22 - Preliminary Project Report

Prashanth N
ai19btech11010@iith.ac.in

Praveen M
ai19btech11018@iith.ac.in

Jayanth Bhaskar M
ai19btech11008@iith.ac.in

Abstract

*This is a thorough visit back to problem of **Document Enhancement** and analyzing various methods that were proposed to address it ranging from simple Binarization [1]to GANs [3] for cleaning up the image for producing more robust results from the already matured OCR especially using the visibility detection [2] as the preprocessing step and drawing observations from various approaches and recording the results to gain interesting observations which could potentially bud a new technique which could be flexible, more accurate and faster load and inference times for the Document Enhancement problem*

1. Introduction

There is huge interest in people to make all kinds of knowledge accessible online but there are many challenges which hinder that process one of the main one being documenting the written texts which are often occluded or comes with damaged versions, maybe coffee stains, overtime degradation and many other natural scenarios, So if we want to have our knowledge repository filled with newer ways of obtaining documents like scanned images from smart-phones problem persists there too in forms such as shadow, blur, different light conditions and distorted images and also other well-known problems of rolling-shutter. Even after the process of obtaining the image digitally there is no guarantee that the image is free from damage there might be some watermark, annotations and stamps which might hinder in obtaining the complete data through these images, So there is a need to adopt to various kinds of algorithms based on different use cases

1. Mobile: Document scanning - Quick inference even though a little accuracy is compromised

2. Knowledge base: Preserving for online libraries - The prediction should be more accurate even though time intense

The main idea is to analyze various existing solutions and contrasting various algorithms with and without the preprocessing step of image visibility detection which costs $O(nlogn)$ for n pixels [2] and proposing which model combination suits best for a specific use case

2. Problem Statement

The problem boils down to classifying various combinations of models with [2] and without and aiming to provide ready made models suited for different needs of the document enhancer

3. Literature Review

[3] studies the same problem and giving a very heavy solution of GAN called DE-GAN [3] which would train up a Generator which tries to fool the Discriminator by producing images as close to the original as the Generator learns to remove the damage in the conditioned image while preserving the required text This paper mainly address two key issues in depth

1. Image blurring
2. Watermark removal

[2] mainly identifies a novel way of representing images to remove the occlusions that are formed due to intermediary intensities in the image and then binarizing with a new technique called Lowlights Map which provides much better results compared to standard one step separation of foreground with background

References

[1] Sheng He and Lambert Schomaker. Deepotsu: Document enhancement and binarization using iterative deep learning. *CoRR*, abs/1901.06081, 2019. 1

[2] Netanel Kligler, Sagi Katz, and Ayellet Tal. Document enhancement using visibility detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2374–2382, 2018. 1

[3] Mohamed Ali Souibgui and Yousri Kessentini. Degan: A conditional generative adversarial network for document enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–1, 2021. 1