DC-SegNet: A discretely constrained deep network for weakly supervised segmentation

Jizong Peng
ETS Montreal, Canada

Christian Desrosiers
ETS Montreal, Canada

Marco Pedersoli
ETS Montreal, Canada

Abstract

An efficient strategy for semi- and weakly-supervised segmentation is to impose constraints or regularization priors on target regions. Recent efforts have focused on incorporating such constraints in the training of convolution neural networks (CNN), however this has so far been done within a continuous optimization framework. Yet, various segmentation constraints and regularization can be modeled and optimized more efficiently in a discrete formulation. This paper proposes a method, based on the alternating direction method of multipliers (ADMM) algorithm, to train a CNN with discrete constraints and regularization priors. This method is applied to the segmentation of medical images with weak annotations, where both size constraints and boundary length regularization are enforced. Experiments on a benchmark cardiac segmentation dataset show our method to yield a performance near to full supervision.

1 Introduction

Higher-level (or weak) annotations, such as image-level tags [1, 2, 3, 4], bounding boxes [5, 6] and scribbles [3, 7], can generally be obtained more efficiently than pixel-wise annotations. This type of weak supervision has therefore been intensively investigated for segmentation. Another valuable strategy in a semi-supervised or weakly-supervised setting is to use prior knowledge to guide the segmentation of unlabeled or weakly-labeled images during training [4, 3]. This is particularly useful for medical segmentation problems, where information about the target region is often known beforehand.

In [4], Pathak et al. use a latent distribution and KL-divergence to constrain the output of a segmentation network. The proposed method allows decoupling the optimization of the network parameters with stochastic gradient descent (SGD) and updating the latent distribution under constraints. They employ their method in a semi-supervised setting to impose size constraints and image level tags (i.e., force the presence or absence of given labels) on the regions of unlabeled images. More recently, Kervadec et al. [3] proposed an L_2-norm penalty to guide the segmentation of weakly-labeled images toward the satisfaction of similar constraints. Results show their method to provide better accuracy and stability than the one of Pathak et al., when few pixels of an image are labeled.

While these works have reduced the performance gap between CNNs trained with fully-annotated images and those trained in a semi-supervised or weakly-supervised setting, they have explored constrained optimization from a continuous perspective. However, many constraints, including region size, are better expressed discretely. Hence, limiting the sum of pixel probabilities over an image does not guarantee that, after thresholding these probabilities to obtain labels, size constraints will be met. Similarly, regularization priors for segmentation, such as boundary length [8], star-shapeness [9] and compactness [10], are usually discrete and optimizing them in a continuous framework is hindered by local minima. Lastly, solving sub-problems in a discrete manner, instead of using gradient descent, benefits from globally-optimal algorithms which can significantly improve the current solution in a single update step.

The contribution of this work is twofold. 1) We present a first method to train a CNN with discrete constraints and regularization priors. The proposed method uses an efficient strategy, based on

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

the alternating direction method of multipliers (ADMM) algorithm, to separate the optimization of network parameters with SGD from optimizing discretely-constrained segmentation labels. We show that updating these discrete labels can be carried out to optimality, in polynomial time. 2) We apply the proposed method for the segmentation of medical images with weak annotations. While previous works have considered either size constraints [3, 4] or boundary length regularization [11] for segmentation, our method combines these two priors in a single efficient model. Experiments on a benchmark cardiac segmentation dataset show our method to yield a performance near to full supervision.

2 The proposed method

We focus on the following weakly-supervised segmentation problem. Given a training dataset $\mathcal{D} = \{(X^i, \Omega^i, Y^i)\}_{i=1}^{|\mathcal{D}|}$ where X^i is an image, Ω^i is the subset of labeled pixels, and Y^i the corresponding labels, we want to learn a segmentation model parameterized by θ, such that $S(X^i, \theta)$ gives the label probabilities at each pixel. For simplicity, we suppose a two-class segmentation problem and, using $S^i = S(X^i, \theta)$ as short-hand notation, denote as $S^i_p \in [0, 1]$ the foreground probability predicted for a pixel p.

To learn parameters θ in this setting, we impose three constraints on the network output: 1) it should respect labeled pixel in Ω^i, 2) it should minimize a regularization term $\mathcal{L}_{\rm reg}$, and 3) it should satisfy global segmentation constraints $\mathcal{C}_j$, $j = 1, \ldots, M$. We formulate this task as the following optimization problem:

$$\operatorname*{argmin}_{\theta} \sum_{i \in \mathcal{D}} \mathcal{L}_{\rm lab}(S^i, Y^i) \, + \, \lambda \mathcal{L}_{\rm reg}(S^i), \quad \text{s.t. } \mathcal{C}_j(S^i) \, \leq \, 0, \ \ \forall i \in \mathcal{D}, \ j = 1, \ldots, M, \tag{1}$$

where λ controls the trade-off between between label satisfaction and regularization. Following [3], partial cross-entropy is used for the former:

$$\mathcal{L}_{\rm lab}(S^i, Y^i) \, = \, -\sum_{p \in \Omega^i} Y^i_p \log S^i_p \, + \, (1 - Y^i_p) \log (1 - S^i_p). \tag{2}$$

While any regularization prior can be used, for this work, we considered the boundary length (i.e., Potts model) [8]: $\mathcal{L}_{\rm reg}(S^i) = \sum_{(p,q) \in \mathcal{E}} W^i_{p,q} \cdot |S^i_p - S^i_q|$, where $\mathcal{E}$ is the set of neighbor pixels, and $W^i_{p,q} = \exp\big(-\frac{1}{2\sigma^2}(X^i_p - X^i_q)^2\big)$ measures the similarity of pixel intensity or color. As in [3] and [4], we impose global size constraints on the segmentation: $A_{\min} \leq \sum_p S_p \leq A_{\max}$.

Because the regularization prior and size constraints are discrete, the formulation in Eq. (1) cannot be optimized directly. To alleviate this problem, we use the following ADMM-based approach. We introduce two binary segmentation vectors $\widehat{Y}^i$ and $\widetilde{Y}^i$, which serve as ancillary variables for the regularization prior and size constraints, respectively, and constrain the network output S^i to be equal to these vectors. Using an augmented Lagrangian formulation, Eq. (1) becomes

$$\begin{aligned} &\operatorname*{argmax}_{\widehat{U}^i, \widetilde{U}^i} \operatorname*{argmin}_{\theta, \widehat{Y}^i, \widetilde{Y}^i} \sum_{i \in \mathcal{D}} \mathcal{L}_{\rm lab}(S^i, Y^i) \, + \, \lambda \mathcal{L}_{\rm reg}(\widehat{Y}^i) \, + \, \frac{\widehat{\mu}}{2} \|S^i - \widehat{Y}^i + \widehat{U}^i\|_2^2 + \frac{\widetilde{\mu}}{2} \|S^i - \widetilde{Y}^i + \widetilde{U}^i\|_2^2 \\ &\qquad\qquad \text{s.t. } A_{\min} \, \leq \, \sum_p \widetilde{Y}^i_p \, \leq \, A_{\max}, \ \ \widehat{Y}^i, \widetilde{Y}^i \, \in \, \{0, 1\}^{N_i}, \ \forall i \in \mathcal{D}. \end{aligned} \tag{3}$$

Here, $\widehat{U}^i$, $\widetilde{U}^i$ are Lagrange multipliers, and $\widehat{\mu}$, $\widetilde{\mu}$ are the corresponding penalty parameters. Since network parameters θ and discrete segmentation $\widehat{Y}^i, \widetilde{Y}^i$ are now only coupled via L_2 terms, we can optimize them independently as follows.

To update network parameters, we use a gradient descent technique, with the gradient given by

$$\sum_{i \in \mathcal{D}} \frac{\partial}{\partial \theta} \mathcal{L}_{\rm lab}(S^i, Y^i) \, + \, \Big(\widehat{\mu}(S^i - \widehat{Y}^i + \widehat{U}^i) \, + \, \widetilde{\mu}(S^i - \widetilde{Y}^i + \widetilde{U}^i) \Big) \frac{\partial S^i}{\partial \theta} \tag{4}$$

Using the property that $(\widehat{Y}^i_p)^2 \, = \, \widehat{Y}^i_p$ for binary variables, updating discrete segmentation $\widehat{Y}^i$ is equivalent to

$$\operatorname*{argmin}_{\widehat{Y}^i \in \{0,1\}^{N_i}} \ \sum_p \big(\tfrac{1}{2} - S^i_p - \widehat{U}^i_p\big) \, \widehat{Y}^i_p \, + \, \frac{\lambda}{\widehat{\mu}} \mathcal{L}_{\rm reg}(\widehat{Y}^i), \tag{5}$$

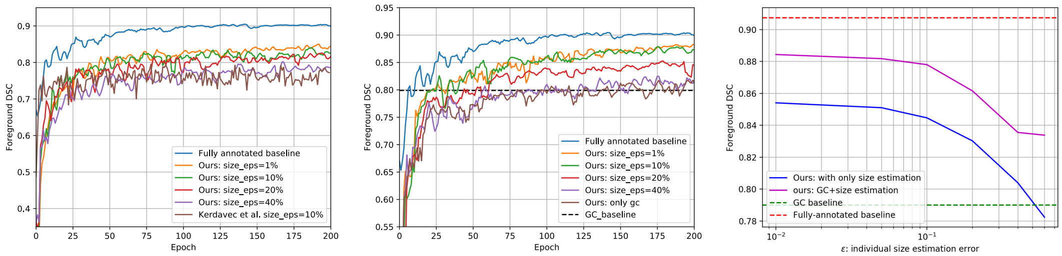

Figure 1: Left: Validation DSC for our proposed method with only image-specific size constraint. The baselines are [3] with global size constraint and fully-supervised baseline. Middle: Validation DSC for our proposed method with image-specific size constraint and regularization prior. Right: DSC versus ϵ for the above two cases.

for which the global optimum can be obtained in polynomial time using a max-flow algorithm [8]. Likewise, we update $\widetilde{Y}^i$ by solving

$$\underset{\widetilde{Y}^i \in \{0,1\}^{N_i}}{\text{argmax}} \sum_p (S^i_p - \widetilde{U}^i_p - \tfrac{1}{2})\, \widetilde{Y}^i_p, \quad \text{s.t. } A_{\min} \leq \sum_p \widetilde{Y}^i_p \leq A_{\max}, \tag{6}$$

which corresponds to an equally-weighted knapsack problem [12] with lower and upper bounds on the total weight. The optimal solution to this problem can be obtained via a simple ranking, setting $\widetilde{Y}^i_p = 1$ for pixels with the highest utility $S^i_p - \widetilde{U}^i_p - \frac{1}{2}$, while respecting the minimum and maximum size constraints. Finally, we update the Lagrange multipliers as follows: $\widehat{U}^i := \widehat{U}^i + (S^i - \widehat{Y}^i)$, $\widetilde{U}^i := \widetilde{U}^i + (S^i - \widetilde{Y}^i)$.

3 Experiments and results

Our experiments focus on left ventricular endocardium segmentation from the 2017 ACDC Challenge[1], consisting of 100 cine magnetic resonance (MR) exams. As weak annotations (i.e. Ω^i), we used the same labeled pixels as in [3], which represent only 0.1% of the original ground truth. To measure the impact of size constraints, we compute size intervals by adding or substracting a relative percentage $\epsilon \in \{1\%, 5\%, 10\%, 20\%, 40\%, 60\%\}$ of the real foreground size of each image.: $[A_{\min}, A_{\max}] = [\text{TrueSize} \times (1-\epsilon),\, \text{TrueSize} \times (1+\epsilon)]$. For all experiments, we used ENet [13] as our segmentation architecture and trained the network from scratch using Adam optimizer with a batch size of 1. The initial learning rate was set to 1×10^{-3} and decreased by 5% for every 20 epochs. Tested models were evaluated using Dice similarity coefficient (DSC).

We first verify the improvements brought by introducing the size constraints. Fig. 1 plots the validation DSC for our method using image-specific size constraints with different ϵ, or a global size constraint as in [3]. For this experiment (Fig. 1, left), no regularization prior is used. We also report the performance of the same model trained with fully-annotated images. Compared to using a global size constraint, our method can reach a DSC score of 80% with an ϵ of 40%, corresponding to an improvement of 10%. For $\epsilon = 1\%$, DSC increases to 85%, demonstrating that an accurate size estimation can help to improve performance in a weakly-supervised setting.

We then examine in Fig. 1 (middle) the impact of imposing a regularization prior on the segmentation. As comparison baseline, we consider the weak annotations as foreground seeds and apply a graphcut algorithm. Using this baseline, we achieve a best DSC score of 79.9%. One can see from Fig. 1 (middle) that the regularization prior boots the DSC score by 3.03%, 3.84%, 3.13%, 3.16%, and 5.14% for ϵ values of 1%, 10%, 20%, 40% and 60%, respectively. This improvement results from guiding the network toward actual ground truth regions, in initial training epochs, and then refining the segmentation boundary toward edges in the images. Moreover, from Fig. 1 (right) we make the following two observations: 1) a gap of 2.31% is found between our best model and the fully-supervised baseline, underlining the effectiveness of the proposed method; 2) even when the size estimation is coarse (ϵ of 40%), the DSC can still be improved by our method to surpass the graphcut baseline.

[1] `https://www.creatis.insa-lyon.fr/Challenge/acdc/`

References

[1] Pedro O Pinheiro and Ronan Collobert. Weakly supervised semantic segmentation with convolutional networks. In *CVPR*, volume 2, page 6. Citeseer, 2015.

[2] George Papandreou, Liang-Chieh Chen, Kevin Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep CNN for semantic image segmentation. *arXiv preprint arXiv:1502.02734*, 2015.

[3] Hoel Kervadec, Jose Dolz, Meng Tang, Eric Granger, Yuri Boykov, and Ismail Ben Ayed. Constrained-cnn losses forweakly supervised segmentation. *arXiv preprint arXiv:1805.04628*, 2018.

[4] Deepak Pathak, Philipp Krahenbuhl, and Trevor Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1796–1804, 2015.

[5] Jifeng Dai, Kaiming He, and Jian Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1635–1643, 2015.

[6] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE transactions on medical imaging*, 36(2):674–683, 2017.

[7] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3159–3167, 2016.

[8] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

[9] Olga Veksler. Star shape prior for graph-cut image segmentation. In *European Conference on Computer Vision*, pages 454–467. Springer, 2008.

[10] Jose Dolz, Ismail Ben Ayed, and Christian Desrosiers. Unbiased shape compactness for segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 755–763. Springer, 2017.

[11] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2017.

[12] Paul C Chu and John E Beasley. A genetic algorithm for the multidimensional knapsack problem. *Journal of heuristics*, 4(1):63–86, 1998.

[13] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.