal amount vocabulary which could be used to handle mixed cursive and discrete legal amount. Pre-processing and feature extraction of mixed cursive and discrete features are used to improve the recognition accuracy. The accuracies using the approach given in this paper varied from 75% - 81%. This approach is unable to handle the spatial orientation features of the handwritten legal amounts. Previous works also utilize multidimensional LSTMs to recognize handwriting [19] however more recent works suggest that similar performances can be achieved with a 1D-LSTM[20]. Briefly, [20] utilized a CNN for image feature extraction and fed the features into a bidirectional LSTM and trained the network to optimize the Connectionist Temporal Classification (CTC) loss.

3 Implementation

3.1 Preprocessing

Preprocessing is used to remove unnecessary details from the cheques and locate and extract zones of interest. The pre-printed straight lines on the cheque also called guidelines are used to locate interest zones for detecting the legal amount on the cheques.

Fig. 1. Preprocessing phases

Implementation details.

1. Conversion of the colored bank cheque image to a grayscale image using thresholding followed by the binarization of the grayscale image.
2. Then pass the binarized gray image to a gaussian blur filter to remove the noise and smoothen the sharp baselines. Then 3 iterations of a dilation using the Morph cross operator of size (5,2) are applied. Based on the layout analysis of the bank cheque, baselines whose length is less than (cheque image length)/5 using a horizontal kernel are eliminated. The output of these operations leave us with payee and legal amount baselines on the cheque.
3. The edges of the payee and legal amount baselines were detected using the Canny detector. Hough Transform was used to isolate the baselines of a particular length from the image to obtain the coordinates of the baselines.
4. After the detection of baselines, the baselines are sorted according to the y-coordinates. The bottom and top sections of the cheque are truncated. The area between the extracted baseline contours are the zones of interest which contain handwritten fields. Height h is the distance between two adjacent contours. But the height of the first line of handwritten field is taken as the distance between the top of sliced image and first baseline contour. The final contours extracted are of height $h' = h + (2 *\Delta)$ where Δ is the margin of error field. The optimal value of delta should be in the range [0.1* h', 0.2* h']. The contours obtained in this step are stored in set C. Set C denotes a set of contours $[c_1,c_2...c_k]$
5. Create a 2d matrix of zeros whose size corresponds to the size of the sliced cheque image. The set C contains contours c_i of the previous step. Apply the algorithm as in Code Snippet 1.
6. For every contour, find all the lines whose length is less than cheque image length/2 using a horizontal kernel and subtract those from the original contour image.

Code Snippet. 1.

```
Consider two lists previous P and next N and set
C = [c1,c2...ck].
P → c1
Foreach Contour ci ϵC i ←1 to i ←n-1 step do
  N = Ci
  If (overlap(P, N)) then
    i = i + 1
  Else
    P = N
```

```
Overlap function
  Input: Two contours C1 and C2
  Each Ci is four-tuple(x,y,w,h) where x: top left contour
x coordinate of Contour i
  Y: top left contour y coordinate of Contour i
  w: width of the contour of Contour i
  h: h height of the x contour of Contour i

Matrix m is of size (W, H) where W is the width of sliced
cheque and height of sliced cheque.
m[0:W][0:H]=0
Foreach Contour ci ϵC i ←1 to i ←2 step do
  m[Ci.x : Ci.x + Ci.w][Ci.y : Ci.y + Ci.h] += 1
  number_of_2s = Σnumber of 2s in m
  number_of_1s = Σnumber of 1s in m
  If (number_of_2s/ number_of_1s > threshold)
    Return true
  Else
    Return false
  Where threshold is taken as the ratio of overlapping
regions and non-overlapping regions
```

The payee and legal amount baselines found in the previous step can be discontinuous in nature. Therefore, these discontinuous lines are merged into a single smooth baseline based on the coordinates of the legal amount baseline. The mechanism to merge the discontinuous baselines is as follows. Consider a bank cheque image I

1. GB a gaussian blur filter is used on this grayscale image to remove the noise smoothen the sharp baselines. $I_2 \rightarrow GB(I)$
2. Apply 3 iterations of dilation with a (5,1) kernel to merge the discontinuous baselines.
3. Apply horizontal kernel on the image to find all long horizontal lines. Eliminate all the horizontal lines by converting the all pixels of horizontal lines to black.

The below terminologies defined would be used ahead

Perfectly segmented. A word is said to be perfectly segmented if the bounding box contains exactly one word. Fig.2.1.

Under segmented. A word is said to be under segmented if the bounding box contains one word along with auxiliary content (characters/lining/dots/ligatures). Fig.2.2.

Over segmented. A word is said to be over segmented if the bounding box does not contain all the characters of the word on which segmentation was applied. Fig.2.3.

Threshold. Threshold is the maximum horizontal distance of pixels between two bounding boxes on an image matrix.

Fig. 2. 1.Perfectly segmented 2.Under-segmented 3.Over-segmented

3.2 Segmentation

This is the phase wherein the extracted zones of interest are broken down into subunits. These subunits are then provided as input for recognition.

Implementation

In this phase, the legal amount is segmented into words. The main concern of this phase is the writer's handwriting which varies in size, height, width, skewness and style of the content written. The main steps followed to produce the optimal results out of the workflow as per Fig.3:

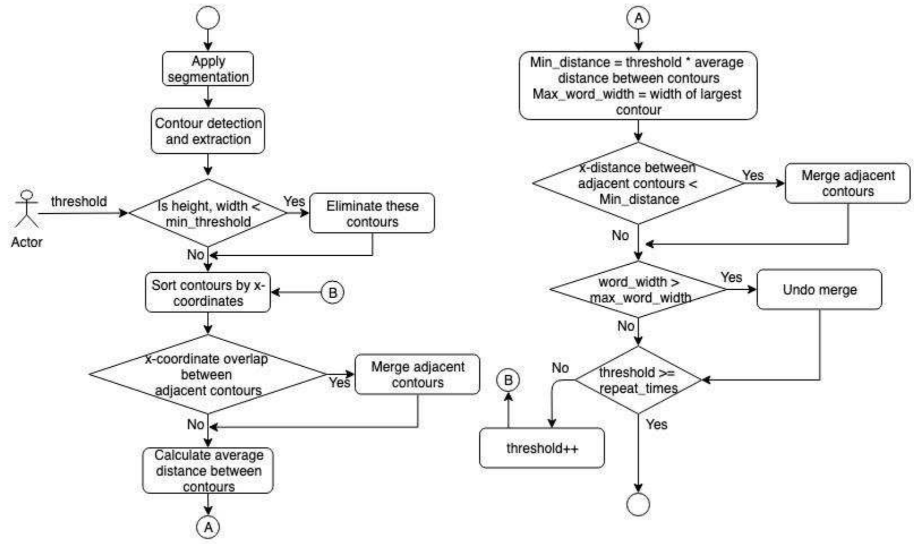

Fig. 3. Segmentation steps

1. Contour detection technique is used for extracting out the handwritten fields within the legal amount image. Based on the handwriting of the writer, the detected content varies from a single stroke of an alphabet to multiple words of legal amount leading to under-segmented and over-segmented contours. Eg. Fig.2
2. The approach used to handle these highly varying contours is to merge or split the contours based on the horizontal distance between the contours.
3. The next step is to eliminate all contours, which are detected due to noise within the image. Minute contours are not required and hence, are eliminated.

4. The contours are then sorted based on x coordinates.
5. Next step is to merge all the adjacent contours with overlapping x-coordinates so that they do not interfere in minimum distance calculation between words.
6. Minimum distance(Min_distance) is the product of threshold and the average distance between 2 adjacent contours(avg_distance). Take the width of the largest contour as the maximum possible width(Max_word_width) of any possible word.
7. The next step is to compare the distance between adjacent contours. If the distance is less than min_distance (threshold * avg_distance), then merge the adjacent contours.
8. However, if the width of the newly formed contour is greater than the maximum possible width of a word(Max_word_width), then undo the merge of contours.
9. Repeat step 1 to 8 for n iterations where n is defined by the user.

Table 1. IDRBT[11] DATASET LEGAL AMOUNT SEGMENTATION RESULTS

Threshold	Perfectly segmented (Number ‖ %)	Under-segmented (Number ‖ %)	Over-segmented (Number ‖ %)	Total
2	382 ‖ 75.94	70 ‖ 13.92	51 ‖ 10.14	503
3	383 ‖ 76.14	79 ‖ 15.71	41 ‖ 8.15	503
4	382 ‖ 75.94	82‖16.3	39 ‖7.76	503

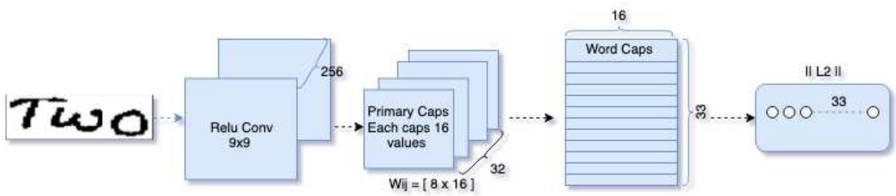

Fig. 4. CapsNet architecture

3.3 Capsule networks

Convolutional Neural Network(CNN)[21] is considered state-of-the-art for any image classification task. However, the use of max pooling in CNNs lead to loss of necessary info and make CNNs invariant to rotations and translations which has been tackled by Capsule Networks. The Capsule Networks architecture is composed of capsules[22] and each capsule is a group of neurons. Each capsule represents an activity vector that encodes entity properties like position, shape, rotation, stroke thickness, skew, width, etc. also called as instantiation parameters. This architecture uses Routing by Agreement algorithm[8] to train the capsules. For legal amount recognition, this paper uses

Capsule networks to take advantage of the positional information of various handwritten legal amount words within the input image and provide a better recognition rate. Architecture used in this paper is as per Fig.4.

Architecture

1. Dimensions of the input provided: 30*100*1.
2. First convolution layer is a convolutional layer(Conv1) which has 256, 9 × 9 convolution kernels with a stride of 1 and ReLU activation. This layer converts pixel intensities to the activities of local feature detectors. These are provided as inputs to the primary capsules.
3. The second layer is a convolution capsule layer called the Primary capsule layer having 32 channels of 8D capsules. Each capsule contains 8 convolutional units with 9 x 9 kernel and a stride of 2. The inputs are 256 × 81 Conv1 units from the previous layer whose receptive fields overlap with the location of the center of the capsule. The output of the primary capsule layer consists of [32 × 6 × 6] capsule outputs (each output is an 8D vector) and each capsule in the [6 × 6] grid is sharing their weights with each other.
4. The final layer (WordCaps) consists of a single 16D capsule per class. These capsules receive input from all the capsules in the previous layer.

The Capsule Network architecture uses a **squashing function**[23] as a non-linear activation function which provides a probability whether a particular feature is present within the image or not. This function shrinks the length of the output vector into a value between 0 and 1. Greater the length of the vector greater the probability of that particular capsule related to that object. The squashing function is given by Fig.5 where s_j=Total input, v_j=Vector Output of capsule j.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \qquad v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}$$

Fig. 5. 1.Summation matrix calculation 2.Squash function

Capsule networks use **routing by agreement algorithm**[24]. Unlike max-pooling, routing-by-agreement algorithm passes only useful information and throws away the data that would add noise to the results. Routing algorithm takes care of selection and routing relevant results to further layers. The capsules which are at a higher level represent complex entities having more degrees of freedom. Refer to Fig.5 where S_j=summation matrix, $u_{j|i}$=prediction vector, and c_{ij}=coupling coefficients determined by iterative dynamic routing.

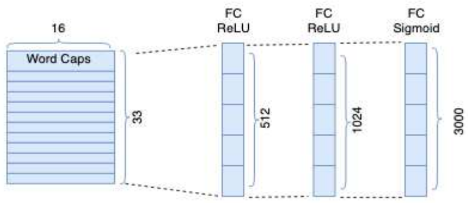

Fig. 6. Reconstruction architecture for regularization

Reconstruction as a regularization method. Reconstruction loss is used to encourage the word capsules to encode the instantiation parameters of the input. During the training phase, the output of the word capsule is classified as a specific word, and this activity vector is used to reconstruct the input image. The reconstruction architecture consists of 3 fully connected layers and uses minimization of the sum of the squared error loss function. Reconstruction loss is scale down by 0.0005 so that it does not dominate the margin loss during training. The reconstructions are done using the 16D output vector from the CapsNet.

Margin loss. Longer the output vector greater the probability of the presence of capsule entity. The model uses margin loss for correctly classifying the capsule outputs. For different classes, separate margin loss is used, L_k for each class capsule k:

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda (1 - T_k) \max(0, \|v_k\| - m^-)^2$$

where $T_k = 1$ if a word of class k is present and $m^+ = 0.9$, $m^- = 0.1$ and $\lambda = 0.5$. λ is a regularization parameter which helps prevent shrinking of the activity vectors during the learning phase. The total loss is the summation of the losses of all capsules.

For training of the CapsNet[22] model, ISIHWD[10] dataset has been used. Capitalized words and lower case words have been taken into a single category. The dataset consists of 31124 handwritten words split into a training set of 24924 words and a test set of 6200 words to provide a common platform in writer independent word recognition.

This model has been trained for a total of 30 epochs and the Adam optimizer has been used with a learning rate of 0.001 and learning rate decay factor of 0.9 after every epoch. Mean squared error (MSE) has been used as the reconstruction loss and the coefficient for the loss is loss_recon=0.0005 x (Image dimensions) = 0.0005 x 30 x 100 = 1.5 which is equivalent to using SSE (sum squared error) and loss_recon=0.0005

Capsule network result.

The trained Capsnet Model has achieved 98.05% test accuracy and 98.13% training accuracy after 29 epochs on ISIHWD dataset. Fig.7 shows the epochs vs accuracy graph for training phase and test phase. The training accuracy converged at around 29 epochs with an accuracy of 98.13%. The testing accuracy achieved is 98.05% on ISIHWD dataset.

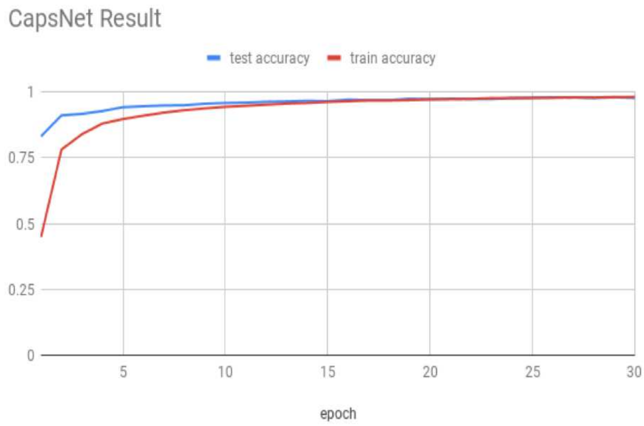

Fig. 7. CapsNet result on ISIHWD dataset

Table 2. ISIHWD DATASET [14] Performance Comparison.

	Approach	Classes	Training Size	Test Size	Training Epochs	Recognition Accuracy
Das Gupta et al. [14]	DCNN+SVM	33	24925	6200	40	95.35
Proposed System	CapsNet [8]	33	24738	6386	29	98.06

[10] uses a DCNN+SVM architecture which is a deep convolutional neural network and classifier approach. This approach is able to achieve a recognition accuracy of 95.35% over 40 training epochs. The capsule network architecture is able to achieve a recognition accuracy of 98.06% over just 29 epochs. The table provides a comparison of word wise accuracy achieved on ISIHWD dataset for CapsNet and the DCNN-SVM approach mentioned in [10].

Fig.8 provides the accuracy with which each of the 33 words are predicted on test data. Words like hundred, thousand, lakh, five are predicted with the highest accuracy. Certain words like forty and crore are predicted with an accuracy of 93%. This is due to ambiguity between pair of words like (forty, fifty) and (one, crore) written in cursive handwriting.

Fig.9 compares the word wise accuracy of the CapsNet architecture vs DCNN-SVM approach on ISIHWD dataset. Based on the results, certain words are predicted better by DCNN-SVM and certain words are predicted better by CapsNet. The model performs better in case which have common word parts like 'teen' in thirteen, fourteen, fifteen and so on.

Fig. 8. CapsNet on ISIHWD dataset

The trained model has been used on the extracted words from the segmentation phase Table.3 shows the results of CapsNet on IDRBT bank for perfectly segmented and over-segmented category of words. On analysis, the experimentation proves that under segmented category words could not be recognized by CapsNet as it contained more than one legal amount word.

4 Improvements and future scope.

This paper proposes a solution which is relevant for Indian bank cheques written in the English language and could be improved to provide support for language and layout agnostic legal amount bank cheque detection. The extraction process can be simplified by using the template matching techniques. A way to extract the desired handwritten fields is by using a blank cheque template and subtracting it from the original cheque image. With the absence of guidelines in the above approach, it will be challenging to identify the handwritten field.

This paper relies on the user for multiple threshold values for the legal amount extraction and segmentation phase. This user dependency provides fine-grained control over-extraction and segmentation phases presently. However, for complete automation of the algorithm, the solution could compute accuracies for an optimal range of values for these thresholds. Data Quality and variance plays a key role in deciding the accuracy of the computation. This solution's classification accuracy can be fine-tuned further by training on more extensive and diversified bank cheque databases which consist of several bank cheque layouts.

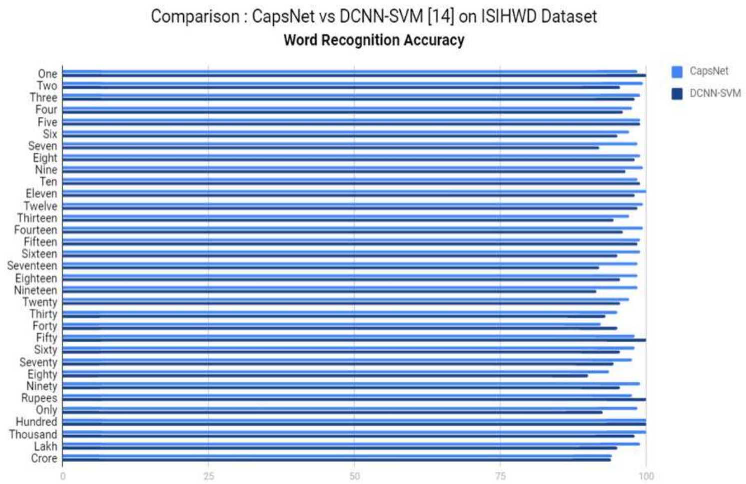

Fig. 9. CapsNet vs DCNN+SVM

Table 3. IDRBT BANK CHEQUE DATASET[11] LEGAL AMOUNT RECOGNITION RESULTS

Threshold	Perfectly segmented recognition accuracy (Percentage)	Over segmented recognition accuracy (Percentage)
Default	82	16.8
1	84.5	30
2	84.7	31
3	85	28
4	84.7	23.8

5 Conclusion

This paper provides an end to end legal amount extraction solution for Indian bank cheques written in English. This paper has emphasized its experimentation on the segmentation and preprocessing algorithms used for extracting the legal amounts. The root cause of the decrease in detection accuracy is identified as inaccuracies in the segmentation algorithms used previous works. Thus, the paper has attempted to improve the segmentation algorithms and has obtained around 76 % accuracy on the IDRBT dataset. This paper uses a promising deep learning architecture known as CapsNet for detecting legal amounts and has achieved 98.05 on training datasets and around 85% accuracy on the testing datasets. The further goals of this paper would be to provide language agnostic mechanisms to detect legal amount and to use myriad deep learning architectures to improve the detection accuracy.

References

1. Puneet P, Garg N (2013) Binarization Techniques used for Grey Scale Images. Int J Comput Appl. https://doi.org/10.5120/12320-8533
2. Arora S, Acharya J, Verma A, Panigrahi PK (2008) Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. Pattern Recognit Lett. https://doi.org/10.1016/j.patrec.2007.09.005
3. Wu L, Deng W, Zhang J, He D (2009) Arnold transformation algorithm and anti-Arnold transformation algorithm. In: 2009 1st International Conference on Information Science and Engineering, ICISE 2009
4. Mukhopadhyay P, Chaudhuri BB (2015) A survey of Hough Transform. Pattern Recognit. https://doi.org/10.1016/j.patcog.2014.08.027
5. Gedraite ES, Hadad M (2011) Investigation on the effect of a Gaussian Blur in image filtering and segmentation. In: Proceedings Elmar - International Symposium Electronics in Marine
6. Wu Y, Liu Y, Li J, et al (2013) Traffic sign detection based on convolutional neural networks. In: Proceedings of the International Joint Conference on Neural Networks
7. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K (2015) Spatial transformer networks. In: Advances in Neural Information Processing Systems
8. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. In: Advances in Neural Information Processing Systems
9. Haghighi PJ, Nobile N, He CL, Suen CY (2009) A new large-scale multi-purpose handwritten farsi database. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
10. Gupta J Das, Samanta S (2018) ISIHWD: A Database for Off-line Handwritten Word Recognition and Writer Identification. In: 2017 9th International Conference on Advances in Pattern Recognition, ICAPR 2017
11. Dansena P, Bag S, Pal R (2017) Differentiating Pen Inks in Handwritten Bank Cheques Using Multi-layer Perceptron. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
12. Liu KE, Suen CY, Cheriet M, et al (1997) Automatic extraction of baselines and data from check images. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/S0218001497000299
13. Liu K, Suen CY, Nadal C (1996) Automatic extraction of items from cheque images for payment recognition. In: Proceedings - International Conference on Pattern Recognition
14. Dimauro G, Impedovo S, Pirlo G, Salzo A (1997) Automatic bankcheck processing: A new engineered system. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/S0218001497000214
15. Kim KK, Kim JH, Chung YK, Suen CY (2001) Legal amount recognition based on the segmentation hypotheses for bank check processing. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR
16. De Almendra Freitas CO, El Yacoubi A, Bortolozzi F, Sabourin R (2000) Brazilian bank

check handwritten legal amount recognition. In: Brazilian Symposium of Computer Graphic and Image Processing
17. Dzuba G, Filatov A, Gershuny D, et al (1997) Check amount recognition based on the cross validation of courtesy and legal amount fields. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/S0218001497000275
18. Han K, Sethi IK (1997) An off-line cursive handwritten word recognition system and ITS application to legal amount interpretation. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/s0218001497000330
19. Graves A, Schmidhuber J (2009) Offline handwriting recognition with multidimensional recurrent neural networks. In: Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference
20. Puigcerver J (2017) Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition? In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR
21. Jmour N, Zayen S, Abdelkrim A (2018) Convolutional neural networks for image classification. In: 2018 International Conference on Advanced Systems and Electric Technologies, IC_ASET 2018
22. Xiang C, Zhang L, Tang Y, et al (2018) MS-CapsNet: A Novel Multi-Scale Capsule Network. IEEE Signal Process Lett. https://doi.org/10.1109/LSP.2018.2873892
23. Hinton G, Sabour S, Frosst N (2018) Matrix capsules with EM routing. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings
24. Peer D, Stabinger S, Rodriguez-Sanchez A (2018) Training Deep Capsule Networks. arXiv