Efficient Learning for Discriminative Segmentation with Supermodular Losses

Jiaqian Yu
jiaqian.yu@centralesupelec.fr

CentraleSupélec
Université Paris-Saclay & Inria
Châtenay-Malabry, France

Matthew B. Blaschko
matthew.blaschko@esat.kuleuven.be

Center for Proc. Speech and Images
Dept. Elektrotechniek - ESAT
KU Leuven, Belgium

(a) Our loss is based on an 8-connected neighborhood.

$$E = - \underbrace{\begin{pmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \end{pmatrix}}_{\text{inference pairwise potential}} - \underbrace{\begin{pmatrix} 0 & \gamma \\ 0 & 0 \end{pmatrix}}_{\text{loss pairwise potential}}$$

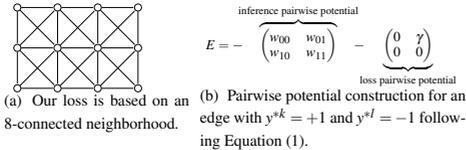(b) Pairwise potential construction for an edge with $y^{*k} = +1$ and $y^{*l} = -1$ following Equation (1).

Figure 1: Non-submodularity of the joint loss augmented inference using the same mapping to a set function for inference and loss functions.

Algorithm 1 ADMM in scaled form for finding a saddle point of the Lagrangian $\mathcal{L}(y_a, y_b, \lambda) = -\langle w, \phi(x, y_a) \rangle - \Delta(y^*, y_b) + \lambda^T (y_a - y_b) + \frac{\rho}{2} \|y_a - y_b\|_2^2$

1: Initialization $u^0 = 0$
2: **repeat**
3: $\quad y_a^{t+1} = \arg\min_{y_a} -\langle w, \phi(x, y_a) \rangle + \frac{\rho}{2} (\|y_a - y_b^t + u^t\|_2^2)$
4: $\quad y_b^{t+1} = \arg\min_{y_b} -\Delta(y^*, y_b) + \frac{\rho}{2} (\|y_a^{t+1} - y_b + u^t\|_2^2)$
5: $\quad u^{t+1} = u^t + (y_a^{t+1} - y_b^{t+1})$
6: $\quad t = t + 1$
7: **until** stopping criterion satisfied

Several non-modular loss functions have been considered in the context of image segmentation. These loss functions do not necessarily have the same structure as the segmentation inference algorithm, and in general, we may have to resort to generic submodular minimization algorithms for loss augmented inference. Although these come with polynomial time guarantees, they are not practical to apply to image scale data.

In this work, we first propose a supermodular loss function that is itself optimizable with graph cuts. It counts the number of incorrect pixels plus the number of pairs of neighboring pixels that both have incorrect labels

$$\Delta(y^*, \tilde{y}) = \sum_{j=1}^p [y^{*j} \neq \tilde{y}^j] + \sum_{(k,l) \in \mathcal{E}_\ell} \gamma [y^{*k} \neq \tilde{y}^k \wedge y^{*l} \neq \tilde{y}^l] \quad (1)$$

where $[\cdot]$ is Iverson bracket, $\mathcal{E}_\ell$ is a loss specific edge set and γ is a positive weight. We may identify this function with a set function to which the argument is the set of mispredicted pixels.

While being incorporated in a joint loss-augmented inference leads to non-submodular potentials, we therefore use the alternating direction method of multipliers (ADMM) based decomposition strategy (Algorithm 1). It consists of alternatingly optimizing the loss function and performing MAP inference, with each process augmented by a quadratic term enforcing the labeling determined by each to converge to the optimum of the sum. In this way, we gain computational efficiency, making new choices of loss functions practical, while simultaneously making the inference algorithm employed dur-

(a) Hamming (b) 8-connected (c) Hamming (d) 8-connected

(e) Hamming (f) 8-connected (g) Hamming (h) 8-connected

Figure 2: The segmentation results of prediction trained with Hamming loss and our supermodular loss.

ing training closer to the test time procedure.

We show improvement both in accuracy and computational performance on the MR Grabcut database (Fig. 2) and a brain structure segmentation task, empirically validating the use of a supermodular loss during training and the improved computational properties of the proposed ADMM approach over the Fujishige-Wolfe minimum norm point algorithm. We envision that this can be of use in a wide range of application settings, and an open source general purpose toolbox for this efficient segmentation framework with supermodular losses is available for download from https://github.com/yjq8812/efficientSegmentation.