Gaussian Conditional Random Field Network for Semantic Segmentation

Raviteja Vemulapalli, Rama Chellappa
University of Maryland, College Park

Oncel Tuzel, Ming-Yu Liu
Mitsubishi Electric Research Laboratories, Cambridge

Main contribution: An end-to-end trainable deep network architecture that combines CNNs with a Gaussian conditional random field model.

Motivation

- CNNs do not explicitly model the interactions between output variables which is very important for structured prediction tasks such as semantic segmentation.
- Various recent approaches [1,2] combine CNNs with discrete CRFs.
- Inference techniques in the case of discrete CRFs do not have optimality guarantees upon convergence.
- In contrast, Gaussian mean field inference gives optimal solution upon convergence in the case of a Gaussian CRF.

Gaussian CRF Network

- We use a Gaussian CRF model on top of a CNN.
- Each discrete output variable is replaced by a vector of $K$ continuous variables $y_i = [y_{i1}, \ldots, y_{iK}] \in \mathbb{R}^K$.
- $y_{ik}$ represents the score for $k^{th}$ class at $i^{th}$ pixel.
- $X$: Input image
- $P(Y|X) \propto e^{-\frac{1}{2}y^T \Sigma y}$, where $\Sigma$ represents the similarity measure between pixels $i$ and $j$.
- $F \succ 0$ is a parameter matrix that defines a Mahalanobis distance function.
- $F_{ij}$ is computed as $F_{ij} = \sum_{k=1}^{K} (y_{ik} - y_{jk})^2$.

Gaussian CRF Inference

- We use the iterative Gaussian mean field (GMF) inference approach.
- We unroll the GMF inference steps into a deep network.
- GMF update equation (optimal coordinate descent step):

$$y_{ik}^{t+1} = \left( I + \sum_{j} W_{ij} y_{jk}^{t} \right)^{-1} \left( r_{ik} + \sum_{j} W_{ij} y_{jk}^{t} \right)$$

References: