Skew Correction for Document Images

Team-19
Thribhuvan S, EC21B1093
Jaswanth, EC21B1073
Anandhakumar, EC21B1087

April 26, 2024

Abstract

Skew correction of document images is a crucial preprocessing step in document analysis and recognition systems. Document images captured using cameras or scanners often suffer from skew due to various factors such as perspective distortion, uneven document placement, or mechanical misalignment. Correcting the skew of document images is essential for improving the accuracy of subsequent processing tasks, such as optical character recognition (OCR) and document understanding. This project aims to come up with a skew correction technique for document images using deep learning-based methods, through RotNet architectures. This deep learning model leverages Convolutional Neural Networks (CNNs) to automatically learn and correct skew angles.

1 RotNet

RotNet, developed by Facebook AI, is a convolutional neural network architecture designed for predicting the rotation angle of input images. Unlike traditional supervised approaches, RotNet adopts a self-supervised learning strategy by leveraging unlabeled data and generating rotated versions of input images during training. By classifying images into specified rotation classes, RotNet learns rotation-invariant representations, improving generalization and robustness to rotation variations. It introduces auto-augmentation, dynamically adjusting the rotation angle distribution of training samples, enhancing performance further.

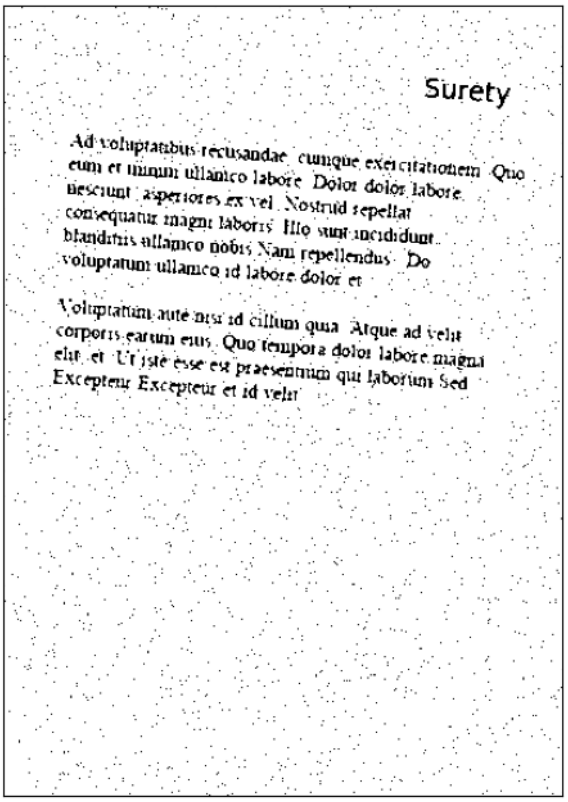
In practical applications, RotNet offers deep learning engineers a powerful tool for pretraining CNN backbones in tasks such as image classification, object detection, and segmentation. Its ability to learn rotation-invariant representations makes it particularly valuable in scenarios with images captured from arbitrary viewpoints or orientations, such as autonomous driving and robotics. RotNet's self-supervised approach and innovative data augmentation strategy contribute to improved model robustness and performance across various real-world applications.

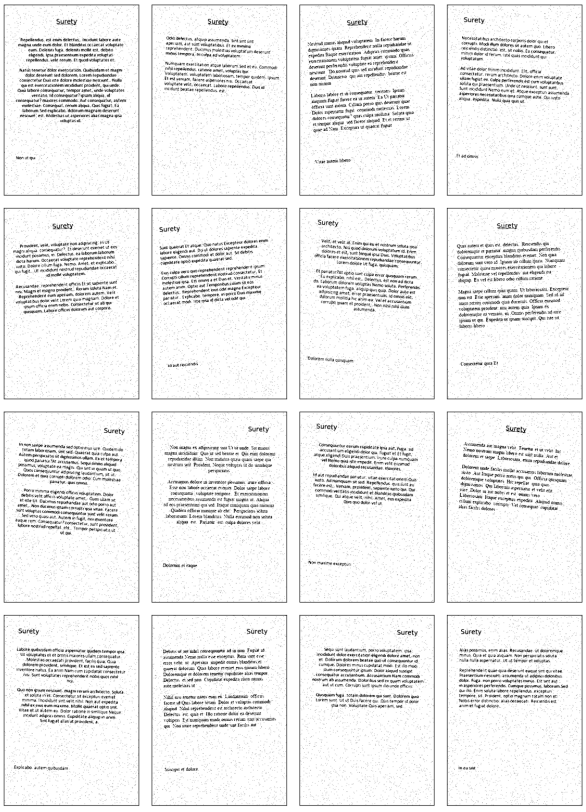
2 Input

Input Dataset is imported from Kaggle, which contains about 1000 skewed text images. The images and their corresponding labels are provided in two separate directories.
Link

2.1 Raw Input Visualisation

The original data has a skew range of only upto [-5, 5] degrees.





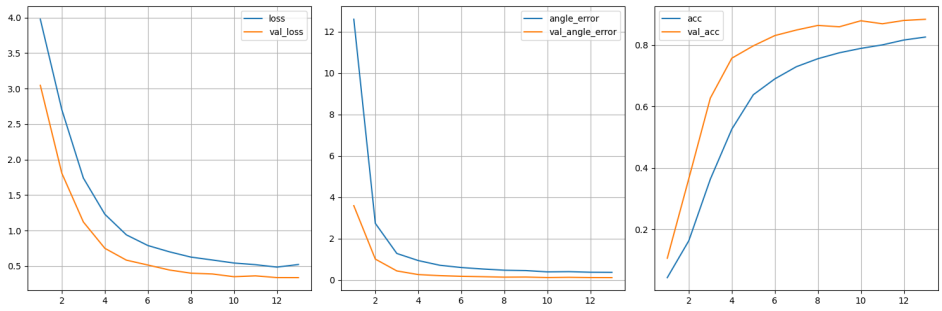
2.2 Data Augmentation

Inorder to make the dataset more complex so that the model could understand the skewness better, we increase the skew range from [-5, 5] to [-45, 45] degrees.

3 Output

3.1 Model

After training the model using the augmented data for 4 hours, we achieved an accuracy of 82%.



3.2 Corrected Images

Though the text is not perfectly horizontal, most of the images are corrected in a satisfactory way.