A WORKFLOW FOR DOCUMENT ENHANCEMENT THROUGH CONTENT SEGMENTATION AND MULTIPLE ENHANCEMENTS

HP INC

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
https://www.tdcommons.org/dpubs_series/3119

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
A Workflow for Document Enhancement through Content Segmentation and Multiple Enhancements

Abstract

Enhancing a mobile scanned document is a very difficult task that the majority of state-of-the-art solutions fail. This is addressed mainly because the algorithms cannot correctly distinguish the document’s content from the background. To overcome this issue, this work proposes a workflow that uses a sequence of machine learning models to address content segmentation, background removal, and content enhancement.

Introduction

Using the smartphone camera to perform scanning is a great solution to reduce the need for a dedicated scanner on each printer. However, while in a flatbed scanner the environmental conditions, such as lighting and document perspective, are highly controlled, when using a mobile camera those same artifacts can vary drastically. Many traditional computer vision algorithms are capable of enhancing mobile scanned documents by removing the lighting surface variance or binarizing its content. Yet, those algorithms end up losing some of the content information during this process, such as colorization, mainly because of the lack of knowledge about the semantic properties of the content.

The proposed solution solves the problem of recognizing the content semantically, resulting in a better segmentation for enhancement. Additionally, we use a sequence of feature extraction and super-resolution models for enhancing the document based on the semantic segmentation.

Proposed Workflow

The proposed workflow to enhance a document based on its content is illustrated in Figure 1. The raw scanned document is fed to a content segmentation model that will lift the content from background information. Using this content mask, we apply a sequence of models to finally achieve the enhanced document. The
number of models used in the workflow can be parametrized, and for now, we combined autoencoders and super-resolution machine learning models. Each of these models can use the content mask and/or the raw image as inputs.

**Figure 1:** Proposed workflow.

**Proof of Concept**

We based our proof of concept for the workflow using three neural network models. The first one is a content segmentation network based on the U-Net architecture [1], and the other two are, a super-resolution network based on the EDSR architecture [2] and an autoencoder based again on the U-Net architecture, respectively. The whole workflow is presented in Figure 2.

**Figure 2:** Full tested network architecture based on Figure 1 workflow.

The predicted segmentation mask from the first model is concatenated with the input scanned document as a fourth channel and fed into the first enhancement model that uses the content displacement information from this to clean the background of the document. Sequentially, the second enhancement network can fix little mistakes from the first enhancement network by applying content enhancements, such as color correction and augmented text contrast. Notice that (for the tested workflow) the input raw image is only used on the segmentation model, and the content mask is only used in the first enhancement model. Figure 3 presents
results from this process and compares our results to the current Smart App [3] outputs.

**Figure 3:** Comparison among scanned document, ground truth image, our proposed solution, and current Smart App enhancement solution, respectively.

### References

