Names : Shuo Cheng, Mingxiang Cai, Siman Wang, Dingcheng Hu

Title: Structure preserved semantic image segmentation

Outline

1. **Motivation**:
Semantic image segmentation plays a very important role in a wide variety of computer vision applications (e.g., robotics, autonomous driving, human-computer interaction). Previous methods (such as graph-cut and energy minimization) only consider hand-craft features (e.g., the color information, the affinity of the pixels), thus can not reach satisfying results. Recent years, with the popularity of deep learning, the performance of semantic image segmentation has been boosted a lot. However, due to the limited resolution of the prediction, the results are still unsatisfactory. In this paper, we innovatively integrate deep learning into conditional random field framework. We treat the prediction of the neural nets as a unified term, and we construct the pairwise term using the semantic features from the neural nets. Based on minimizing the global energy, we can obtain more smooth and accurate results. Exhaustive experiments demonstrate the effectiveness of our proposed method.

2. **Statement of problem**

**Semantic Image Segmentation**
Semantic segmentation is a pixel level classification problem. In other words, the algorithms need to label the category of each pixel in the image. It is a very challenging problem because of the inherited ambiguity (e.g., various shapes, colors, occlusions) and the noise in the image space (i,e., lighting, blurring). Traditional methods usually consider hand-craft features, and obtain the final prediction via minimizing energy function. These methods, often fail as they ignore the semantic relationships within different parts of the objects and the layout of the whole scene. Deep learning based methods bring new revolution to the computer vision fields in recent years. Current state of the art algorithms typically use convolution neural network to extract per-pixel features, and then do the classification in an end-to-end manner. Although these methods have already outperform previous methods by a large margin, the results still have some drawbacks (e.g., low-resolution, incorrect boundaries).

**Conditional random fields**
Conditional random fields (CRF) is a probabilistic framework for inferring structured and sequential data. It is widely used in natural language processing problems and have achieved a certain level of success. In the pipeline of CRF learning based image segmentation, finding a good feature representation is of great significance, and can have a profound impact on the segmentation accuracy. Most previous studies rely on hand-crafted features, e.g., using color histograms, HOG or SIFT descriptors to construct bag-of-words features, these methods have strong limitations when applied in complicated scenes. Based on these observations, we innovatively integrate deep learning method into CRF framework.

**Deep learning in CRF**
Recently, feature learning and especially deep learning methods have gained great popularity in machine learning and related fields, This type of methods typically takes raw images as an input and learn a (deep) representation of the images, and have found phenomenal success in various tasks such as classification, object detection, and tracking, etc. Deep learning methods attempt to model high-level abstractions in data at multiple layers, inspired from the cognitive processes of human brains, which generally starts from simpler concepts to more abstract ones. The advantages of deep learning methods shed light on us to integrate it into the CRF framework. On one hand, we want to take merits of the powerful features learned through the data-driven method, on the other hand, we rely on graph model to guarantee structural completeness. We innovatively incorporate the learned features to model the co-occurrence pairwise potentials, and minimize the energy function to achieve smooth and structured predictions.

3. **Wishlists**:

- Data

  Cityscape Dataset (Stereo video sequences recorded in street scenes, with pixel-level annotations.)

- Method :
  - Step 1 : train CNN for semantic segmentation

- Step 2 : train CRF model with features extracted from pre-trained CNN using self-defined optimization function
- Step 3 : compare performance on the benchmark dataset.

- Innovative part :

  The advantages of deep learning methods shed light on us to integrate it into the CRF framework. On one hand, we want to take merits of the powerful features learned through the data-driven method, on the other hand, we rely on graph model to guarantee structural completeness. We innovatively incorporate the learned features to model the co-occurrence pairwise potentials, and minimize the designed new energy function to achieve smooth and structured predictions.

4. **Task assignment of each member**

We work together and contribute to those four sections equally:

1. Paper survey
2. Coding
3. Experiment
4. Write up

5. **References**

CRF learning with CNN features for image segmentation https://www.sciencedirect.com/science/article/abs/pii/S0031320315001582
Segmentation-Aware Convolutional Networks Using Local Attention Masks https://arxiv.org/abs/1708.04607
Conditional Random Field and Deep Feature Learning for Hyperspectral Image Segmentation https://arxiv.org/pdf/1711.04483.pdf
Conditional Random Fields as Recurrent Neural Networks http://www.robots.ox.ac.uk/~szheng/papers/CRFasRNN.pdf
Multiscale Conditional Random Fields for Image Labeling http://www.cs.toronto.edu/pub/zemel/Papers/cvpr04.pdf
Multi-class image segmentation using Conditional Random Fields and Global Classification