SUNet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal MRI

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

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

Abstract

Acute stroke lesion segmentation and prediction tasks are of great clinical interest as they can help doctors make better informed time-critical treatment decisions. Automatic segmentation of these lesions is a complex task due to their heterogeneous appearance, dynamic evolution and inter-patient differences. Typically, acute stroke lesion tasks are approached with methods developed for chronic stroke or other brain lesions. However, the pathophysiology and anatomy of acute stroke lesions establishes an inherently different problem that needs special consideration. In this work, we propose a novel deep learning architecture specially designed for acute stroke tasks that involve approximating complex non-linear functions with reduced data. Within our strategy, class imbalance is tackled using a hybrid strategy based on state-of-the-art train sampling strategies designed for other brain lesion related tasks, which is more suited to the anatomy and pathophysiology of acute stroke lesions. The proposed method is evaluated on three unrelated public international challenge datasets (ISLES) without any dataset specific hyper-parameter tuning. These involve the tasks of sub-acute stroke lesion segmentation, acute stroke penumbra estimation and chronic extent prediction from acute MR images. The performance of the proposed architecture is analysed both against similar deep learning architectures from chronic stroke and related biomedical tasks and also by submitting the segmented test images for blind online evaluation on each of the challenges. When compared with the rest of submitted strategies, our method achieves top-rank performance among the best submitted entries in all the three challenges, showing its capability to deal with different unrelated tasks without hyper-parameter tuning. In order to promote the reproducibility of our results, a public version of the proposed method has been released.

Keywords: Brain, MRI, acute ischaemic stroke, automatic lesion segmentation, convolutional neural networks

1. Introduction

Stroke is a medical condition by which an abnormal blood flow in the brain causes the death of cerebral tissue. Stroke is the third cause of morbidity worldwide, after myocardial infarction and cancer, and the most prevalent cause of acquired disability (Redon et al., 2011). The affected tissue in the acute phase can be divided into three concentric regions depending on the potential for recovery, also referred as salvageability: core, penumbra and benign oligemia (Rekik et al., 2012). The core, located at the centre, is formed by irreversibly damaged tissue from a fatally low blood supply. The penumbra, located around the core, represents tissue at risk but that can still be recovered if blood flow is restored. Finally, the benign oligemia is the outter most ring whose vascularity has been altered but is not at risk of damage. Once the symptoms of stroke have been identified, a shorter time to treatment is highly correlated with a positive outcome (Sheth et al., 2015). Mechanical thrombectomy is a strongly recommended option for elegible patients (Campbell et al., 2017). However, this surgery is not free of risks, an overall complication rate of 15.3% was observed in a year long study (Singh et al., 2017). In the treatment decision context, acute penumbra segmentation can be used as a quantitative estimate of the tissue that could be salvaged with treatment. On the other hand, lesion outcome prediction methods can provide a quantitative estimate of the tissue that could be lost without any treatment. These two methods can provide doctors with the estimated out-

*Corresponding author. A. Clèrigues, Ed. P-IV, Campus Montilivi, University of Girona, 17003 Girona (Spain). e-mail: albert.clerigues@udg.edu. Phone: +34 683645681; Fax: +34 972 418976.

come of both scenarios and allow for more informed treatment decisions.

Automating tasks for stroke lesions is complex due to its relation with the vascular system and a highly dynamic evolution. Stroke lesions can occur anywhere in the brain, may not appear as homogeneous regions and its appearance varies significantly over time (Maier et al., 2017). Additionally, imaging of acute stroke lesions requires the use of several MRI modalities since the distinct lesion parts have different appearance depending on the imaging principle (Rekik et al., 2012). The complexity of the task, clinical nuances and the absence of quality public datasets have been factors that contributed to a sparse and diffuse state of the art for acute stroke methods.

Recently, the Ischemic Stroke Lesion Segmentation (ISLES) challenge (Maier et al., 2017) started in 2015 to provide a platform for a fair and direct comparison of automated methods for stroke. Its first edition included the sub-acute ischemic stroke lesion segmentation (SISS) and the acute stroke outcome/penumbra estimation (SPES) subtasks. The ISLES challenge in 2016 and 2017 focused on prediction of chronic lesion outcome from acute images. Until recently, Random Decision Forests (RDFs) (Ho, 1995) were the state-of-the-art methods for stroke lesion segmentation due to their excellent generalisation properties, which make them well suited for difficult tasks with few training samples (Maier et al., 2015a). The two best methods on SPES (Maier et al., 2015b; Mckinley et al., 2015) and the third best on the SISS (Halme et al., 2015) were based on RDFs. More recently, Convolutional neural networks (CNNs) (Lecun et al., 2015), which enable the learning of optimal features for each task, have quickly become the state of the art and are replacing RDF based methods. In the 2016 edition, among the top three methods one was based on RDFs and two on CNNs. In the 2017 edition, only CNNs were present among the top three methods (Winzeck et al., 2018). Even though CNNs can learn optimal features for the task at hand, they are still restricted by the architectural design, the amount and quality of available data and the training procedure. Recently, advances in regularisation techniques and data imbalance handling allow for increased CNN generalisation performance in brain lesion segmentation that rivals that of RDFs. More specifically, the U-Net architecture (Ronneberger et al., 2015) is very well suited for segmentation tasks and methods based on it have quickly become state of the art for stroke segmentation. This is clearly seen in the submissions for the ISLES 2017 challenge, where 10 out of the 14 participating methods, including the top three, were based on this kind of architectures (Winzeck et al., 2018).

In this work, we propose a novel deep learning architecture, the Stroke U-Net (SUNet), specifically designed for different acute stroke tasks such as segmentation and prediction. The architecture is an asymmetric encoder-decoder network with global and local residual connections that allows the use of a higher number of parameters without increased overfitting. Additionally, we review two recently proposed training patch sampling strategies originally designed for chronic stroke and brain lesions aiming to solve the data imbalance problem. We also combine the best of both into a hybrid strategy specifically designed for acute stroke tasks. The SUNet architecture and the reviewed patch sampling strategies are evaluated by cross-validation on three public ISLES challenge datasets, encompassing the tasks of stroke lesion and penumbra segmentation and outcome prediction. The SUNet architecture is compared against three baseline U-Net architectures, including one 2D model and two 3D models with and without residual connections, to assess the effect of different architectural design elements and its performance in acute stroke. The proposed methodology is evaluated by submission to the ISLES 2015, SISS and SPES, and 2017 challenge editions, using the provided web platform for an external and direct comparison to state-of-the-art methods. Furthermore, we make the development framework used for evaluation of the proposed and reviewed methods publicly available.

2. Methodology

2.1. Proposed architecture

Automated methods for acute stroke need to deal with a highly complex non-linear function between input intensities and output probabilities. From a deep learning point of view, the proposed architecture needs to have a large number of parameters to capture the complexity of the function. However, if the number of parameters is high enough, the network might have enough capability to *memorise* the whole training set without distilling meaningful and generalisable features. In this case, the network would have overfitted the training data and would generalise poorly for other images. The typically small amount of samples in public stroke datasets, due to the clinical nuances of acute stroke imaging, increases the risk of overfitting. To obtain the desired generalisability properties, a strong regularisation can be applied to the training images with data augmentation and during the training procedure with

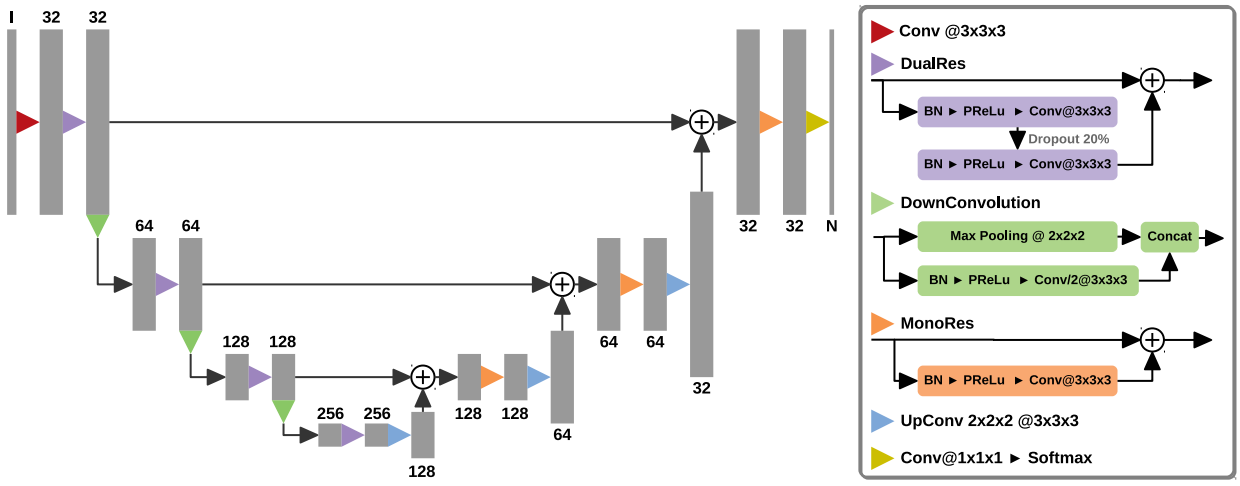


Figure 1: Proposed SUNet architecture with 3D convolutions, 4 resolution steps and 32 base filters. The architecture consists of an asymmetrical encoder-decoder network optimised for acute stroke tasks. The number of channels is indicated above or under each feature map. In the input and output feature maps, I and N denote the number of image modalities and segmentation classes respectively.

techniques such as dropout (Srivastava et al., 2014), adversarial training (Goodfellow et al., 2014) or ℓ_1 or ℓ_2 weight penalty. However, regularisation hinders network training and increases convergence times. For instance, Srivastava et al. (2014) noted that when using dropout typical training times were doubled or even tripled. Consequently, since strong regularisation techniques have to be used to promote generalisability, architectural features that ease convergence as much as possible are needed. The SUNet architecture, depicted in Figure 1, was designed based on these premises, using a high number of parameters to capture the high complexity of the task while maximising the representational power to promote generalisability.

The proposed architecture is an encoder-decoder network based on the original 2D U-Net (Ronneberger et al., 2015). More recently, the uResNet architecture (Guerrero et al., 2018), using short and long residual connections, achieved state-of-the-art performance with a fraction of the parameters in chronic stroke lesion segmentation. Furthermore, volumetric information for acute stroke is vital due to the wide spatial correlations involved in tissue differentiation or capturing the collateral blood supply for prediction. Hence, we also take design elements from the uResNet and 3D U-Net (Çiçek et al., 2016), which is the extension of the 2D U-Net made for volumetric dense biomedical segmentation. For the convolutional and transposed convolutional layers we only employ small 3×3 or $3 \times 3 \times 3$ kernels to improve the capacity of the network while using a lower number of parameters for the same receptive field (Simonyan and Zisserman, 2014).

The use of local residual connections (He et al., 2016), with the use of residual blocks, results in architectures that are easier to optimise and can gain accuracy from considerably increased depth. Residual connections ease backpropagation flow, improving convergence properties, and the identity mapping facilitates the learning of better fitted features. The work of Zagoruyko and Komodakis (2016) shows that local residual blocks using two convolutional layers are the most optimal in terms of parameters per residual connections. Furthermore, global residual connections with the use of summation skip connections have shown state-of-the-art results in chronic stroke lesion segmentation while using a reduced number of parameters (Guerrero et al., 2018). For the design of SUNet we use summation skip connection and residual blocks featuring two convolutions per residual connection. We explicitly avoid changing the number of channels inside residual blocks and instead perform this change in the downsampling and upsampling blocks. This avoids the use of a $1 \times 1 \times 1$ convolution in the identity path, that would increase parameter count, to adapt the number of channels for the summation connection.

Convolutional neural networks for real-time semantic segmentation improve inference times through the use of less parameters while keeping the same representational power. Since this paradigm is nearly identical to the one desired for stroke segmentation, we have successfully imported many design cues

from this kind of networks. The SegNet architecture (Badrinarayanan et al., 2017) achieved state-of-the-art performance in indoor and outdoor semantic object segmentation with a three fold reduction in inference times with respect to previous methods. The ENet architecture (Paszke et al., 2016) achieved comparable results to SegNet while using 98.7% less parameters, which allows for 20 times faster inference times. In the following, we summarise the most important design choices from this kind of networks for the SUNet architecture.

Asymmetric design. Typical U shaped architectures for biomedical image segmentation have symmetric encoder-decoder designs. However, it has been shown that the decoder's role is not as crucial as it may seem, mainly upsampling the work of the encoder and fine-tuning the details (Paszke et al., 2016). In other words, the amount of parameters in the decoder branch can be reduced without a decrease in representational power. For our design, we use residual blocks with two convolutional layers in the encoder and with a single convolutional for the decoder.

Information preserving dimensionality changes. We perform downsampling by concatenating the result of a $2 \times 2 \times 2$ max pooling operation in parallel with a convolutional using a stride of $2 \times 2 \times 2$ as proposed by Szegedy et al. (2015). This strategy avoids representational bottlenecks while keeping the number of parameters contained. Finally, upsampling in the decoder branch is performed with the use of transposed convolutions with a stride of $2 \times 2 \times 2$.

Non-linear projections. Instead of the more typical rectified linear unit (ReLU) (Nair and Hinton, 2010) we use a parametric version, the PReLU non-linearity (He et al., 2015), in our residual blocks as suggested by Paszke et al. (2016). This parametric version of the ReLu learns the slope of the non-linear projection of the activation map at training time. In this way, more suited activation slopes for each feature can be found and irrelevant information can be filtered out quicker. This is especially beneficial in networks with limited depth that do not have enough room to adapt the features to the identity transfer function of the conventional ReLu.

2.2. Class imbalance

Class imbalance is an issue for training automatic lesion segmentation methods since these are typically only a small part of the brain. Using uniform patch sampling would result in a training set where only a small fraction of the patches present lesion voxels. Not providing enough examples of the appearance of lesions at training time would induce a bias towards the healthy class and reduce the performance of the network. A recent approach in brain lesion segmentation is the use of patch based networks combined with deliberate patch sampling strategies to have a balanced class representation in the training set (Wang et al., 2016; Havaei et al., 2017; Kamnitsas et al., 2017; Chen et al., 2017; Guerrero et al., 2018). Additionally, patch sampling strategies can be combined with the use of weighted loss functions, like the Generalised Dice Loss (Sudre et al., 2017) or the Generalised Wasserstein Dice Score (Fidon et al., 2018), or multi-step training (Havaei et al., 2017).

Another important factor influencing class imbalance is the patch size. Since there are much fewer lesion voxels than healthy ones, bigger patches tend to include more healthy class voxels and further worsen class imbalance in the training set. The effect of patch size is quite significant and the average overlap quickly degrades as bigger sizes are considered. Hence, the patch sizes need to make a compromise between providing a big enough receptive field for effective segmentation and a sufficiently small patch size so that class imbalance does not negatively affect performance. In this work, we use small patch sizes of $24 \times 24 \times 8$ for the 3D models, given the typically lower axial resolution of stroke datasets, and 48×48 for the 2D ones. These patch sizes were determined by empirical tests and make a good compromise between receptive field and the effect of class imbalance in the three considered datasets.

2.2.1. Training patch sampling

In this work, we review two recently proposed training patch sampling strategies for chronic stroke and brain lesions and evaluate how they perform in each of the three considered acute and sub-acute stroke tasks. More specifically, we review a balanced strategy (Kamnitsas et al., 2017) and a lesion centred strategy (Guerrero et al., 2018). In the balanced strategy, training patches are extracted with 50% probability of being centred either on lesion or healthy voxels. In contrast, the lesion centred strategy exclusively uses lesion patch sampling, where all training patches are extracted centred on a lesion voxel. Additionally, a random offset added to the sampling point avoids location bias, where a lesion voxel is always expected at the patch centre, while also providing some degree of data augmentation.

We find these strategies can be further optimised for acute stroke imaging by taking into account the anatomy and pathophysiolgy of stroke. In the acute case, a lesion centred strategy makes the implicit assumption that

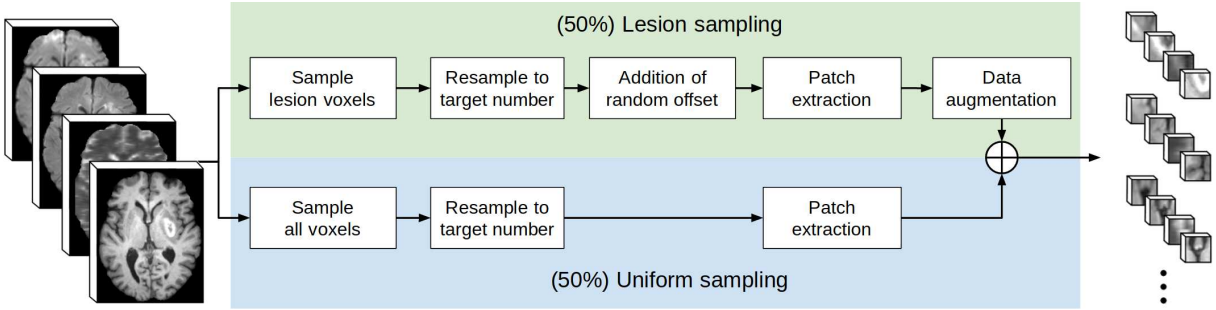


Figure 2: Diagram of the balanced with offset training patch sampling strategy for acute stroke related tasks that considers the anatomy and pathophysiology of stroke lesions. The size of extracted patches is either $24 \times 24 \times 8$ for 3D architectures or 48×48 for 2D ones.

the tissue surrounding the lesion is enough representation of the healthy class and that it can be generalised for segmentation of the rest of the brain. However, the tissue surrounding the lesion has altered vascular and structural properties, forming the area known as benign oligemia. This area is affected by the physiological reperfusion response of the brain by protraction of neighbouring vessels that induces an abnormal vascularity (Rekik et al., 2012). The balanced strategy implicitly assumes a binary classification problem with two tissue types, healthy or lesion. However, the anatomy of stroke lesions involves up to four different tissue types: healthy, benign oligemia, penumbra and core. We combine elements of both into a hybrid training patch sampling strategy, the balanced with offset strategy depicted in Figure 2, for acute stroke imaging tasks.

In the balanced with offset strategy, class imbalance is addressed in a similar way as in the balanced strategy where, for each patient, 50% of the training patches are extracted with uniform sampling and the other 50% with lesion sampling. We aim to have a balanced patch representation of each patient lesion by ensuring the same number of patches is extracted from each case. For patients with smaller lesions, a combination of several patch extractions from the same lesion voxel and data augmentation is done to ensure the desired number is reached. Since lesion voxels will have a random offset applied, the central voxel of the extracted patch will be different for each of the repetitions of a voxel. The voxels sampled from the lesion class have a random offset added of up to half of the patch size, to ensure the originally sampled voxel remains in the patch. This will increase the representation of the benign oligemia, the region surrounding the lesion, which would otherwise be an underrepresented part of the healthy class. The voxels are then resampled, removing or repeating them to reach the desired amount, at regular spatial steps to ensure that all parts of the brain are equally represented.

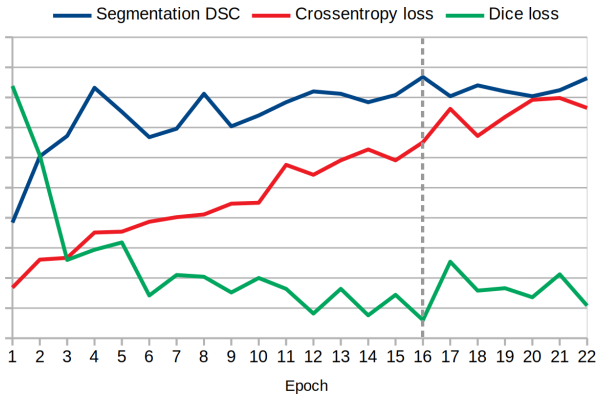


Figure 3: Evolution of segmentation DSC in the ISLES 2015 (SISS) dataset and the validation crossentropy and Dice loss as reported by the optimiser during training. The segmentation DSC was obtained by performing whole volume predictions on the validation set between each epoch. With an early stopping patience of 6 epochs, the dashed line marks the point where training would be interrupted. At this point, the highest segmentation DSC is achieved and the monitored loss for early stopping should be at its minimum.

Finally, patches are extracted centred on these voxels. Additionally, for the lesion sampled patches, data augmentation is applied with five anatomically feasible operations including sagittal reflections and 90°, 180° and 270° axial rotations.

2.3. *Network training*

For training of deep learning architectures in stroke related tasks, crossentropy is employed as the loss function and a *soft* Dice loss (Sudre et al., 2017) based on the Dice similarity coefficient (DSC) is used as the monitored metric for early stopping with a patience of 6 epochs. It was observed that, when training with a reduced amount of data, a lower validation crossentropy loss is not always correlated with a better segmentation overlap. However, the soft Dice loss is much more correlated and serves as a good metric to detect overfitting

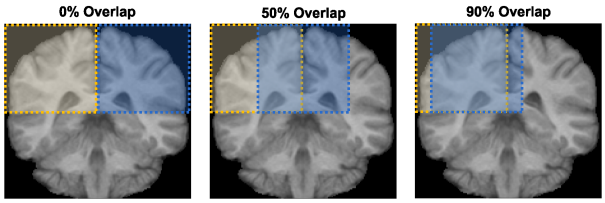


Figure 4: Examples of patches extracted with different overlap levels. Yellow and blue box respectively correspond to the first and second extracted patches.

and interrupt the training when using early stopping. Figure 3 shows the average segmentation overlap of validation cases with respect to the validation crossentropy and soft Dice loss metrics, the monitored ones for early stopping, as reported by the optimiser during training.

2.4. Segmentation and post-processing

To segment a volume with patch-based methods first the patches are extracted and forward passed through the net individually. The segmented patches are then combined in a common space with the same dimensions as the original volume, preserving their original spatial location, to produce the final segmentation. In our case, the combination is performed per voxel by averaging the class probabilities of the various segmentations. Patches are sampled uniformly with a regular extraction step of $6 \times 6 \times 1$ to make sure all the parts of the volume are forward passed through the network. In this way, some degree of overlap between the extracted patches is used, as depicted in Figure 4, since the extraction step is smaller than the patch size. Therefore, the same voxel is labelled seen in different neighbourhoods and the resulting class probabilities are averaged. This technique reduces the need for post-processing steps as it improves spatial label coherence and it minimises the number of small segmentation errors, i.e. holes, block boundary artefacts, etc. Furthermore, the employed extraction step combined with the full patch prediction offered by U-Net architectures means segmentation of a whole volume typically takes less than a minute.

Finally, the post-processing step involves a thresholding of the class probabilities followed by a connected component filtering by lesion volume. The variable threshold can compensate over/under confident networks while the minimum lesion size takes advantage of lesion priors to minimise false positives. In practice, the probability maps are binarised using an empirically computed threshold and minimum lesion size for each evaluation that maximises the desired metrics.

2.5. Implementation details

The proposed and reviewed methods have been implemented in Python, using the research oriented Keras high-level neural networks API (Chollet and Others, 2015). All experiments have been run on a GNU/Linux machine running Ubuntu 18.04 with 32GB of RAM memory and an Intel® Core™ i7-7800X CPU. The training of reviewed and proposed architectures has been done with an NVIDIA TITAN X GPU (NVIDIA corp, United States) with 12GB G5X memory. The proposed method is currently available for download[1].

3. Evaluation and results

3.1. Data

For evaluation of the proposed methodology we will use three public datasets from the 2015 and 2017 editions of the ISLES challenge. Each of these datasets represents a different task including penumbra and whole lesion segmentation, in the ISLES 2015 SISS and SPES dataset respectively, and lesion outcome prediction for the ISLES 2017 dataset, which is an extension of the 2016 with some additional cases.

ISLES 2015 (SISS). This subtask consisting on subacute lesion segmentation included 28 training images composed of 4 co-registered modalities including anatomical (T1, T2, FLAIR) and diffusion (DWI) MRI acquired in the first week after onset. All four provided MRI modalities are used for evaluation. The provided ground truth, the lesion extent, was manually segmented by an experienced medical doctor using both FLAIR and DWI images.

ISLES 2015 (SPES). This subtask focusing on penumbra segmentation included 30 training images composed of 7 co-registered modalities including anatomical (T1 contrast, T2), diffusion (DWI) and perfusion (CBF, CBV, TTP, Tmax) MRI acquired in the acute stage. All seven provided MRI modalities are used for evaluation. The whole lesion extent and the core had to be segmented in perfusion and diffusion images respectively to obtain the penumbra label, the gold standard (Maier et al., 2017). Manual segmentations were essentially generated by applying currently accepted linear thresholds on these two kinds of images. Despite the semi manual intervention and post processing, the semi systematic way in which the ground truth was generated by thresholding establishes a relatively simple numerical correlation between the intensities and output label.

[1] https://github.com/NIC-VICOROB/SUNet-architecture

ISLES 2017. The ISLES challenge 2016 and 2017 editions focused on prediction, consisting on segmentation of the chronic lesion extent from acute images. The 2017 dataset includes 43 training images composed of 6 co-registered modalities including diffusion (ADC) and perfusion (CBF, CBV, MTT, TTP, Tmax) MRI acquired in the first 24h after onset. All six provided MRI modalities are used for evaluation. An anatomical sequence (T2 or FLAIR) was acquired when the lesion had stabilised and the provided ground-truth, the chronic lesion extent, was manually drawn on those scans.

3.2. Analysis

3.2.1. Training patch sampling strategies

With the purpose of evaluating the performance of the reviewed patch sampling strategies for each of the considered tasks we implemented the lesion centred, balanced and balanced with offset strategies. These were then used to extract the training set patches for training the SUNet network depicted in Figure 1. To obtain the evaluation results, cross-validation is performed in 5 folds on all three datasets, adjusting the amount of cases per fold accordingly, with each strategy. The metrics used for evaluation of the strategies are the Dice similarity coefficient (DSC) (Dice, 1945), sensitivity, specificity and Hausdorff distance (HD). The DSC, defined in Equation (1), measures the relative overlap of the segmentation with the ground truth and is used as a measure of segmentation performance. The sensitivity and specificity, defined in Equation (2) and (3) measure the segmentation accuracy for the lesion and healthy classes respectively and are used to study the effects of different strategies for each class independently.

$$\text{DSC} = \frac{2 \cdot TP}{FN + FP + 2 \cdot TP} \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

where TP and FP denote the number of voxels correctly and incorrectly classified as lesion, respectively, and TN and FN denote the number of voxels correctly and incorrectly classified as non-lesion, respectively.

The HD, defined in Equation (4), can be intuitively seen as a measure of the distance of the *largest* error between the segmentation (A) and ground truth (B) and is a measure of segmentation quality and consistency.

$$\begin{aligned} &HD(A, B) = \max(h(A, B), h(B, A)), \\ &\text{where } h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \end{aligned} \tag{4}$$

3.2.2. Proposed architecture

With the purpose of evaluating the effects of different design elements, the proposed architecture is compared with three baseline U-Net based architectures using 2D patches and 3D patches with and without residual connections. The baseline models are based in networks for either chronic stroke lesion segmentation, the uResNet (Guerrero et al., 2018), or other related biomedical tasks, 2D U-Net (Ronneberger et al., 2015) and 3D U-Net (Çiçek et al., 2016). In the case of Guerrero et al. (2018) a 3D extension of the uResNet architecture is used. The SUNet model used for comparison is the one depicted in Figure 1, with 4 resolution steps and 32 base filters, to match the number of steps and features in the latent space of the implemented 3D U-Net and 3D uResNet architecture. The evaluation is performed by cross-validation in 5 folds on all three datasets, adjusting the amount of cases per fold accordingly. The models were trained with patches extracted according to the balanced with offset strategy, since it had a good overall performance in all three tasks. The metrics used for evaluation of the different architectures are the Dice Similarity Coefficient (DSC) and Hausdorff distance (HD) between the predicted segmentation and the provided ground truth to compare the overall segmentation performance and quality respectively.

3.2.3. Network training

All deep learning networks have been trained with the same training hyper-parameters to offer a fair comparison between them. To avoid costly grid search, the Adadelta optimiser (Zeiler, 2012) is used. This optimiser requires no manual tuning of a learning rate and appears robust to noisy gradient information, different model architecture choices, various data modalities and selection of hyper-parameters. A batch size of 32 is used, as smaller batch sizes help capture finer variabilities between the given samples. As stated in Section 2.2, we use patches of size $24 \times 24 \times 8$ for the 3D models and 48×48 for the 2D ones. A global goal of 250 000 patches for training is set for each cross-validation fold.

3.2.4. Challenge evaluation

With the purpose of evaluating the proposed methodology against state-of-the-art methods for acute stroke we submit it for external evaluation with the three challenges testing set. The web platform used to hold the ISLES 2015[2] and 2017[3] challenges remains open for

[2] https://www.smir.ch/ISLES/Start2015
[3] https://www.smir.ch/ISLES/Start2017

Table 1: Cross-validation metrics with different patch sampling strategies used in patch training set creation for the tasks of lesion/penumbra segmentation (SISS and SPES datasets) and lesion outcome prediction (ISLES 2017 dataset). The best results for each metric and dataset are highlighted in bold.

Dataset	Sampling strategy	DSC (%)	Sens. (%)	Spec. (%)	HD
ISLES 2015 (SISS)	Lesion centred	39.1 ± 29.3	53.6 ± 30.7	99.56 ± 0.48	78.8 ± 25.6
	Balanced	57.0 ± 25.4	57.2 ± 22.2	99.88 ± 0.19	**44.5 ± 31.9**
	Balanced w/ offset	**58.4 ± 25.4**	**59.4 ± 21.3**	**99.88 ± 0.17**	45.2 ± 33.5
ISLES 2015 (SPES)	Lesion centred	76.2 ± 14.6	**80.3 ± 17.0**	99.31 ± 0.97	19.2 ± 12.8
	Balanced	75.9 ± 17.3	76.6 ± 20.4	99.53 ± 0.66	**14.5 ± 9.4**
	Balanced w/ offset	**78.0 ± 17.3**	79.2 ± 18.5	**99.54 ± 0.66**	15.8 ± 10.9
ISLES 2017	Lesion centred	36.0 ± 23.2	47.7 ± 26.3	99.71 ± 0.35	28.7 ± 19.6
	Balanced	38.0 ± 23.5	51.5 ± 29.4	99.68 ± 0.45	22.8 ± 16.9
	Balanced w/ offset	**39.5 ± 21.6**	**54.0 ± 24.5**	**99.73 ± 0.30**	**21.0 ± 15.9**

later submission and maintains an ongoing challenge leaderboard where the testing set results are publicly displayed (Kistler et al., 2013; Maier et al., 2017). In this way, a fair and direct method comparison is possible. For evaluation in the challenge framework of ISLES we use the proposed training patch sampling strategy for training five models of the SUNet architecture, one for each cross-validation fold, with 4 resolution steps and 32 base filters for all three tasks. Then, an ensemble is made with the five models that averages the predicted probabilities of each to produce a single output. The cross-validation models used for the submission to the SPES and ISLES 2017 challenges are the ones trained for the comparison to baseline architectures. For the SISS challenge submission, an additional cross-validation was performed increasing the number of training patches to 700 000 and using a smaller batch size of 16, all other hyper-parameters were kept equal.

3.3. Training patch sampling results

Table 1 shows the evaluation results of the different strategies in each of the datasets. In general, the balanced with offset strategy shows consistent higher DSC values in all three datasets as compared with the other strategies. In all three tasks, the lesion centred strategy obtains the worst HD while the balanced strategies achieve comparable values.

Lesion/Penumbra segmentation. In the SISS and SPES dataset, the addition of a random offset to the balanced strategy improves the average DSC, mainly due to the better sensitivity, while slightly worsening HD values. In the SISS dataset, the lesion centred strategy obtains a much lower average overlap, sensitivity and specificity as compared with the balanced strategies. In the SPES dataset, the lesion centred strategy achieves the best sensitivity but also the worst specificity, resulting in a lower DSC as compared with the balanced with offset strategy.

Lesion outcome prediction. In the ISLES 2017 dataset, the addition of a random offset to the balanced strategy improves the results of all evaluated metrics. The lesion centred strategy obtains similar specificity values with respect to the balanced strategies but worse DSC, HD and sensitivity.

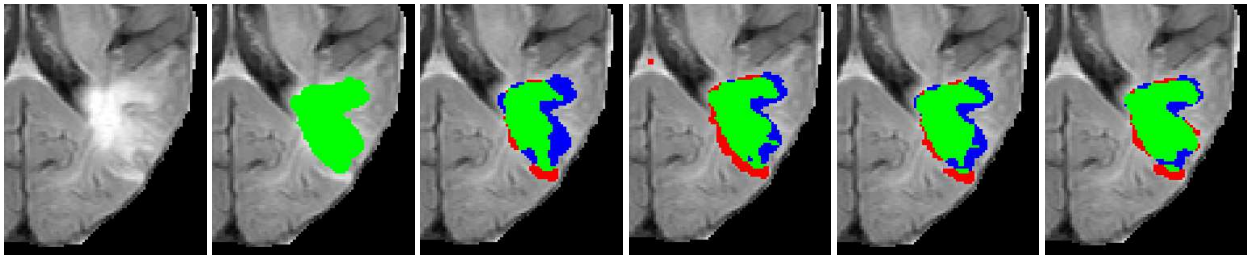
3.4. Proposed architecture results

Table 2 shows the evaluation results of the proposed and reviewed deep learning architectures. In general, the SUNet architecture has a higher average DSC in all three datasets with respect to the second best method.

Lesion/penumbra segmentation. The results show consistently improved average overlap of the SUNet architecture than the rest in both the SISS and SPES datasets. The other two 3D architectures, 3D U-Net and 3D uResNet, obtain really similar overlap and Haussdorf distance values in both tasks. Finally, the two dimensional U-Net architecture obtains a much better Haussdorf distance on both datasets as compared with the 3D architectures. In the SISS dataset, our architecture achieves the best average overlap while obtaining a similar Haussdorf distance to the rest. In the SPES dataset, SUNet obtains the best average overlap value with the lowest deviation as compared against the 3D architectures, the 3D U-Net and uResNet. Qualitative results of representative cases from the SISS and SPES dataset can be found on Figure 5 and 6 respectively.

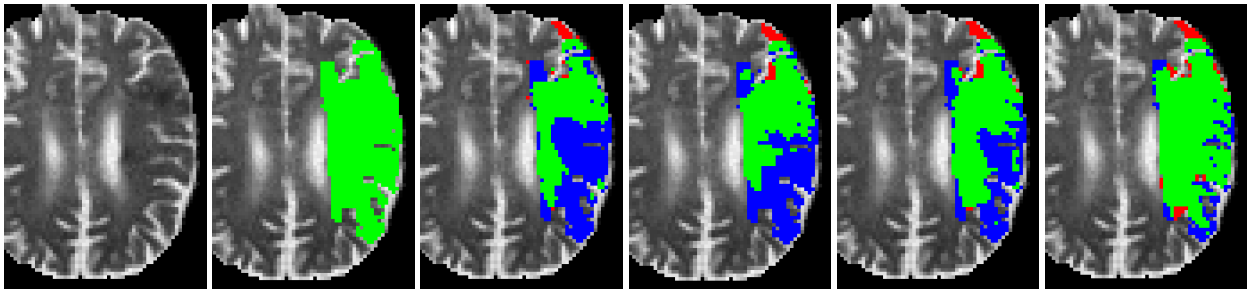
Table 2: Cross-validation metrics of all evaluated deep learning architectures for the tasks of lesion/penumbra segmentation (SISS and SPES datasets) and lesion outcome prediction (ISLES 2017 dataset). The best results for each metric and dataset are highlighted in bold.

	ISLES 2015 (SISS)		ISLES 2015 (SPES)		ISLES 2017	
Architecture	DSC (%)	HD	DSC (%)	HD	DSC (%)	HD
2D U-Net	60.1 ± 27.4	**32.2** ± **28.0**	78.2 ± 15.3	**11.7** ± **6.1**	33.6 ± 22.4	34.8 ± 24.2
3D U-Net	62.1 ± 25.5	38.1 ± 26.8	77.8 ± 17.8	12.2 ± 5.6	38.8 ± 23.0	26.0 ± 20.0
3D uResNet	61.5 ± 25.3	36.2 ± 27.6	77.6 ± 17.6	14.1 ± 10.3	38.7 ± 22.2	26.2 ± 20.9
SUNet	**65.0** ± **26.0**	35.1 ± 29.7	**78.5** ± **14.9**	16.0 ± 10.1	**40.1** ± **23.0**	**21.8** ± **17.0**



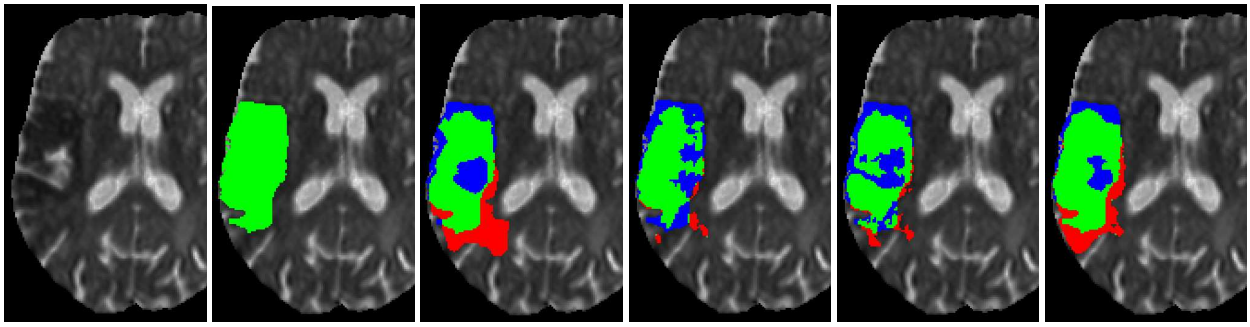
(a) FLAIR (b) Ground truth (c) 2D U-Net (d) 3D U-Net (e) 3D uResNet (f) SUNet

Figure 5: Output segmentation masks of sub-acute lesion extent from training case 11 from ISLES 2015 SISS dataset for each of the evaluated architectures. On all images, true positives are denoted in green, false positives in red and false negatives in blue.



(a) DWI (b) Ground truth (c) 2D U-Net (d) 3D U-Net (e) 3D uResNet (f) SUNet

Figure 6: Output segmentation masks of acute penumbra from training case 4 from ISLES 2015 SPES dataset for each of the evaluated architectures. On all images, true positives are denoted in green, false positives in red and false negatives in blue.



(a) ADC (b) Ground truth (c) 2D U-Net (d) 3D U-Net (e) 3D uResNet (f) SUNet

Figure 7: Output segmentation masks of the predicted chronic lesion extent from training case 12 from ISLES 2017 dataset for each of the evaluated architectures. On all images, true positives are denoted in green, false positives in red and false negatives in blue.

Table 3: Top entries ranked by average DSC in the ongoing testing leaderboard of the ISLES 2015 (SISS and SPES) and 2017 challenges. Our approach, highlighted in bold, ranks 6th out of 43 in SISS, 1st out of 19 in SPES and 2nd out of 32 in ISLES 2017.

ISLES 2015 (SISS)

Rank	Username	DSC	HD
1	kamnk1	0.59 ± 0.31	39.6 ± 30.7
2	zhanr6	0.58 ± 0.31	38.9 ± 35.3
3	lianl1	0.57 ± 0.29	43.0 ± 30.5
4	monkf1	0.56 ± 0.30	38.4 ± 29.4
5	wangj8	0.56 ± 0.30	39.6 ± 30.0
6	clera2 **(ours)**	0.55 ± 0.30	45.0 ± 30.1

ISLES 2015 (SPES)

Rank	Username	DSC	HD
1	clera2 **(ours)**	0.82 ± 0.09	23.9 ± 13.5
2	mckir1	0.82 ± 0.08	29.0 ± 16.3
3	maieo1	0.81 ± 0.09	23.6 ± 13.0
4	clera1	0.80 ± 0.10	25.7 ± 17.1
5	peres2	0.80 ± 0.11	22.2 ± 11.9

ISLES 2017

Rank	Username	DSC	HD
1	blumj1	0.32 ± 0.22	38.2 ± 18.7
2	clera2 **(ours)**	0.32 ± 0.22	36.8 ± 15.9
3	mokmc2	0.32 ± 0.23	40.7 ± 27.2
4	kwony1	0.31 ± 0.23	45.3 ± 21.0
5	pinta2	0.31 ± 0.21	35.7 ± 17.2

Lesion outcome prediction. Table 2 shows the evaluation metrics for the ISLES 2017 dataset while Figure 7 shows the segmentation results of an example case. The results show a better average overlap and Haussdorf distance of the SUNet architecture in the ISLES 2017 dataset than any of the other architectures. The two 3D architectures, 3D U-Net and uResNet, obtain really similar overlap and Haussdorf distance values. The two dimensional U-Net obtains much worse performance in this task than the other architectures in all evaluated metrics compared to the lesion/penumbra segmentation task, where it had similar and better values.

3.5. Challenge results

Table 3 shows the best entries at the time of writing this paper, ranked by average DSC, from the SISS, SPES and ISLES 2017 ongoing testing leaderboards. The proposed methodology ranks 6th out of 43 in the SISS challenge, 1st out of 19 in the SPES challenge and 2nd out of 32 in ISLES 2017. In the SPES challenge, our proposed methodology is the first based on deep learning to surpass the original challenge winners, all using RDF based methods (Maier et al., 2017).

4. Discussion

In this paper, we have presented a novel deep learning architecture tailored to acute stroke segmentation and prediction tasks. Additionally, three training patch sampling strategies have been reviewed to solve the problem of data imbalance with acute stroke lesions. The proposed methodology achieves state-of-the-art performance in acute and sub-acute stroke segmentation and prediction as shown in the results of the three submitted ISLES challenges. It is worth noting that we obtain these results in three different challenges while performing minimal changes to the training hyper-parameters and procedure. This performance is achieved thanks to the balanced with offset training patch sampling strategy, more suited for acute stroke lesions, and the higher parameter count of SUNet combined with its better implicit generalisation.

The lower rank obtained in the SISS challenge is due to the sub-acute nature of the images, acquired in the first week after onset. At this point, the lesion has largely stabilised and the regions of penumbra and benign oligemia are less relevant for lesion segmentation. In this case, the higher parameter count of SUNet, originally designed for acute stroke, combined with the lower number of relevant tissue types means the network is likely overfitting the data. Despite this, our method still achieves competitive performance and ranks 6th out of the 43 entries in the online leaderboard.

4.1. Training patch sampling strategy

The balanced with offset strategy shows a higher overlap, sensitivity and specificity with respect to the balanced strategy in all three tasks. This can only be due to the only difference between them, the addition of a random offset to the lesion sampled voxels before patch extraction. In practice, the added random offset increases the number of voxels from the region surrounding the lesion in the training set. In segmentation tasks, this means an increased representation of the benign oligemia is present that allows for learning improved lesion parts differentiation. In prediction tasks, the probability of spontaneous reperfusion, and hence positive outcome, is strongly correlated with the amount of collateral blood supply. The addition of the random offset increases the representation of adjacent regions and helps to better learn the correlation between these and lesion fate.

Lesion/penumbra segmentation. The addition of a random offset to the balanced strategy improves the DSC values but slightly increases HD in both segmentation tasks, which is due to a worse lesion border segmentation accuracy. However, it seems counter-intuitive that an increased representation of lesion border regions in the training set is causing worse performance in these same regions. With the more balanced representation of lesion parts, the training loss is better optimised by improving tissue differentiation than by ensuring more confident classification. The higher number of characterised tissues might cause spurious activations in border regions where the tissue has an ambiguous appearance.

The lower specificity values of the lesion centred strategy in the SISS and SPES dataset as compared with the balanced strategies suggests that the area surrounding the lesion is not enough representation of the healthy class. Solely sampling from lesion voxels results in a training set where most of the healthy class voxels correspond to the benign oligemia and few to truly healthy tissue with unaltered vascular properties. This means a part of the healthy class is under-represented in the training set and the network cannot learn to differentiate it, which causes a reduced specificity and worse overlap. We can conclude that some degree of healthy voxel sampling distant from the lesion is needed to have a complete representation of the healthy class in segmentation tasks.

Lesion outcome prediction. While in the SISS and SPES datasets the specificity obtained by the lesion centred strategy is quite worse than that of the other strategies, it achieves comparable values in the prediction task. This is due to a more varied representation of tissue under the lesion class in prediction images. Given the time passed since image acquisition and labelling, combined with the dynamic lesion evolution, solely sampling from lesion labelled voxels would still yield an increased representation of healthy tissue with unaltered vascularity. The lesion centred strategy obtains a much worse sensitivity compared with the balanced strategies. Exclusively sampling from lesion labelled voxels does not consider the cases where the tissue at risk was finally salvaged and not labelled as lesion in the chronic stage. Without examples from those cases, a bias towards fatal tissue fate is established. Consequently, a bigger number of false positives is obtained that lowers the sensitivity of the network. The HD, sensitivity and specificity improvement seen when a random offset is added to the balanced strategy would also be explained by a similar mechanism. Increasing the representation of areas surrounding the lesion also raises the number of examples of tissue at risk that was eventually salvaged. This offers a more balanced representation of tissue with good and bad outcome and allow the network to better learn correlations for tissue fate, achieving better overall performance.

4.2. Proposed architecture

In all three tasks, the SUNet achieves higher DSC values due to a combination of the higher parameter count and the implicit generalisation as a result of the network design. Despite the increased number of parameters of SUNet (7 million) with respect to 3D U-Net (5.5 million) and 3D uResNet (2.7 million) the average overlap is increased. Since the training procedure is the same for all architectures, the results suggest the SUNet architecture can use more parameters without an increase in overfitting that would lead to worse results.

Lesion/penumbra segmentation. The higher DSC obtained by the SUNet architecture in the SISS and SPES datasets shows more tissue overall is correctly identified. However, despite the overlap improvement, the slight increase in HD values in the SPES dataset suggests worse performance on segmentation of borders. This is a similar case to the one observed with the addition of a random offset to the balanced strategy. In this case, a bigger amount of captured correlations is making accurate border segmentation harder as the transition between tissue types is generating spurious activations.

The high similarity between the DSC values of all evaluated methods in the SPES dataset is most probably due to the semi systematic way in which the ground truth was generated by thresholding. In this way, a mainly linear numerical correlation is established between the intensities and output label that all architectures are able to approximate to a similar degree. Still, SUNet achieves a higher average DSC with less variability as compared with the other 3D architectures. Furthermore, in the qualitative results in Figure 6 it can be seen that, while all other architectures tend to undersegment the lesion, SUNet is able to segment the whole region with minimal false positives.

The 2D U-Net achieves the best Haussdorf distance in the SISS and SPES dataset probably due to the lower dimensionality of the patch, which allows it to find more robust and less confounding correlations. However, by not being able to take advantage of three dimensional information it cannot reach the higher overlap values achieved by the 3D architectures. This disadvantage can be clearly seen in the results of the SISS dataset

but not fully seen in the ones of SPES. Due to the semi systematic way in which the ground truth was generated in the SPES dataset, 2D correlations are enough and there is no information gain in the richer 3D correlations. Hence, despite the lower dimensionality neighbourhood, for this specific dataset the 2D U-Net is still able to obtain higher average overlap.

Lesion outcome prediction. In the ISLES 2017 dataset, the SUNet architecture has a higher overlap and lower Haussdorf distance as compared to the other three architectures. The increase in segmentation quality and consistency is probably due to the higher parameter count which allows for capturing a bigger number of correlations. This is especially important for lesion outcome prediction, due to the highly chaotic nature of stroke lesion evolution, that requires approximating a highly complex non-linear function with many different and subtle interactions. The high parameter count combined with the implicit generalisation capabilities of the network mean a bigger number of robust features can be distilled from the learned correlations resulting in more consistent segmentations. In the qualitative results (see Figure 7), it can be observed that the 3D U-Net and 3D uResNet tend to under-segment the lesion and produce irregular boundaries while the 2D U-Net offers more defined boundaries but clearly over-segments the bottom part of the lesion. In this case, the SUNet architecture is able to achieve both defined boundaries and a higher overlap with minimal over-segmentation at the bottom.

The 2D U-Net obtains the worst Haussdorf distance compared with the other evaluated architectures in the ISLES 2017 dataset while obtaining the best values in the segmentation tasks. This data suggests that 2D correlations are insufficient for lesion outcome prediction due to the spatial influence of the collateral blood supply in the 3D space around the lesion for tissue fate. The low DSC and HD values obtained by the two dimensional U-Net underline the importance of a 3D context for the task of lesion outcome prediction.

5. Conclusion

Acute stroke related tasks set up an inherently different machine learning problem than other brain lesions or even chronic stroke. The high complexity combined with the typically small datasets means stroke related tasks need to be approached with specialised methods that take into account the particular clinical and machine learning considerations. In this work, we explored a design principle for deep learning architectures dealing with acute stroke tasks and proposed a novel deep learning architecture, SUNet, based on these premises. Additionally, we have reviewed three training patch sampling strategies and evaluated their application for acute stroke tasks. We have proven that an increased representation of the surrounding region in the training set leads to better results in both segmentation and prediction tasks. The same proposed methodology performs well in three different acute stroke tasks with no tuning of training hyper-parameters. It achieves state-of-the-art performance and ranks among the top performing methods in three different ISLES challenge tasks dealing with segmentation and prediction tasks. Furthermore, we make the development framework used for evaluation of the proposed and reviewed methods publicly available. We strongly believe that the contribution in this work may be beneficial in future clinical scenarios to inform treatment decisions and improve patient outcome.

Acknowledgements

Jose Bernal holds an FI-DGR2017 grant from the Catalan Government with reference number 2017FI_B00476. 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

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (12), 2481–2495.

Campbell, B. C., Mitchell, P. J., Churilov, L., Keshtkaran, M., Hong, K.-S., Kleinig, T. J., Dewey, H. M., Yassi, N., Yan, B., Dowling, R. J., 2017. Endovascular thrombectomy for ischemic stroke increases disability-free survival, quality of life, and life expectancy and reduces cost. Frontiers in Neurology 8, 657.

Chen, L., Bentley, P., Rueckert, D., 2017. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. NeuroImage: Clinical 15, 633–643.

Chollet, F., Others, 2015. Keras. https://keras.io.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O., 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp. 424–432.

Dice, L. R., 1945. Measures of the Amount of Ecologic Association Between Species. Ecology 26 (3), 297–302.

Fidon, L., Li, W., Garcia-Peraza-Herrera, L. C., Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T., 2018. Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation Using Holistic Convolutional Networks. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, pp. 64–76.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. MIT Press, pp. 2672–2680.

Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., Wolz, R., Valdés-Hernández, M., Dickie, D., Wardlaw, J., Rueckert, D., 2018. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. NeuroImage: Clinical 17, 918–934.

Halme, H.-L., Korvenoja, A., Salli, E., 2015. ISLES (SISS) challenge 2015: Segmentation of stroke lesions using spatial normalization, Random Forest classification and contextual clustering. In: Proceedings of ISLES 2015 challenge. pp. 31–34.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis 35, 18–31.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

Ho, T. K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. Vol. 1. IEEE Comput. Soc. Press, pp. 278–282.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61–78.

Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P., 2013. The virtual skeleton database: an open access repository for biomedical research and collaboration. Journal of medical Internet research 15 (11), e245.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Maier, O., Menze, B. H., von der Gablentz, J., Häni, L., Heinrich, M. P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Götz, M., Haeck, T., Halme, H.-L., Havaei, M., Iftekharuddin, K. M., Jodoin, P.-M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochelle, H., Ledig, C., Lee, J.-H., Maes, F., Mahmood, Q., Maier-Hein, K. H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J. R., Reza, S. M., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.-W., Wilms, M., Kirschke, J. S., Krämer, U. M., Münte, T. F., Schramm, P., Wiest, R., Handels, H., Reyes, M., 2017. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis 35, 250–269.

Maier, O., Schröder, C., Forkert, N. D., Martinetz, T., Handels, H., 2015a. Classifiers for Ischemic Stroke Lesion Segmentation: A Comparison Study. PLOS ONE 10 (12), e0145118.

Maier, O., Wilms, M., Handels, H., 2015b. Random forests for acute stroke penumbra estimation. In: Proceedings of ISLES 2015 challenge. pp. 77–80.

Mckinley, R., Häni, L., Wiest, R., Reyes, M., 2015. Segmenting the ischemic penumbra: a spatial Random Forest approach with automatic threshold finding. In: Proceedings of ISLES 2015 challenge. pp. 69–73.

Nair, V., Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10). pp. 807–814.

Paszke, A., Chaurasia, A., Kim, S., Culurciello, E., 2016. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. arXiv preprint arXiv:1606.02147.

Redon, J., Olsen, M. H., Cooper, R. S., Zurriaga, O., Martinez-Beneito, M. A., Laurent, S., Cifkova, R., Coca, A., Mancia, G., 2011. Stroke mortality and trends from 1990 to 2006 in 39 countries from Europe and Central Asia: implications for control of high blood pressure. European Heart Journal 32 (11), 1424–1431.

Rekik, I., Allassonnière, S., Carpenter, T. K., Wardlaw, J. M., 2012. Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. NeuroImage: Clinical 1 (1), 164–178.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.

Sheth, S. A., Jahan, R., Gralla, J., Pereira, V. M., Nogueira, R. G., Levy, E. I., Zaidat, O. O., Saver, J. L., 2015. Time to endovascular reperfusion and degree of disability in acute stroke. Annals of Neurology 78 (4), 584–593.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

Singh, R. K., Chafale, V. A., Lalla, R. S., Panchal, K. C., Karapurkar, A. P., Khadilkar, S. V., Ojha, P. K., Godge, Y., Singh, R. K., Benny, R., 2017. Acute Ischemic Stroke Treatment Using Mechanical Thrombectomy: A Study of 137 Patients. Annals of Indian Academy of Neurology 20 (3), 211–216.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15, 1929–1958.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. S., Cardoso, M. J., Jorge Cardoso, M., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Lecture Notes in Computer Science 10553 LNCS, 240–248.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826.

Wang, Y., Katsaggelos, A. K., Wang, X., Parrish, T. B., 2016. A deep symmetry convnet for stroke lesion segmentation. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 111–115.

Winzeck, S., Hakim, A., McKinley, R., Pinto, J. A., Alves, V., Silva, C., Pisov, M., Krivov, E., Belyaev, M., Monteiro, M., Oliveira, A., Choi, Y., Paik, M. C., Kwon, Y., Lee, H., Kim, B. J., Won, J.-H., Islam, M., Ren, H., Robben, D., Suetens, P., Gong, E., Niu, Y., Xu, J., Pauly, J. M., Lucas, C., Heinrich, M. P., Rivera, L. C., Castillo, L. S., Daza, L. A., Beers, A. L., Arbelaezs, P., Maier, O., Chang, K., Brown, J. M., Kalpathy-Cramer, J., Zaharchuk, G., Wiest, R., Reyes, M., 2018. ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI. Frontiers in Neurology 9.

Zagoruyko, S., Komodakis, N., 2016. Wide Residual Networks. arXiv preprint arXiv:1605.07146.

Zeiler, M. D., 2012. ADADELTA: An Adaptive Learning Rate Method. arXiv preprint arXiv:1212.5701 abs/1212.5.