DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement

Abstract

Documents often exhibit various forms of degradation, which make it hard to be read and substantially deteriorate the performance of an OCR system. The system proposed is an effective end-to-end framework named Document Enhancement Generative Adversarial Networks (DE-GAN) that uses the conditional GANs (cGANs) to restore severely degraded document images. We demonstrate that, in different tasks (document clean up, binarization, deblurring and watermark removal), DE-GAN can produce an enhanced version of the degraded document with a high quality. In addition, our approach provides consistent improvements compared to state-of-the-art methods over the widely used DIBCO 2013, DIBCO 2017 and H-DIBCO 2018 datasets, proving its ability to restore a degraded document image to its ideal condition. The obtained results on a wide variety of degradation reveal the flexibility of the proposed model to be exploited in other document enhancement problems

Introduction

AUTOMATIC document processing consists in the transformation into a form that is comprehensible by a computer vision system or by a human. Thanks to the development of several public databases, document processing has made great progress in recent years. However, this processing is not always effective when documents are degraded. Lot of damage could be done to a document paper. For example: Wrinkles, dust, coffee or food stains, faded sun spots and lot of real-life scenarios.Degradation could be presented also in the scanned documents because of the bad conditions of digitization like using the smart-phones cameras (shadow, blur, variation of light conditions, distortion, etc.). Moreover, some documents could contain watermarks, stamps or annotations.
The recovery is even harder when certain types of these later take the text place for instance in cases where the stains color is the same or darker than the document font color . Hence, by using the C-GAN approach we will try to recover a clean version of the degraded document.

Motivation

In this work, we focus on the enhancement of degraded document images by addressing different kinds of degradation which are document clean up, binarization, and watermark removal. From a document analyst viewpoint, this recovery of a clean version from the degraded document falls in the research field called document enhancement. It is necessary because Document enhancement addresses the problems involved in improving the perceptual quality of document images and in removing degradations and artifacts present in images, with the aim of restoring their original look.

Problem Statement

Document images are often obtained by digitizing paper documents like books or manuscripts. They could be poor in appearance due to degradation of paper quality, spreading and flaking of ink toner, imaging artifacts etc. All the above phenomena lead to different types of noise at the word level including boundary erosion, dilation, cuts/breaks and merges of characters.The purpose is to look for a restoration approach that does not perform a explicit character segmentation, but still uses the repetitive component nature of document images.

Literature Survey

Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections paper proposed a very deep fully convolutional encoding-decoding framework for image restoration such as denoising and super-resolution. High Performance Convolutional Neural Networks for Document Processing proposes Convolutional neural networks (CNNs) are well suited for solving visual document tasks that rely on recognition and classification.

Proposed Work - Algorithm;

We consider the problems of document enhancement as an image to image translation task where the goal is to generate clean document images given the degraded one GANs were initially proposed and consist in two neural networks, a generator G and a discriminator D

Generator:

The generator is performing an image-to-image translation task. The generator have the goal of learning a mapping from a random noise vector z to an image y,While the discriminator has the function of distinguishing between the image generated by G and the ground truth one. Hence, given y, D should be able to tell if it is fake or real by outputting a probability value

Discriminator:

The discriminator is a simple Fully Convolutional Network (FCN), composed of 6 convolutional layers, that gives an output containing probabilities of the generated image being real. the discriminator receives two input images which are the degraded image and its clean version (ground truth or cleaned by the generator).

Given a degraded image, we only use the generative network to enhance it. But, this discriminator shall force the generator during training to produce better.

Justification

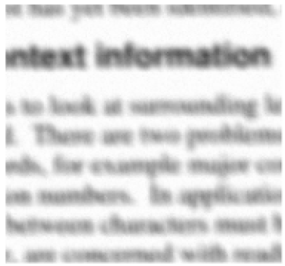
Although GAN was able to generate some good examples of data points, it is not able to generate the data point with the target label and the dataset generated from it lacks diversity.

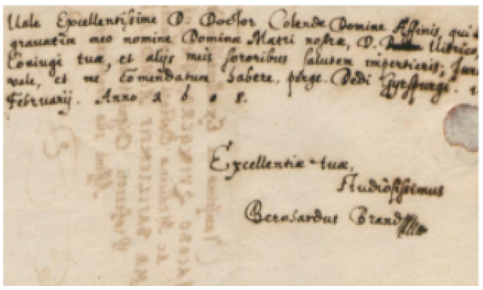
In cGANs, a conditional setting is applied, meaning that both the generator and discriminator are conditioned on some sort of auxiliary information (such as class labels or data)
As a result, the ideal model can learn multi-modal mapping from inputs to outputs by being fed with different information.
There is a modification in the architecture by adding the label y as a parameter to the input of the generator and try to generate the corresponding data point. It also adds labels to the discriminator input to distinguish real data better.

Experimental Setup :

Dataset:

https://archive.ics.uci.edu/ml/machine-learning-databases/00318/NoisyOffice.zip





theory holds that the fungus was spread by North American bullfrogs which have been introduced—sometimes accidentally, sometimes purposefully—into Europe, Asia, and South America, and which are often exported for human consumption. North American bullfrogs, too, are widely infected with Bd but do not seem to be harmed by it. The first has become known as the "Out of Africa" and the second might be called the "frog-leg soup" hypothesis.

Either way, the etiology is the same. Without being loaded by someone onto a boat or a plane, it would have been impossible for a frog carrying Bd to get from Africa to Australia or from North America to Europe. This sort of intercontinental reshuffling, which nowadays we find totally unremarkable, is probably unprecedented in the three-and-a-half-billion-year history of life.

* * *

Even though Bd has swept through most of Panama by now, Griffith still occasionally goes out collecting for EVACC, looking for survivors. I scheduled my visit to coincide with one of these collecting trips, and one evening I set out with him and two of the American volunteers who were working on the waterfall. We headed east, across the Panama Canal, and spent the night in a region known as Cerro Azul, in a guesthouse ringed by an eight-foot-tall iron fence. At dawn, we drove to the ranger station at the entrance to Chagres National Park. Griffith was hoping to find females of two species that EVACC is short of. He pulled out his government-issued collecting permit and presented it to the sleepy officials manning the station. Some underfed dogs came out to sniff around the truck.

Beyond the ranger station, the road turned into a series of craters connected by deep ruts. Griffith put Jimi Hendrix on the truck's CD player, and we bounced along to the throbbing beat. Frog collecting requires a lot of supplies, so Griffith had hired two men to help with the

Software - JUPYTER NOTEBOOK

ML/DL Model -

MODEL Name: DE-GAN

DE-GAN is a conditional generative adversarial network designed to enhance the document quality before the recognition process. It could be used for document cleaning, binarization, deblurring and watermark removal. The weights are available to test the enhancement.

Pre-processing

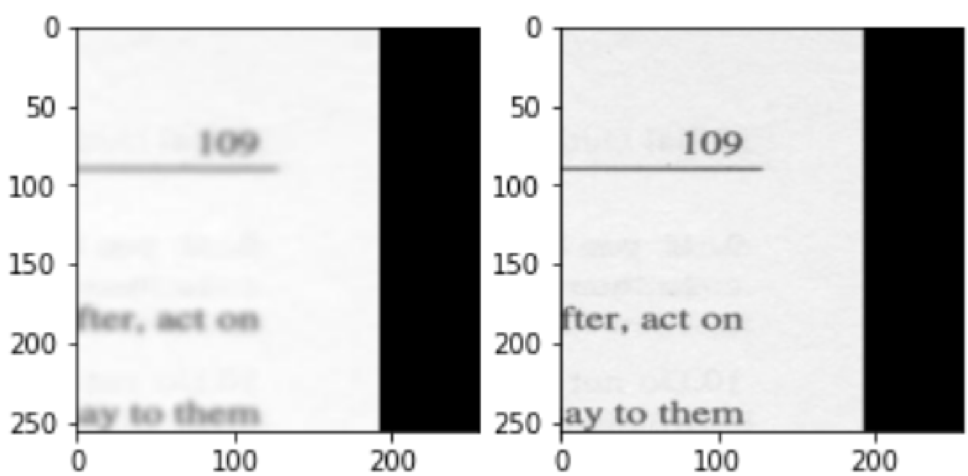
Each image was preprocessed using python scripts into a numpy file in a ".NPY" format. This reduced the dataset from 8.9GB to 1.3GB. Moreover while looping during training, the images in NPY format are loaded faster and thus the training process is faster. This was learnt and used by past experience of team members while training networks.
So, after preprocessing, all our files were present as 256x256x3 numpy arrays arranged suitably in files and folders to facilitate an appropriate train-test divide.

In this model first we train the discriminator and generate the fake data and then reset the gradients and then train the real data, calculate the error and backpropagate then update weights with gradients and train generator and implement the model and use checkpoints to see the further progress.

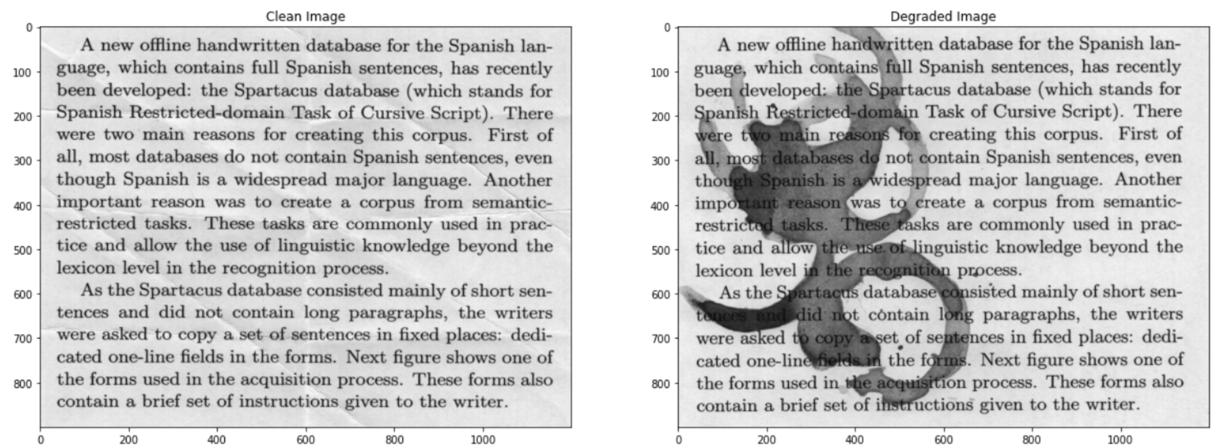
We trained for 200 epochs (original implementation was 300 epochs)
We stopped after 200 because of time constraints and more importantly because of early stopping to prevent overfitting. The number of images is only 2000. The dataset, although large, is not very diverse.

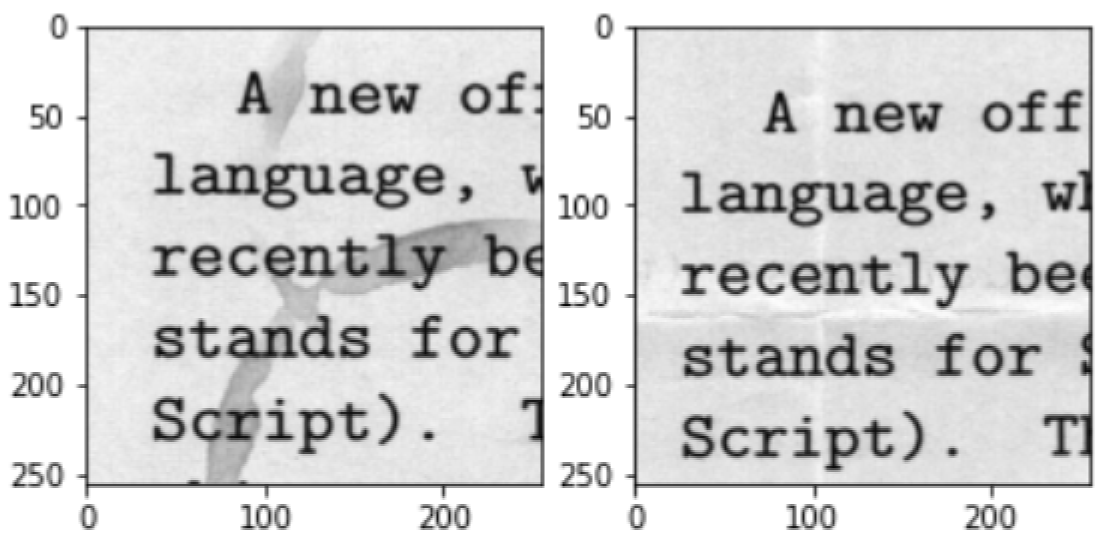
Results obtained

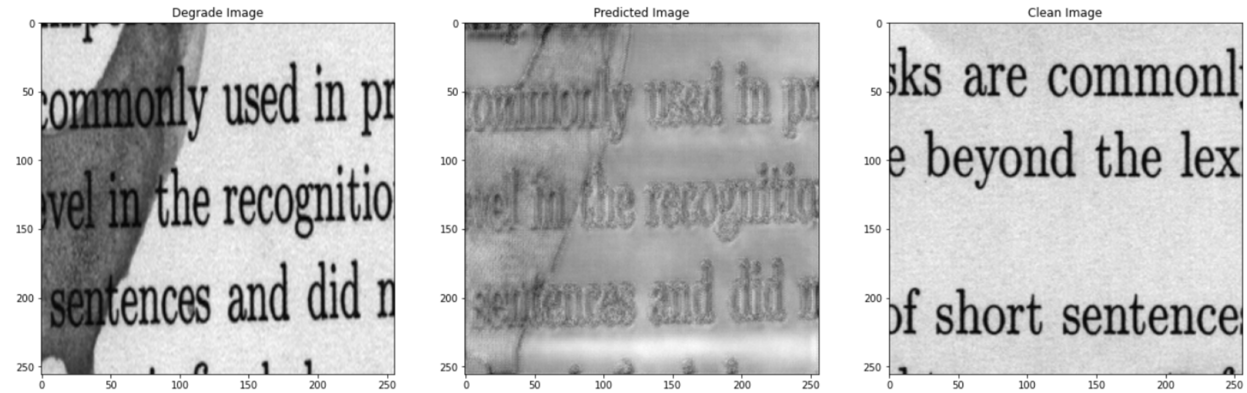
<u>DEBLURRING OF IMAGES:</u>



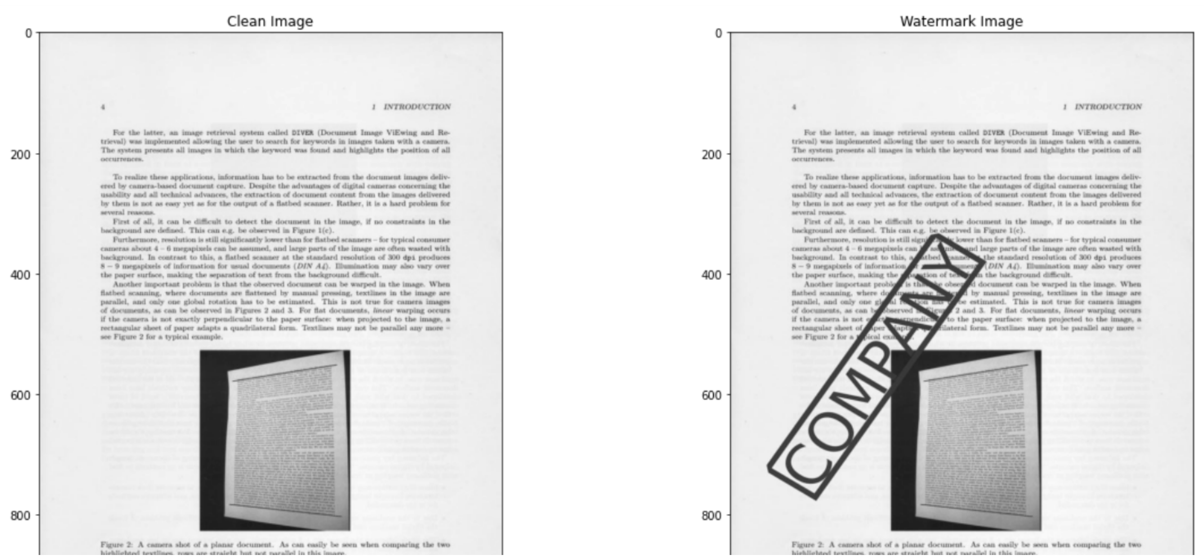
DE-GAN DEGRADE:

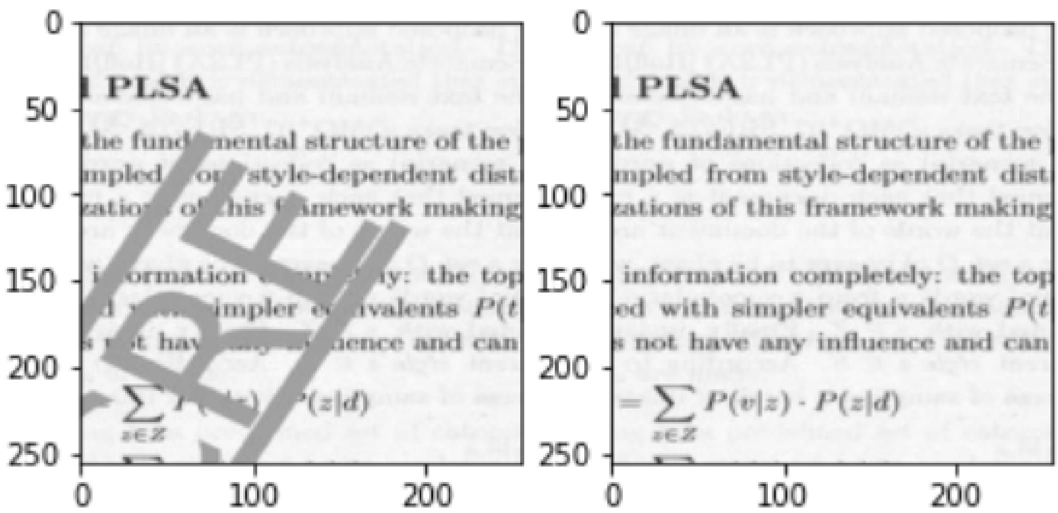






DE-GAN WATERMARK REMOVAL





Conclusion

We showed that the proposed enhancement method boosts the baseline OCR performance by a large margin. Hence, as an immediate future work, we plan to add the OCR evaluation in the discriminator part. Thus, we can give the discriminator the ability of reading the text to decide if it is real or fake, which will force it to generate more readable images. We intend, also, to test the performance of the DEGAN on mobile captured documents which present many problems like shadow, real blur, low resolution, distortion,etc

REFERENCES:

https://arxiv.org/pdf/2010.08764.pdf

C. Dong, Y. Deng, C. C. Loy, and X. Tang. Compression artifacts reduction by a deep convolutional network. In Proc. IEEE Int. Conf. Comp. Vis., pages 576–584, 2015.

C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell., 38(2):295–307, 2016.

W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process., 22(4):1620–1630, 2013.

S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 2862–2869, 2014.

S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang. Convolutional sparse coding for image super-resolution. In Proc. IEEE Int. Conf. Comp. Vis., pages 1823–1831, 2015.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., volume abs/1512.03385, 2016.