Visual saliency is the ability of a vision system to promptly select the most relevant data in the scene and reduce the amount of visual data that needs to be processed. We propose a novel unsupervised approach to detect visual saliency in videos of natural scenes. We employ a hierarchical segmentation technique to obtain super-voxels of a video and simultaneously we build a dictionary from cuboids of the video. Then we create a feature matrix from coefficients of dictionary elements. Next, we decompose this matrix into sparse and redundant parts and obtain salient regions using group lasso. The applicability of our method is examined on four video data sets of natural scenes. Saliency detection only highlights important regions, and there is no notion of classes in saliency. In Semantic Segmentation, the aim is to assign a semantic label to each pixel in the image. In typical approaches, general context information beyond the simple smoothness is not considered. In this dissertation, two supervised approaches to address this problem are proposed. First, an approach to discover interactions between labels and regions using a sparse estimation of the precision matrix using graphical lasso. Second, a knowledge-based method to incorporate dependencies among regions in the image during inference. High-level knowledge rules - such as co-occurrence, spatial relations, and mutual exclusivity - are extracted from training data and transformed into constraints in Integer Programming formulation. A difficulty which most supervised semantic segmentation approaches are confronted with is the lack of enough training data. Annotated data should be at the pixel-level, which is highly expensive to achieve. To address this limitation, next a semi-supervised learning approach to exploit the plentiful amount of available unlabeled, as well as synthetic images generated via Generative Adversarial Networks (GAN) is presented.