PREFACE

Semantic image labelling, also known as pixel-based classification, is a task of segmenting the objects within an image by pixel-level spectral similarity. Semantic image labelling has also referred to as image parsing, which is a process of decomposing the image in different regions and constructing a structured input. In this thesis, we summarise the four aspects of research in the semantic labelling, i.e., classical machine learning(ML), feature engineering, deep learning(DL), and relaxation labelling(RL).

The above mentioned four aspects of semantic labelling lead to realizing that it is not a separate domain but a natural step in moving from coarse to fine interpretation. The original procedure could have been derived from a classification scheme, which predicts the label for a complete input. This process is also known as image category classification in the literature. This process has various applications such as bio-metric image classification, classification of tumor in a different grade, classification of different classes of species, digit classification using digit databases, emotion detection using face databases, and category classification using features from CBIR systems. A vast set of image databases, such as MNIST, CIFAR, ORL, YALE, etc., are available to validate the category classification methods.

In the previous classification methods, i.e., coarse grain, the classifiers predict the objects or provide a rank list in case of many objects. The next step is to localize and detect the objects within the image, which is a fine grain inference. The main aim of such inference is not only to provide classes but also some specific information such as the spatial location of classes, centroids, and bounding boxes. These kinds of fine-grain classification processes have performed on pixel-level, not image level. Therefore, they are computationally costly. Such inferences in various literature have denoted semantic labelling, pixel-based semantic classification/segmentation, or semantic image parsing. Consequently, we can summarise the semantic parsing as an image-based method to achieve fine-grain predictions. The goal is to make a dense prediction for the label of each pixel in such a way that each pixel is labelled with its class of enclosed regions. These kinds of fine-grained studies are the main interest of this thesis.

Another important aspect is to select the complex data-sets for semantic labelling. We have used low dimensional RGB images such as sift-flow, pascal-voc data-sets, and high

dimensional hyperspectral(HD-HSI) data-sets to perform the semantic labelling based experiments. The facial expression based images in ORL, YALE-A, and B, COIL data-sets have been used for category prediction also.

We have discussed the classical machine learning methods, some feature engineering, and knowledge embedding techniques to develop accurate and efficient frameworks. The salient and CNN feature-based approaches have been adapted to achieve effective and robust features from raw data. Feature selection and dimension embedding have also played a pivotal role in proposed frameworks. Deep learning and CNN based approaches have been used to design the custom CNN architecture and exploited the significant results from the image. Some relaxation labelling-based methods have also detailed to improve the CNN and classical ML-based probabilistic outcomes significantly. The advantages and drawbacks of the proposed frameworks have discussed. The comparative analysis for benchmark image sets and evaluation matrices have also been performed. Finally, some encouraging future works have drawn out, and the conclusion has drawn for pixel-wise semantic scene labelling or image parsing.