

22 - Mid Term Report

Prashanth N

ail9btech11010@iith.ac.in

Praveen M

ail9btech11018@iith.ac.in

Jayanth Bhaskar M

ail9btech11008@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 [2] to GANs [6] for cleaning up the image for producing more robust results from the already matured OCR especially using the visibility detection [5] 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 possibly be flexible and adaptable to various kinds of issues and artifacts, 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, over-time 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, bleed-through, different lighting conditions (occlusions and overexposure)?? 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 $O(n \log n)$ for n pixels [5] and proposing which model combination suits best for a specific use case

2. Previous works

Although the problem has been studied in various magnifications of scientific lenses some narrowed to specific artifacts like shadows. For example, Bako et al. [1] which showed very appreciable results owing to their very simple method of eliminating shadows by scaling down by a factor of α_i which is the ratio of local and global background. They made use of the fact that the background stays the same color

In contrast there are very novel latest methods which take even the information given by **text** present in the document to help its enhancement. For example, Khamekhem Jemni et al. [4] used the UNet and encoded Convolutional Recurrent Neural Nets (CRNNs) embeddings into Gated Recurrent Units (GRUs) (for text recognition) and made the unavailable degraded datasets of KHATT and IAM (with strokes and other artifacts added) All these approach the problem of Document Enhancement in a fantastic manner but for an amateur interested in enhancing certain documents with some set of constraints they bring in the paradox of choice and tremendous load of research

So, our work mainly is to reduce the burden and let the amateur steer towards a method well-suited to his use-case

3. Problem Statement

[width=0.3]images/issues-and-artifacts.jpg

The problem boils down to classifying various combinations of models with [5] and without and aiming to provide ready-made models suited for different needs of the document enhancer

4. Literature Review

[6] studies the same problem and giving a very heavy solution of GAN called DE-GAN [6] 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

[5] 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

5. Preliminary Results

We are implementing visibility detection(VD) [5] on certain damaged images whose processed images are used as pre-processed images for DE-GAN method.

We have found the pretrained model of DE-GAN from the authors Souibgui et al. where we tried to take the DIBCO 2018 Dataset [3] and tested the degraded documents and tested the inference time although the accuracy was laudable the enhancing time was on an average *10-12 seconds* and the added overhead of VD will increase the document enhancing time

References

- [1] Steve Bako, Soheil Darabi, Eli Shechtman, Jue Wang, Kalyan Sunkavalli, and Pradeep Sen. Removing shadows from images of documents. *Asian Conference on Computer Vision (ACCV 2016)*, 2016. 1
- [2] Sheng He and Lambert Schomaker. Deepotsu: Document enhancement and binarization using iterative deep learning. *CoRR*, abs/1901.06081, 2019. 1
- [3] Xenofon Karagiannis Ioannis Pratikakis, Konstantinos Zagoris. Icdar 2019. <https://vc.ee.duth.gr/dibco2019/>, 2019. 2
- [4] Sana Khamekhem Jemni, Mohamed Ali Souibgui, Yousri Kessentini, and Alicia Fornés. Enhance to read better: A multi-task adversarial network for handwritten document image enhancement. *Pattern Recognition*, 123:108370, Mar 2022. 1
- [5] 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, 2
- [6] Mohamed Ali Souibgui and Yousri Kessentini. De-gan: A conditional generative adversarial network for document enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–1, 2021. 1