V-Net and U-Net for Ischemic Stroke Lesion Segmentation in a Small Dataset of Perfusion Data

Gustavo Retuci Pinheiro[1]([✉]), Raphael Voltoline[1], Mariana Bento[2], and Leticia Rittner[1]

[1] School of Electrical and Computing Engineering (FEEC), University of Campinas (UNICAMP), Campinas, SP, Brazil
gustavorp@dca.fee.unicamp.br
[2] Calgary Image Processing and Analysis Centre (CIPAC), Department of Radiology and Clinical Neuroscience, University of Calgary, Calgary, Canada

Abstract. Ischemic stroke is the result of an obstruction within a brain blood vessel, blocking the fresh blood flow, resulting in a tissue lesion. Early prediction of the ischemic stroke lesion region is important because it can help to choose the most suitable treatment. However, that is not trivial since current medical data, such as CT and MRI, have no explicit information about the future extension of the permanent lesion. A step towards efficiently using these data to predict the lesions is the use of Deep Convolutional Neural Networks as they are able to extract "hidden" information from the data when a reasonable labeled dataset is available and the deep networks are used properly. In order to try to extract this information, we have tested two different deep network architectures that are the state of the art in segmentation problems: V-net and U-net. In both networks, we tried different configurations, such as depth variations, pixel interpolations, MRI image combinations, among others. Experiments showed the following: normalizing the voxels sizes results in better training and predictions; deeper U-Net performs slightly better than the shallower U-Net, however it requires much more computation for only a small gain in accuracy; the inclusion of CT modality improved slightly the results; the use of only perfusion maps brought much better results than the use of raw perfusion data; smaller lesions are harder to detect properly.

Keywords: Ischemic stroke · Lesion segmentation · Deep learning · U-Net · V-Net · Perfusion

1 Introduction

Ischemic stroke occurs when there is an obstruction of a brain blood vessel, usually of small caliber, that irrigates the brain. This obstruction is named

microangiopathy and it causes a decrease or cessation of blood circulation, fast degeneration of brain tissue and tissue lesion. These brain lesions observed in ischemic stroke work as biomarkers of the disease, aiding in the diagnosis and treatment.

The most suitable imaging modality to detect and analyze these brain lesions is MRI because it presents an excellent contrast in soft issues, allowing the detection of subtle abnormalities in early stages of the disease [1]. Early prediction of the ischemic stroke lesion region is relevant to the early patient diagnosis and selection of the most suitable treatment strategy [2].

We developed an automatic approach for ischemic stroke lesion by using two deep net architectures that are the state of the art regarding medical image segmentation: V-Net [3] and U-Net [4].

This paper is organized as follows: The dataset is presented and explained at Sect. 2. The methods, including the architectures and parameters description, are described in Sect. 3. The experimental setting and results are presented in Sect. 4 and discussed in Sect. 5. Our conclusions are presented in Sect. 6.

2   Dataset

Two datasets were used in the development of this project, both from ISLES challenge: ISLES2017 and ISLES2018 [5,6].

In the first dataset (ISLES2017), the training set comprises data and ischemic lesion segmentation masks of 43 patients, data from another 32 patients with no ground truth was also available as testing set in the challenge. This dataset is composed of Apparent Diffusion Coefficient (ADC), Perfusion Weighted Images (PWI), and Perfusion maps, including Cerebral Blood Volume (CBV), Cerebral Blood Flow (CBF), Mean Transit Time (MTT), Time to Peak Concentration of the contrast agent (TTP), and the time need at which the residue function reaches its maximum value (Tmax).

The second dataset (ISLES2018) contains 63 patients (split in 94 cases/volumes) to train and 40 patients (split in 62 cases/volumes) to the test. Different from the ISLES2017 data, this data has no ADC and PWI, and it has the addition of CT Perfusion (CTP) data. Another difference is that the ISLES2018 dataset has kept only slices that contain lesions, thus, some subjects could have more than one slab to cover the lesion.

All the data were acquired during the ischemic stroke acute stage (within 8 h of the stroke). The ground-truth was manually drawn on $T2$ or FLAIR, when the stroke lesion had stabilized, for ISLES2017 and on DWI for ISLES2018.

Both datasets are provided in NIfTI format and already pre-processed with skull stripping, anonymization and co-registration for each subject individually.

3   Methods

The initial step in our proposed method is to create patches with 64×64 pixels. All the patches must contain lesions, at least partially (Fig. 1) in order to the dataset not be unbalanced.
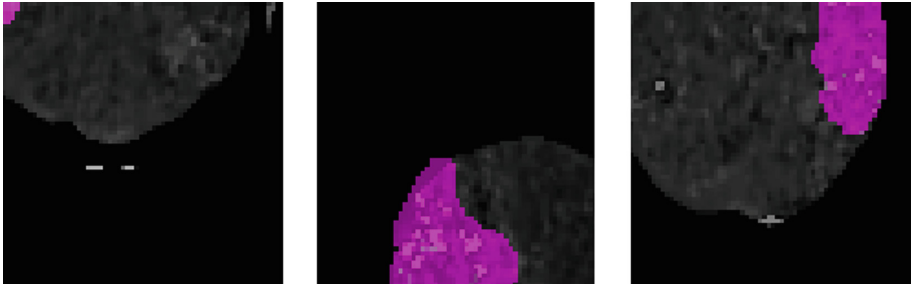
Fig. 1. Example of 64×64 pixels resolution patches; ground truth of the ischemic stroke lesion highlighted over Perfusion MTT map. (Color figure online)

After patches are ready, two deep networks were applied: V-Net and U-Net. V-Net (Fig. 2(a)) is a fully Convolutional Neural Networks (CNN) for volumetric medical image segmentation, therefore, it is originally a $3D$ deep net architecture. U-Net (Fig. 2(b)) is a fully CNN that was developed for biomedical image segmentation. In its original configuration, it works only in $2D$ images, requiring an independent prediction for every slice of the volume.

(a) V-net [3]   (b) U-net [4]
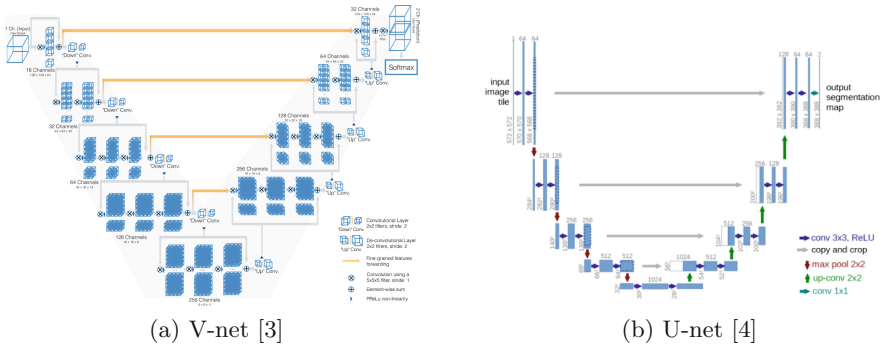
Fig. 2. Used deep neural network architectures for our proposed segmentation approach

Initially these CNNs were trained using a leave-out-out approach [7], in which the training dataset was randomly splitted into training and validation sets in a 80–20 ratio to avoid overfitting issues. After this initial experiment, we applied a different approach: k-fold cross validation, using k equal to 4. This change was made to have all the data in the training group, thus, increasing the accuracy in the test dataset. As only the prediction done in the test data was required to be submitted to the challenge platform (ISLES2018), the k-fold approach was done only in this data.

Another important step in the training was data augmentation. In addition to the patching, we have done flipping in the training patches in 50% of the

cases. This flip means that half of the times a patch and its respective mask enter in the train batch, they are horizontally mirrored, thus, inputting a "new" valid data in the training.

In both training methods, the used parameters were: optimizer RMSprop [8]; learning rate 0.0005, momentum 0.9; up to 300 epochs.

4   Experiments and Results

As shown in Sect. 2, the datasets have a variety of available image modalities, including CT and MRI. We have tested different combinations to achieve the best result. Each image modality or measure was a channel in the image for both CNNs.

In the ISLES2017 dataset, we have separately tested the networks with Perfusion Weighted Images (PWI) and Perfusion Maps. Since PWI is a $4D$ image that has up to 40 volumes, this dimension is taken as channels of the image in the CNNs. In the case of Perfusion Maps, the channels are each different measure (CBV, CBF, MTT, TTP, TMAX) plus ADC map.

In the ISLES2018 dataset, we had a similar approach, however the channels were CBF, MTT, CBV, TMAX, and CTP. We also tested the CNNs without the CTP.

The results show that the inclusion of CTP (Table 1) slightly improves the performance of the CNN. Differently, in the CNN trained using only PWI, the prediction were worse than the same CNN with Perfusion maps (Table 2).

Table 1. Comparison of CNNs performance for ISLES2018 dataset: Perfusion Maps only and Perfusion Maps plus CT Perfusion.

CNN	Data	$n°$ of channels	DICE
32U-NET	CBF, MTT, CBV, TMAX	4	0.324
17U-NET	CBF, MTT, CBV, TMAX	4	0.341
32U-NET	CBF, MTT, CBV, TMAX, CTP	5	0.351
17U-NET	CBF, MTT, CBV, TMAX, CTP	5	0.333

4.1   Architectures Variations

V-net and U-net were originally too deep to be used to analyze our patches with ischemic stroke lesions, since the pooling layers reduce the image size, completely eliminating it before reaching the middle layer. To overcome this issue, we have trimmed these CNNs by removing a few layers.

In this scenario, we used 2 different depth for each CNN: 32U-Net and 17U-Net; 10V-Net and 6V-Net, where the number refers to the amount of convolutional layer in the CNN. We also have add another dimension to the U-Net in order to have it in $3D$ form, allowing the direct comparison with V-Net.

The experiments with the CNN architectures (Table 2) showed that the $3D$ U-Net always outperforms the V-Net. We also can see that the $2D$ U-Net performs better than its $3D$ version and that the depth variation in the U-Net have only a small effect in the prediction accuracy.

Table 2. Comparison of architecture and data type combinations (ISLES2017 dataset): average DICE value for V-Net and U-Net on raw Perfusion images (PWI) and Perfusion Maps.

CNN	Data	$\overline{\text{DICE}}$	CNN	Data	$\overline{\text{DICE}}$
32U-NET 3D	PWI	0.359	**10V-NET 3D**	PWI	0.336
17U-NET 3D	PWI	0.357	**6V-NET 3D**	PWI	0.203
32U-NET 3D	P Maps	**0.526**	**10V-NET 3D**	P Maps	0.479
17U-NET 3D	P Maps	**0.518**	**6V-NET 3D**	P Maps	0.510
32U-NET 2D	PWI	0.414	-	-	-
17U-NET 2D	PWI	0.429	-	-	-
32U-NET 2D	P Maps	**0.609**	-	-	-
17U-NET 2D	P Maps	**0.560**	-	-	-

4.2 Voxel Interpolation

The dataset is not consistent regarding the voxel size, thus, the CNNs have to deal with different voxels resolution. One of our findings based on previous experiments is that the prediction results on testing images with the same voxel size as the majority of the train data is better than results achieved in testing images with different size. Our approach to minimize this effect was to normalize the size of the voxels among the whole dataset. By doing this, we have added another parameter to tune, but were able to improve the results.

We used a trilinear interpolation of the voxels to the sizes of $0.5 \times 0.5 \times 6$ mm, $1.0 \times 1.0 \times 6$ mm, $2.0 \times 2.0 \times 6$ mm, and $2.5 \times 2.5 \times 6$ mm. We have not tested variation of the Z axis because the datasets were more consistent in this dimension and about 6 mm in height.

The results (Table 3) showed that the voxel size of $2.5 \times 2.5 \times 6$ mm presented the best results. It is also shown, by the standard deviation, that normalizing the voxel size reduces the variability in the prediction quality.

4.3 Computational Environment

The experiments were performed using Python 3.6 and PyTorch on Jupyter Notebook. They were locally run on a machine with Intel I7 3.3 GHz processor, 8 GB RAM, and Nvidia GeForce GTX TITAN with 6 GB GDDR5.

Table 3. Comparison of different voxel size interpolations on the 17U-Net on ISLES2018 dataset: voxel interpolation size, average DICE, and standard deviation for the whole dataset.

Voxel size (mm)	CNN	$\overline{\text{DICE}}$	σ
Original size (no interpolation)	**17U-NET 2D**	0.341	0.272
$\mathbf{0.5 \times 0.5 \times 6}$	**17U-NET 2D**	0.369	0.254
$\mathbf{1.0 \times 1.0 \times 6}$	**17U-NET 2D**	0.438	0.237
$\mathbf{2.0 \times 2.0 \times 6}$	**17U-NET 2D**	0.448	0.254
$\mathbf{2.5 \times 2.5 \times 6}$	**17U-NET 2D**	**0.524**	0.247

4.4   Training Time

Different data combination and changes in the CNN architecture have a considerable influence in the training time (Table 4). For example, the 32U-Net $2D$ consumes more than twice the time for each epoch when compared to the 17U-Net $2D$. Moreover, the use of PWI instead Perfusion Maps increases the time by more than 5 times. The same CNN spends more than 10 times in the $3D$ version than in its $2D$ version.

Table 4. Epoch duration for each CNN architecture

CNN	Data	Time	CNN	Data	Time
32U-NET 3D	PWI	205.8	**10V-NET 3D**	PWI	265.7
17U-NET 3D	PWI	203.6	**6V-NET 3D**	PWI	238.8
32U-NET 3D	P Maps	42.3	**10V-NET 3D**	P Maps	83.1
17U-NET 3D	P Maps	34.2	**6V-NET 3D**	P Maps	55.5
32U-NET 2D	PWI	17.5	-	-	-
17U-NET 2D	PWI	13.4	-	-	-
32U-NET 2D	P Maps	7.0	-	-	-
17U-NET 2D	P Maps	3.3	-	-	-

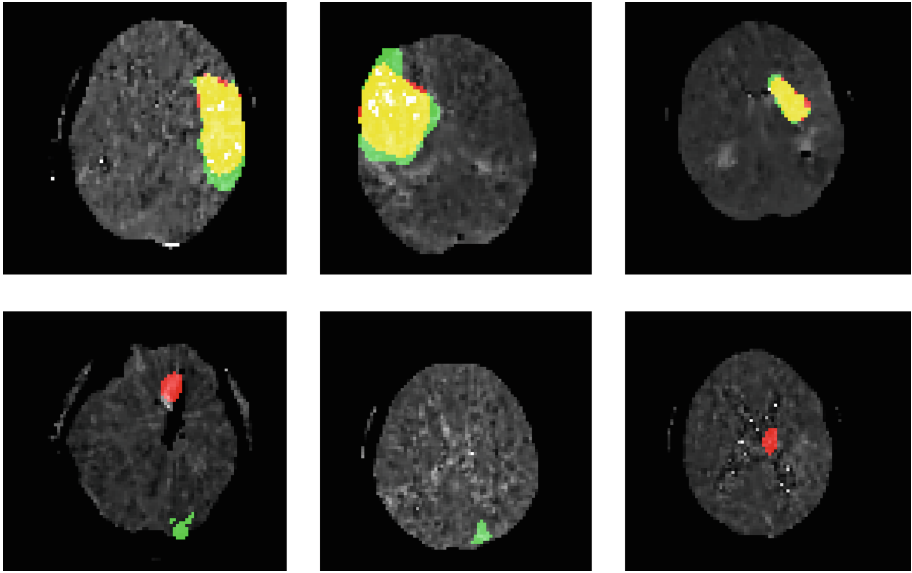## 4.5   Prediction

Since the networks are fully convolutional, the slices or volumes can simply be processed at once, independently of the size of the patch and the images used during training.

At this point we have defined the training method, the best data arrangement, the deep net architecture, the voxel size, and the prediction method, therefore, we are able to train the net to make the prediction on the test dataset.

We have defined the number of epochs to avoid overfitting by using the leave-one-out experimental approach. Then, we changed to k-fold training in order to increase the amount of data in the training while ensuring a non overfit, improving the generalization capability of the model.

With the k-fold model trained, we did the prediction on the training dataset to analyze the results qualitatively (Fig. 3). The prediction showed that for medium to larger lesion the CNNs are performing very well, with a slight tendency of overestimation. On the other hand, the predictor makes mistakes when segmenting small lesions, either in position or in extension.

Fig. 3. Example of Ischemic Stroke Predictions, in axial slices, on validation group: 3 best predictions (top) and 3 worst prediction (bottom); correct prediction (yellow), false negative (green), and false positive (red), all over Perfusion MTT map. (Color figure online)

4.6   Challenge Results

The ISLES2018 challenge had 24 participating teams that were ranked by averaging the segmentation rank for every subject. With the results achieve by our team (Table 5), we were in the 8th global position.

Table 5. Results of the segmentation done on test dataset (ISLES2018) by the selected model: 17U-Net 2D with 2.5 mm interpolation using k-fold training method.

Dice		Precision		Recall		AVD	
avg	σ	avg	σ	avg	σ	avg	σ
0.43	0.32	0.50	0.38	0.49	0.34	14.69	16.83

5   Discussion

The dataset is very complex when compared to other medical image segmentation problems, given the best achieved Dice was 0.51[1]. Besides, with the limited amount of data, we restricted our model in terms of complexity and depth. A larger dataset would allow us to train more complex or deeper models. And although data augmentation was applied and improved the results, there is a limitation on what can be achieved by using such techniques.

Another relevant finding is related with the normalization of voxels size. This step plays an important role in our segmentation solution. This probably has to deal with the scale that the convolutional windows analyze the data. For example, if the features that the filters are extracting from the image are textures, they may not be valid in a different scale, thus, confusing the predictor.

When discussing about the models architecture, our experiments had shown that $3D$ architectures requires much more computational power than $2D$ with no significant gain in the Dice coefficient, so at least for our segmentation solution is not recommended. Regarding the amount of convolutional layers in the networks, it was verified that shallower versions of the CNNs are comparable in performance and present an expressive gain in computational efficiency against the deeper version.

6   Conclusion

In this paper, we have proposed the investigation of the V-Net and the U-Net in the context of ischemic stroke lesion segmentation. We have concluded that U-Net on MRI Perfusion maps plus CT Perfusion and voxel normalization $(2.5 \times 2.5 \times 6 \,\text{mm})$ is the best combination to estimate the extension of the stroke lesion. However, the use of CT Perfusion must be further investigate in order to determine its role in the results.

The use of raw Perfusion Weighted Images led to poor results. As this data is too complex, there is a need for a much larger dataset in order to the CNN to be able to extract the necessary features. When we compute the Perfusion Maps from PWI, we need a simpler CNN because we are, analogously, already extracting and feeding the net with relevant features by classical methods.

---

[1] http://www.isles-challenge.org/.

Additionally, voxel size standardization is crucial to improve the performance of the predictor. Furthermore, downsampling the images has improved the performance of the trained model. If this step is not done, the CNN would also have to deal with scale.

Finally, U-Net always outperform V-Net in this particular problem. Even in the case where the U-Net is in the $3D$ for directly comparison, V-Net performs worse.

References

1. Vernooij, M.W., et al.: Incidental findings on brain MRI in the general population. N. Engl. J. Med. **357**(18), 1821–1828 (2007). https://doi.org/10.1056/NEJMoa070972
2. Ginsberg, M.D.: Neuroprotection for ischemic stroke: past, present and future. Neuropharmacology **55**, 363–389 (2008). PubMed
3. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of 2016 4th International Conference on 3D Vision, 3DV, pp. 565–571 (2016)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
5. Maier, O., et al.: ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. **35**, 250–269 (2017). https://doi.org/10.1016/j.media.2016.07.009
6. Kistler, M., et al.: The virtual skeleton database: an open access repository for biomedical research and collaboration. JMIR **15**(11), e245 (2013). https://doi.org/10.2196/jmir.2930
7. Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. ArXiv e-prints (2009)
8. Hinton, G., Srivastava, N., Swersky K.: Neural networks for machine learning. http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. Accessed 10 Oct 2018