Combining Good Old Random Forest and DeepLabv3+ for ISLES 2018 CT-Based Stroke Segmentation

Lasse Böhme[1], Frederic Madesta[1,2], Thilo Sentker[1,2], and René Werner[1,2(✉)]

[1] Department of Computational Neuroscience,
University Medical Center Hamburg-Eppendorf, Hamburg, Germany
`lasse.boehme@stud.uke.uni-hamburg.de`, `r.werner@uke.de`
[2] DAISYlabs, Forschungszentrum Medizintechnik Hamburg (fmthh),
Hamburg, Germany

Abstract. Recent years' segmentation challenges on Ischemic Stroke Lesion Segmentation (ISLES) attracted great interest in the medical image computing domain, reflected in >80 citations of the 2017 summary article of the initial ISLES 2015 challenge [1]. While 2015–2017 ISLES challenges focussed on MRI images, the 2018 challenge takes into account clinical relevance of (perfusion) CT to triage stroke patients. Thus, from a methodological point of view, it is now to be analyzed whether and to what extent the 2015–2017 methods can be adapted to automated core lesion segmentation using acute stroke CT perfusion imaging.

We strive to deliver a baseline for ISLES 2018 by using two well established machine learning-based segmentation approaches already applied for the initial ISLES 2015 challenge: random forest (RF) with classical hand-crafted image features (i.e. the most frequently used type of algorithm in ISLES 2015) and encoder-decoder-style convolutional neuronal networks (CNNs). In detail, for CNN-based segmentation, we employ the DeepLabv3+ architecture. The performance of the individual as well as a combination of the segmentation approaches is evaluated based on the ISLES 2018 training data set, and respective results are presented. Aiming at an ISLES 2018-specific performance baseline, we do neither make use of additional data other than the provided challenge data nor perform extensive data augmentation. The results highlight the potential to improve stroke lesion segmentation accuracy by combining RF and CNN information.

Keywords: ISLES challenge · Stroke segmentation · Random forest · Convolutional neural networks (CNN)

1 Introduction

Ischemic stroke is one of the most prevalent causes for death and disability worldwide [2]. It is caused by an occlusion of the cerebral arteries, which results

in hypoxia and finally in death of the affected brain tissue. Brain imaging plays a crucial part for diagnosis of ischemic stroke and respective therapeutic decisions. Detection and evaluation of stroke lesions require, however, a significant amount of time from the radiologist. To overcome or at least alleviate this problem, reliable methods to automatically segment the lesions are strongly desired.

The Ischemic Stroke Lesion Segmentation (ISLES) [3] strives to provide open benchmark data to objectively evaluate and compare respective algorithms. The first ISLES challenges and evaluation data sets focussed on magnetic resonance imaging (MRI) data [1,4]. Compared to computed tomography (CT), MRI comes with the advantages of better visibility of stroke lesions and the absence of radiation; yet, CT has higher availability in most hospitals [5]. Taking related clinical relevance of (perfusion) CT to triage stroke patients into account, the 2018 ISLES challenge therefore now focuses on methods to segment acute ischemic stroke lesions in native CT images and contrast medium-enhanced perfusion maps. Thus, it is now to be answered whether and to what extent the 2015–2017 methods can be adapted to core lesion segmentation using such data. Therefore, a training dataset including radiologists' ground-truth segmentations and a testing dataset without known ground truth were provided by the challenge organizers. The participants were asked to upload their stroke segmentation results for the testing data set.

In this article, we describe our participation in the ISLES 2018 challenge. The final choice of the applied method(s) was based on the original MRI-based 2015 ISLES challenge [1]. At that time, the majority of participants applied (either as single solution or as part of a pipeline) random forest for segmentation [6–15]. In addition, first convolutional neural network (CNN)-based approaches were also applied [16,17], and it seems to be the case that (at least at the moment) encoder-decoder-style CNNs [18] manifest as quasi-standard for medical image segmentation. We therefore decided to implement both approaches and to report on their performance on the ISLES 2018 training data set – either applied individually or in combination. Being interested in the pure algorithm performance, we did neither make use of additional data other than the provided challenge training data nor performed extensive data augmentation. Algorithmic details are described in depth in succeeding sections.

2 Materials and Methods

2.1 Image Data Description

The ISLES 2018 training data contained 94 individual data sets (subsequently called 'cases') from 63 patients, i.e. image data from patients with separated lesions were split into different cases. The individual cases consisted of only slices around the stroke lesions, with the number of slices per case ranging from 2 to 16. For each case, native CT image data as well as contrast medium enhanced 4D CT image data were provided. From the latter, cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and time-to-maximum flow (T_{max}) perfusion maps were calculated and also provided.

Ground truth data were binary masks manually drawn by radiologists on the basis of co-registered diffusion weighted imaging (DWI)-MRI data.

The DWI-MRI data were not available for the ISLES 2018 challenge testing data; we therefore did not consider them for training purposes and the experiments described in the present paper. As dealing with the 4D perfusion image series is also challenging from a computational perspective (e.g. GPU memory handling), we did not make use of the respective explicit temporal information; we only computed a corresponding maximum intensity projection (MIP) along the temporal axis as additional image data.

2.2 Image Preprocessing

Image preprocessing consisted of two parts: intensity normalization and fast brain segmentation. Image intensity normalization was applied independently for each case and image sequence by rescaling the original image intensity dynamics to the range [0; 1]. Fast brain segmentation was based on a 3D-connected component (CC) analysis of the CBV images. The largest non-zero CC was considered an initial brain segmentation, which was further refined for the individual axial slices by filtering out small 2D-CCs based on their size and position (CC centroid should be rather posterior than anterior; the latter corresponded, e.g., to close-to-eyes CCs in inferior slices).

2.3 Random Forest (RF) Implementation

For random forest (RF) implementation, we used the Scikit-learn Python library [19], Gini impurity as impurity measure, and 150 trees. For RF feature extraction, we applied the Python medical image processing library MedPy [20]. Features were computed for each case and voxel inside the case-specific brain mask (see Sect. 2.2); class imbalance due to small lesions (compared to healthy brain tissue) was not accounted for.

As voxel features, we used the image intensity as well as Gaussian-weighted local mean intensity values (standard deviations $\sigma = 3, 7, 13$). These intensity features were computed for each of the rescaled image sequences (i.e. native CT, CBF, CBV, MTT, and T_{max}), resulting in 20 features. Despite the fact that the MedPy implementation to derive hemispheric difference values may lead to partially erroneous values due to the images not being normalized/registered to symmetric atlas space, we also computed voxel-wise hemispheric difference values for the different image sequences, yielding additional five features. Thirdly, we extracted a 10-bin local histogram (5×5 neighborhood) from the T_{max} and used the ten frequency values as additional features. Fourth and finally, we computed the distance of the voxel to the brain mask boundary and the distance of the voxel to the image center. Thus, in total, we extracted 37 features for each voxel.

For classification of test cases, i.e. application of the trained RF, case-specific RF probability maps were generated based on the voxel-wise probability of belonging to the lesion class. The probability map was rescaled to the range

[0; 1] in such a way that the 99th percentile (and values above) of the probability map values corresponded to a value of 1. For lesion segmentation, the resulting map was thresholded by a value of 0.5. Given the resulting binary map, we excluded all connected areas with size smaller than 70 voxels to avoid false positives related to noise or image artifacts. Finally, a hole-filling algorithm was applied for each axial slice.

2.4   CNN Architecture and Workflow

Different CNN architectures were proposed in recent years for semantic image segmentation via deep learning. In this study, we decided to employ the currently popular DeepLabv3+ network that is designed to learn multi-scale contextual image features and controlling signal decimation [21]. For this purpose, it utilizes three key concepts:

Feature extraction. Feature extraction is conducted by a ResNet, pre-trained on the ImageNet data set, which employs residual learning to allow for efficient training of deeper neural networks [22].

Atrous convolutions. Efficient computation of feature responses by adding larger context information without increasing the number of parameters, i.e. sparse convolution kernels, is achieved using atrous convolutional layers.

Atrous spatial pyramid pooling. In order to combine the multi-scale information yielded by different atrous convolution layers, the atrous spatial pyramid pooling is applied.

For the problem defined by the ISLES 2018 challenge, the original DeepLabv3+ structure had to be slightly modified. To convert the network output into a binary segmentation, a convolutional filter was randomly chosen from the 21 pre-trained filters regarding the last convolutional layer. Further, we had to change the number of input color channels from 3 to 6 as we concatenated the given native CT, CBF, CBV, MTT, T_{max} and the computed MIP image data, yielding a tensor for every case and slice of shape $n_x \times n_y \times 6$. To allow for a robust and efficient training, 256 patches of shape $64 \times 64 \times 6$ of each (n_x, n_y)-slice were randomly selected and used as input data set. Based on this data, the model was re-trained in 5 epochs.

2.5   Combining RF and CNN Results

To analyze potential segmentation improvement by combining RF and CNN outputs, we averaged the RF probability map after rescaling by the 99th percentile and the CNN probability map. Thresholding and post-processing of the binary maps remained unchanged compared to previous explanations.

2.6   Experiments

To evaluate the implemented stroke segmentation approaches, we performed a 5-fold cross validation based on the underlying patient collective of the ISLES

Table 1. RF-only, CNN-only and results (mean ± standard deviation over cases) after combining respective probability maps for the individual folds and corresponding average data. HD and ASSD values refers multiples of the in-plane pixel side length.

Approach	Fold	Dice	HD	ASSD
RF-only	Fold 1	0.42 ± 0.25	50.59 ± 29.89	7.15 ± 5.33
	Fold 2	0.43 ± 0.26	34.98 ± 21.07	6.18 ± 6.52
	Fold 3	0.31 ± 0.24	56.25 ± 23.82	11.18 ± 8.57
	Fold 4	0.52 ± 0.23	41.47 ± 30.69	6.04 ± 6.73
	Fold 5	0.47 ± 0.24	55.53 ± 32.57	7.12 ± 7.23
	Avg	**0.42 ± 0.26**	**47.33 ± 28.94**	**7.52 ± 7.29**
CNN-only	Fold 1	0.51 ± 0.16	50.52 ± 27.77	6.58 ± 5.51
	Fold 2	0.41 ± 0.26	36.42 ± 24.93	5.70 ± 6.09
	Fold 3	0.31 ± 0.29	37.03 ± 23.99	6.23 ± 8.82
	Fold 4	0.49 ± 0.29	47.20 ± 29.68	8.07 ± 10.22
	Fold 5	0.48 ± 0.25	44.57 ± 33.40	6.62 ± 5.85
	Avg	**0.44 ± 0.27**	**43.30 ± 28.82**	**6.60 ± 7.46**
Combination	Fold 1	0.51 ± 0.21	44.63 ± 28.02	5.73 ± 5.45
	Fold 2	0.49 ± 0.27	31.45 ± 19.47	6.02 ± 7.82
	Fold 3	0.39 ± 0.26	48.21 ± 25.53	10.09 ± 10.34
	Fold 4	0.58 ± 0.22	39.76 ± 31.23	5.21 ± 6.05
	Fold 5	0.54 ± 0.24	42.83 ± 36.36	6.05 ± 6.93
	Avg	**0.50 ± 0.26**	**41.12 ± 28.85**	**6.69 ± 7.86**

training dataset. Splitting the 63 patients instead of the 94 cases ensured that the same patient was not part of both the training and testing data, and thereby prevented a related bias and overestimation of the segmentation performance.

The same folds were used for RF-only training and evaluation, CNN-only training and evaluation, and evaluation of the combination of RF and CNN output. Following previous ISLES contributions, we focused our evaluation on three metrics: Dice coefficient, Hausdorff distance (HD), and average symmetric surface distance (ASSD).

3 Results

The results for RF-only, CNN-only, and combined RF and CNN stroke lesion segmentation are summarized in Table 1. RF-only segmentation resulted in a mean Dice coefficient of 0.42 ± 0.26 (averaged over all testing cases and folds). The average CNN-only Dice coefficient was slightly higher (0.44 ± 0.27); however, this was consistently observed for all folds. In contrast, the Dice coefficient after combining RF and CNN outputs was for all folds at least as high as the higher single-approach value, indicating the potential of improving segmentation
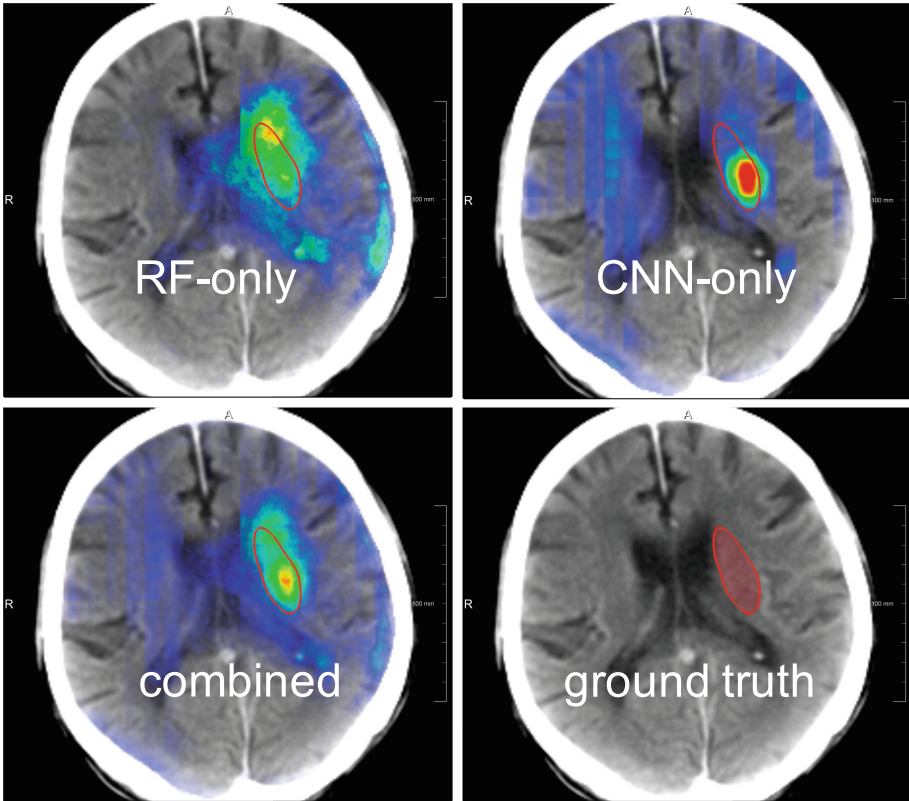
Fig. 1. Illustration of complementary information by RF-only and CNN-only stroke lesion segmentation, and the potential by combining respective outputs. All images are shown for case 40 and the fifth fold. Top row, left: probability map (blue: low probability; ref: max. values) and ground truth segmentation contours for RF-only (Dice 0.44); top row, right: same information for CNN-only (Dice 0.31). Bottom row, left: combined probability map (Dice 0.66); bottom row, right: ground truth segmentation. (Color figure online)

quality by combining the two approaches. Similar observations hold true for the Hausdorff distance values. ASSD was, in turn, on average lower for CNN-only than for RF-only and the combined approach. An illustration of the complementary information of the individual approaches can be found in Fig. 1.

4   Discussion and Conclusion

We applied and evaluated a classical random forest classifier employing hand-crafted image features commonly used in the context of stroke lesion segmentation [1], a state-of-the-art CNN architecture for image segmentation, and a combination of the two algorithms. Given the still relatively small amount of training

data especially for deep learning purposes, the underlying hypothesis was that the two methodically different approaches extract complimentary information and that, as a consequence thereof, the performance of the joint approach leads to an improved segmentation performance. The presented results (especially in terms of the Dice coefficient) support the hypothesis. Further improvement of segmentation performance could, for instance, be reached by using additional image data during training, exploiting data augmentation strategies and extensive use of additional methods (i.e. consideration of larger ensembles); nevertheless, the presented baseline performance is considered promising.

References

1. Maier, O., Menze, B.H., von der Gablentz, J., Hani, L., Heinrich, M.P., et al.: ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. **35**, 250–269 (2017)
2. Thrift, A.G., Thayabaranathan, T., Howard, G., Howard, V.J., Rothwell, P.M., et al.: Global stroke statistics. Int. J. Stroke **12**(121), 13–32 (2017)
3. ISLES challenge(s). http://www.isles-challenge.org
4. Winzeck, S., Hakim, A., McKinley, R., Pinto, J., Alves, V., et al.: ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Front. Neurol. **9**, 679 (2018)
5. Tatlisumak, T.: Is CT or MRI the method of choice for imaging patients with acute stroke? Why should men divide if fate has united? Stroke **33**, 2144–2145 (2002)
6. Wang, C.-W., Lee, J.-H.: Stroke lesion segmentation of 3D brain MRI using multiple random forests and 3D registration. In: Proceedings of MICCAI-ISLES 2015, pp. 35–38 (2015)
7. Chen, L., Bentley, P., Rueckert, D.: A novel framework for sub-acute stroke lesion segmentation based on random forest. In: Proceedings of MICCAI-ISLES 2015, pp. 9–12 (2015)
8. Maier, O., Wilms, M., Handels, H.: Random forests with selected features for stroke lesion segmentation. In: Proceedings of MICCAI-ISLES 2015, pp. 17–22 (2015)
9. Reza, S.M.S., Pei, L., Iftekharuddin, K.M.: Ischemic stroke lesion segmentation using local gradient and texture features. In: Proceedings of MICCAI-ISLES 2015, pp. 23–26 (2015)
10. Robben, D., Christiaens, D., Rangarajan, J.R., Gelderblom, J., Joris, P., et al.: ISLES challenge 2015: a voxel-wise, cascaded classification approach to stroke lesion segmentation. In: Proceedings of MICCAI-ISLES 2015, pp. 27–30 (2015)
11. Halme, H.-L., Korvenoja, A., Salli, E.: ISLES (SISS) challenge 2015: segmentation of stroke lesions using spatial normalization, random forest classification and contextual clustering. In: Proceedings of MICCAI-ISLES 2015, pp. 31–34 (2015)
12. Goetz, M., Weber, C., Maier-Hein, K.: Input data adaptive learning (IDAL) for sub-acute ischemic stroke lesion segmentation. In: Proceedings of MICCAI-ISLES 2015, pp. 39–42 (2015)

13. Mahmood, Q., Basit, A.: Automatic ischemic stroke lesion segmentation in multi-spectral MRI images using random forests classifier. In: Proceedings of MICCAI-ISLES 2015, pp. 43–46 (2015)
14. Jesson, A., Arbel, T.: Hierarchical segmentation of normal and lesional structures combining an ensemble of probabilistic local classifiers and regional random forest classification. In: Proceedings of MICCAI-ISLES 2015, pp. 57–62 (2015)
15. Muschelli, J.: Prediction of ischemic lesions using local image properties and random forests. In: Proceedings of MICCAI-ISLES 2015, pp. 63–65 (2015)
16. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. In: Proceedings of MICCAI-ISLES 2015, pp. 13–16 (2015)
17. Dutil, F., Havaei, M., Pal, C., Larochelle, H., Jodoin, P.-M.: A convolutional neural network approach to brain lesion segmentation. In: Proceedings of MICCAI-ISLES 2015, pp. 51–56 (2015)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
20. Maier, O.: MedPy – medical image processing in Python. https://pypi.org/project/MedPy
21. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 833–851. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)