

POPCORN: Progressive Pseudo-labeling with Consistency Regularization and Neighboring

Reda Abdellah Kamraoui, Vinh-Thong Ta, Nicolas Papadakis, Fanny Compaire, José V Manjon, Pierrick Coupé

▶ **To cite this version:**

Reda Abdellah Kamraoui, Vinh-Thong Ta, Nicolas Papadakis, Fanny Compaire, José V Manjon, et al.. POPCORN: Progressive Pseudo-labeling with Consistency Regularization and Neighboring. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'21), Sep 2021, Strasbourg (virtual), France. pp.373-382, 10.1007/978-3-030-87196-3_35 . hal-03365185

HAL Id: hal-03365185
https://hal.science/hal-03365185

Submitted on 14 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

POPCORN: PROGRESSIVE PSEUDO-LABELING WITH CONSISTENCY REGULARIZATION AND NEIGHBORING

A PREPRINT

Reda Abdellah Kamraoui[1], Vinh-Thong Ta[1], Nicolas Papadakis[2],
Fanny Compaire[1,2], José V Manjon[3], and Pierrick Coupé[1]

[1]Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI,, UMR5800, PICTURA, F-33400 Talence, France
[2]Univ. Bordeaux, Bordeaux INP, CNRS, IMB,, UMR5251, F-33400 Talence, France
[3]ITACA, Universitat Politècnica de València,, 46022 Valencia, Spain,

October 14, 2021

ABSTRACT

Semi-supervised learning (SSL) uses unlabeled data to compensate for the scarcity of annotated images and the lack of method generalization to unseen domains, two usual problems in medical segmentation tasks. In this work, we propose POPCORN, a novel method combining consistency regularization and pseudo-labeling designed for image segmentation. The proposed framework uses high-level regularization to constrain our segmentation model to use similar latent features for images with similar segmentations. POPCORN estimates a proximity graph to select data from easiest ones to more difficult ones, in order to ensure accurate pseudo-labeling and to limit confirmation bias. Applied to multiple sclerosis lesion segmentation, our method demonstrates competitive results compared to other state-of-the-art SSL strategies.

Keywords: Semi-supervised Learning, Pseudo-labeling, Consistency regularization, MS lesion segmentation

1 Introduction

Semi-Supervised Learning (SSL) is a promising field which aims to exploit unlabeled data in order to enhance the performance achieved using only labeled data. SSL is explored to mitigate both problems of the limited availability of labeled data and the lack of model generalization to unseen domains. Among SSL works proposed for medical image segmentation tasks, we can distinguish three main categories:

Consistency Regularization (CR) constrains the model to give consistent predictions for the same unlabeled input under different perturbations. Bortsova *et al.* [4] constrained the model to produce similar segmentation when applying different elastic transformations to the same unlabeled images. Similarly, Perone *et al.* [18] used a mean teacher strategy where the consistency loss constrained teacher and student predictions to be consistent. Orbes *et al.* [16] designed an adversarial loss to minimize the amount of information for a specific domain and to maximize segmentation consistency. CR offers interesting consistency properties on the learned features but it is usually trained under unrealistic scenarios (e.g., using the same input data under different perturbations). Such oversimplification does not guarantee a good generalization of the learned features. Besides, some works showed that consistency regularization using perturbation on input data is not adapted for segmentation [10, 17].

Pseudo-Labeling (PL) strategies automatically assign labels to unlabeled data in order to use them during training in combination with labeled data. Pseudo-labels are generally assigned by a model trained on labeled data. Uncertainty can be used to measure the confidence of the predictions. For example, Sedai *et al.* [19] employed prediction uncertainty for estimating segmentation confidence on soft labels. Cao *et al.* [5] considered an uncertainty aware temporal ensembling strategy. Xia *et al.* [20] used uncertainty-weighted mechanism for the pseudo-label fusion of multiple networks predictions. PL is a simple way to use unlabeled data. PL is nevertheless prone to confirmation bias (*i.e.*, error propagation) [2]. So far, this is the main limitation of PL.

Auxiliary Tasks (AT) are secondary objectives combined with the main segmentation task which do not require ground truth annotations. Using unlabeled data, in such a way, implicitly extracts relevant features for the primary segmentation task. Li *et al.* [14] proposed the prediction of surface distance maps to capture more effectively shape-aware features. Kervadec *et al.* [12] predicted the size of the target segmentation as an intermediate task. Alternatively, Chen *et al.* [7] combined supervised segmentation and unsupervised input reconstruction. Finally, Luo *et al.* [15] proposed to predict geometry-aware level set representation of the transformed ground truth annotations. AT demonstrated good performance, but the choice of the AT highly depends on the addressed problem which limits the method generalization for other segmentation tasks.

In this work, our main contribution is threefold:

- We propose a novel framework that combines consistency regularization and pseudo-labeling for segmentation.
- We propose a consistency regularization strategy that ensures proximity in latent space of images with similar segmentations. This allows us to produce meaningful feature representation and accurate predictions.
- We propose a new pseudo-labeling strategy which selects progressively unlabeled samples according to their similarity with training data, in order to limit confirmation bias.

2 Method

2.1 Method overview

The proposed strategy is a PrOgressive Pseudo-labeling with COnsistency Regularization and Neighboring (POPCORN) for semi-supervised learning in segmentation (see Fig.1). First, the training is performed with a new CR ensuring that: i) augmented versions of the same image have identical feature maps, and ii) images with similar segmentations have similar feature maps. Second, PL of the unlabeled data is performed gradually. At each selection step, the proximity graph is used to select new unlabeled samples. The pseudo-labels of the chosen data-points are estimated with the current segmentation model and incorporated in the training set.

The main intuition is that our segmentation model is able to produce more accurate pseudo-labels for images similar to our training set. Since our CR ensures close features for similar data, features extracted from the model are used to select new samples.

2.2 Bottleneck consistency regularization

In POPCORN, the model architecture is based on 3D U-Net composed of an encoder and a decoder, linked by a bottleneck and skip connections at different scales (see Fig.1). For an image X, $F(X)$ represents the prediction of the segmentation network, and $h(X)$ represents the latent features of X extracted at the bottleneck level. Our method is based on a dual/hybrid loss ensuring segmentation quality and consistency relevance.

2.2.1 Segmentation loss:

As traditionally done in supervised learning, we use the Dice similarity loss as the first element of our global loss. This loss ensures the similarity of the produced output with the expected one.

$$L_{seg}(F(X), y) = 1 - Dice(F(X), y), \tag{1}$$

where y is either the expert segmentation of X when available, or the pseudo-label otherwise.

2.2.2 Consistency regularization loss:

Alongside L_{seg}, a regularization loss on the bottleneck is used:

$$L_{reg}[\,(X_i, y_i), (X_j, y_j)\,] = Distance(\,h(X_i),\, h(X_j)\,) \times Similarity(y_i, y_j), \tag{2}$$

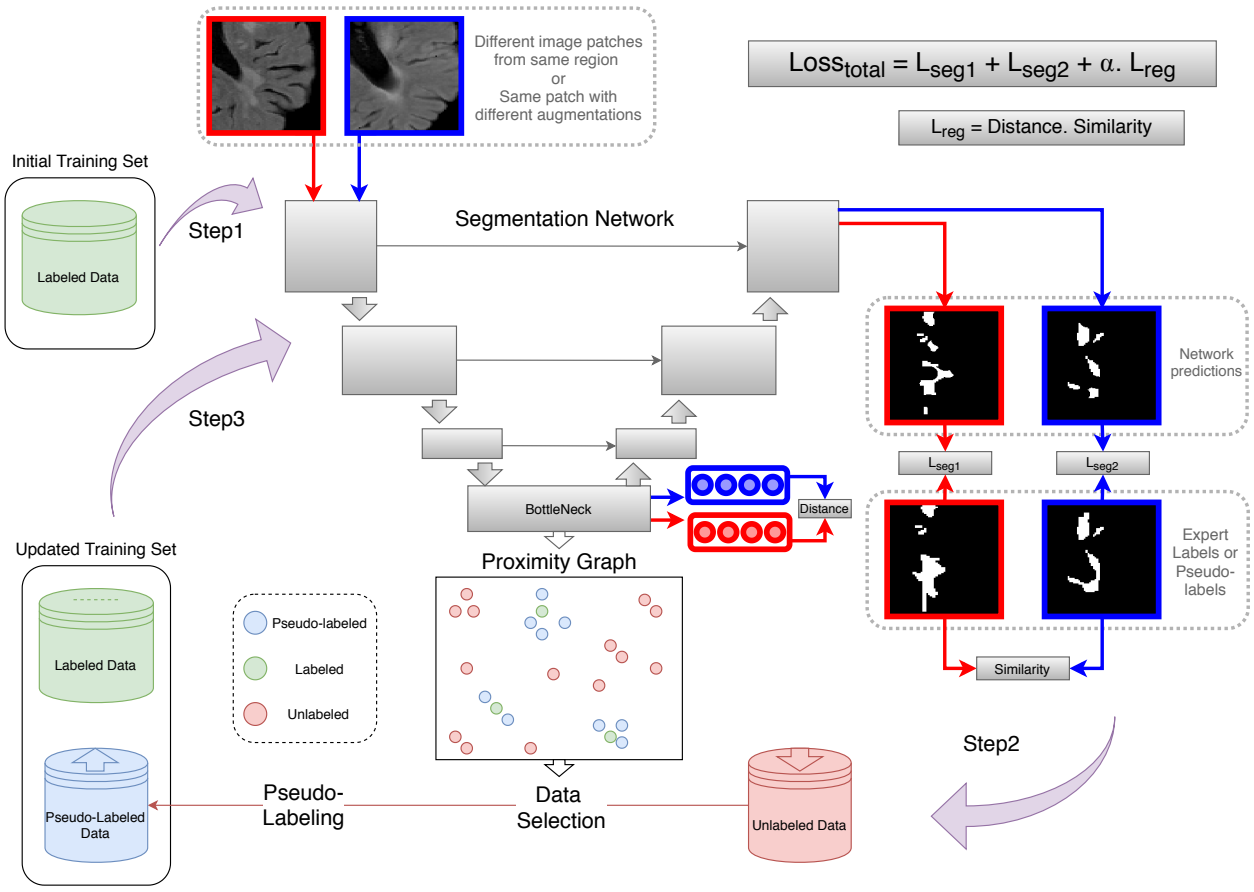


Figure 1: The training process of POPCORN

where X_i, X_j are image patches randomly selected as either different augmented versions of the same patch, or patches from different images extracted from the same region and with the same orientation. Moreover, let us define:

$$Similarity(y_i, y_j) = e^{-mse(y_i, y_j)}, \tag{3}$$

$$Distance(h_i, h_j) = \frac{2\,||h_i - h_j||^2}{||h_i||^2 + ||h_j||^2}, \tag{4}$$

where mse is the mean squared error, and $||\,.\,||$ is the euclidean distance.

The total loss is a combination of (1) and (2) with a weighting coefficient α:

$$L_{total}[(X_i, y_i), (X_j, y_j)] = L_{seg}(F(X_i), y_i) + L_{seg}(F(X_j), y_j) + \alpha . L_{reg}[(X_i, y_i), (X_j, y_j)]\,. \tag{5}$$

2.3 Pseudo-labeling data selection

Curriculum learning [3] showed that presenting data with an increasing difficulty can lead to a better learning process. We consider unlabeled data close to the training data as easy examples to be incorporated first in the training process, whereas distant samples are considered more challenging. Indeed, the latent distance between unlabeled and training data can be viewed as a measure of similarity. Thus, pseudo-labeling using the trained model is more accurate for unlabeled samples which are similar to training data.

Our data selection is performed in three steps (see Fig.1). Step1: the training set is limited to labeled data. Once the segmentation model is trained until convergence, it is used to extract latent space representation for each unlabeled datapoint. Step2: the proximity graph is used to select K unlabeled data that guarantee a smooth learning (as described in 2.4). For each selected unlabeled datapoint, a pseudo-label (segmentation) is assigned by the trained model. Step3:

the model is trained for N epochs with the new training set containing both labeled and pseudo-labeled data. Step two and three are repeated every N epochs by picking each time K new data points, their respective pseudo-labels are being computed with the newly updated segmentation model. The process is maintained until all unlabeled data are integrated into the training set.

2.4 Proximity graph

The proximity graph represents the euclidean distance between the training and the unlabeled samples latent representations:

$$P_{i \in U, j \in T} = ||h(X_i) - h(X_j)||^2 \tag{6}$$

where T and U represent respectively the training set (labeled and pseudo-labeled), and unlabeled data. For data selection, we propose the following criteria to select K elements of U close to T. For each datapoint of U, the proximity with T is defined by the sum of the p closest elements of T to the datapoint:

$$[i_1^\star, i_2^\star \ldots, i_K^\star] = Argmin^K(\sum_{j \in Argmin^p(P_i)} P_{i,j}), \tag{7}$$

where $[i_1^\star, i_2^\star \ldots, i_K^\star]$ represent the K indices of the data selected. $Argmin^k(V)$ returns the indices of the k smallest values of the vector V.

3 Experiments

3.1 Dataset

Labeled Data: For labeled data, the ISBI training dataset [6] is used. It consists of 21 longitudinal multimodal images (including FLAIR modality) from only five different subjects with Multiple Sclerosis (MS). The images have been acquired on the same MRI scanner. MS lesions were delineated by two expert raters. This dataset has limited image quality diversity (all the images were acquired with the same protocol on a single site) and inter-subject variability (only 5 subjects).

Unlabeled Data: The unlabeled dataset consists of 2901 FLAIR MRI (large inter-subject variability) with white matter hyperintensities. It does not only contain MS, which increases pathology diversity. MRI have been collected across multiple acquisition sites based on different manufacturers, 1.5T and 3T scanners, 2D and 3D sequences. This dataset covers a large diversity in terms of image quality, pathology and inter-subject variability.

Testing Data: For assessing our results, the dataset described in [9] is used. It contains 3D multimodal MRI from 43 subjects diagnosed with MS. The images have been acquired with three different scanners and different acquisition protocols. Consequently this dataset proposes a larger diversity than the labeled dataset. Lesion masks have been obtained by expert manual delineation.

All images have been pre-processed using the same pipeline [9].

3.2 Reference Methods

POPCORN is compared to state-of-the-art strategies [7], [19] and [4]. The following strategies have been implemented based on their published works and adapted to MS lesion segmentation. First, the multi-task attention-based SSL [7] is an AT strategy. It combines supervised segmentation and unsupervised reconstruction objectives. The reconstruction task uses an attention mechanism to predict input image regions of different classes. Second, the uncertainty guided pseudo-labeling [19] is a PL strategy. The teacher model, trained only on labeled data, generates soft segmentation (pseudo-labels) and uncertainty maps for all the unlabeled data at once. The uncertainty is used for estimating segmentation confidence of the generated segmentation when training the student model. Finally, the semi-supervised transformation consistency [4] is based on CR. In addition to the primary loss, a consistency loss ensures that the prediction of the same images under transformations are consistent.

3.3 Implementation details

The method hyperparameters were chosen empirically according to the size of labeled and unlabeled datasets. First, 200 from the $M = 2901$ unlabeled images were chosen after each training cycle that ran for 2 epochs ($K = 200$, $N = 2$) to limit computational burden. Second, the number of neighbors $p = 5$ was selected considering the initial training data

Table 1: The table details the impact of each contribution: the consistency regularization (CR), the proximity graph, and using labeled/pseudo-labeled (lab/pseudo) data. Best result is displayed in bold, and the second best result is underlined.

Strategy	Trained on	CR on	Dice	Precision	Sensitivity
Our method	Lab + Pseudo	Lab + Pseudo	**73,09%**	73,33%	74,29%
Ours with half selection steps ($M = 1400$)	Lab + Pseudo	Lab + Pseudo	70,59%	68,26%	**75,91%**
Ours without CR	Lab + Pseudo	None	69,13%	70,49%	70,58%
Ours without proximity graph	Lab + Pseudo	Lab + Pseudo	68,06%	65,14%	74,40%
Baseline with CR	Lab	Lab	68,08%	**77,77%**	61,94%
Baseline	Lab	None	64,41%	61,80%	69,70%

of 21 labeled images. We suggest that this value is a good compromise in order to consider relevant near neighbors while avoiding far neighbors which mislead data selection.

In addition, we used the architecture proposed by [11] with a patch size of $[64 \times 64 \times 64]$ and a threshold of 0.5 to obtain the binary segmentation. Moreover, image quality data augmentation was used to introduce realistic perturbations, where blur, edge enhancement, and other augmentations simulated image quality heterogeneity [11]. Furthermore, the coefficients for the regularization part of the loss have been set to 0.2 ($\alpha = 0.2$). Finally, the experiments have been performed with Keras 2.2.4 [8] and Tensorflow 1.12.0 [1] on Python 3.6. The model was optimized with Adam [13] using a learning rate of 0.0001 and a momentum of 0.9.

3.4 Statistical Analysis

To assert the advantage of a technique obtaining the highest average score, we conducted a Wilcoxon test over the lists of Dice scores measured at image level. The significance of the test is established for a p-value below 0.05.

4 Results

4.1 Ablation study

To evaluate our contributions, we compare POPCORN with other versions of our strategy when isolating key elements. As shown in Table 1, our full method achieves the highest Dice and the second best result in terms of precision. First, when comparing POPCORN without consistency regularization (corresponds to $\alpha = 0$ in (5)) and our full method, we notice a decrease in both precision and sensitivity. This suggests that without CR, the latent space is less meaningful for our selection process of unlabeled data. Second, to underline the impact of the proximity graph, we consider another progressive PL strategy where pseudo labels are randomly selected. Although the strategy without proximity graph is slightly more sensitive, we observe an important drop in both Dice and precision compared to our full method. This demonstrates that the proposed progressive selection based on image proximity in latent space is more robust to confirmation bias than random selection. Next, when running only half the selection steps ($M = 1400$), our method obtained the second best Dice score. This shows that POPCORN with nearly half unlabeled data can achieve better performance than the other variations and methods with full dataset (see also 4.2). Finally, when combining the proposed CR (on labeled data only) with the baseline (supervised learning), the precision is considerably improved. This shows the importance of our CR on segmentation accuracy, beyond data selection. Overall, the statistical analysis shows that our full method has a significantly higher Dice than the baseline, the version without CR, baseline with CR, and Ours without proximity graph.

4.2 Comparison with state-of-the-art approaches

Table 2 shows the results of POPCORN compared to the reference methods presented in section 3.2. First, all the SSL strategies obtain a significantly better Dice scores compared to the baseline. Second, POPCORN obtains the highest Dice followed by Uncertainty guided Pseudo-labeling [19]. Next, the multi-task attention-based SSL [7] and the semi-supervised transformation consistency [4] respectively obtain the highest precision and sensitivity rates. Finally, POPCORN obtains the best balance between precision and sensitivity, as opposed to the other strategies which are more prone to FP [4, 19] and FN [7]. Overall, POPCORN has a significantly higher Dice compared to the other methods according to our Wilcoxon test.

Fig.2 shows image segmentations produced by POPCORN and the compared strategies. A, B, and C are images from the testing dataset, specifically chosen to showcase acquisition and lesion diversity. For A, we observe that POPCORN

Table 2: The table represents results of POPCORN (our method) compared to other state-of-the-art strategies on the testing dataset (see 4.2 for complementary details).

Strategy	Dice	Precision	Sensitivity
POPCORN	**73,09%**	73,33%	74,29%
Multi-task Attention-based SSL [7]	67,23%	**75,72%**	61,99%
Uncertainty guided Pseudo-labeling [19]	68,31%	67,93%	71,95%
Semi-supervised transformation consistency [4]	66,75%	61,52%	**78,79%**
Baseline (labeled data only)	64,41%	61,80%	69,70%

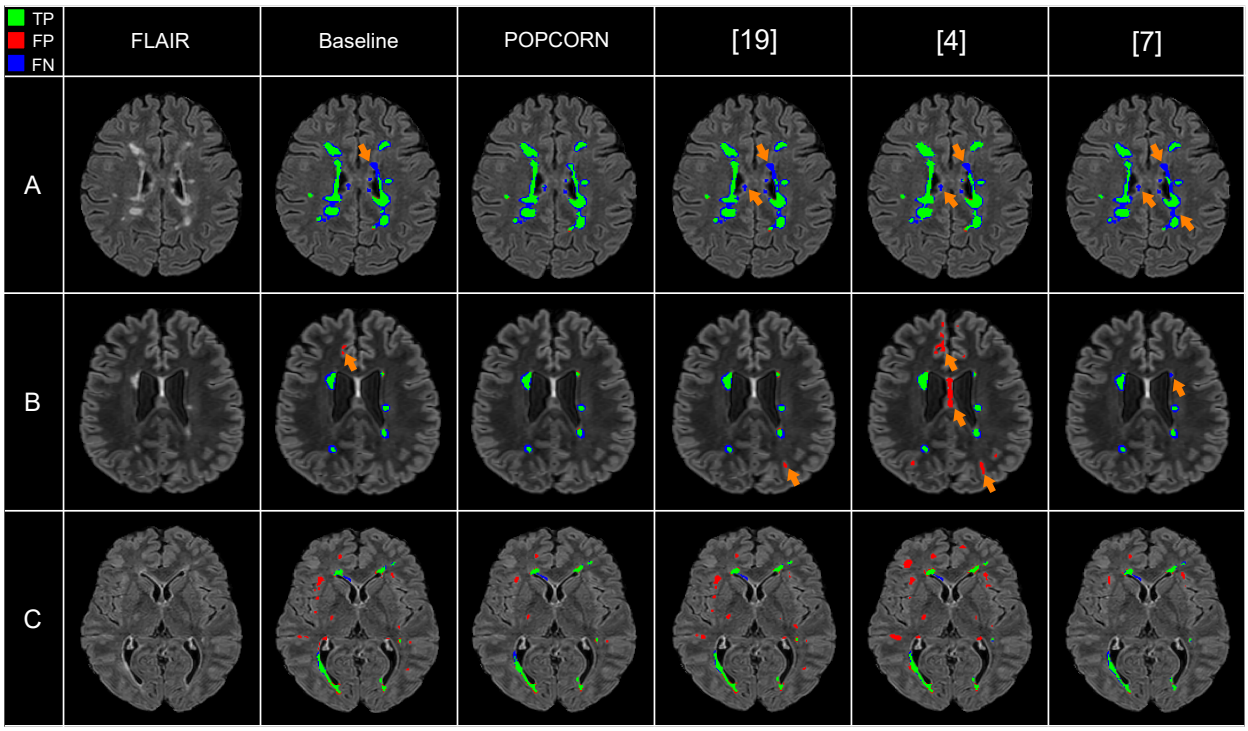


Figure 2: Comparison of POPCORN, Uncertainty guided Pseudo-labeling [19], multi-task attention-based SSL [7], and the semi-supervised transformation consistency [4] lesion segmentations. Orange arrows indicate key segmentation differences.

segmentation is the most accurate. On the contrary, [7, 19] are the least sensitive with high volumes of false negative. Similarly, the segmentations obtained with the baseline and [4] do not cover all lesions. On image B, the segmentation provided by [4] contains several false positive lesions, compared to the other strategies. Both the baseline and [19] only include one or two false detections. POPCORN proposes an accurate segmentation. Last, the method [7] misses a small lesion. For C, we notice that [4, 19] and the baseline detect many false positive lesions. POPCORN and [7] produce fewer false detection on this challenging sample. To conclude, our strategy segments accurately most lesions while minimizing false detection. Compared to the other strategies, POPCORN maintains the best balance between the sensitivity and the precision of lesion segmentation.

5 Conclusion

We propose a novel strategy for SSL segmentation. Our method combines consistency regularization and pseudo-labeling. POPCORN progressively selects unlabeled samples with an increasing difficulty using a proximity graph. Overall, we have shown the improvement of using POPCORN compared to other state-of-the-art strategies, as well as the impact of each of our contributions.

6 Acknowledgements

This work benefited from the support of the project DeepvolBrain of the French National Research Agency (ANR-18-CE45-0013). This study was achieved within the context of the Laboratory of Excellence TRAIL ANR-10-LABX-57 for the BigDataBrain project. Moreover, we thank the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02, HL-MRI Project), Cluster of excellence CPU and the CNRS/INSERM for the DeepMultiBrain project. This study has been also supported by the DPI2017-87743-R grant from the Spanish Ministerio de Economia, Industria Competitividad. The authors gratefully acknowledge the support of NVIDIA Corporation with their donation of the TITAN Xp GPU used in this research.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[3] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[4] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne. Semi-supervised medical image segmentation via learning consistency under transformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 810–818. Springer, 2019.

[5] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, and L. Cheng. Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation. *IEEE Transactions on Medical Imaging*, 40(1):431–443, 2020.

[6] A. Carass, S. Roy, A. Jog, J. L. Cuzzocreo, E. Magrath, A. Gherman, J. Button, J. Nguyen, F. Prados, C. H. Sudre, et al. Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *NeuroImage*, 148:77–102, 2017.

[7] S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne. Multi-task attention-based semi-supervised learning for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 457–465. Springer, 2019.

[8] F. Chollet et al. Keras. `https://keras.io`, 2015.

[9] P. Coupé, T. Tourdias, P. Linck, J. E. Romero, and J. V. Manjón. Lesionbrain: an online tool for white matter lesion segmentation. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 95–103. Springer, 2018.

[10] G. French, T. Aila, S. Laine, M. Mackiewicz, and G. Finlayson. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations.

[11] R. A. Kamraoui, V.-T. Ta, T. Tourdias, B. Mansencal, J. V. Manjon, and P. Coupé. Towards broader generalization of deep learning methods for multiple sclerosis lesion segmentation. *arXiv preprint arXiv:2012.07950*, 2020.

[12] H. Kervadec, J. Dolz, É. Granger, and I. B. Ayed. Curriculum semi-supervised segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 568–576. Springer, 2019.

[13] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[14] S. Li, C. Zhang, and X. He. Shape-aware semi-supervised 3D semantic segmentation for medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 552–561. Springer, 2020.

[15] X. Luo, J. Chen, T. Song, Y. Chen, G. Wang, and S. Zhang. Semi-supervised medical image segmentation through dual-task consistency. *arXiv preprint arXiv:2009.04448*, 2020.

[16] M. Orbes-Arteaga, T. Varsavsky, C. H. Sudre, Z. Eaton-Rosen, L. J. Haddow, L. Sørensen, M. Nielsen, A. Pai, S. Ourselin, M. Modat, et al. Multi-domain adaptation in brain mri through paired consistency and adversarial learning. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 54–62. Springer, 2019.

[17] Y. Ouali, C. Hudelot, and M. Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020.

[18] C. S. Perone and J. Cohen-Adad. Deep semi-supervised segmentation with weight-averaged consistency targets. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 12–19. Springer, 2018.

[19] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, and R. Garnavi. Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 282–290. Springer, 2019.

[20] Y. Xia, F. Liu, D. Yang, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth. 3D semi-supervised learning with uncertainty-aware multi-view co-training. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3646–3655, 2020.