Chapter 1

Introduction

1.1 Introduction

In computer vision, the ability to capture the overall and object-specific gist of a complex image scene has become a critical research domain. The complex scenes have analysed by two major ways (1) gist recognition or scene categorization and (2) pixel-wise semantic classification or scene parsing. The gist recognition has used to categorize the scene in indoor, outdoor, busy roads and clean beaches, etc. [28]. In semantic classification, we try to capture the objects in deferential shapes or textures. The computation complexity of semantic classification is very high since the empirical process works at each pixel level. Therefore, this process also termed as pixel-based scene labelling in various literature. The pixel-based scene labelling and semantic classification have become major research lines for medical, transportation, and engineering research. Many benchmark image-sets are available to verify the proposed algorithms. In our work, we have primarily focused on pixel-based semantic parsing of images, which has also termed as scene labelling. The inception of deep learning and other machine learning paradigms has given an extra thrust in scene labelling algorithms. In a mixed coarse grain and fine grain environment, image-specific tasks such as category classification and object detection annotate the image in various categories. Category classification puts the images in some predefined categories according to their feature descriptors. Detection locates certain objects of interest in the image. The semantic classification has manifold applications in road signal detection, crypto colon clustering, remote sensing, and brain-tumor detection in medical images. An auto-driving system for driver assistance completely relies upon semantic classification and scene labelling. The emergence of artificial intelligence and deep neural networks(DNN) has given a strong momentum towards efficient learning paradigms. Before the arrival of DNNs, machine learning(ML) based, classical methods have applied for scene labelling. Therefore, the semantic classification has been divided into two main categories: (1) traditional and (2) recent methods [14]. For both categories, various issues like colour to grayscale conversion, feature extraction, and objective quality assessment for image decoloring has been proposed in [20].

1.1.1 Major Issues

Pixel-wise semantic image labelling is a challenging task in machine learning and computer vision domain. Several methods and architectures are in progress that requires an accurate precision with high efficiency(low computation time). The higher number of training samples leads to highly accurate predictions up-to a certain extent. The data collection cost of training samples, to obtain pixel-label train sets, is very high, especially in high dimensional images. Therefore, we have an insufficiency problem of training samples in semantic labelling. In this domain, there is a requirement for a robust model that trains on minimal samples with minimal computation time and performs accurate testing on large testing samples. Therefore, obtaining a trade-off between high accuracy and low computational time is a significant challenge. Several methods and architectures have proposed for efficient scene labelling, yet the problem has not solved efficiently [32]. There are many more challenges that occur in semantic segmentation based scene labelling that we have described as:

- Identification of specific class patterns in an image is a significant issue in the labelling problems in the semantic scene. Existing problems have to compromise on either accuracy or deal with huge computation time.
- The intensive feature extraction causes an adverse effect in semantic predictions in RGB, gray, and optical images. In sensor-based, high dimensional datasets such as hyperspectral images, both feature extraction and machine learning prediction lead

to shallow performances as well as worst prediction accuracy due to the class mixing at each pixel.

- Most of the machine learning algorithms directly learn from spectral data of an image. The feature extraction and fusion techniques are required to improve the prediction accuracy and performance. The features information in the fusion process has also called as prior knowledge. The redundant and wrong prior knowledge can mislead to obtain the improper patterns in data hence increasing the false positive rate.

- Unlike the general and optical images, the high dimensional sensor-based images face the curse of dimensionality in some pixels. Therefore, robust feature extraction methods or subspace projection is required to detect object patterns and subsequently perform the semantic labelling.

- Some label relaxation-based post-processing methods are highly effective in improving the prediction performance in machine learning approaches. The label relaxation methods generate and fused the similarity-based external information in an unsupervised way to predict the output. Therefore, we have introduced some similarity-based label relaxation frameworks to increase label prediction.

- In machine learning(ML), feature extraction has to lead to robust handcrafted features, which results in excellent accuracy, but the features are not good enough to improve the performance after a certain threshold. It has been found that the deep neural network(DNN) based features have performed with a significantly higher performance than ML features. But the DNN based features are massive in amount, and computational complexity is very high. Therefore, it is crucial to construct a deep network-based architecture that can perform in high-dimensional and low-dimensional images for semantic classification.

- The process of finding exclusive and robust features to reduce redundant computation is a significant issue in computing techniques. Recently, in some studies, it has found that wavelet and morphological operations based methods can equally perform and some times outperform the DNN based features in terms of robust information. In DNNs, we do not have any control over the computation of the feature. Therefore, some duplicity can be found in features space. However, wavelet and morphological features have calculated using the iterative mathematical process that turns out exclusive and less computationally redundant features.

1.2 Objectives

The thesis work is focused on seven objectives that belong to four categories.

1.2.1 Exploring the DNN feature for scene labelling

Generally, the scene labelling based methods use the direct spectral values in deep convolution neural networks(CNN) to extract the local features and further uses the multi-layer perceptron(MLP) layer for scene classification. In deep networks, MLP layer generally uses the features obtained from the last convolution layer, which have features in a minimal spatial resolution. The features from intermediate layers do not play any role in label prediction. Therefore, our approach has extracted the features in all the resolutions from each convolution pooling pair. These features have used in the following objectives as:

Objective-1 Investigate the role of DNN features and their manifold embedding in the semantic label prediction process by using SVM(support vector machine) and regression-based classifier. Demonstrate the importance of manifold embedding in extreme features space.
For evaluation, we have used the standard evaluation matrices like overall accuracy(OA), average accuracy(AA), kappa value(K), class-wise accuracy, and computation time(T). The high dimensional hyperspectral images(HSI) have used for experimental validation. High dimensionality and low availability of training samples create the hurdle in the prediction process of semantic labeling. Therefore, the DNN features have used in the prediction process as the extra dimensions to prevent the low sample problem. Manifold embedding has used as a feature embedding procedure since DNN based features have added a new high dimensional feature space that resulted in gigantic image data-cube. The classical dimensionality reduction methods such as PCA and SVD can not be used

because they lead to a significant loss in feature information. Therefore, an information preservation-based dimension reduction is required to convert and store the data-cube in an optimum feature space. Finally, the comparative advantage of our approach over the direct spectral value-based techniques needs to be addressed.

Objective-2 In contrast to Manifold embedding, design a gradient-based method that selects the informative dimensions from the feature space by using salient features based approach and performs the prediction with SVM.

1.2.2 Exploring the Label relaxation and feature expansion based information generation

In various state-of-art works, the post-processing methods plays an essential role in the improvement of classifier prediction. In sensor-based high dimensional images, the image pixels are noisy due to large feature space. Therefore, most of the classifiers are unable to predict the pixel labels semantically accurate. In some state-of-art, this issue has resolved by using post-processing based label relaxation methods such as graph-cut integer optimization and variational optimization.

Objective-3 Design a post relaxation labelling based optimizer that fuses the external spatial information in the probabilistic prediction of semantic image labelling. The strength of this optimizer is depended upon the external data.

Objective-4 Propose a framework for the classical machine learning process that generates useful knowledge by expanding the feature space. Subsequently, design a robust feature selector method based on some shared properties like pairwise correlation.

1.2.3 Develop a deep learning architecture, study of manifold learning and convex optimization

Various state-of-art architectures have proposed for semantic labelling of an image. Still, these models are paired with some post-processing methods such as a conditional random field or Markov random filed to achieve high accuracy and efficiency. Therefore, we have designed a highly efficient DNN model that can extract the features from both high dimensional (HD) image as well as low dimensional image sets such as pascal-voc and sift-flow data-sets.

Objective-5 Examine the application of manifold learning on high dimensional image data-cube reduction using spatial potential and design the DNN architecture for semantic labelling.

Objective-6 Develop a five-layer architecture of DNN that predicts the pixel probability for each label and subsequently applied a convex optimization strategy to improve the pixel prediction probability.

1.2.4 Exploring the multi-scale morphological features and predicate based feature merging

Various types of morphological operations have introduced for feature extraction. In morphological image processing, the objects of an image are captured in different frames by applying some iterative criteria. This generates the high dimensional feature space, which again forms a feature data-cube. The iterative morphology is an uncontrolled operation. It creates highly informative features, but some features are redundant by some nature-inspired criteria such as correlation, structural similarity, and entropy. Therefore, our secondary aim is to reduce the feature space by using these criteria as a decision factor.

Objective-7 Investigate the iterative morphological features and develop a predicate based technique to select the feature based on some criteria and perform scene labelling on chosen features.

1.3 Contribution

The main contribution of our work is divided into five major sections that resolve the objectives mentioned above. Covering the first and second objectives, we have investigated the CNN features in different resolutions and proposed the manifold learning based frame-work for scene labelling using SVM. For the second objective, we have uncovered some gradient-based determinant and trace operations to find salient objects from deep features. Dimension expansion experiments have also been conducted in this section. Further, these salient objects have used for label prediction. In objectives three and four, the label relaxation-based methods have introduced. We have optimized the probabilistic values by using a filter-based spatial fusion method considering objective three. For objective four, we have applied a cluster energy-based optimization technique for label relaxation. To cover the deep CNN architecture based objectives, i.e., objective five and six, the novel patch-wise CNNs have proposed with manifold learning and label relaxation. Some other quality matrices have also introduced to measure performance. In the seventh objective, we have proposed a mathematical morphology based feature extraction method followed by an unsupervised feature selection method to discard the redundant features for semantic labelling.

1.3.1 Deep CNN feature and manifold reduction based scene labelling

To examine the first objective, we have explored the deep CNN based spatial features. CNN's typically reduces the spatial resolution of image features. At the end of convolution pooling operations, the features collect better spectral knowledge but worse spatial learning. Therefore, we have performed a de-convolution on each layer of CNN. Features have obtained at each resolution to avoid any loss of spatial information. Further, the considerable feature space has generated, and another challenge is to reduce the feature-space for the convenient computation. The feature space has been reduced by using a knowledge embedding method. The main focus of this work is to develop a feature extraction, feature embedding, and label prediction for the segmentation and scene labelling. The dimension of features space is very high, and the worst-case complexity of the algorithm is $o(N^2) + O(d^3)$. The experiments have performed on publicly available high dimensional hyperspectral images and it has found that this framework has outrun the performance of some existing methods.

1.3.2 Extracting the salient features from Deep CNN and investigate the feature expansion

Following the second and third objective, salient feature extraction has considered in addition to deep features. The mathematical concept of derivative, trace, determinant has applied for extracting only salient features. The main advantage of this scheme is to generate informative knowledge first and then use some saliency extraction methods to obtain the most suitable features from extended features space. These salient features have also denoted as feature descriptors. For the empirical study of the saliency selection approach, we have expanded the feature space by using two different methods (1): by de-convolution of CNN layers and (2): by the expansion of features by statistical properties such as correlation.

1.3.3 Label relaxation for scene labelling

To address the fourth and fifth objectives, we have implemented the label relaxation techniques to improve the label prediction probabilities by the convex optimization of some label relaxation function. Optimization-based techniques are very much prevalent in scene relaxation. In this section, we have used some standard function that optimizes the parsed input image clusters to better-relaxed labels. The convex optimization has been used for better convergence of objective functions. Following the third objective, a variational optimization-based objective function, previously used as total variation, has been used for label relaxation. Considering the fifth objective, we have used a cluster energy function in which we have minimized the cluster energy with a graph-cut optimization process to find the energy of the improvised cluster. Subsequently, we have performed a comparative analysis of proposed schemes with some state-of-art methods.

1.3.4 CNN architecture design and relaxation labelling

The feature extraction in scene parsing(labelling) is an important task. Considering the sixth and seven objectives, we have demonstrated designing a customized CNN architecture that takes our image in the form of patches or complete data sets and produces the scene parsing probability of each pixel after the soft-max layer. Considering the seventh objective, we have applied a jacobian method based convex optimization to improve the scene parsing probabilistic output. Following the sixth objective, we can use a manifold reduction method to efficiently reduce the feature dimensions before passing them into our scene parsing based CNN. These approaches have been compared with recently proposed CNN methods and found a significant improvement in accuracy and computation time.

1.3.5 Morphological features and predicate based feature merging for scene labelling

To address the eighth objective, we have applied a morphological learning-based feature extraction method with some iterative policies. We have considered some geometrical attributes to extract the features. To accurately identify the natural objects in the image, we have used some tree pruning strategy to obtain non-redundant and informative features. These features are more informative than a deep feature because they have calculated in a controlled environment. Due to high information content, these features can be reduced to improve computation time. Therefore, a natural criteria based feature selection has applied and has tested for both semantic classifications as well as categorical label prediction. The complexity analysis and comparison with state-of-art have performed to investigate the performance of scene parsing with recent methods.

1.4 Thesis Organization

The thesis has organized into eight chapters.

Chapter two details the literature survey for machine learning, deep learning, image processing, semantic scene parsing/labelling, labelling relaxation, and evaluation techniques.

Chapter three has studied CNN based feature learning, manifold learning, and logistic regression techniques for semantic scene labelling.

Chapter four study two experiments: (1) A feature expansion method by generating the informative knowledge using correlation followed by an efficient feature merging technique. (2) A gradient-based salient feature extraction and their application in scene parsing. In this chapter, some feature engineering methods have been applied to enhance and compress the feature quality to increase the accuracy and efficiency of the algorithm.

Chapter five investigate two experiments: (1) the total-variation based spatial relaxation methods with some subspace reduction methods and (2) A graph energy-based two-phase label optimization technique to improve the semantic classification results. In this chapter, some robust fusion mechanisms have been demonstrated to enhance classification accuracy.

Chapter six studies the optimization of edge features in a CNN probabilistic output by using jacobian optimization techniques. Some other evaluation parameters and statistical tests have introduced to evaluate the proposed methods. In the case of HD image, manifold embedding can be applied, as discussed in objective-5, before passing the training features in Deep CNN.

Chapter seven determine the iterative morphological feature extraction methods and multi-criteria feature pooling to evaluate a semantic and categorical target prediction.

Chapter eight concludes the thesis with a contribution summary and future research lines.