Computers in Biology and Medicine 115 (2019) 103487

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks

Albert Clèrigues *, Sergi Valverde, Jose Bernal, Jordi Freixenet, Arnau Oliver, Xavier Lladó

Institute of Computer Vision and Robotics, University of Girona, Spain

ARTICLE INFO

Keywords:
Acute ischemic stroke
Brain
CT
CT perfusion
Automatic lesion segmentation
Convolutional neural networks
CNN

ABSTRACT

The use of Computed Tomography (CT) imaging for patients with stroke symptoms is an essential step for triaging and diagnosis in many hospitals. However, the subtle expression of ischemia in acute CT images has made it hard for automated methods to extract potentially quantifiable information. In this work, we present and evaluate an automated deep learning tool for acute stroke lesion core segmentation from CT and CT perfusion images. For evaluation, the Ischemic Stroke Lesion Segmentation (ISLES) 2018 challenge dataset is used that includes 94 cases for training and 62 for testing. The presented method is an improved version of our workshop challenge approach that was ranked among the workshop challenge finalists. The introduced contributions include a more regularized network training procedure, symmetric modality augmentation and uncertainty filtering. Each of these steps is quantitatively evaluated by cross-validation on the training set. Moreover, our proposal is evaluated against other state-of-the-art methods with a blind testing set evaluation using the challenge website, which maintains an ongoing leaderboard for fair and direct method comparison. The tool reaches competitive performance ranking among the top performing methods of the ISLES 2018 testing leaderboard with an average Dice similarity coefficient of 49%. In the clinical setting, this method can provide an estimate of lesion core size and location without performing time costly magnetic resonance imaging. The presented tool is made publicly available for the research community.

1. Introduction

Stroke is the third largest cause of death and the biggest source of acquired disability worldwide [1]. This condition is caused by a fatally low blood supply in a region of the brain. A shorter time to treatment since onset is strongly linked to a better outcome [2]. The stroke lesion is initially divided in two areas: the infarct core, composed of irreversibly damaged tissue, and the penumbra, tissue at risk that can still be recovered if blood flow is restored. Localization and quantification of the acute core or penumbra is of great clinical interest since it can help evaluate the amount of tissue that could be recovered with different treatments and take better informed decisions.

Non-contrast computed tomography (CT) imaging is fast, inexpensive, ubiquitous and is already used by clinicians as an essential first step for triage, diagnosis and treatment assessment of acute ischemic stroke [3]. Additionally, the information in these images has good prognostic potential, but are difficult to interpret. The infarct core is seen through subtle texture and intensity changes, also called parenchymal hypoattenuation, often masked by artifacts, noise or other tissue abnormalities [4]. Additionally, CT perfusion (CTP) can be used to assess the blood perfusion in the brain. To acquire CTP images, first an intra-venous contrast agent is injected and then repeated scans are made as it spreads through the brain. While CT shows the lesion core, CTP more clearly shows all areas with abnormal perfusion including both core and penumbra. The combination of both is also fast to acquire and might provide enough reliable information for automatic analysis.

Early work on supervised methods for acute stroke detection and segmentation using exclusively CT images relied on hand-crafted features exploiting texture and intensity [5–8].

Recent developments on Convolutional Neural Networks (CNN) [9] have given rise to methods with superior results that are present in the majority of state-of-the-art biomedical segmentation frameworks [10–13]. This trend can also be seen in the most recent methods for stroke lesion segmentation from MR images [14–16]. More specifically, U-shaped architectures based on the U-Net [10] are well suited for dense semantic segmentation. These kind of architectures have seen a number of recent improvements such as their extension for 3D volumetric segmentation [11,17] or the introduction of long and short residual skip connections [15,18]. Stroke lesion segmentation on CT images shares many of the same challenges as MR imaging, but still poses an

* Correspondence to: Ed. P-IV, Campus Montilivi, University of Girona, 17003 Girona, Spain.
E-mail address: albert.clerigues@udg.edu (A. Clèrigues).

https://doi.org/10.1016/j.compbiomed.2019.103487
Received 2 July 2019; Received in revised form 30 September 2019; Accepted 5 October 2019
Available online 9 October 2019

inherently different learning problem. Despite the promising results of deep learning applied to brain lesion segmentation, it still presents limitations for real world scenarios that severely limit its applicability. The most critical issues include typically small size of annotated datasets to train, domain and task dependent training procedures, highly unbalanced class extent (i.e. much less lesion tissue than healthy) and overfitting to the training images.

Deep learning has only been recently applied to CT imaging for acute stroke with the 2018 edition of the Ischemic Stroke Lesion Segmentation (ISLES) challenge. This challenge started in 2015 to provide a platform for a fair and direct comparison of automated methods for stroke imaging. The fourth edition in 2018 provides the first public acute stroke dataset using CT and CTP images. From the five challenge finalists, all deep learning based methods, four report the use of CNNs based on the U-Net architecture [10], one of which corresponds to our workshop challenge approach [19]. In these works, the issue of class imbalance was alleviated mainly with the use of cost sensitive loss functions, either class weighting [20,21] or difficulty weighting [22], or using patches with balanced sampling strategies [19].

In this work, we present and evaluate an automated deep learning tool for acute stroke lesion core segmentation from CT and CTP images. The presented tool is a simpler and improved version of the method initially submitted to the ISLES 2018 challenge, which already ranked among the challenge finalists, referred to as the workshop challenge approach. It achieves state-of-the-art performance while offering an easy training procedure and fast inference times. For alleviating class imbalance, both a patch based method with a balanced sampling strategy and a hybrid class weighted loss function are used. The deep learning architecture is an asymmetric encoder–decoder using long and short residual connections as done in recent state-of-the-art networks for dense segmentation [15,23]. Additionally, symmetric modality augmentation is performed that allows to exploit the brain symmetry property between hemispheres to find more robust image features. The introduced improvements with respect to our workshop challenge submission are quantified by crossvalidation on the ISLES 2018 training set. The proposed methodology is evaluated against other state-of-the-art methods with the blind challenge testing set submission, ranking among the top out of 41 entries. In the treatment decision workflow, this tool could provide a fast estimate of the lesion core location and volume without having to perform costly MR imaging. We release this tool to the community, available at https://github.com/NIC-VICOROB/stroke-core-ct-segmentation.

2. Materials

2.1. Data

The ISLES 2018 challenge tackled the segmentation of stroke lesion core from acute CT scans, taken within 8 h of stroke onset. The provided dataset (Kistler, 2013; Maier, 2017) includes 94 labeled training images and 62 unlabeled testing images. For each case, a CT scan, a raw CT perfusion time series (CT-PWI) and four derived perfusion maps (CBF, CBV, MTT and Tmax) are provided. The images were acquired as slabs with a variable number of axial slices, ranging from 2 to 22 depending on the patient, with 5 mm spacing and a resolution of 256×256. The raw perfusion time series include between 40 and 63 volumes, acquired 1–2 s apart, of the same dimensions as the CT for each patient. The provided gold standard was manually drawn on additional magnetic resonance DWI trace images not included in the challenge testing set, where the infarct core is seen more clearly, taken within 3 h of the initial CT scan.

2.1.1. Pre-processing

From the provided modalities, we only consider the use of CT and the four derived CT perfusion maps (CBF, CBV, MTT and Tmax), omitting the raw CT-PWI time series. Image pre-processing is then applied to the provided images in two steps: Firstly, the CT image is skull stripped and, secondly, a modality augmentation to exploit the symmetry of brain hemispheres is performed.

CT Skull stripping. The brain mask for skull stripping is obtained from the non-zero values of the sum of the four provided perfusion images, which did not take any value on the skull. Finally, the mask is multiplied with the CT image, which leaves only the desired brain tissue.

Symmetric modality augmentation. The use of the symmetry property showed significant improvements on chronic stroke lesion segmentation in MR images [24]. Since typically only one hemisphere of the brain is affected by the stroke, the brain mid-sagittal symmetry can be exploited to assess differences between both hemispheres and locate the lesion more accurately. In our case, we take advantage of the symmetry property by creating a symmetric version of each provided modality. In this way, a single patch will include information from the same spatial location of both hemispheres. To generate the symmetric modalities we first flip the CT images by the mid-sagittal axis. Since the images are not perfectly centered in the volume and some are slightly rotated, the opposing hemispheres might not be correctly aligned after the flip. Hence, we use FSL FLIRT [25] constrained to an axial affine transformation to linearly register both images and roughly align opposing hemispheres. In this case, a linear registration is sufficient since the symmetry features are not expected to rely on fine differences but rather on overall differences of patch intensity, parenchyma and/or perfusion statistics. Finally, the provided modality volumes are merged with the symmetrically augmented and used together for segmentation as an image with ten modality volumes. In this way, a single patch will additionally include bilateral information of all modalities. Fig. 1 shows an example case with the provided and augmented modalities.

3. Method

The proposed approach is a 2D patch based deep learning approach for segmentation of the acute stroke lesion core from CT perfusion images. Since the lesion core class represents around 5% of the brain tissue in the training set, class imbalance is an issue that needs to be dealt with. If no deliberate action is taken, the training set would include fewer examples of lesion than healthy tissue, which would bias the learning and worsen segmentation performance. Additionally, overfitting to the training set is likely, considering the small quantity of data, which would cause bad performance for other images. To minimize this effect, the training is regularized by using: (a) data augmentation with elastic deformation fields, (b) dropout layers that introduce noisy updates during training and (c) early stopping that interrupts training when no more generalizable knowledge can be learned. Finally, a combination of classification uncertainty estimation and use of highly overlapping patches further reduces outliers and segmentation artifacts.

3.1. Class imbalance

The most common techniques to alleviate this issue for deep learning methods are three: cost sensitive loss functions, which assign different cost to misclassification of examples from different classes [26]; the use of patches with deliberate sampling, typically aiming to overrepresent the minority class, or multi-phase training, where a part of the network is retrained with a different class distribution. In this work, we propose the use of both a balanced patch sampling and a cost sensitive loss function to alleviate the imbalance.

The employed sampling strategy is an extension of a recent proposal for brain lesions in general [12]. The strategy has been extended to take into account the anatomy and pathophysiology of acute stroke. In practice, a target number of patches is set for each patient. Then, half of the patches are extracted centered on lesion voxels and the other half on healthy ones. These are sampled in regular spatial steps to ensure all parts of the volume are uniformly represented. For the lesion class sampled voxels have a random offset applied in the x and y axis before

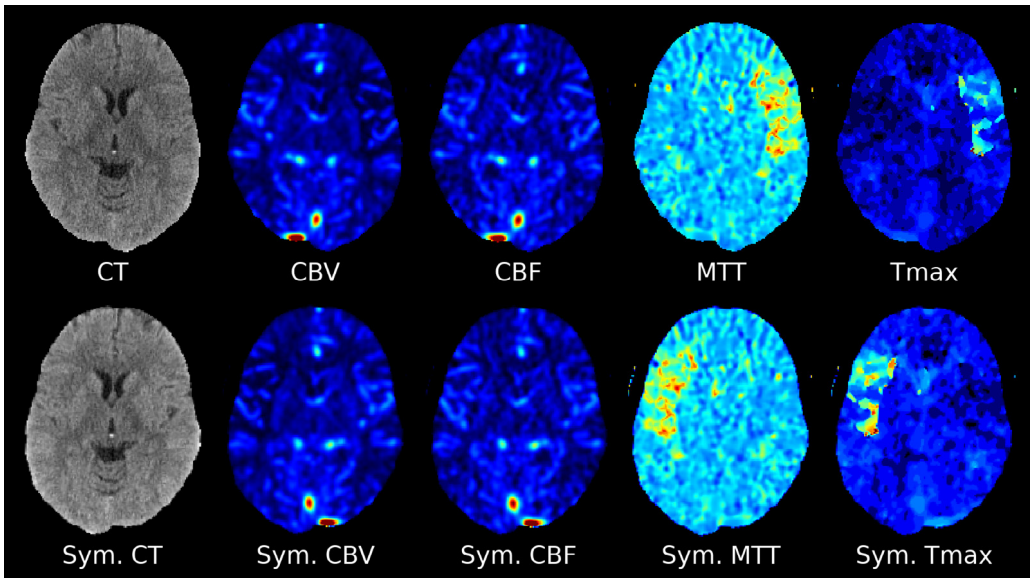

Fig. 1. Top row: Provided CT and derived CT perfusion maps. Bottom row: Resulting symmetrically augmented modalities.

patch extraction, as done by Guerrero et al. [15]. This offset is sampled from a random uniform distribution and is limited to half of the patch size to ensure the originally sampled voxel is inside the finally extracted patch. This increases the representation of areas adjacent to the core label, the penumbra region, while providing a degree of translational data augmentation. The patches will be extracted centered on these voxels. For patients with smaller lesions, several patch extractions from the same lesion voxel and data augmentation are applied, using the elastic deformation described in [27] with parameters $\alpha = 2.5$ and $\gamma = 0.12$, to reach the target number of patches per patient. In this way, only if the number of lesion voxels is smaller than the target number, they will be repeated and augmented using elastic deformation. On average, the augmented patches amount to 5% of the training set. The use of this patch sampling strategy raised the lesion voxel fraction in the training set from 5% to 12%.

Additionally, we use a cost sensitive loss function that is the sum of the Generalized Dice Loss (GDL) [28] and the crossentropy loss to further minimize the effects of class imbalance. While the crossentropy loss is minimized with correct confident predictions, the GDL is minimized by maximizing the relative overlap between prediction and ground truth. In practice, jointly minimizing both terms provides the crossentropy convergence properties with the balancing class weighting of the GDL.

However, despite the use of both techniques, the overlap segmentation is decreased when bigger patch sizes are considered due to worsened imbalance, since larger patch sizes will tend to include a bigger ratio of healthy to lesion voxels. After empirical testing with several patch sizes ranging from 16×16 to 96×96, we choose a patch size of 64×64 that offers the best compromise between a large receptive field and worsened class imbalance.

3.2. Deep learning architecture

The employed network, depicted in Fig. 2, is a 2D asymmetric residual encoder–decoder that produces whole patch predictions. It is based on recent state-of-the-art networks for chronic stroke [15] and related biomedical tasks [11]. The network has five resolution steps with 8 base filters, which are doubled in each step, resulting in a latent space with 128 feature maps of 4×4 resolution. It has long and short residual connections to ease gradient flow, which improves convergence properties and allows for better accuracy [29]. The asymmetry comes from the reduced number of parameters found in the decoder branch. It has been shown that the role of the decoder is not as critical and its complexity can be reduced without damaging the performance [23]. In this way, the residual blocks have two convolutional layers in the encoder and one in the decoder, resulting in 75% and 25% of the parameters in each respectively. Additionally, it includes prediction dropout layers that will be used for estimating the uncertainty in classification to minimize outliers.

3.3. Pipeline overview

In this section we will briefly describe the different parts of the training and testing pipeline to train the network and use it to segment the desired images.

Training. In the training phase, the randomly initialized network weights are trained with patch training and validation sets built from the provided images. A total of 376,000 patches, 4000 from each case, of size 64×64 are extracted using the sampling strategy described in Section 3.1 to create the training set. The sum of the Generalized Dice and Crossentropy loss is used as the objective function. During training, the weights are updated with the Adadelta optimizer [30], which requires no manual tuning of learning rate. After several empirical tests, we use a batch size of 64 patches during training since it provides a good compromise between sensitivity and overfitting. The batch size determines the number patches whose gradients will be averaged before a network weight update during training. A bigger batch size averages the gradients of more patches, which improves the overall accuracy while giving less weight to errors in individual samples. To further minimize overfitting, early stopping with a patience of ten epochs is performed when the sum of error rate and L1 loss on the validation set reaches a global minimum. We set the low number of ten patience epochs to avoid excessive overfitting to the validation set, given the small size of the dataset. Although further training might still improve the validation metrics it could be at the cost of overfitting to the validation images and worsening the performance with testing images. In practice, the networks are trained for a maximum of 100 epochs or until the early stopping condition has been met, storing the network weights with the best validation metrics. The number of training epochs ranges from 20 to 40 for the reported experiments.

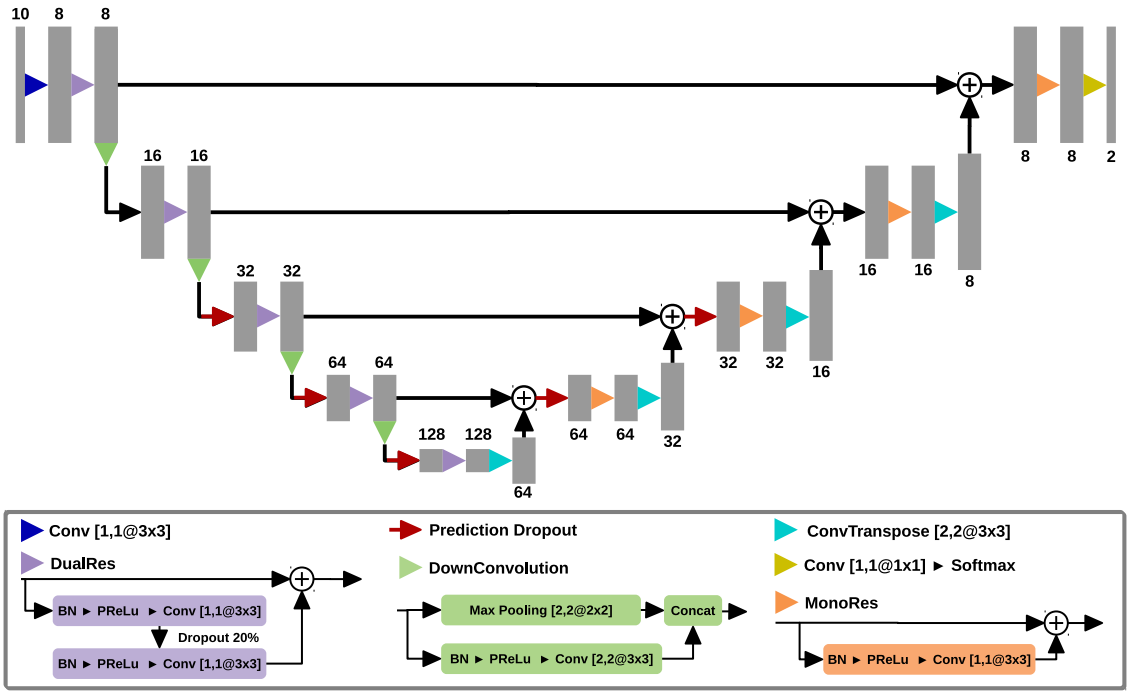

Fig. 2. Diagram of the employed deep learning architecture, an asymmetrical residual encoder–decoder CNN. Gray rectangles represent feature maps with the number of features indicated either on top or bottom. For the convolutional layers, $[S_x,S_y@K_x\times K_y]$ indicates the strides and kernel size in each axis respectively. Red arrows mark the location where dropout is applied at prediction time for average uncertainty filtering.

Testing. In the testing phase we predict the class probability distribution for each voxel of a given image with the trained network. Firstly, patches are extracted from the whole image at regular spatial intervals to make sure all parts of the volume are represented. Furthermore, a degree of overlap is considered to improve spatial label coherence and minimize boundary artifacts. Each extracted patch is forward passed through the network and its predicted probabilities accumulated into a common space, preserving its spatial location. Finally, the average of accumulated probabilities in each voxel is made. Additionally, uncertainty filtering by averaging is applied to each patch forwarded through the network. It has been shown that a patch predicted while using dropout can be considered a Monte Carlo sample from the unknown classification probability distribution [31]. In our case, for each patch, 3 forward passes are performed with a voxel-wise prediction dropout rate of 10%. As suggested by [32], dropout in prediction is only performed in the deepest resolution steps as seen in Fig. 2. Finally, the probability distribution is computed as the voxel-wise average of the three noisy predictions.

Post-processing. In the post-processing phase, a binary segmentation is produced from the predicted probability maps. It is performed in two steps, first the probabilities are binarized according to a threshold T and then a connected component filtering removes lesions smaller than S_{min} voxels. The parameters T and Smin that optimize the DSC and HD of the tool are found through grid search for each evaluation. More specifically, we test 9 different thresholds T, from 0.1 to 0.9 in 0.1 steps, and 6 minimum lesion sizes (Smin) ranging from 10 to 500 voxels. Each combination of these parameters is then used to binarize the predicted probability maps and compute segmentation metrics. We select the T and Smin that jointly optimize the DSC and HD metrics, the ones used to rank the ISLES 2018 challenge workshop participants.

3.4. Implementation details

The proposed method has been implemented with Python, using the Torch scientific computing framework [33]. All experiments have been run on a GNU/Linux machine running Ubuntu 18.04 with 64 GB of RAM memory and an Intel® Core™ i7-7800X CPU. The network training and testing has been done with an NVIDIA TITAN X GPU (NVIDIA corp, United States) with 12 GB G5X memory.

4. Evaluation and results

The proposed methodology is evaluated with a crossvalidation experiment showing the improvements against our initial workshop challenge approach and with an external blind evaluation against state-of-the-art methods using the testing set. The evaluation metrics for both experiments include the Dice similarity coefficient (DSC) [34] and Hausdorff distance (HD), the ones considered to rank the workshop challenge participants. Additionally, we also consider other metrics more relevant to the clinical setting such as positive predictive value (PPV), sensitivity and coefficient of determination (COD), also called R^2, between the predicted and true core volume. Finally, we consider the dependent t-test for paired samples to assess the statistical significance of differences between the evaluation results.

4.1. Crossvalidation experiment

The purpose of this experiment is to quantitatively assess the improvements introduced to the proposed method with respect to our workshop challenge approach (the baseline). Mainly, the improvements come from a more regularized network training procedure, symmetric modality augmentation and uncertainty filtering. Additionally, a single network is used in contrast with the two networks in cascade configuration of the baseline. Thanks to the added improvements we can avoid the use of the second model, which simplifies the training procedure and reduces inference times. The current more regularized training procedure uses the sum of GDL and crossentropy as loss function and the sum of L1 loss and error rate for early stopping. However, for the baseline approach [19] the networks were trained using crossentropy as loss function and a probabilistic Dice loss [28] for early stopping. Additionally, we are able to use bigger 64 × 64 patches without a decrease in segmentation performance as it happened with the baseline,

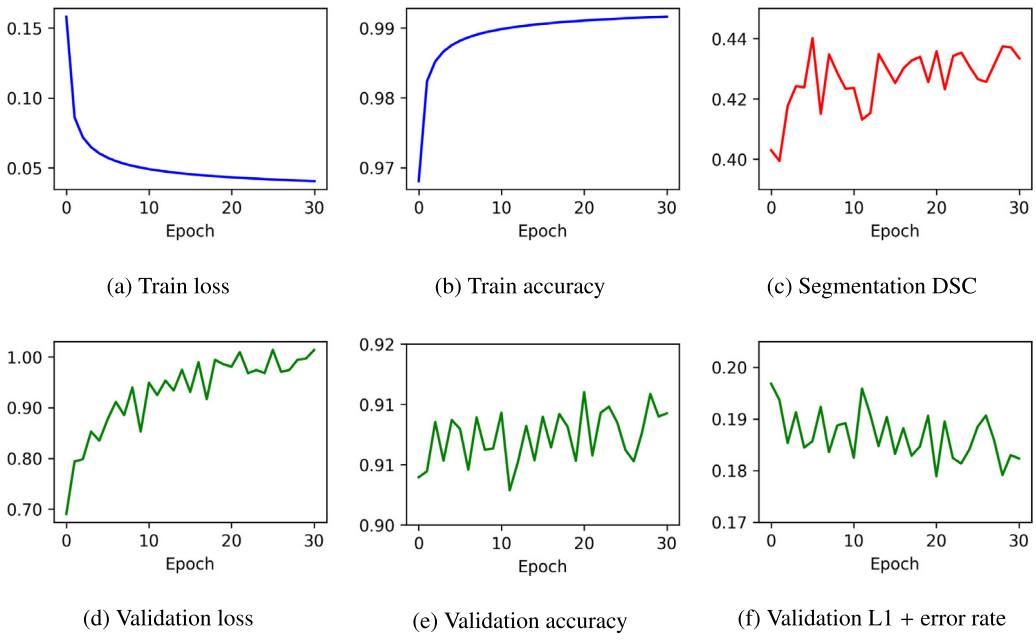

Fig. 3. Loss and accuracy plots for a single cross-validation fold.

where using 64 × 64 patches resulted in 4% lower DSC than using 48 × 48. The bigger patch size of 64 × 64 offers a bigger receptive field from which to learn features.

The experiment consisted of four evaluations, first the baseline and then three with the incremental improvements that comprise the proposed method. Each evaluation is performed in 5 crossvalidation folds across the 94 labeled images of the ISLES 2018 dataset, having 75 training and 19 validation images for each fold. Since some scans correspond to different regions of the same patient, we ensure that all the same patient scans are within the same set. In each fold, a single network is trained with the training patches and then the validation volumes are predicted, resulting in a probability map for each case. After all five folds have finished there will be one predicted probability map for each of the 94 training images. Finally, the probability maps are post-processed using T and Smin, found through grid search, that achieve the best segmentation metrics across all folds of the crossvalidation.

Fig. 3 shows loss and accuracy plots for a single cross-validation fold, since the other folds were of similar nature. Additionally, it shows the early stopping metric value, the L1 loss plus error rate, and the segmentation DSC of the validation images. The figure shows how the loss function evaluated on the validation set increases, instead of decreasing, while the validation accuracy improves. For this reason, we do not use the validation loss and instead use the sum of L1 loss and Error rate as a monitored metric on the patch validation set for early stopping, since it is more correlated with segmentation DSC of the validation images. In this case, the early stopping metric reaches a global minimum in epoch 20 where the segmentation DSC begins to stabilize. Although further training might still improve the validation metrics it might be at the cost of overfitting to it and worsening the performance with testing images.

Table 1 shows the evaluation metrics obtained from the baseline and incremental improvements. Compared with the baseline, the regularized training procedure with a single model significantly improves the DSC and sensitivity ($p < 0.02$). When augmented modalities are additionally considered, the PPV significantly improves although the sensitivity is reduced ($p < 0.05$). Moreover, when uncertainty filtering is considered the HD is significantly reduced at the expense of a lower sensitivity ($p < 0.03$). In general, all introduced improvements raise the COD, meaning that the estimated volume is closer to the gold standard. In summary, the proposed tool provides significantly better DSC, HD and PPV ($p < 0.05$) than the baseline with a marginal higher sensitivity.

Fig. 4 shows qualitative evaluations of the incremental improvements for three representative cases. As compared with the baseline, the regularized training achieves better sensitivity and specificity in all cases, reducing the amount of false positives and negatives. The addition of symmetric modalities overall improves lesion localization but can reduce the sensitivity for some samples. For instance, the use of symmetric modalities increases the false positives in the middle row case. Finally, the bottom row is a good example of the effect of uncertainty filtering in the majority of cases, improving lesion localization and estimated volume. However, in some cases it may also introduce additional outliers as seen in the top row case, where false positives appear in the upper part of the lesion.

4.2. ISLES 2018 testing evaluation

For segmentation of the 62 unlabeled testing images from the ISLES 2018 dataset, we used all five networks, one from each fold, that were trained for the crossvalidation evaluation with all improvements. An averaging approach is used where each patch is passed through the five trained models and the five predictions are averaged together to produce a single patch prediction. In this way, bootstrap aggregation [35] is performed, where each network is trained with a different subset of training data. Finally, the resulting class probability maps of the testing images are binarized using the previously computed optimal parameters $T = 0.2$ and $S_{min} = 200$ from the crossvalidation experiment. Table 2 shows ongoing benchmark leaderboard of the ISLES 2018 testing set sorted by average DSC, where the proposed methodology ranks among the top entries out of 41 participants.

5. Discussion

The results of the ISLES 2018 testing set evaluation show that the proposed methodology achieves state-of-the-art performance ranking 2nd in the ongoing benchmark leaderboard among 41 submissions. The approach by Song et al. [20] manages to achieve a 2% higher DSC by additionally using the 40 or more volumes that comprise each raw perfusion time series (CT-PWI) to further extract features for segmentation. The use of the raw perfusion time series would involve an increase in memory requirements and processing time, additionally making the training procedure more complex. In our case, we still use some of the

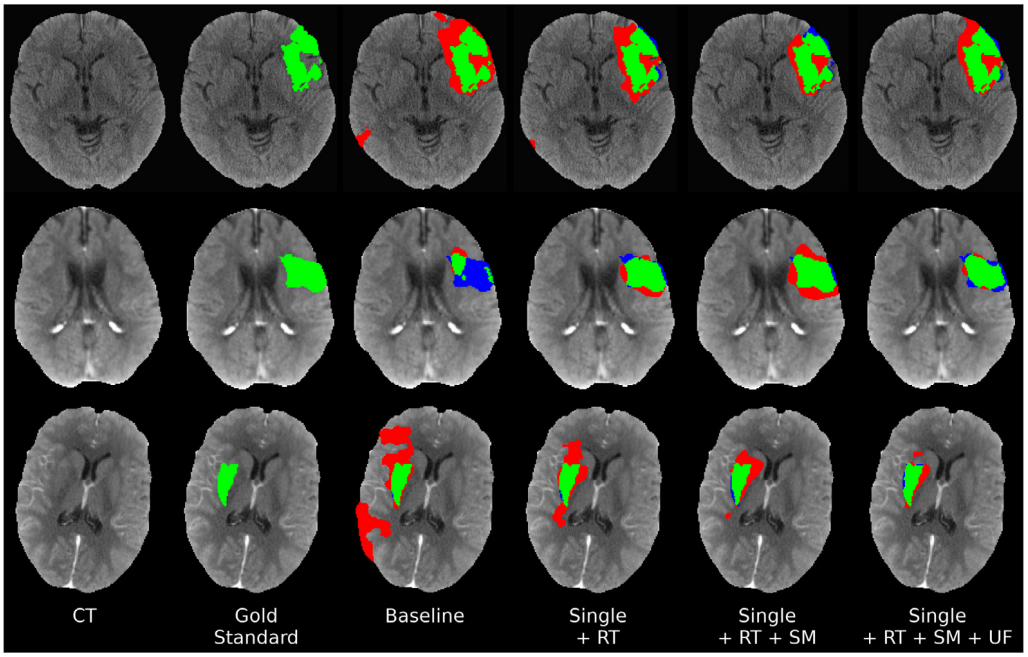

Fig. 4. Lesion core segmentation masks of the baseline and incremental improvements. True positives are denoted in green, false positives in red and false negatives in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Evaluation metrics of the crossvalidation experiment in the ISLES 2018 training set. The baseline results correspond to our workshop challenge approach while Single refers to the current approach using a single network. The evaluated improvements are three: the regularized training (RT), symmetric modality augmentation (SM) and the uncertainty filtering (UF).

Method	T	S_{min}	DSC (%)	PPV (%)	Sens. (%)	HD	R^2
Baseline	.4	200	49.0 ± 23.6	46.9 ± 29.5	57.2 ± 26.7	29.5 ± 18.9	.67
Single +RT	.1	300	53.5 ± 24.6	51.8 ± 28.9	66.0 ± 25.5	29.8 ± 23.9	.74
Single +RT +SM	.2	200	54.8 ± 24.8	58.3 ± 29.8	63.7 ± 25.3	26.6 ± 19.9	.78
Single +RT +SM +UF	.2	200	54.7 ± 24.2	57.8 ± 29.1	60.9 ± 25.0	23.5 ± 15.8	.82

Table 2
Top 10 entries of the ongoing benchmark leaderboard, last accessed 25/06/2019, of ISLES 2018 testing set as ranked by average DSC. The entry of the presented tool is highlighted in bold. *Values divided by 1,000,000.

Rank	User	DSC	PPV	Sensitivity	HD*
1	songt1 [20]	0.51 ± 0.31	0.55 ± 0.36	0.55 ± 0.34	19.4 ± 39.5
2	**clera2 (ours)**	**0.49 ± 0.31**	**0.51 ± 0.36**	**0.57 ± 0.35**	**11.3 ± 31.6**
3	pengl1 [21]	0.49 ± 0.31	0.56 ± 0.37	0.53 ± 0.33	19.4 ± 39.5
4	zhans10	0.49 ± 0.32	0.53 ± 0.35	0.54 ± 0.35	17.7 ± 38.2
5	cheny11 [36]	0.48 ± 0.32	0.59 ± 0.38	0.46 ± 0.33	9.7 ± 29.6
6	lilic2	0.48 ± 0.32	0.48 ± 0.34	0.6 ± 0.36	17.7 ± 38.2
7	liliy8	0.48 ± 0.31	0.5 ± 0.36	0.55 ± 0.34	19.4 ± 39.5
8	liliy2	0.47 ± 0.32	0.53 ± 0.36	0.47 ± 0.32	16.1 ± 36.8
9	xiaoh3 [22]	0.47 ± 0.31	0.56 ± 0.37	0.49 ± 0.33	19.4 ± 39.5
10	zhuoj2	0.47 ± 0.32	0.51 ± 0.36	0.54 ± 0.36	11.3 ± 31.6

information obtained from the absorption curve parametrization of the raw perfusion time series in the 4 perfusion parameter maps (CBF, CBV, MTT and Tmax). Despite the potential performance improvement of also processing the raw time series as shown by Song et al. [20], we avoid it in favor of reducing the training complexity and provide faster inference times.

The crossvalidation experiment shows the big influence that class imbalance and training regularization can have on segmentation performance. For instance, the class weighting properties of the focal loss allow the use of bigger 64 × 64 patches without worsened imbalance and provides a DSC improvement of 4.5% over the baseline. However, this patch size is too small to fit both brain hemispheres simultaneously and makes implausible exploiting symmetrical features. The use of symmetric modality augmentation allows learning of these features without having to use bigger patches that would worsen class imbalance. Despite the overall improvement from augmented modalities, some cases are actually worsened, as seen in the middle row of Fig. 4 with a lower PPV that increases false positives. Finally, we noted that the use of uncertainty filtering significantly reduced outliers but also harmed segmentation performance with bigger dropout rates. We found that averaging the output of several passes with a low dropout rate of 10% in prediction was enough to reduce outliers without significantly harming the overlap performance. Despite the marginally worsened DSC, PPV and sensitivity that uncertainty filtering provides, we believe the significantly reduced HD and better estimation of the core volume are more desirable properties in the clinical setting. Additionally, since each patch will require the average of three noisy predictions, this effectively triples the network inference time. However, even when considering the pre-processing step, segmentation of the largest images typically takes under two minutes in our system.

6. Conclusions

In this work, we presented and evaluated an automated method for acute stroke lesion core segmentation from CT and CTP images. The presented tool achieves state-of-the-art performance while using a simple training procedure with a single network. The training requires minimal tuning of parameters thanks to the Adadelta optimizer and a robust class imbalance handling using balanced patch sampling and a class weighting loss function. We improve segmentation performance with a novel way of using the symmetry property of brain hemispheres in patch based methods. We also explore the use of prediction dropout layers to reduce outliers and improve lesion core volume estimation, a

predictor of clinical severity and outcome in ischemic stroke [37]. This tool can provide with an estimate of core location and volume without acquiring time costly MR images. In the clinical setting, this estimate can be used to guide treatment decisions or help assess the need for further MR imaging. A trainable implementation of the presented tool is freely released for the research community.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Albert Clèrigues holds an FPI grant from the Ministerio de Ciencia, Innovación y Universidades with reference number PRE2018-083507. This work has been partially supported by Retos de Investigación TIN2015-73563-JIN and DPI2017-86696-R from the Ministerio de Ciencia, Innovación y Universidades. The authors gratefully acknowledge the support of the NVIDIA Corporation with their donation of the TITAN X GPU used in this research.

References

[1] C.L. Sudlow, C.P. Warlow, Comparable studies of the incidence of stroke and its pathological types: Results from an international collaboration. International stroke incidence collaboration., Stroke 28 (3) (1997) 491–499.

[2] Sunil A. Sheth, Reza Jahan, Jan Gralla, Vitor M. Pereira, Raul G. Nogueira, Elad I. Levy, Osama O. Zaidat, Jeffrey L. Saver, Time to endovascular reperfusion and degree of disability in acute stroke, Ann. Neurol. 78 (4) (2015) 584–593, http://dx.doi.org/10.1002/ana.24474.

[3] Michael H. Lev, Jeffrey Farkas, Joseph J. Gemmete, Syeda T. Hossain, George J. Hunter, Walter J. Koroshetz, R. Gilberto Gonzalez, Acute stroke: Improved nonenhanced CT detection—Benefits of soft-copy interpretation by using variable window width and center level settings, Radiology 213 (1) (1999) 150–155, http://dx.doi.org/10.1148/radiology.213.1.r99oc10150.

[4] Rafał Józwiak, Artur Przelaskowski, Grzegorz Ostrek, Conceptual improvements in computer-aided diagnosis of acute stroke, J. Med. Inf. Technol. 17 (2011).

[5] Andrius Ušinskas, Romualdas Dobrovolskis, Bernd F. Tomandl, Ischemic stroke segmentation on CT images using joint features, Informatica 15 (2) (2004) 283–290.

[6] M. Chawla, S. Sharma, J. Sivaswamy, L.T. Kishore, A method for automatic detection and classification of stroke from brain CT images, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 3581–3584, http://dx.doi.org/10.1109/IEMBS.2009.5335289.

[7] Fuk-hay Tang, Douglas K.S. Ng, Daniel H.K. Chow, An image feature approach for computer-aided detection of ischemic stroke, Comput. Biol. Med. 41 (7) (2011) 529–536, http://dx.doi.org/10.1016/J.COMPBIOMED.2011.05.001.

[8] N. Hema Rajini, R. Bhavani, Computer aided detection of ischemic stroke using segmentation and texture features, Measurement 46 (6) (2013) 1865–1874, http://dx.doi.org/10.1016/J.MEASUREMENT.2013.01.010.

[9] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551, http://dx.doi.org/10.1162/neco.1989.1.4.541.

[10] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241, arXiv:1505.04597.

[11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2016, pp. 424–432, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[12] Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, Ben Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78, http://dx.doi.org/10.1016/J.MEDIA.2016.10.004.

[13] Jose Dolz, Christian Desrosiers, Ismail Ben Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study, NeuroImage 170 (2017) 456–470, http://dx.doi.org/10.1016/J.NEUROIMAGE.2017.04.039.

[14] Liang Chen, Paul Bentley, Daniel Rueckert, Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks, NeuroImage: Clin. 15 (2017) 633–643, http://dx.doi.org/10.1016/J.NICL.2017.06.016.

[15] R. Guerrero, C. Qin, O. Oktay, C. Bowles, L. Chen, R. Joules, R. Wolz, M.C. Valdés-Hernández, D.A. Dickie, J. Wardlaw, D. Rueckert, White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks, NeuroImage: Clin. 17 (2018) 918–934, http://dx.doi.org/10.1016/J.NICL.2017.12.022.

[16] Rongzhao Zhang, Lei Zhao, Wutao Lou, Jill M. Abrigo, Vincent C.T. Mok, Winnie C.W. Chu, Defeng Wang, Lin Shi, Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densenets, IEEE Trans. Med. Imaging 37 (9) (2018) 2149–2160, http://dx.doi.org/10.1109/TMI.2018.2821244.

[17] Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, 2016, arXiv:1606.04797.

[18] Ke Zhang, Miao Sun, Tony X. Han, Xingfang Yuan, Liru Guo, Tao Liu, Residual networks of residual networks: Multilevel residual networks, 2016, arXiv:1608.02908. http://dx.doi.org/10.1109/TCSVT.2017.2654543.

[19] Albert Clèrigues, Sergi Valverde, Jose Bernal, Kaisar Kushibar, Mariano Cabezas, Arnau Oliver, Xavier Lladó, Ensemble of convolutional neural networks for acute stroke anatomy differentiation, in: International MICCAI Brainlesion Workshop, 2018.

[20] Tao Song, 3D Multi-scale U-Net with atrous convolution for ischemic stroke lesion segmentation, in: International MICCAI Brainlesion Workshop, 2018.

[21] Pengbo Liu, Stroke lesion segmentation with 2D convolutional neutral network and novel loss function in: International MICCAI Brainlesion Workshop, 2018.

[22] Xiaojun Hu, Weilin Huang, Sheng Guo, Matthew R. Scott, StrokeNet: 3D Local refinement network for ischemic stroke lesion segmentation, in: International MICCAI Brainlesion Workshop, 2018.

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

[24] Yanran Wang, Aggelos K. Katsaggelos, Xue Wang, Todd B. Parrish, A deep symmetry convnet for stroke lesion segmentation, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 111–115, http://dx.doi.org/10.1109/ICIP.2016.7532329.

[25] Mark Jenkinson, Peter Bannister, Michael Brady, Stephen Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, NeuroImage 17 (2) (2002) 825–841.

[26] Charles Elkan, The foundations of cost-sensitive learning, in: International Joint Conference on Artificial Intelligence, 2001, pp. 973–978.

[27] P.Y. Simard, D. Steinkraus, J.C. Platt, Best practices for convolutional neural networks applied to visual document analysis, in: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, vol. 1, IEEE Comput. Soc, pp. 958–963. http://dx.doi.org/10.1109/ICDAR.2003.1227801.

[28] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Sebastien Ourselin, M. Jorge Cardoso, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, Lecture Notes in Comput. Sci. 10553 LNCS (2017) 240–248, http://dx.doi.org/10.1007/978-3-319-67558-9_28, arXiv:1707.03237.

[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[30] Matthew D. Zeiler, ADADELTA: An adaptive learning rate method, 2012, arXiv preprint arXiv:1212.5701, abs/1212.5. http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503.

[31] Yarin Gal, Zoubin Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, in: International Conference on Machine Learning, 2015, pp. 1050–1059. arXiv:1506.02142.

[32] Tanya Nair, Doina Precup, Douglas L. Arnold, Tal Arbel, Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 655–663. http://dx.doi.org/10.1007/978-3-030-00928-1_74.

[33] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, Adam Lerer, Automatic differentiation in pytorch, in: Neural Information Processing Systems, 2017.

[34] Lee R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302, http://dx.doi.org/10.2307/1932409.

[35] Leo Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, http://dx.doi.org/10.1007/BF00058655.

[36] Yu Chen, Yuexiang Li, Yefeng Zheng, Ensembles of modalities fused model for ischemic stroke lesion segmentation, in: International MICCAI Brainlesion Workshop, 2018.

[37] Karl-Olof Löuvbld, Alison E. Baird, Gottfried Schlaug, Andrew Benfield, Bettina Siewert, Barbara Voetsch, Ann Connor, Cara Burzynski, Robert R. Edelman, Steven Warach, Ischemic lesion volumes in acute stroke by diffusion-weighted magnetic resonance imaging correlate with clinical outcome, Ann. Neurol. 42 (2) (1997) 164–170, http://dx.doi.org/10.1002/ana.410420206.