Semi-supervised Learning for Segmentation Under Semantic Constraint

Pierre-Antoine Ganaye, Michaël Sdika(✉), and Hugues Benoit-Cattin

Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm CREATIS UMR 5220, U1206, 69100 Lyon, France

Abstract. Image segmentation based on convolutional neural networks is proving to be a powerful and efficient solution for medical applications. However, the lack of annotated data, presence of artifacts and variability in appearance can still result in inconsistencies during the inference. We choose to take advantage of the invariant nature of anatomical structures, by enforcing a semantic constraint to improve the robustness of the segmentation. The proposed solution is applied on a brain structures segmentation task, where the output of the network is constrained to satisfy a known adjacency graph of the brain regions. This criteria is introduced during the training through an original penalization loss named NonAdjLoss. With the help of a new metric, we show that the proposed approach significantly reduces abnormalities produced during the segmentation. Additionally, we demonstrate that our framework can be used in a semi-supervised way, opening a path to better generalization to unseen data.

Keywords: Medical image segmentation
Convolutional neural network · Semi-supervised learning
Adjacency graph · Constraint

1 Introduction

In medical imaging, semantic segmentation is of major importance, it helps to quantify the volume and positions of anatomical structures [1,2] and lesions [3]. In the case of brain segmentation, it enables to track the volume of structures over time, providing valuable evidences to hypothesize over a possible malfunction. Multi-atlas methods [4,5] are established solutions for this problem, they are based on the registration and fusion of image atlases, which results in consistent segmentations that preserve the topology of the structures, taking into account the inter-structures relationships. In CNN based approach [1], Moeskops *et. al.* proposed a patch based segmentation architecture which leverages contextual information around the center pixel. Encoder-decoder model [6] has also been used for the same task, by first pre-training on a different dataset annotated with Freesurfer and then fine-tuning with a novel error corrective loss.

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11072, pp. 595–602, 2018.
https://doi.org/10.1007/978-3-030-00931-1_68

These pipelines are optimized to maximize the similarity between the segmentation and the ground truth, based on cost functions like the cross entropy and the dice similarity [6]. The precision and robustness of these methods are bounded by the quality and quantity of the data. At best it should represent the variability in appearance and the segmentations should bring consensus among examiners. However the lack of annotations is a common factor in medical imaging, due to the complexity of the task. Solutions have been explored to harness properties such as anatomical invariance and semantic knowledge, allowing to improve the modeling capacities of CNNs by constraining the loss. It can take the form of an implicit knowledge (contextual information, spatial position) integrated as an input of the model or a soft penalty (volume, shape) specified by an expert.

Previous works on learning under constraints [7–9] demonstrate the interest in such method, Stewart *et al.* [8] suggested to use physics laws as a domain prior and proves its applicability to object tracking. In the medical community, [10] trained a by-patch segmentation model by integrating a spacial representation of the patch, implying that the position is correlated to the label of interest, enforcing an automatically learned spatial constraint. In [9], Oktay *et al.* went further by learning a representation of the label space with an auto-encoder, extracting shape and location priors of the structures. The final model is constrained to minimize the label representation of the segmentation and the ground truth.

In this paper, we investigate how segmentation abnormalities can be reduced by introducing knowledge about the connectivity of anatomical structures. We apply the proposed method to brain structures segmentation on MR T1w images. First, an encoder-decoder model inspired from [6] is trained on a dataset. Second, a labels adjacency prior is extracted from the training set, with the objective of matching the network's output with it. A novel loss function is applied on the trained network, in a simple fine-tuning step. Finally, we take advantage of the semi-supervised nature of this constraint by applying it on an external dataset. Doing so provides better generalization, without compromising the quality.

In Sect. 2 we introduce the segmentation architecture, together with the adjacency constraint term. In Sect. 3 we present the various experiments realized to demonstrate the interest in such method. In Sect. 4 the results of the experiments are commented.

2 Methods

2.1 Encoder-Decoder Architecture

The architecture of our 2D network (Fig. 1) is directly inspired by [6], with some minor changes. This network is composed of an encoding path, followed by a decoding path where the features are upsampled with max-unpooling. During the decoding, features from the encoder are reused via skip-connections and concatenated with the upsampled path, at each resolution level. The input of the network is composed of 7 adjacent slices, while segmenting only the central slice. The other difference is that we used convolution kernels of size 3×3, proving to be more efficient in terms of parameters and with equal performance.

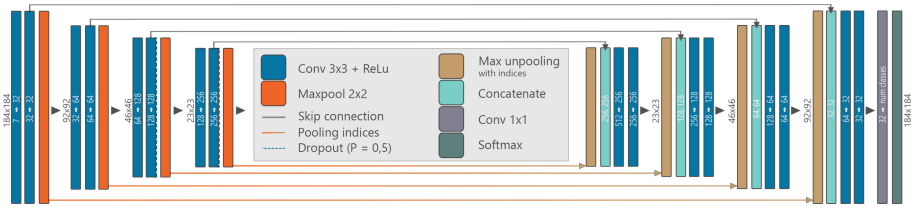

Fig. 1. Encoder-decoder architecture

2.2 Adjacency Graph of the Anatomical Structures

In this work, structural invariance of the segmentation map is assumed, the connectivity between regions should be the same from one subject to another. The adjacency graph of a segmentation map is defined as a graph where each label of the map is a vertex and where edges are weighted by the number of voxels joining two regions. Formally, the weight of the edge connecting the structures i and j for a given subject is defined by:

$$\mathbf{A}_{ij} = \sum_{x} \sum_{v \in V} \delta_{i,s(x)} \delta_{j,s(x-v)}, \tag{1}$$

where x is a voxel, $s(x)$ the mapping to the corresponding label, δ the Kronecker delta and V defines a neighborhood which does not include 0.

The matrix $\mathbf{A}$ ultimately describes how many contours are shared between pairs of structures in the 3D volume. Although, this matrix can vary from one subject to another, it is assumed in this work that its binary version $\tilde{\mathbf{A}} = (\mathbf{A} > 0)$ does not change. One can consequently define the set of forbidden transitions between structures: $F = \{(i,j) | \ \tilde{\mathbf{A}}_{ij} = 0\}$.

2.3 Loss Functions

Constraint Training. Knowing which regions interact together allows to determine which adjacencies should be considered as abnormal and finally not present in the output of an automatic segmentation system. We propose in this work to train the network ϕ such that its weights $\mathbf{w}$ minimize inconsistencies in the output segmentation by solving the constrained optimization problem:

$$\min_{G(\mathbf{w})=0} \frac{1}{|D_S|} \sum_{(\mathbf{x},y) \in D_S} L(\phi(\mathbf{x}, \mathbf{w}), y)) \tag{2}$$

where L is a segmentation loss (Dice or cross-entropy for example) and

$$G(\mathbf{w}) = \sum_{\mathbf{x} \in D_{NA}} \sum_{(i,j) \in F} a_{ij}(\phi(\mathbf{x}, \mathbf{w})). \tag{3}$$

The a_{ij} function measures the adjacency between labels i and j from the network probability output and will be discussed below. D_S and D_{NA} are respectively the training dataset for the segmentation and the Non-Adjacency loss.

Computing the Adjacency Loss. The a_{ij} function takes as input a map $p(.)$ of probability vectors (the output of the network) and should return, as Eq. 1, a measure of connectivity between the i and j labels. A simple idea to build a_{ij} would be to define an approximation f of the $\delta_{.,s(x)}$ function from the p map and plug it into Eq. 1. If f is such an approximation, a_{ij} can be computed as follows:

$$a_{ij}(f) = \sum_x \sum_{v \in V} f_i(x) f_j(x - v), \tag{4}$$

$$= \sum_x f_i(x) \sum_{v \in V} f_j(x - v), \tag{5}$$

$$= \sum_x f_i(x) \tilde{f}_j(x) \tag{6}$$

where $\tilde{f} = f * \mathbb{1}_V$ is the convolution of f and the indicatrix of the neighborhood (a kernel with all one values in V, $0 \notin V$).

Note that if $f_i(x) = \delta_{i,\text{argmax}_k p_k(x)}$, we obtain exactly $a_{ij}(f) = \mathbf{A}_{ij}$. However, as the derivative of the argmax function is 0 a.e., gradient descent algorithms such as SGD would not be possible for the training. We will investigate two families of smooth approximation:

$$f^{\text{p}}(p) = p^\beta \qquad \text{and} \qquad f^{\text{norm}}(p) = \left(\frac{p}{\max_k p_k} \right)^\beta, \tag{7}$$

where the exponent is meant component-wise.

Semi-supervised Learning. Evaluating the loss of the semantic constraint does not require the ground truth segmentation, thus the framework of this method is extended to semi-supervised learning, where the model is simultaneously optimized to minimize the classical segmentation loss on the annotated dataset D_S and constrained by the connectivity term on an external non-annotated dataset D_{NA} complemented by the ground truth dataset D_S.

2.4 Numerical Resolution

Constrained Optimization. The constrained learning is performed by fine-tuning a model trained to solve the original task (without any form of penalization). During this second step, $G(w)$ is weighted by a coefficient λ and added as a penalization to the segmentation loss. The λ coefficient is linearly increased as a function of the iteration index until it reaches a predefined $\lambda_{\max}$.

Multi-objective Model Selection. A standard way to select the best network is to choose the iteration with minimal loss on the validation dataset. In our case, selecting the best model involves balancing between the quality of the segmentation and its fidelity to anatomical properties. To solve this multi-objective problem, we opt to pick the model that maximizes the average graph loss among the five best average dice.

3 Experiments

3.1 Data

The proposed method was evaluated on brain-region segmentation from T1-weighted MR images, using the MICCAI 2012 multi-atlas challenge and IBSRv2 datasets. Each dataset was split into training/validation/test subsets as presented in Table 1. The OASIS dataset [11] was used as the source of unlabelled training data for the semi-supervised experiments, excluding the subjects who also appear in MICCAI 2012. In IBSRv2, 6 of the 39 labels were removed from the segmentation problem (such as Lesions, Blood vessel or Unknown).

Table 1. The three brain MRI datasets used for the experiments.

	Nb subjects	Nb labels	Nb train	Nb validation	Nb test
MICCAI12	35	135	10	5	20
IBSRv2	18	33	10	3	5
OASIS	406		284	122	

All the images were affine registered to a reference atlas in the MNI space. Bias field correction was applied with N4ITK. The mean and standard deviation were estimated on each of the datasets and respectively centered and reduced.

3.2 Implementation Details

The cross entropy and the average dice similarity [6] are the loss functions. The numerical optimization is performed with SGD, with a batch size of 8, the initial learning rate is set to 0.01 and updated following the poly rate policy [12], for 300 epochs. While applying the penalization, λ_{max} (see 2.5) is set to 0.01, the learning rate is lowered to $1e{-}3$, for 50 epochs.

Due to the number of classes and important volume variations between structures, we noticed that class imbalance was causing issues during the optimization of the cross entropy. Following the work of Roy *et al.* in [6], median frequency weighting was applied with success. The dice loss is left unaffected by this problem. We used elastic deformation [13] as the main data augmentation method. The code and the models' parameters will be made available publicly[1].

3.3 Evaluation

To quantify how many abnormalities based on the adjacency prior are present, we introduce two new metrics, CA^{unique} and CA^{volume}:

$$CA^{unique}(a) = 100\frac{|O \cap H|}{|H|} \quad \text{and} \quad CA^{volume}(a) = \frac{\sum_{(i,j)\in(O\cap H)} a_{ij}}{vol_{contour}}$$

[1] https://github.com/trypag/NonAdjLoss.

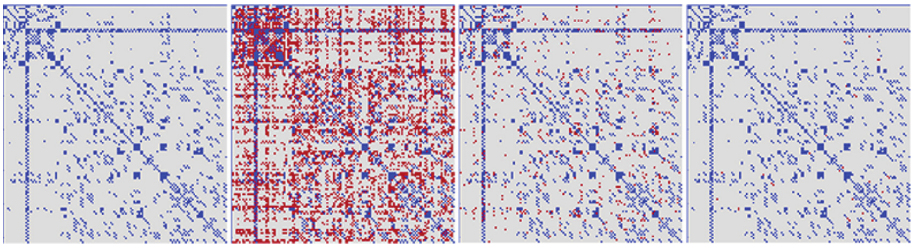

Fig. 2. Binary adjacency graphs matrices. Blue shows correct connections, red shows impossible adjacencies. From left to right: ground truth $\tilde{A}$ for the MICCAI12 dataset, after training without constraint, after training with NonAdjLoss, after semi-supervised training of NonAdjLoss with 100 images.

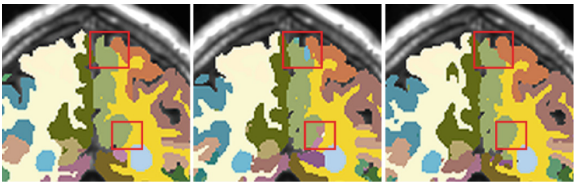

Fig. 3. Segmentation maps of one patient for: ground truth, model without loss, model with NonAdjLoss (from left to right). Red boxes highlight areas where inconsistencies were corrected.

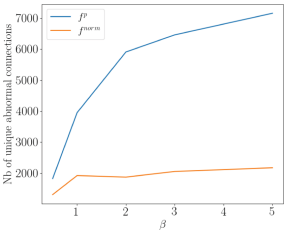

Fig. 4. Number of unique abnormal connections as a function of f^p, f^{norm} and β.

where a is the adjacency graph of a segmentation map, $O = \{(i,j) \mid a_{ij} > 0\}$, $H = \{(i,j) \mid \tilde{A}_{ij} = 0\}$ and $vol_{contour}$ the volume of contours voxels in the inferred segmentation (Fig. 3).

4 Results

To select the best approximation of f, we evaluated several values of β for f^p and f^{norm}. For these models optimized with NonAdjLoss on the train dataset of MICCAI12, we measured the total number of unique abnormal connections on part of the validation set (20 images from OASIS). Figure 4 shows that f^{norm} is significantly better at reducing the number of forbidden transitions than f^p. The optimal value of β lies at 0.5, we opt to use this configuration for the rest of the experiments. The effect of training with NonAdjLoss is demonstrated in Fig. 2, where the adjacency graphs obtained from the segmentation of the MICCAI12 test dataset are presented. After optimizing with the penalty term, the number of unique abnormalities (red dots) considerably decreases, while correct connections (blue dots) are preserved. We further improve the results by performing semi-supervised training on 100 unannotated images from OASIS,

Table 2. Distance, similarity and connectivity metrics for each models. MSD is the mean surface distance. All the metrics are averaged over the test dataset. Our loss (n) = NonAdjLoss(n), where n is the number of images used for the semi-supervised training. (average ± standard deviation).

MICCAI12	Dice	Hausdorff	MSD	CA^{unique}	CA^{volume}
Baseline	0.735 ± 0.11	21.19 ± 9.82	1.25 ± 0.43	2.57 ± 5.7e−2	7.6e−3 ± 2.5e−2
Our loss (0)	0.730 ± 0.10	13.20 ± 4.70	1.13 ± 0.35	0.21 ± 6.1e−3	2.7e−4 ± 1.0e−3
Our loss (20)	0.733 ± 0.10	11.42 ± 4.21	1.08 ± 0.33	0.02 ± 3.9e−4	7.3e−6 ± 2.4e−5
Our loss (100)	0.739 ± 0.10	11.20 ± 4.13	1.05 ± 0.34	0.01 ± 1.5e−4	1.5e−6 ± 2.4e−6
IBSRv2					
Baseline	0.825 ± 0.11	20.86 ± 21.14	0.82 ± 0.36	5.44 ± 1.8e−2	5.4e−4 ± 1.7e−4
Our loss (0)	0.833 ± 0.10	15.45 ± 20.32	0.78 ± 0.31	0 ± 0	0 ± 0
Our loss (20)	0.833 ± 0.10	14.54 ± 19.57	0.77 ± 0.31	0 ± 0	0 ± 0
Our loss (50)	0.835 ± 0.10	15.16 ± 19.26	0.77 ± 0.30	0.12 ± 1.5e−3	2.2e−6 ± 4.2e−6

effectively showing that the true objective of minimizing inconsistent predictions is achieved. It also demonstrates that semi-supervision has the ability to strengthen the generalization of the constraint, by learning from unseen cases.

In Table 2, we evaluate classical metrics to measure the quality of segmentation and quantify abnormalities. For the MICCAI12 dataset, we can see that the proposed methodology does not harm the dice similarity, keeping a steady level, while considerably lowering the Hausdorff distance (significantly better than the baseline for all the proposed models, with 95% confidence). This means that training with the NonAdjLoss enables to correct segmentation errors that are spatially far away from their ground truth, thus reducing the level of inconsistency. The CA^{unique} metric provides a percentage of unique abnormal connections in the segmentations, for both datasets we prove to gradually decrease it by applying the proposed loss, sometimes even resulting in no abnormality. The only exception is the model trained with semi-supervision on 50 images of IBSRv2, we suggest that it is due to an optimization problem. CA^{volume} indicates the overall volume of inconsistent segmentations, quantifying how many abnormalities were observed. Again we notice the same pattern as before, gradually diminishing the errors.

5 Conclusion

To our knowledge, this is the first time in the literature that a loss constraint based on a label connectivity prior is proposed. It can be applied to any image segmentation problem where invariance in the label space is ensured, without needing to modify the network's architecture. Furthermore, while no segmentation quality measure was impaired, the Hausdorff and MSD were clearly improved. The higher the number of labels, the more constrained the problem is, which leads to a potentially better efficiency of the method. Not requiring the

ground truth annotation is also a serious advantage to extend to semi-supervised training, enforcing the generalization of the new loss on larger datasets.

Acknowledgments. This work was funded by the CNRS PEPS "APOCS" and was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir"(ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. Also we would like to thank the IN2P3 computing center for sharing their resources.

References

1. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J.N.L., Išgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE TMI **35**(5), 1252–1261 (2016)
2. Roth, H.R., et al.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 556–564. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_68
3. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
4. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage **33**(1), 115–126 (2006)
5. Wang, H., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. Front. Neuroinformatics **7**, 27 (2013)
6. Roy, A.G., Conjeti, S., Sheet, D., Katouzian, A., Navab, N., Wachinger, C.: Error corrective boosting for learning fully convolutional networks with limited data. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 231–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_27
7. Xu, J., Zhang, Z., Friedman, T., Liang, Y., den Broeck, G.V.: A semantic loss function for deep learning with symbolic knowledge (2018)
8. Stewart, R., Ermon, S.: Label-free supervision of neural networks with physics and domain knowledge (2017)
9. Oktay, O., et al.: Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE TMI **37**(2), 384–395 (2018)
10. de Brebisson, A., Montana, G.: Deep Neural Networks for Anatomical Brain Segmentation. arXiv:1502.02445 [cs, stat], February 2015
11. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J. Cogn. Neurosci. **22**(12), 2677–2684 (2010)
12. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. CoRR abs/1606.00915 (2016)
13. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition, pp. 958–963, August 2003