Degraded Image Segmentation, Global Context Embedding, and Data Balancing in Semantic Segmentation

Dazhou Guo

Abstract
Recently, semantic segmentation -- assign a categorical label to each pixel in an image -- plays an important role in image understanding applications, e.g., autonomous driving, human-machine interaction, and medical imaging. Semantic segmentation has made progress by using the deep convolutional neural networks, which are surpassing the traditional methods by a large margin. Despite the success of the deep convolutional neural networks (CNNs), there remain three major challenges.

The first challenge is how to segment the degraded images semantically -- degraded image semantic segmentation. In general, image degradations increase the difficulty of semantic segmentation, usually leading to decreased semantic segmentation accuracy. While the use of supervised deep learning has substantially improved the state-of-the-art of semantic image segmentation, the gap between the feature distribution learned using the clean images and the feature distribution learned using the degraded images poses a major obstacle to improve the degraded image semantic segmentation performance. We propose a novel Dense-Gram Network to more effectively reduce the gap than the conventional strategies and segment degraded images. Extensive experiments demonstrate that the proposed Dense-Gram Network yields state-of-the-art semantic segmentation performance on degraded images synthesized using PASCAL VOC 2012, SUNRGBD, CamVid, and CityScapes datasets.

The second challenge is how to embed the global context into the segmentation network. As the existing semantic segmentation networks usually exploit the local context information for inferring the label of a single-pixel or patch, without the global context, the CNNs could miss-classify the objects with similar color and shapes. In this dissertation, we propose to embed the global context into the segmentation network using the object's spatial relationship. In particular, we introduce a boundary-based metric that measures the level of spatial adjacency between each pair of object classes and find that this metric is robust against object size induced biases. We develop a new method to enforce this metric into
the segmentation loss. We propose a network, which starts with a segmentation network, followed by a new encoder to compute the proposed boundary-based metric, and then trains this network in an end-to-end fashion. We evaluate the proposed method using CamVid and CityScapes datasets and achieve favorable overall performance and a substantial improvement in segmenting small objects.

The third challenge of the existing semantic segmentation network is how to address the problem of imbalanced data induced performance decrease. Contemporary methods based on the CNNs typically follow classic strategies such as class re-sampling or cost-sensitive training. However, for a multi-label segmentation problem, this becomes a non-trivial task. At the image level, one semantic class may occur in more images than another. At the pixel level, one semantic class may show larger size than another. Here, we propose a selective-weighting strategy to consider the image- and pixel-level data balancing simultaneously when a batch of images are fed into the network. The experimental results on the CityScapes and BRATS2015 benchmark datasets show that the proposed method can effectively improve the performance.