Generative Pipeline for Data Augmentation of Unconstrained Document Images with Structural and Textural Degradation (Student Abstract)

Arnab Poddar[1], Abhishek Kumar Sah[2] , Soumyadeep Dey[3], Pratik Jawanpuriya[3], Jayanta Mukhopadhyay[2], Prabir Kumar Biswas[1]

[1]Dept. of Electronics & Electrical Communication Engg, Indian Institute of Technology Kharagpur, 721302 India
[2]Dept. of Computer Science & Engg, Indian Institute of Technology Kharagpur, 721302 India
[3]Microsoft R&D India, Hyderabad
arnabpoddar@iitkgp.ac.in, abhishekkumar2046@gmail.com, soumyadeep.dey@microsoft.com, pratik.jawanpuria@microsoft.com, jay.cse@iitkgp.ac.in, pkb.ece@iitkgp.ac.in

Abstract

Computer vision applications for document image understanding (DIU) such as optical character recognition, word spotting, enhancement etc. suffer from structural deformations like strike-outs and unconstrained strokes, to name a few. They also suffer from texture degradation due to blurring, aging, or blotting-spots etc. The DIU applications with deep networks are limited to constrained environment and lack diverse data with text-level and pixel-level annotation simultaneously. In this work, we propose a generative framework to produce realistic synthetic handwritten document images with simultaneous annotation of text and corresponding pixel-level spatial foreground information. The proposed approach generates realistic backgrounds with artificial handwritten texts which supplements data-augmentation in multiple unconstrained DIU systems. The proposed framework is an early work to facilitate DIU system-evaluation in both image quality and recognition performance at a go.

Introduction

The computer vision applications for DIU transform the visual written information into recognisable formats like text, language, etc. The state-of-the-art vision systems for DIU primarily deal with constrained scanning environment. (Tensmeyer et al. 2019; Dey and Jawanpuria 2021).

The performance of DIUs are affected by both structural deformation of written content and textural degradation (Fig. 1). However, in real world, there are out-of-distribution (OOD) instances arising due to non-uniform illumination, free movements and motions, unavailability of scanners, shadows and also uneven surface etc. as captured by handheld and mobile devices. It degrades the readability and intelligibility and textural content. In free-form unconstrained documents, the structural deformation like strike-out words (STW), underlines, line-marking etc causes performance deterioration. Consequently, both the textually degraded and structurally deformed images are expected to downgrade the performance of constrained DIU systems (Fig. 1).

To address both structural deformations and textural diversity of backgrounds due to aging, smudging, blurring, motion-artifacts, shadows, non-linear illuminations, uneven-surface warping etc, the DIU systems need significant data

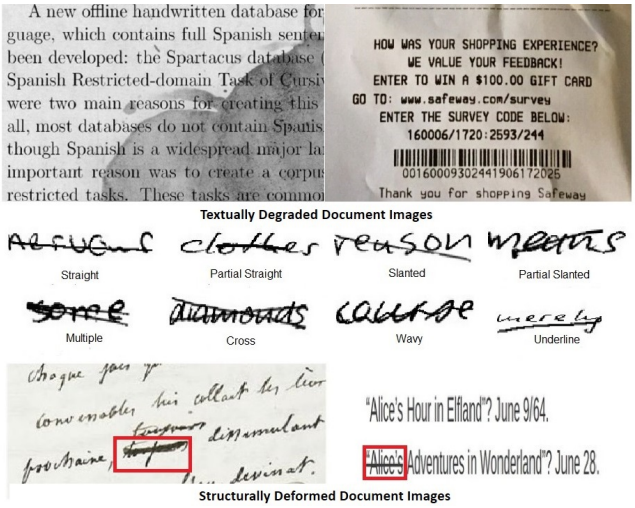

Figure 1: Illustration for unconstrained document images with textural degradation and structural deformations.

with both text and pixel-level ground-truth at one go. The lack of research addressing the real-world challenges like degraded texture and shape based deformations altogether is due to unavailability of data with annotation of text and corresponding pixel-level ground-truth simultaneously.

However, document image enhancement systems mostly concentrate on the intelligibility and texture of scanned images and shadow in terms of image quality (SSIM and SNR) (Tensmeyer et al. 2019; Dey and Jawanpuria 2021). Few generative networks for data augmentation of document image enhancement systems are applied successfully in (Tensmeyer et al. 2019; Lee, Hong, and Kim 2021).

Here we propose a generative pipeline for realistic handwritten images with annotation of text and corresponding foreground-pixels both. The proposed system generates variable realistic background textures with artificial handwritten texts. The synthesized images can be used for data-augmentation of various DIU like optical character recognition (OCR), language recognition systems (LRS), word spotting systems (WSS), word retrieval systems (WRS), enhancements, binarisation etc. Furthermore, an end-to-end

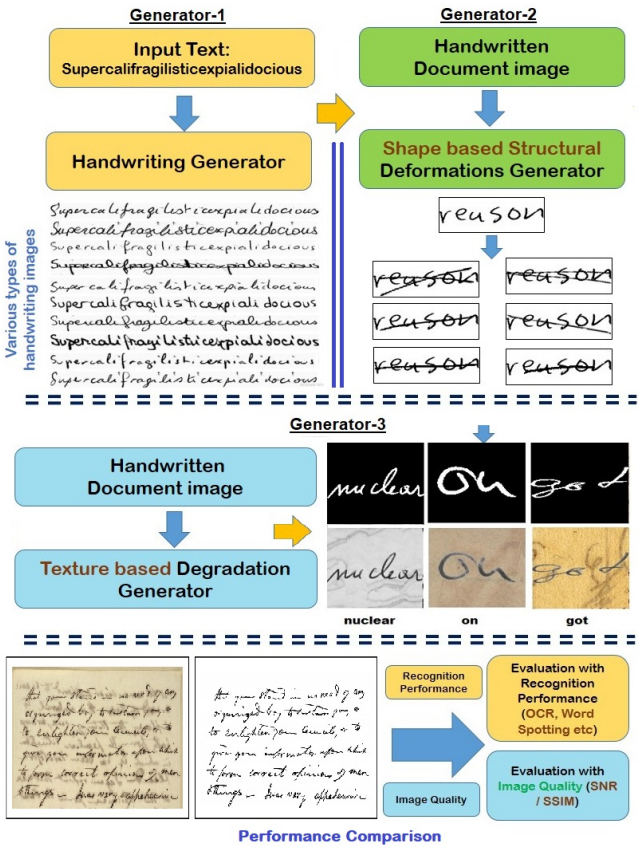

Figure 2: Illustration of Generative framework. Generator-1 produce artificial handwritten texts with text input. Generator 2 adds the structural deformations like STW etc. Generator 3 produces handwritten image with background texture.

.

system can be developed for robust information extraction from degraded and deformed document images.

Generative Framework with Three Modules

The proposed generative pipeline outputs document images with diverse background texture containing annotation of the spatial text-pixels and text-content simultaneously. It can be used for augmentation of OCR, LRS, WRS, enhancement systems etc altogether. It is an early attempt to enable DIU system-evaluation with image-quality and recognition assessment at a go. The proposed framework endorses three generative modules (Fig.2). First module generates handwritten text image with white background in various writing-style for a given input text. We extend the GanWriting framework (Kang et al. 2020) with losses addressing handwriting style and overall similarity. Secondly, for generation of structural deformation, we collect a pool of 3000 types of real stroke-template and induce randomisation of position, rotation and size as in (Poddar et al. 2021). Thirdly, for background texture, we propose generator architecture extending pix2pix with Resnet (9 blocks) (Poddar et al. 2021).

The generated images of various stages of the generators are evaluated with recognition performance of OCR (Shi, Bai, and Yao 2016). The character error rate (CER %) of OCR degrades to 45.11% from 10.12% on clean words from IAM dataset against structural deformation only. Whereas it degrades up-to 24.63% from 10.12% with only texture degradation. The illustrations of generated images are depicted for all three generator modules in Fig. 2. The proposed framework holds both text and foreground pixel annotation. It enables for the first time to evaluate both image quality enhancement indexes (SSIM, SNR etc.) as well as recognition performance (CER, WER etc.) of multiple unconstrained DIU systems.

Conclusion

Unconstrained DIU systems suffer from both structural and texture degradation. Simultaneous text and pixel level ground-truth is needed to address both at a go. The benchmark databases of DIU applications do not contain text and pixel-level annotation with diverse background texture. Here we propose a generative pipeline with three GAN based modules to produce realistic synthetic handwritten documents with simultaneous annotation of text and corresponding pixel-level annotation. Realistic synthetic data with both text and pixel-level annotation can supplement unconstrained data-augmentation in a wide range of DIU systems like OCR, LRS, WSS, binarisation, enhancement etc. It is an early attempt to synthesize data for DIU system evaluation in terms of both image quality metrics and recognition scores simultaneously.

References

Dey, S.; and Jawanpuria, P. 2021. Light-Weight Document Image Cleanup Using Perceptual Loss. In *International Conference on Document Analysis and Recognition*, 238–253. Springer.

Kang, L.; Riba, P.; Wang, Y.; Rusiñol, M.; Fornés, A.; and Villegas, M. 2020. GANwriting: content-conditioned generation of styled handwritten word images. In *European Conference on Computer Vision*, 273–289. Springer.

Lee, Y.; Hong, T.; and Kim, S. 2021. Data Augmentations for Document Images. In *SDU@ AAAI*.

Poddar, A.; Chakraborty, A.; Mukhopadhyay, J.; and Biswas, P. K. 2021. Detection and Localisation of Struck-Out-Strokes in Handwritten Manuscripts. In *International Conference on Document Analysis and Recognition*, 98–112. Springer.

Shi, B.; Bai, X.; and Yao, C. 2016. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE T-PAMI*, 39(11): 2298–2304.

Tensmeyer, C.; Brodie, M.; Saunders, D.; and Martinez, T. 2019. Generating realistic binarization data with generative adversarial networks. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 172–177. IEEE.