Using ML To Improve Image Quality

TEAM - 3



Introduction



Identifying Problems



ML Techniques & Uses



Solutions

Table of contents



GLOBALCERT TEAM - 3

INTRODUCTION

Document scans don't always come out the way we expect. A common problem encountered when scanning documents is 'noise' which can occur in an image because of paper quality, the typing machine used, or it can be created by scanners during the scanning process.

Overview



When scanning pages, we sometimes scan the photos at odd angles, and such images are frequently found to be rotated and may lead to loss of texts.

When the scanning screen has not been adequately wiped, they can be noisy and full of speckle dirt markings.



We ensured that such flaws are eliminated and that the photos are readable by the reader in order to improve image quality.

To avoid speckle dirt stains across the images, it would have been more beneficial if the scanner had used optical character recognition (OCR).

What Is Image Noise?

Noise is an additional pixel values added to an image causing the loss of information.

Impulse Noise

Pixel values are completely different from the surrounding pixel values

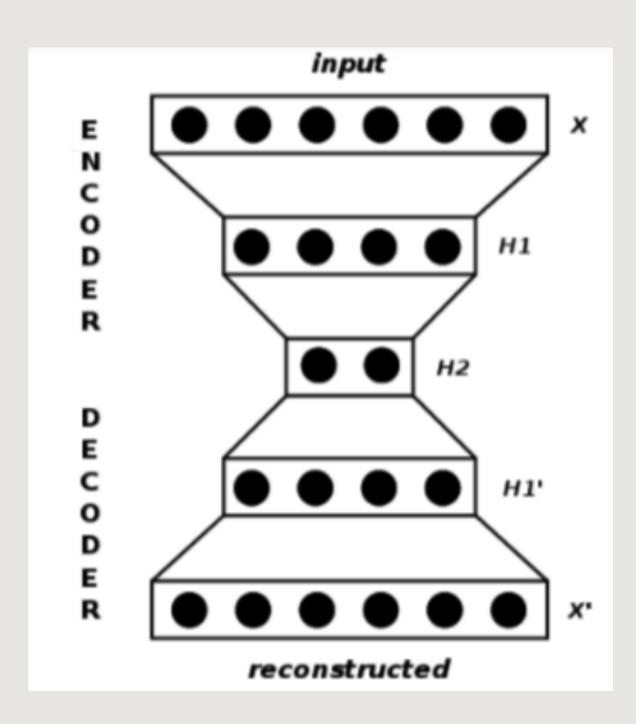
IMAGE NOISE

Additive White Gaussian Noise

Each pixel in the image will be changed from its original value by a small amount.

Image Denoising Using Autoencoders

Autoencoders are a class of Neural Networks that attempt to recreate their target using Back Propagation



ML Techniques

Number of different ways we can solve this problem:

- Denoising [In R we can use EBImage and Magick packages for denoising.]
- Blurring the image to some extent can help us to remove some noise, but it will also remove some important information.
- Gaussian Threshold method using openCV [Python]

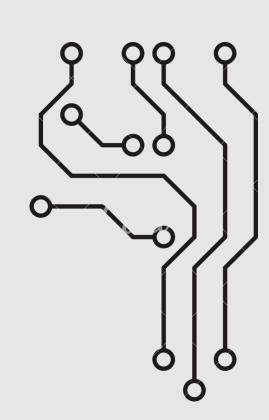


Image 1

Original Image 136.png

A new offline handwritten databas language, which contains full Spani recently been developed: the Spart stands for Spanish Restricted-domai Script). There were two main reason this corpus. First of all, most day contain Spanish sentences, even the a widespread major language. Anoth was to create a corpus from semanti These tasks are commonly used in prothe use of linguistic knowledge bey

Processed Image Manually 136.png
A new offline handwritten databas
language, which contains full Spani
recently been developed: the Spani
stands for Spanish Restricted-domai
Script). There were two main reaso
this corpus. First of all, most da
contain Spanish sentences, even the
a widespread major language. Anoth
was to create a corpus from semanti

These tasks are commonly used in pr

the use of linguistic knowledge bey

Image 2

Original Image 16.png

A new offline bandwritten database for the Spanish lang Spanish sentences, has recently been developed: the Spartacu for Spanish Bestricted-domain Task of Cursive Script). There e creating this corpus. First of all, most databases do not contai though Spanish is a widespread major language. Another impos a corpus from semantic-restricted tasks. These tasks are commallow the use of linguistic knowledge beyond the lexicon level in

As the Spartacus database consisted mainly of short sents long paragraphs, the writers were asked to copy a set of sentences one-line fields in the forms. Next figure shows one of the form process. These forms also contain a brief set of instructions give

Processed Image Manually 16.png

A new offline handwritten database for the Spanish long Spanish scutences, has recently been developed: the Spartacu for Spanish Restricted domain Task of Cursive Script). There is creating this corpus. First of all, most databases do not contain though Spanish is a widespread major language. Another import a corpus from semantic-restricted tasks. These tasks are communicallow the use of linguistic knowledge beyond the lexicon level in

As the Spartacus database consisted mainly of short sente long paragraphs, the writers were asked to copy a set of sentences one-line fields in the forms. Next figure shows one of the form process. These forms also contain a brief set of instructions give

IMPROVED IMAGES

After removing noisiness from the scanned images.

Mark/ Moll: