The emerging recent large scale visual recognition methods, and in particular the deep Convolutional Neural Networks (CNN), are promising to revolutionize many computer vision based artificial intelligent applications, such as autonomous driving and online image retrieval systems. One of the main challenges in large scale visual recognition is the complexity of the corresponding algorithms. This is further exacerbated by the fact that in most real-world scenarios they need to run in real time and on platforms that have limited computational resources. This dissertation focuses on improving the efficiency of such large scale visual recognition algorithms from several perspectives.

First, to reduce the complexity of large-scale classification to sub-linear with the number of classes, a probabilistic label tree framework is proposed. A test sample is classified by traversing the label tree from the root node. Each node in the tree is associated with a probabilistic estimation of all the labels. The tree is learned recursively with iterative maximum likelihood optimization. Comparing to the hard label partition proposed previously, the probabilistic framework performs classification more accurately with similar efficiency.

Second, we explore the redundancy of parameters in Convolutional Neural Networks (CNN) and employ sparse decomposition to significantly reduce both the amount of parameters and computational complexity. Both inter-channel and inner-channel redundancy is exploit to achieve more than 90% sparsity with approximately 1% drop of classification accuracy. We also propose a CPU based efficient sparse matrix multiplication algorithm to reduce the actual running time of CNN models with sparse convolutional kernels.

Third, we propose a multi-stage framework based on CNN to achieve better efficiency than a single traditional CNN model. With a combination of cascade model and the label tree framework, the proposed method divides the input images in both the image space and the label space and processes each image with CNN models that are most suitable and efficient. The average complexity of the framework is significantly reduced, while the overall accuracy remains the same as in the single complex model.