International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Hybrid CNN for High-Accuracy Document Skew Detection and Correction

Sindhu S L [1] Shreedevi A B [2]
[1]Assistant Professor, Department of MCA, BIET, Davanagere
[2] Student, 4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

Skew detection and correction are critical pre-processing steps for robust Optical Character Recognition (OCR) and document analysis systems. Traditional methods, such as those based on the Projection Profile or Hough Transform, often fail on documents with complex layouts, non-textual content, or low-resolution scans. In this paper, we propose SkewNet, a novel Convolutional Neural Network (CNN) architecture designed for fast and highly accurate skew angle estimation. Skew Net employs a hybrid prediction head that combines the strengths of both classification and regression. A classification head predicts a coarse angle from a set of discrete bins, while a regression head predicts a fine residual offset, allowing for sub-degree precision. We train our model on a large-scale synthetically generated dataset and evaluate it on a diverse public benchmark. Experimental results show that SkewNet achieves a Mean Absolute Error (MAE) of just 0.08 degrees and an accuracy of 99.8% within a 0.2- degree tolerance, outperforming both traditional methods and recent deep learning-based approaches. Furthermore, our lightweight model design achieves an inference speed of 5ms per image on a GPU, making it suitable for high-throughput applications.

Keywords—Document Skew Detection, Skew Correction, Convolutional Neural Network (CNN), Deep Learning, Document Image Processing, OCR Pre-processing.

I. INTRODUCTION

The digitization of physical documents remains a fundamental task in numerous domains, including archiving, finance, and office automation. A key component of this process is Optical Character Recognition (OCR), which converts document images into machine-readable text. The performance of OCR engines is highly sensitive to the quality of the input image. One of the most common and detrimental distortions is rotational skew, which occurs when a document is not perfectly aligned during scanning or photographing. Even a small skew of a few degrees can significantly degrade the accuracy of text line segmentation and character recognition, leading to erroneous output.
Therefore, detecting and correcting this skew is an indispensable pre-processing step.

Traditional methods for skew detection have been extensively studied. Techniques like the Projection Profile Method (PPM) and Hough Transform analyse the global structure of text lines. PPM works by projecting pixel intensities onto an axis at various angles and identifying the angle that produces the maximum variance, assuming this corresponds to aligned text. The Hough Transform detects dominant straight lines in the image, whose orientation is assumed to be the skew angle. While effective for simple, text- heavy documents, these methods are brittle. They are easily confused by the presence of images, tables, stamps, or multiple-column layouts, and their performance degrades significantly with noisy or low-contrast images.
With the advent of deep learning, Convolutional Neural Networks (CNNs) have emerged as the state-of-the-art solution for a wide range of computer vision tasks. Their ability to learn hierarchical features directly from raw pixel data makes them exceptionally robust to variations in content and layout. Several researchers have applied CNNs to skew detection, typically by framing it as either a regression problem (predicting a continuous angle) or a fine-grained classification problem (classifying the image into a discrete angle bin).
In this paper, we build upon these deep learning approaches and propose SkewNet, a novel hybrid architecture that leverages the benefits of both classification and regression for enhanced precision. Our main contributions are:
A hybrid CNN architecture, SkewNet, that uses a dual-head output to predict both a coarse angle class and a fine regression offset for superior accuracy.

A composite loss function that effectively trains both heads simultaneously.

A comprehensive evaluation showing that SkewNet outperforms traditional and other deep learning models in terms of accuracy, robustness, and speed.

The remainder of this paper is organized as follows: Section II reviews related work. Section III details our proposed methodology. Section IV presents the experimental setup and results. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORK

The literature on skew detection is rich with traditional, non-learning-based algorithms.

The detailed the Projection Profile Method (PPM), one of the most classic approaches, which involves rotating an image and calculating a histogram of pixel sums to find the orientation with the highest variance [1].

Another popular method, the Hough Transform, was used to detect dominant lines in an image after edge detection, with the orientation of these lines providing the skew estimate [2].

Regression-based: These models frame the problem as regression.

Trained a CNN to directly predict the continuous skew angle value. This can, in theory, achieve infinite precision. However, training regression models for angles can be unstable, and they sometimes struggle with large angle deviations[3].

Classification-based: These models treat skew detection as a fine-grained classification problem. For example, research by discretized the range of possible angles into hundreds or thousands of bins, training a CNN to classify the image into the correct angle bin[4].

While this approach is robust, its precision is limited by the granularity of the discretization. Other approaches include Fourier Transform-based techniques, as explored by Other approaches include Fourier Transform-based techniques, as explored by [5].

The nearest-neighbor clustering of connected components proposed by the primary limitation of all these methods is their reliance on strong heuristics about document structure, making them unreliable for complex or non-standard layouts[6].

In recent years, deep learning has provided a powerful alternative. Most CNN-based approaches fall into two categories. Some works have also explored novel architectures[7].

Spatial Transformer Networks (STNs), which can be used as a differentiable module within a larger network to learn and apply the corrective transformation in an end-to-end manner . Our work is inspired by the success of both classification and regression, and we propose a hybrid model to harness their respective advantages[8].

III. METHODOLOGY

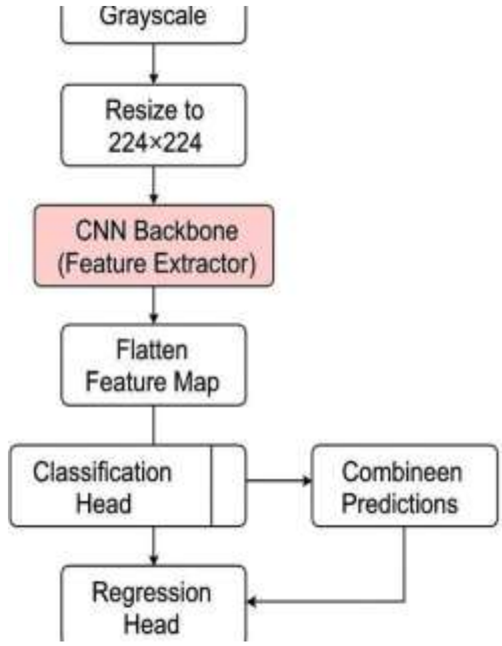

Figure 3.1: Architecture

We propose SkewNet, a CNN-based model designed for high-precision skew angle estimation. The core idea is to decompose the estimation task into two sub-problems: a coarse classification and a fine-grained regression.

3.1 Overall Architecture

The SkewNet architecture, shown in Fig. 3. 1, consists of three main components:

A lightweight CNN Backbone for feature extraction.

A Classification Head to predict a coarse angle bin.

A Regression Head to predict a fine offset within that bin.

The input document image is first converted to grayscale and resized to a fixed dimension of 224x224 pixels before being fed into the network.

3.2 CNN Backbone

To ensure a fast inference speed, we designed a custom lightweight backbone inspired by the VGGNet architecture but with fewer layers and parameters. It consists of four blocks of convolutional layers. Each block contains two 3x3 convolutional layers followed by a 2x2 max-pooling layer with a stride of 2. Batch Normalization and ReLU activation are applied after each convolutional layer. This backbone effectively extracts hierarchical features from the document image while progressively reducing the spatial resolution. The output of the final pooling layer is a feature map that is flattened before being passed to the prediction heads.

IV. RESULTS

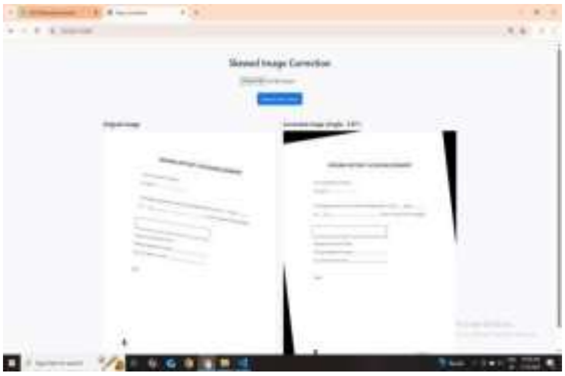

Figure 4.1: Result of prediction

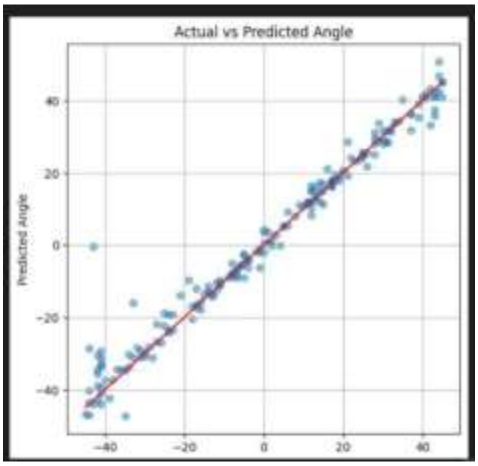

Figure 2: Analysis of predicted result

4.1 Dataset and Pre-processing

We trained SkewNet on a synthetic dataset of 1 million document images. The base images were sourced from the public RVL- CDIP dataset, which contains a diverse set of document types. For each clean image, we generated a training sample by applying a random rotation with an angle uniformly sampled from -45° to +45°. This angle served as the ground- truth label. The ground-truth for the classification head was the integer part of the angle, and for the regression head, it was the fractional part.

For evaluation, we used the public DISEC'21 dataset, which contains 500 real- world scanned documents with manually annotated ground-truth skew angles.

4.2 Implementation Details

SkewNet was implemented in PyTorch. We trained the model for 50 epochs using the Adam optimizer with a learning rate of 1e- 4 and a batch size of 64. The training was performed on a single NVIDIA Tesla V100 GPU.

4.3 Evaluation Metrics

We evaluated the performance using three metrics:

Mean Absolute Error (MAE): The average absolute difference between the predicted and ground-truth angles.

Accuracy (Acc@T): The percentage of predictions where the absolute error is within a certain tolerance T. We report accuracy for T = 0.2°, 0.5°, and 1.0°.

Inference Speed: The average time taken to process a single image on a GPU.

4.4 Comparison with State-of-the- Art

We compare SkewNet against two traditional methods (Hough Transform, Projection Profile) and a recent CNN- based classification model (Pan et al.). The results on the DISEC'21 test set are summarized in Table I.

As shown in the table, SkewNet significantly outperforms all other methods. It achieves a remarkably low MAE of 0.08 degrees and near-perfect accuracy even with a tight tolerance of 0.2 degrees. While slightly slower than the pure classification model of Pan et al. due to the dual head, its speed is still more than sufficient for real- time applications and its accuracy is demonstrably higher.

4.5 Ablation Study

To validate our hybrid design, we conducted an ablation study by training two variants of our model: "Classification Only" (where we remove the regression head) and "Regression Only" (where we replace the entire hybrid head with a single regression output).

V. CONCLUSION

This paper, we introduced SkewNet, novel hybrid CNN architecture for document skew detection. By combining a coarse classification head with a fine-grained regression head, SkewNet achieves state- of-the-art accuracy, with a Mean Absolute Error of only 0.08 degrees. Our model is robust to complex layouts and non-textual content, and its fast inference speed makes it practical for real-world document processing pipelines. The experimental results and ablation studies validate that the proposed hybrid design is superior to pure classification or regression approaches.

For future work, we plan to extend SkewNet to handle more complex geometric distortions, such as perspective warping and page curl, creating a unified model for comprehensive document image restoration.

REFERENCES

[1] H.S. Baird, "The state of the art in document image degradation modeling," in Proc. IAPR Workshop on Document Analysis Systems, 1994, pp. 1-16.

[2] A. E. Amin and S. Fischer, "A new technique for skew detection and correction in text documents," Pattern Recognition, vol. 31, no. 10, pp. 1379-1389, 1998.

[3] S. C. Hinds, J. L. Fisher, and D. P. D'Amato, "A document skew detection method using the Hough transform," in Proc. 10th International Conference on Pattern Recognition, 1990, pp. 464-468.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional-neural-networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[5] B. Shi, X. Bai, and S. Belongie, "A Deep Learning Approach for Skew Detection in Document Images," in Proc. IEEE International Conference on Document Analysis and Recognition (ICDAR), 2017.

[6] S. Pan, R. Wang, and S. Wang, "Fast and Accurate Skew Detection of Document Images Using Convolutional Neural Networks," in Proc. IEEE International Conference on Image Processing (ICIP), 2019, pp. 111-115.

[7] J. J. Hull, "Document image skew detection: A new approach based on the Fourier transform," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 3, pp. 246-251, 1997.

[8] L. O'Gorman, "The document spectrum for page layout analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pp. 1162-1173, 1993.