Skew Detection of Bills using Deep Learning

P S Sai Ganesh · Siva Prakash Ramesh · Gnanadeep M · Vinayak K

Abstract The use of deep learning in day-to-day tasks is rapidly increasing. Different domains related to the field of capturing information uses DL to make the processes automatic. One such case is the capturing of crucial information from receipts using mobile cameras. We propose a method that can be used to identify the skew in captured images of bills. It is not always possible to obtain an image with almost perfect lighting or zoom, and the origin of the receipt can be from any country. Current solutions, using machine learning, have only been made to identify the rotation in bills that use the English language. They also fail to work in the case of a large skew in the image (>90). We aim to compare ten different state-of-the-art learning architectures like MobileNetV2, ConvNext etc. and provide the best working model, that will be able to successfully identify angles in the complete range of 0-360 degrees. Also, we consider this to be a classification problem rather than a regression problem, hence we create our own dataset with five different resolutions using the publicly available wildreceipts dataset. Upon comparison, the transformer models seem to perform the best among the ten models followed closely by DenseNet.

P S Sai Ganesh
Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
E-mail: cb.en.u4aie19044@cb.students.amrita.edu

Siva Prakash Ramesh
Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
E-mail: cb.en.u4aie19060@cb.students.amrita.edu

Gnanadeep M
Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
E-mail: cb.en.u4aie19041@cb.students.amrita.edu

Vinayak K
Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
E-mail: cb.en.u4aie19032@cb.students.amrita.edu

1 Related Works

Table 2 Literature Review

Papers Referred	Author (Year)	Title	Methodology	Dataset	Results	Major Findings	Scope for Improvement
Paper #1	Costin-Anton Boiangiu, Alex Dinu, Cornel Popescu, Nicolae Constantin and Petrescu Catalin-Dumitru (2020)	Voting-Based Document Image Skew Detection	Combination of FFT, projection profile and Hough transform	Private Dataset	The voting policy had an uptime of 99.5% by only misclassifying 2 images in the dataset.	Combination of different methods makes sure that if one algorithm fails, the others back it up	Changing the confidence measure of each algorithm can result in better results
Paper #2	Bezmaternykh P.V., Nikolaev D.P. (2019)	A Document Skew Detection Method Using Fast Hough Transform	Fast Hough transform	DISEC ‘13	The average error deviation is 0.547 for the range of angles –15 to 15 degrees	This method does not require any preprocessing step	Applying Fourier analysis before FHT will greatly improve results
Paper #3	Yue, Alex. (2018)	Automated Receipt Image Identification, Cropping, and Parsing	Line segment detector, probabilistic Hough transform, holistically nested edge detection	Private dataset	100% accuracy on images with skew of 0 to 90 degrees	Using probabilistic Hough transform instead of normal Hough transform saves time in computation and combination of multiple methods to detect lines increases accuracy	HED architecture takes 10 seconds for each image, the architecture size can be reduced
Paper #4	Shafii, M., Sid-Ahmed, M (2015)	Skew detection and correction based on an axes-parallel bounding box	Minimization of area made by axes-parallel bounding box	University of Maryland Tobacco800	There is no angle limitation and works for text, images and diagrams	Geometrical features of the document itself can be used to find the skew of the text	The method fails if text lines are not parallel to edges.
Paper #5	Jonathan Fabrizio (2014)	A Precise Skew Estimation Algorithm for Document Images Using KNN Clustering and Fourier Transform	Fourier transform	DISEC	The average error deviation is 0.072 and has more than 77% correct estimations	Performs better than Hough transform	Improve resource consumption without losing precision in the clustering step
Paper #6	Kumar, Deepak & Singh, Dalwinder (2012)	Modified Approach of Hough Transform for Skew Detection and Correction in Documented Images	Basic Hough transform	N/A	Speed of Hough transform increased for angles 0 to 45 degrees	Hough transform is slower compared to conventional methods	The range of angle detected can be increased
Paper #7	Panwar, Subhash & Nain, Neeta (2012)	A Novel Approach of Skew Normalization for Handwritten Text Lines and Words	Orthogonal projection of segmented line with respect to x-axis	Handwritten text	Works for all angles between 0 and 360 degrees with 98% accuracy	The baseline does not need to calculate explicitly as the orthographic projection with maximum length is the horizontal axis or the baseline itself	Work has been done only on a dataset containing only one handwritten word
Paper #8	A. Papandreou and B. Gatos (2011)	A Novel Skew Detection Technique Based on Vertical Projections	Vertical projection profile	French Textbooks	The average error deviation is 0.14 with 58% correct estimations for images with rotation of –5 to 5 degrees	Gives better results when compared to horizontal projection profile method. Also, the method is noise resistant	Increase range of angles detected

2 Introduction

Currently, the public prefers to capture key information such as the total amount, number of items purchased, items purchased, etc. from their receipts or bills automatically through their cameras as it is time consuming to manually keep track of their expenses. This is possible due to a process known as Optical Character Recognition (OCR)(Kumar et al., 2020)[9] which converts the textual information in the image to a machine-processable text that can be accessed by the computers. Though, the accuracy of OCR has improved significantly, it still fails to produce accurate results when there is any skew present in the image.

When the skew angle in the input image increases (both in the positive and negative side), the accuracy of OCR decreases (Jirasuwankul, 2011)[10]. A considerable amount of text (95%) is returned when the degree of skew in the image is less than 7 degrees, but any angle greater than that will produce illegible results that cannot be read. To extract the exact information from the bills, we need to correct the skew of the image such that the rotation angle of the image is between –5 degrees to 5 degrees. The existing methods to correct the orientation of the images do not work perfectly in all the scenarios.

The current regression algorithms that are commonly used for this task provide considerable results when the angle of skew is between –45 degrees to 45 degrees. One such method is the Hough transform (Bezmaternykh and Nikolaev, 2019)[2] (Kumar et al., 2012)[6], which detects the straight lines in the edge filtered input image. The algorithm assigns a score for each angle based on the number of lines present that are rotated by that specific angle, if the image is rotated in the opposite direction by the angle with the highest score, then the orientation of the bill is corrected. The second method is the projection profile analysis (Panwar et al., 2011)[8] which rotates the image by a small degree of angle in a loop and constructs a histogram plot for all the foreground pixels until a histogram is obtained where there is maximum difference in the peaks of the histogram. This corresponds to the image which is at zero- degree alignment. The Fourier Transform (Fabrizio, 2014)[5] algorithm finds the center point in the image around which an even symmetry exists, the idea is that if the image is rotated then the center line is also rotated along with it. So, finding the rotation angle of the center line will give us the orientation of the entire document. All the above-mentioned methods, work flawlessly when combined based on a voting system (Boiangiu et al., 2020)[1], the system has an uptime of 99.5% because if one method fails, then the other two methods make sure to back it up. But these methods fail to work when the skew angle is large (90, 180, 270) which is possible when the bills are captured through a camera due to human error. Also, the problem has always been formulated as a regression problem and not as a classification problem using deep learning.
This paper aims to

- Formulate the skew correction as a classification problem rather than a regression problem.
- Create the required datasets for training the models and perform classification task.
- Explore the usage of Deep Learning in Skew Correction and document the performance of various standard architectures
- To build a generalized solution which can identify and correct a wide range of skew angles.

We propose a method for training and comparing 10 different state of the art architectures that are chosen over a broad spectrum of lightweight models such as MobileNet and heavy models such as SWIN transformers. For this we create 5 different datasets with different degrees of rotation (180, 90, 30, 15, 5). The dataset with the highest degree of resolution (180) has only 2 classes while the dataset with the lowest degree of resolution has 72 classes. We also aim for low resource training i.e.; we use only 33% of the original dataset for training the models while the remaining 67% is left for testing the models. We also explore the effects of data augmentation and compare it with the results produced without data augmentation. The goal is to train a model that can identify the skew angles from 0 to 360 degrees in all conditions, which includes low resolution images or images with low lighting.

The experimental outcomes are as follows:

- Overall, the accuracies of all 10 models are higher during the case of without data augmentation when compared to with data augmentation case.
- But the best performing model with an average accuracy of 99% across all datasets is swin_base_patch4_window7_224_in22k that is obtained using data augmentation.
- When viewed as a classification problem, we can provide a model that works in almost all cases when compared to the regression algorithms that works only with a small skew.

3 Proposed Methodology

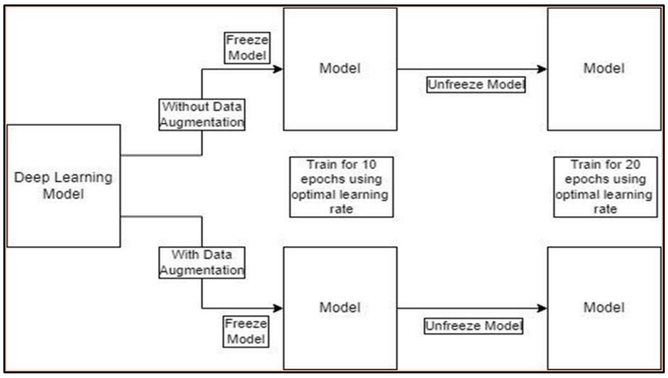

Fig. 1 Pipeline of proposed solution

The models chosen are over a broad spectrum of lightweight models such as MobileNet and heavy models such as Swin Transformers. We benchmark the performance of all models across the 5 different datasets with different resolutions. Our aim is to compare the different models and find out which one performs better for the task given. The models chosen are:

- DenseNet161 (densenet161)
- MobileNet (mobilenetv3_large_100_miil_in21k)
- Vgg19 (vgg19_bn)
- ConvNext (convnext_base_in22k)
- Inceptionv3 (inception_v3)
- ResNet50 (resnet50)
- InceptionResNet (inception_resnet_v2)
- ResNetv2 (resnetv2_50)
- Vision Transformer (vit_base_patch16_224_in21k)
- Swin Transformer (swin_base_patch4_window7_224_in22k)

Initially we train the model for 10 epochs after freezing the top layers, that is all the layers before the fully connected layer. This applies the concept of transfer learning and allows the model to adapt to our dataset quickly. Then we unfreeze the frozen layers and train for another 20 epochs. This produces results that are specific to our dataset.

4 Results and Discussion:

4.1 Dataset Description:

The datasets that are available publicly for image skew detection are prepared using images of documents. There is no publicly available dataset for the problem of skew detection in bill images. For this reason, we have synthesized our own dataset by using the images from the wildreceipts dataset as the base images. The wildreceipts dataset contains images of bills taken under various conditions.

Out of the 1765 images in the wildreceipts dataset, we took 1478 images which were properly aligned initially (0-degree skew). We use 33 percent of this for training and the rest for testing. The train dataset is created using 370 images and the test dataset is created using 1118 images.

We consider 5 different resolutions: 180, 90, 30, 15 and 5. For each resolution we keep rotating the base image by that resolution until we reach 360 degrees. In each rotation in each resolution a random error value between –2 degrees and 2 degrees is introduced.

Fig. 2 Receipts rotated by a) 60 degrees b) 0 degrees c) 180 degrees

Table 2 Dataset Description

Resolution	Number of Images		
	Training	Validation	Testing
180	564	176	2236
90	1127	352	4472
30	3380	1056	13416
15	6759	2112	26832
5	20276	6336	80496

4.2 Results

We get the accuracy of the model on the validation dataset during training as well as the final accuracy of prediction in the test dataset. We evaluate each model in this manner both before and after performing data augmentation. We further compare the performance of the best model before data augmentation with the performance of the best model after data augmentation based on accuracy and F1 score metrics.

Table 3 Train and Test accuracy of 10 models before data augmentation

Models	180		90		30		15		5	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
Vgg19	95%	97%	97%	98%	97%	99%	97%	99%	98%	99%
Dense161	**97%**	**97%**	**99%**	**99%**	**99%**	**99%**	**98%**	**99%**	**97%**	**98%**
ConvNext	98%	99%	95%	95%	98%	99%	97%	99%	98%	99%
ResNet50	96%	97%	97%	98%	97%	99%	98%	99%	99%	99%
InceptionV3	94%	95%	92%	95%	95%	96%	96%	98%	97%	98%
InceptionResnet	84%	85%	96%	97%	98%	98%	98%	99%	96%	97%
ResnetV2	66%	64%	98%	99%	99%	98%	99%	99%	98%	97%
MobileNet	97%	98%	98%	99%	97%	99%	98%	99%	98%	98%
Vit	95%	96%	97%	98%	98%	99%	98%	99%	97%	99%
Swin	97%	99%	88%	88%	97%	99%	97%	99%	98%	99%

(Column group header: **Degrees of Resolution**)

The best performing model for without data augmentation is Dense161 with an average accuracy of 98% while SWIN performs well for classifying above 30-degree resolution.

Table 4 Train and Test accuracy of 10 models after data augmentation

Models	180		90		30		15		5	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
Vgg19	94%	96%	95%	97%	95%	97%	94%	97%	90%	92%
Dense161	96%	97%	91%	93%	92%	94%	96%	98%	92%	93%
ConvNext	82%	74%	97%	99%	98%	99%	97%	98%	92%	95%
ResNet50	47%	49%	74%	77%	71%	71%	95%	97%	86%	89%
InceptionV3	89%	92%	86%	89%	95%	97%	95%	98%	87%	90%
InceptionResnet	88%	93%	98%	99%	91%	93%	96%	99%	86%	89%
ResnetV2	56%	59%	61%	63%	94%	96%	96%	97%	90%	93%
MobileNet	94%	97%	89%	92%	93%	95%	89%	92%	79%	82%
Vit	98%	99%	97%	99%	97%	99%	96%	99%	90%	94%
Swin	**98%**	**99%**	**98%**	**99%**	**98%**	**99%**	**97%**	**99%**	**96%**	**98%**

(Column group header: **Degrees of Resolution**)

Table 5 Comparision of best model before (Densenet) and after (Swin Transformer) data augmentation

Resolutions	Densenet				Swin Transformer			
	Without Data Augmentation		With Data Augmentation		Without Data Augmentation		With Data Augmentation	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
180	97%	97%	97%	97%	99%	99%	**99%**	**99%**
90	99%	99%	93%	93%	88%	88%	**99%**	**99%**
30	99%	99%	94%	94%	99%	99%	**99%**	**99%**
15	99%	99%	98%	98%	99%	99%	**99%**	**99%**
5	98%	98%	93%	93%	99%	99%	**98%**	**98%**

It can be observed that in almost all cases test accuracy is higher than validation accuracy. Densenet has the best performance when we train the models before performing data augmentation. Resnet V2 has the worst performance before performing data augmentation. After performing data augmentation Swin Transformers have the best performance. A decrease in accuracy can be observed in most all cases after performing data augmentation. On comparing the performance of DenseNet and Swin transformer it can be observed that Swin transformer after data augmentation has a better performance both in terms of accuracy as well as F1 score.

5 Conclusion

We have proposed a low resource learning solution to the problem of skew detection in bill images and converted the regression problem into a classifcation problem by creating our own dataset by manually rotating for each orientation. The performance comparison has been done across various deep learning architectures and the effects of data augmentation has been studied. Out of all the models, SWIN transformer after data augmentation gives the best results. There are a few times where misclassification occurs due to the presence of noise such as fingers, faces and hand-written text.

This method produces a result faster than the other regression methods such as Hough Transform or the Projection Profile analysis method. It also covers all the angles from 0 to 360 degrees instead of the normal range of 0 to 45 degrees. By using this skew correction method on the bills before OCR, the results can be greatly improved. We have also been able to train the models with only 30% of the original dataset.

The lowest degree considered in our dataset is 5 degrees, but when this is applied to real life problems, there might be a requirement for a much lower degree of correction (1-2 degrees). Furthermore, the absence of clear and visible text in the images causes the model to misclassify due to not being able to identify the features required. This happens even there are extra features present in the image, the extra features include hands holding the receipts, text written with a pen or pencil, highlighted text etc.

6 References

[I] Voting-Based Document Image Skew Detection, Costin-Anton Boiangiu, Ovidiu Alexandru Dinu, Cornel Popescu, Nicolae Constantin and Cătălin Petrescu (2020)

[II] A Document Skew Detection Method Using Fast Hough Transform, Bezmaternykh P.V. Nikolaev D.P (2019).

[III] Yue, Alex. "Automated Receipt Image Identification, Cropping, and Parsing." (2018).

[IV] Shafii, M., Sid-Ahmed, M. Skew detection and correction based on an axes-parallel bounding box. (2015).

[V] A Precise Skew Estimation Algorithm for Document Images Using KNN Clustering and Fourier Transform, Jonathan Fabrizio (2014)

[VI] Kumar, Deepak & Singh, Dalwinder. Modified Approach of Hough Transform for Skew Detection and Correction in Documented Images. International Journal of Research in Computer Science. (2012).

[VII] Papandreou and B. Gatos, "A Novel Skew Detection Technique Based on Vertical Projections,"International Conference on Document Analysis and Recognition. (2012)

[VIII] Panwar, Subhash & Nain, Neeta. A Novel Approach of Skew Normalization for Handwritten Text Lines and Words. 8th International Conference on Signal Image Technology and Internet Based Systems. (2011).

[IX] V. Kumar, P. Kaware, P. Singh, R. Sonkusare and S. Kumar, "Extraction of information from bill receipts using optical character recognition," 2020 International Conference on Smart Electronics and Communication (ICOSEC). (2020)

[X] N. Jirasuwankul, "Effect of text orientation to OCR error and anti-skew of text using projective transform technique," IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). (2011).