

Abstract

Our thesis explores the field of document image analysis, outlining its significance and the challenges it faces. We progress from basic text extraction (word spotting) to understanding complex structures (cell extraction and article segmentation, addressing various challenges found in this progression. Our first work (Srivastava and Harit [2020b]) navigates the complexity of Word Spotting in Cluttered Environments where a word is cluttered by a strike-through with a line stroke. We present a comprehensive approach to word spotting in cluttered environments, focusing on the use of Vertical Projection Profile (VPP) feature and its modified version, the combinatorics Vertical Projection Profile (cVPP). We compare our method with (Rath and Manmatha [2003]) and PHOCNET (Sudholt and Fink [2016]) for handwritten word spotting in the presence of strike-through, achieving better results.

We then explore structural insights of document images, focusing on cell extraction and horizontal-scale correction in handwritten form images (Srivastava and Harit [2020a]). Our focus laid on structured documents like forms and cheques, where there is a predefined space called frame field/cell for the user to fill the entry. We address the non-uniformity of inter-character spacing while writing by extracting cells using the modified Region growing method (Gonzalez [2009]) and applying horizontal scale correction on the extracted form fields. This system results in reduced error rates when applied as a preprocessing step in a recognition system (Almazán et al. [2014]).

In continuation to handwritten form images, we propose a graph-based deep network for predicting the associations pertaining to field labels and field values in heterogeneous handwritten form images (Divya and Gaurav [2019]). We consider forms in which the field label comprises printed text and field value can be the handwritten text. We have used a Graph Autoencoder (Kipf and Welling [2016a]) to perform the intended field label to field value association in a given form image. It was the first attempt to perform label-value association in a handwritten form image using a machine learning approach.

Simultaneous super resolution and denoising in document image (Divya and Gaurav [2023]). The approach is a one shot unpaired technique where a single unpaired example is used as reference for training a SinGAN model (Shaham et al. [2019]), where first, a clean reference image is used to train a SinGAN to learn the characteristics of the clean image. Then, super resolution and denoising of the given test image is carried out using another SinGAN. The formulation of the loss function helps in this task by prompting the generated images to have characteristics similar to the reference clean image.

Selective Image denoising (Divya and Gaurav [2024]) deals with removal of unwanted noise from images. The existing methods process an image in its entirety, assuming that the noise uniformly affects the entire image. For inputs where the noise affects a localised part of the image, applying methods that attempt to denoise the entire image can adversely affect the clean portions. To address this issue, the authors propose a deep reinforcement learning-based framework for selective image denoising. The framework uses a two-step procedure that first identifies the noisy patch and then denoises the extracted patch. The authors use reinforcement learning for noise localization and PixelRL (Furuta et al. [2019]) for noise removal.

Next, we perform an article segmentation model for noisy degraded old newspaper images as a downstream task. These are Bangla newspaper images taken from the era of 1937-1980. We

propose a deep model based on convolution, dilated convolution, and skip connections to handle noise elements in these images. In this work we faced two major challenges: 1. noise was getting super resolved while enhancing document, 2. there were noise patches denoising which lead to text removal from clean areas. The focus of the work was to demonstrate the benefits of the preprocessing steps of denoising and super resolution on a downstream task.

...