

In this thesis we presented our work on certain problems pertaining to preprocessing in document images. We contributed by developing algorithms to process noisy degraded documents. We started with handcrafted techniques and gradually shifted towards deep learning approaches. We hope our work will fit as effective preprocessing steps for higher-level tasks such as, text recognition, semantic analysis and classification. In this chapter, we summarize our work and highlight the key findings of our thesis.

9.1 CONCLUSION

Document image processing primarily deals with digitizing the document image which in turn makes information indexing and retrieval faster and easier. There are various processing steps that are performed to achieve digitization. These tasks involve entity segmentation, recognition, interpretation, and identifying semantic associations between entities. Handwritten text images pose a higher challenge in comparison to machine-printed text images. In this thesis, we work towards degraded document processing and automated form processing. The contributions made in this thesis are summarized below.

1. In chapter 3, we applied the word spotting problem in cluttered scenario where some words were crossed out using line strokes. These line strokes could be made accidentally by human or could be formed due to the word getting overlapped with borders of form fields, tables, etc. Word spotting done by previous researchers have not considered cluttered scenarios. We came up with this problem while dealing with handwritten forms, where handwritten text was overlapping with frame fields. To address this problem, we have proposed combinatorics Vertical Projection Profile (cVPP) feature of cluttered word. This feature vector was matched with Vertical Projection Profile(VPP) of uncluttered words using the proposed modified Dynamic Time Warping algorithm. This allowed matching with words cluttered with a strike-through.
2. In chapter 4, we presented an algorithm for preprocessing printed and handwritten content in form images. We first addressed the task of form segmentation and we extracted the field cells present in a form image. To segment a form image, we proposed a modified region growing algorithm. This algorithm also considers the broken frame boundaries which sometimes occur due to degradation during scanning of forms. These broken borders erroneously prompt traditional region growing algorithms to merge the cells. Our modified region growing algorithm avoids this problem. We also worked on preprocessing steps to counter irregular spacing between the handwritten text letters in a word. This problem occurs when a writer writes with more inter-character spacing at the beginning of the word but reduces the spacing towards the end to fit the text within the confines of the framed border..
3. In Chapter 5, we proposed a graph autoencoder model to associate form field elements in heterogeneous handwritten form images. The form image is transformed into its graphical representation, which serves as input for an autoencoder model. The autoencoder predicts

the associations among the form field components. We provide a new dataset of handwritten form image.

4. In chapter 6, we present a unified framework for super resolution and denoising using reference-based unsupervised learning. This not only solves the problem of noise getting enhanced while super-resolving an image but also mitigates the need of a large dataset for training. In this work we have extended SinGAN model for denoising purpose. Our two phase training approach involved training the SinGAN to learn patch characteristics from a clean image and then use the learned discriminator to provide adversarial loss to achieve denoising and super resolution.
5. In chapter 7, we proposed selective denoising which involves localizing the noisy patch and selectively denoising that patch, while leaving the other regions unchanged. This approach, when contrasted against whole image denoising, prevents information loss which happens when denoising algorithms modify pixel values in the clean portions of the image. We used a reinforcement learning-based framework for noise patch localization and PixelRL for denoising the noise patch.
6. In chapter 8, we show the application of the developed selective denoising and super-resolution algorithms on a dataset of degraded scanned images of old newspapers. The benefits of the pre-processing steps are shown in terms of improvement in the performance of the downstream task of segmenting newspaper pages into constituent articles.