Chapter 7

Conclusion and Future Work

This chapter recalls the context of the thesis, provides a summary of the main contributions of this dissertation, and finally outlines the future research directions.

Conclusion

In this thesis, we proposed scene text detectors that work efficiently in an unconstrained environment. Our proposed network detects text instances, signs, doorplates, license plate numbers, and markings with high accuracy. We mitigate the problem caused by the partial occlusion, truncation, blur, camera shake, poor contrast, and faint edges text in scene images. For occlusion, we focus on local and global structural contextual information, whereas for blurred text, we enhance the transformation modeling capability of the features and focus on a multi-attention network based on finer spatial features and inter-channel dependencies. In case of faint edges text, we incorporate semantic edge supervision with non-local attention. We have reduced misclassification by occlusion, blur, and fain edges text by addressing the problem of inter-class interference using Gaussian softmax. We perform an extensive set of experimentation to show the efficacy of our models. We use publicly available benchmark datasets with standard matrices for both detection and recognition. The proposed outperforms the existing

methods in terms of performance in an unconstrained environment. We create a noisy scene text image dataset for detection and recognition of text instances in noisy scene images.

Future Directions

The scene text analysis in an unconstrained environment using deep network presented in this thesis is still an ongoing research problem. The work described here can be extended further, in the following directions:
In future, the work presented in the thesis can be extended for addressing the problem caused by the mixture of multiple noises, label noise, and multi-label text generation.

Mixture of multiple noises: The mixture of multiple noises like partial occlusion, blur, truncation, perspective distortion, fainted edge text cause highly adversarial impact on the text detection process. In case of text spotting, the feature map contains information about edge contrast, scale, and mean luminance at the edge location and more text properties are responsible for detecting text using a deep network. Due to the presence of a mixture of unconstrained noises the text properties preserved in the extracted feature are lost and cause misclassification in text detection. The mixture of noises causes transitions in surface properties of scene images or illumination. The extracted feature map of scene text images are very informative and allows the visual system to take an early decision in real-time application. The presence of a mixture of multiple noises weakens the information present in the extracted feature map and it increases the misclassification problem in scene text detection.

Label noise: Label noise is considered to be the observed label which is classified incorrectly because of imbalanced classes. Label noise increases the deep network model complexity. Label noise can come from several sources like 1) Insufficient information in data on which the model is trained. 2) Errors are caused by the expert in labeling

the data. 3) Training of the model is performed with imbalanced labels in data.

Multi-label text generation: The problem of multi-label text generation causes serious issues in text recognition in terms of multi-label classification. In the causes when a detected scene text is resembled as having two or more labels for a single text instance.

Finally, in real-life applications, the captured images may have a mixture of noises, label noise, or have multi-label text, which is mentioned in the thesis added more challenge in the scene text detection process.