The final authenticated version of this work is published in
MICCAI Part I, LNCS 12261, pp. 116–126, 2020
Available online at https://doi.org/10.1007/978-3-030-59710-8_12

Feature Preserving Smoothing Provides Simple and Effective Data Augmentation for Medical Image Segmentation

Rasha Sheikh[0000−0002−5822−4061] and Thomas Schultz[0000−0002−1200−7248]

University of Bonn
{rasha,schultz}@cs.uni-bonn.de

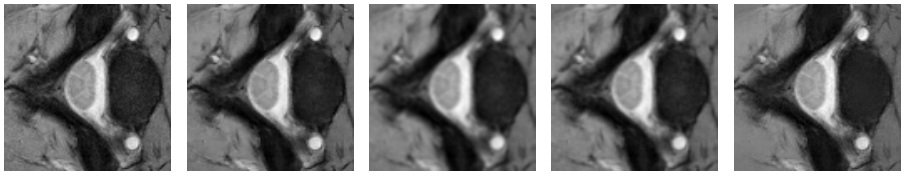
Abstract. CNNs represent the current state of the art for image classification, as well as for image segmentation. Recent work suggests that CNNs for image classification suffer from a bias towards texture, and that reducing it can increase the network's accuracy. We hypothesize that CNNs for medical image segmentation might suffer from a similar bias. We propose to reduce it by augmenting the training data with feature preserving smoothing, which reduces noise and high-frequency textural features, while preserving semantically meaningful boundaries. Experiments on multiple medical image segmentation tasks confirm that, especially when limited training data is available or a domain shift is involved, feature preserving smoothing can indeed serve as a simple and effective augmentation technique.

1 Introduction

Image segmentation is a key problem in medical image analysis. Convolutional neural networks (CNNs) often achieve high accuracy, but only if trained on a sufficient amount of annotated data. In medical applications, the number of images that are available for a given task is often limited, and the time of experts who can provide reliable labels is often scarce and expensive. Therefore, data augmentation is widely used to increase the ability to generalize from limited training data, and it is often indispensable for achieving state-of-the-art results.

Augmentation generates artificial virtual training samples by applying certain transformations to the original training images. Many of these transformations reflect variations that are expected in test images. Examples are geometric transformations such as image flipping, rotations, translations, elastic deformations [19], or cropping, but also intensity or color space transformations, which can be used to simulate changes in illumination or acquisition device characteristics [22, 5]. More complex augmentation has been performed via data-driven generative models that either generate images for augmentation directly [6, 21], or generate spatial and appearance transformations [7, 25]. However, implementing and training these approaches requires a relatively high effort.

In this work, we demonstrate that feature preserving smoothing provides a novel, simple, and effective data augmentation approach for CNN-based semantic segmentation. In particular, we demonstrate the benefit of augmenting the



(a) (b) (c) (d) (e)

Fig. 1: Sample image (a) and the results of smoothing with Total Variation (b), Gaussian (c), bilateral (d), and guided filters (e).

original training images with copies that have been processed with total variation (TV) based denoising [20]. As shown in Figure 1 (b), TV regularization creates piecewise constant images, in which high frequency noise and textural features are removed, but sharp outlines of larger regions are preserved.

Unlike the above-mentioned augmentation techniques, ours does not attempt to generate realistic training samples. Rather, it is inspired by the use of neural style transfer for augmentation. Neural style transfer combines two images with the goal of preserving the semantic contents of one, but the style of the other [8]. It is often used for artistic purposes, but has also been found to be effective as an augmentation technique when training CNNs for image classification [12]. To explain this, Geirhos et al. [9] perform experiments for which they generated images with conflicting shape and texture cues. When interpreting such images, ImageNet-trained CNNs were found to favor texture, while humans favor shape. Since shape is more robust than texture against many image distortions, this might be one factor that allows human vision to generalize better than CNNs.

Our work is based on the hypothesis that CNNs for image segmentation suffer from a similar bias towards texture, and will generalize better when increasing the relative impact of shape. Augmenting with TV regularized images should contribute to this, since it preserves shapes, but smoothes out high frequency textural features. In a similar spirit, Zhang et al. [24] use superpixelization for augmentation, and Ma et al. [15] use lossy image compression. Both argue that the respective transformations discard information that is less relevant to human perception, and thus might prevent the CNN from relying on features that are irrelevant for human interpretation. However, Ma et al. only evaluate JPEG augmentation in one specialized application, the segmentation of sheep in natural images. As part of our experiments, we verify that their idea, which has a similar motivation as ours, carries over to medical image segmentation.

2 Materials and Methods

2.1 Selection of Datasets

We selected datasets that allow us to test whether feature preserving smoothing would be an effective data augmentation technique when dealing with limited

training data in medical image segmentation. Moreover, we hypothesized that this augmentation might reduce the drop in segmentation accuracy that is frequently associated with domain shifts, such as changes in scanners. If differences in noise levels or textural appearance contribute to those problems, increasing a network's robustness towards them by augmenting with smoothed data should lead to better generalization.

As the primary dataset for our experiments, we selected the Spinal Cord Grey Matter Segmentation Challenge [18], because it includes images that differ with respect to scanners and measurement protocols, and provides detailed information about those differences. To verify that our results carry over to other medical image segmentation tasks, we additionally present experiments on the well-known Brain Tumor Segmentation Challenge [16, 3, 4], as well as the White Matter Hyperintensity Segmentation Challenge [13].

2.2 CNN Architecture

We selected the U-Net architecture [19] for our experiments, since it is widely used for medical image segmentation. In particular, our model is based on a variant by Perone et al. [17], who train it with the Adam optimizer, dropout for regularization, and the dice loss

$$\text{dice} = \frac{2\sum pg}{\sum p^2 + \sum g^2}, \tag{1}$$

where p is the predicted probability map and g is the ground-truth mask. Our model is trained to learn 2D segmentation masks from 2D slices. The initial learning rate is 0.001, the dropout rate is 0.5, betas for the Adam optimizer are 0.9 and 0.999, and we train the model for 50, 30, and 100 epochs for the SCGM, BraTS, and WMH datasets respectively. The number of epochs is chosen using cross-validation, the others are the default settings of the framework we use.

2.3 Feature-Preserving Smoothing

We mainly focus on Total Variation based denoising [20], which we consider to be a natural match for augmentation in image segmentation problems due to its piecewise constant, segmentation-like output. The TV regularized version of an n-dimensional image $f : D \subset \mathbb{R}^n \to \mathbb{R}$, with smoothing parameter α, can be defined as the function $u : D \to \mathbb{R}$ that minimizes

$$E(u; \alpha, f) := \int_D \left(\frac{1}{2} (u - f)^2 + \alpha \|\nabla u\| \right) dV, \tag{2}$$

where the integration is performed over the n-dimensional image domain D. Numerically, we find u by introducing an artificial time parameter $t \in [0, \infty)$, setting $u(\mathbf{x}, t = 0) = f$, and evolving it under the Total Variation flow [1]

$$\frac{\partial u}{\partial t} = \text{div} \left(\frac{\nabla_{\mathbf{x}} u}{\|\nabla_{\mathbf{x}} u\|} \right), \tag{3}$$

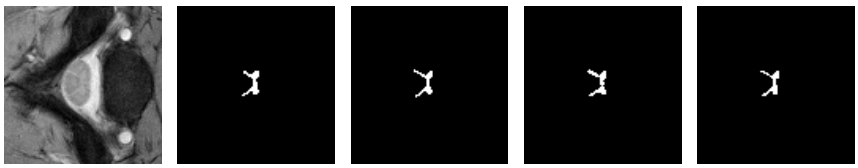


Fig. 2: Spinal cord image and segmentation masks from Site 2.

using an additive operator splitting scheme [23]. The resulting image $u(\mathbf{x}, t)$ at time t approximates a TV regularized version of f with parameter $\alpha = t$. We apply TV smoothing to all images in the training set, and train on the union of both sets, randomly shuffling all images before each epoch. We did not observe a clear difference between this basic strategy and a stratified sampling which ensured that half of the images in each batch were TV smoothed.

For comparison, we also consider two alternative feature preserving filters, the bilateral filter [2] and the guided filter [10], as well as standard, non feature preserving Gaussian smoothing. We hypothesized that feature preserving filters other than TV regularization might also be effective for augmentation, even though maybe not as much, since they do not create a segmentation-like output to the same extent as TV. We expected that Gaussian smoothing would not be helpful, because it does not preserve sharp edges, and therefore does not provide clear shape cues to the network. Moreover, it can be expressed using a simple convolution which, if useful, could easily be learned by the CNN itself.

3 Experiments

3.1 Spinal Cord Grey Matter

This challenge dataset consists of spinal cord MR images of healthy subjects acquired from four different sites with different scanners and acquisition parameters. The task of the challenge is to segment the grey matter in those images. The publicly available training data includes MR images of 10 subjects per site for a total of 40 subjects. Each MR image was annotated by four raters. A sample image from site 2 and its four masks are shown in Figure 2.

Setup The data was split into 80% training and 20% test sets. To evaluate the effect of TV augmentation under domain shift, we train on data from only one site (Montreal) and report results on the test sets from all sites. Since the four sites have different slice resolutions (0.5 mm, 0.5 mm, 0.25 mm, 0.3 mm), we resample images to the highest resolution. We also standardize intensities.

We consider it unusual that each image in this dataset has annotations from four raters. In many other cases, only a single annotation would be available per training image, due to the high cost of creating annotations. We expected that training with multiple annotations per image would provide an additional regularization. To investigate how it interacts with data augmentation, we conducted

Table 1: Dice scores for the model trained on Site 2 using annotations from the first rater, and evaluated by comparing the predictions to annotations made by the same rater (top rows) and annotations made by all raters (bottom rows).

	Original	TV	Flip	Rotate	Elastic	Gauss	Bilateral	Guided	JPEG
Site 1	0.5930	**0.7332**	0.5148	0.4842	0.6484	0.4670	0.5530	0.6050	0.5674
Site 2	0.8337	**0.8610**	0.8254	0.8289	0.8258	0.8341	0.8356	0.8262	0.8567
Site 3	0.5950	0.6466	0.5952	0.6260	0.6260	0.6397	0.6323	0.6207	**0.6605**
Site 4	0.7978	**0.8395**	0.7896	0.8011	0.7960	0.8021	0.8006	0.7909	0.8233
Site 1	0.5695	**0.7044**	0.4934	0.4599	0.6344	0.4533	0.5392	0.5714	0.5366
Site 2	0.8185	**0.8483**	0.8129	0.8158	0.8218	0.8209	0.8245	0.8104	0.8435
Site 3	0.6707	0.7475	0.6664	0.7012	0.7109	0.7182	0.7090	0.6960	**0.7582**
Site 4	0.7776	**0.8184**	0.7751	0.7798	0.7843	0.7813	0.7840	0.7708	0.8046

Table 2: Dice scores for the model trained on Site 2 using annotations from all raters, and evaluated by comparing the predictions to annotations made by the first rater (top rows) and annotations from all raters (bottom rows).

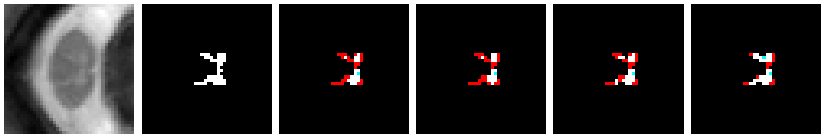
	Original	TV	Flip	Rotate	Elastic	Gauss	Bilateral	Guided	JPEG
Site 1	0.7936	**0.8254**	0.8227	0.7875	0.7823	0.8238	0.8084	0.8057	0.8236
Site 2	0.8710	**0.8819**	0.8771	0.8729	0.8453	0.8758	0.8786	0.8756	0.8674
Site 3	0.6507	0.6511	0.6582	0.6632	0.6501	0.6620	**0.6680**	0.6546	0.6585
Site 4	0.8480	0.8572	0.8535	0.8486	0.8487	0.8583	0.8544	0.8562	**0.8610**
Site 1	0.7827	**0.8303**	0.8083	0.7617	0.7754	0.8096	0.8000	0.7912	0.8048
Site 2	0.8786	**0.8956**	0.8869	0.8750	0.8554	0.8866	0.8846	0.8796	0.8790
Site 3	0.7672	0.7651	0.7734	0.7768	0.7617	0.7763	**0.7790**	0.7689	0.7761
Site 4	0.8430	**0.8562**	0.8510	0.8424	0.8445	0.8549	0.8497	0.8494	0.8561

two sets of experiments. In the first, we train using annotations from the first rater only. In the second, we use annotations from all raters. To facilitate a direct comparison between both, we evaluate each model twice, first by comparing its predictions to annotations from rater one, second by using those from all raters.

Other Augmentation Techniques We compare TV augmentation to random flipping, rotation with a random angle between [-10,10] degrees, and elastic deformations. We also compare against augmentation with Gaussian, bilateral, and guided filters, as well as JPEG compression [15]. We chose filter parameters visually, so that they result in a smoothed image while not distorting the gray matter shape. Examples are shown in Figure 1.

Results Table 1 shows the average dice score of each site's test subjects when training using annotations from the first rater. The top rows compare the predictions to annotations made by the same rater, the bottom rows show the average dice across all four raters.

In Table 2, we train using annotations from all four raters. At training time, each input image is replicated four times with a different mask for each input.



(a) (b) (c) (d) (e) (f)

Fig. 3: Sample image from Site 1 (a), its ground truth (b), the prediction of a model trained with no augmentation (c), or with bilateral (d), jpeg (e), and TV augmentation (f), respectively. White, red, and blue colors indicate TP, FN, and FP, respectively. Quantitative performance is shown in Table 1.

We again evaluate the results by comparing the predictions to the annotations provided by the first rater (top) and annotations from all raters (bottom).

Discussion TV augmentation improved results almost always, by a substantial margin in some of the cases that involved a domain shift. We show qualitative results of a challenging image from Site 1 in Figure 3. Most of the time, TV performed better than any other augmentation technique. In the few cases where it did not, the conceptually similar bilateral or JPEG augmentation worked best. As expected, augmentation with Gaussian smoothing did not lead to competitive results. For Site 3, we observed that TV sometimes smoothes out very fine details in the spinal cord structure. This might explain why JPEG augmentation produced slightly higher dice scores than TV on that site.

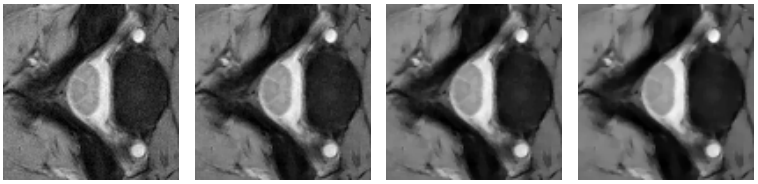
Traditional augmentation techniques such as flipping, rotation, and elastic deformation did not show a clear benefit in this specific task. Even though it is possible to combine them with TV augmentation, our results suggest that it is unlikely to benefit this particular task. It is thus left for future work.

Comparing Table 2 to Table 1, we can see the largest benefits when annotations by only one rater are available at training time. This agrees with our intuition that repeated annotations provide a form of regularization. Despite this, we continue seeing a benefit from data augmentation.

Impact of smoothing parameter and multi scale augmentation The effect of the TV smoothing parameter α that was discussed in Section 2.3 is illustrated in Figure 4. Previous experiments used $\alpha = 5$. Table 3 studies how sensitive TV augmentation is with respect to this parameter. Again, training was on Site 2 using annotations from the first rater, and evaluation compared the predictions to annotations made by the same rater (left) and annotations made by all raters (right). All scales perform quite well, and combining different ones, as shown in the last column, yields the best dice score in nearly all cases.

3.2 BraTS 2019

This challenge data consists of brain MRI scans with high- and low-grade gliomas, collected at different institutions. The publicly available training data includes 3D scans of 335 subjects.



(a) $\alpha = 0$ (b) $\alpha = 5$ (c) $\alpha = 10$ (d) $\alpha = 15$

Fig. 4: Sample image and the results of different TV smoothing parameters.

Table 3: Results with varying smoothing parameters suggest that TV augmentation is robust to its choice, and combining multiple scales is a feasible strategy. Dice scores on the left are from the same rater, on the right from all raters.

	$\alpha = 5$	$\alpha = 10$	$\alpha = 15$	Combined		$\alpha = 5$	$\alpha = 10$	$\alpha = 15$	Combined
Site 1	0.7332	0.7423	0.7573	**0.8187**	Site 1	0.7044	0.7130	0.7305	**0.7848**
Site 2	0.8610	0.8653	0.8599	**0.8757**	Site 2	0.8483	0.8530	0.8536	**0.8749**
Site 3	0.6466	**0.6581**	0.6579	0.6487	Site 3	0.7475	0.7576	0.7574	**0.7633**
Site 4	0.8395	0.8457	0.8428	**0.8525**	Site 4	0.8184	0.8217	0.8215	**0.8378**

Setup We again aimed for an evaluation that would separately assess the benefit with or without a domain shift. Although there was no explicit mapping of individual subjects to their respective institutions, we could identify four groups based on the challenge data description and the filenames: Brats2013 (30 subjects), CBICA (129 subjects), TCGA (167 subjects), and TMC (9 subjects).

We split each group's data into 80% training and 20% test sets. We train on data from only one group (TMC) and report results on the test sets from all groups. An example image with the ground truth and the smoothed augmentation is shown in Figure 5.

As input we use the FLAIR modality and train the model to learn the Whole Tumor label. We follow the preprocessing steps of [11] and standardize intensities to zero mean and unit variance.

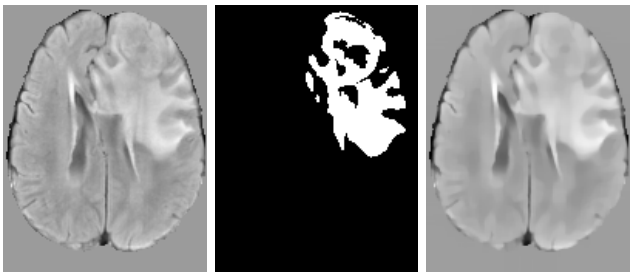


Fig. 5: BraTS (TMC) training image, ground truth, and the TV smoothed image.

Table 4: Dice Score on held-out test sets of different subsets of BraTS 19.

	Br13	CBICA	TCGA	TMC
Original	0.8088	0.7933	**0.7929**	0.7894
TV Smoothed	**0.8570**	**0.8152**	0.7885	**0.8233**

Fig. 6: WMH site 1 training image, ground-truth, and the TV smoothed image.

Results As shown in Table 4, TV augmentation increased segmentation accuracy in three out of the four groups. The largest benefit was observed in the group Br13. This involved a domain shift, as the model was only trained on TMC. On the group where the performance slightly decreases (TCGA), we found that the predicted segmentation sometimes misses some of the finer tumor details. Such fine structures might be more difficult to discern after TV smoothing.

3.3 White Matter Hyperintensity

The publicly available data from the WMH challenge is acquired with different scanners from 3 institutions. 2D FLAIR and T1 MR images and segmentation masks are provided for 60 subjects, 20 from each institution.

Setup The data is split into 80% training and 20% test sets. We combined the training data from all three sites, and report results on the test sets we held out ourselves, as well as on the official test sets, by submitting to the challenge website. To avoid creating a large number of submissions, we do not investigate the impact of training on one compared to all sites in this case. We follow the preprocessing steps of [14] and standardize intensities. An example image with the ground truth and TV smoothed augmentation is shown in Figure 6.

Results Table 5 shows that TV augmentation improved segmentation accuracy on the held-out data in all three sites. Table 6 shows a clear improvement also on the official test set. The evaluation criteria in this challenge are Dice Score (DSC), Hausdorff distance (H95), Average Volume Difference (AVD), Recall for individual lesions, and F1 score for individual lesions.

Table 5: Dice Score on held-out test sets of WMH challenge

	Site 1	Site 2	Site 3
Original	0.6879	0.8348	0.7170
TV Smoothed	**0.7329**	**0.8480**	**0.7685**

Table 6: Results on the official test set of the challenge. Higher values for DSC, Recall, F1 are better, and lower values for H95, AVD are better.

	DSC	H95	AVD	Recall	F1
Original	0.74	9.05	29.73	0.65	0.64
TV Smoothed	**0.77**	**7.42**	**24.97**	**0.76**	**0.67**

Discussion Results on both the held-out test set and the official test set show that TV smoothing improves the segmentation performance, in some cases by a substantial margin. TV-smoothing performs better with respect to all evaluation criteria in the detailed official results. The most pronounced improvement is in the number of lesions that the model detects (recall).

4 Conclusion

Our results indicate a clear benefit from using feature preserving smoothing for data augmentation when training CNN-based medical image segmentation on limited data. Advantages were especially pronounced when using TV smoothing, which creates a piecewise constant, segmentation-like output. TV augmentation also helped when a domain shift between training and test data was involved. Consequently, we propose that TV smoothing can be used as a relatively simple and inexpensive data augmentation method for medical image segmentation.

In the future, we hope to better characterize the exact conditions under which the different augmentation techniques that have been proposed for semantic segmentation work well, and when it makes sense to combine them. We expect that factors such as the nature of differences between training and test images will play a role, as well as characteristics of the images (e.g., noise), and the structures that should be segmented (e.g., presence of fine details).

References

1. Andreu, F., et al.: Minimizing total variation flow. Differential and integral equations **14**(3), 321–360 (2001)
2. Aurich, V., Weule, J.: Non-linear gaussian filters performing edge preserving diffusion. In: Sagerer, G., Posch, S., Kummert, F. (eds.) Mustererkennung. pp. 538–545. Informatik Aktuell, Springer (1995)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data **4**, 170117 (2017)
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. Tech. Rep. 1811.02629, arXiv (2018)

5. Billot, B., Greve, D., Van Leemput, K., Fischl, B., Iglesias, J.E., Dalca, A.V.: A learning strategy for contrast-agnostic MRI segmentation. Tech. Rep. 2003.01995, arXiv (2020)
6. Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R.N., Hammers, A., Dickie, D.A., del C. Valdés Hernández, M., Wardlaw, J.M., Rueckert, D.: GAN augmentation: Augmenting training data using generative adversarial networks. Tech. Rep. 1810.10863, arXiv (2018)
7. Chaitanya, K., Karani, N., Baumgartner, C.F., Becker, A., Donati, O., Konukoglu, E.: Semi-supervised and task-driven data augmentation. In: Int'l Conf. on Information Processing in Medical Imaging (IPMI). pp. 29–41. Springer (2019)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2414–2423 (2016)
9. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In: Int'l Conf. on Learning Representations (ICLR) (2019)
10. He, K., Sun, J., Tang, X.: Guided image filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Proc. European Conf. on Computer Vision (ECCV), Part I. Lecture Notes in Computer Science, vol. 6311, pp. 1–14. Springer (2010)
11. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. In: Int'l MICCAI Brainlesion Workshop. pp. 287–297. Springer (2017)
12. Jackson, P.T.G., Abarghouei, A.A., Bonner, S., Breckon, T.P., Obara, B.: Style augmentation: Data augmentation via style randomization. In: CVPR Deep Vision Workshop. pp. 83–92 (2019)
13. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. IEEE Trans. on Medical Imaging **38**(11), 2556–2568 (2019)
14. Li, H., Jiang, G., Zhang, J., Wang, R., Wang, Z., Zheng, W.S., Menze, B.: Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. NeuroImage **183**, 650–665 (2018)
15. Ma, R., Tao, P., Tang, H.: Optimizing data augmentation for semantic segmentation on small-scale dataset. In: Proc. Int'l Conf. on Control and Computer Vision (ICCCV). pp. 77–81 (2019)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. on Medical Imaging **34**(10), 1993–2024 (2014)
17. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. NeuroImage **194**, 1–11 (2019)
18. Prados, F., Ashburner, J., Blaiotta, C., Brosch, T., Carballido-Gamio, J., Cardoso, M.J., Conrad, B.N., Datta, E., Dávid, G., De Leener, B., et al.: Spinal cord grey matter segmentation challenge. NeuroImage **152**, 312–329 (2017)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015)

20. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1), 259–268 (1992)
21. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports **9**(1), 16884 (2019)
22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 60 (2019)
23. Weickert, J., Romeny, B.t.H., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. on Image Processing **7**(3), 398–410 (1998)
24. Zhang, Y., Yang, L., Zheng, H., Liang, P., Mangold, C., Loreto, R.G., Hughes, D.P., Chen, D.Z.: SPDA: superpixel-based data augmentation for biomedical image segmentation. In: Int'l Conf. on Medical Imaging with Deep Learning (MIDL). Proceedings of Machine Learning Research, vol. 102, pp. 572–587 (2019)
25. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 8543–8553 (2019)