White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning

Élodie Puybareau [1] Yongchao Xu [1,2,3] Joseph Chazalon [1]
Isabelle Bloch [2] Thierry Géraud [1]

elodie.puybareau@lrde.epita.fr

[1] EPITA Research and Development Laboratory (LRDE), Le Kremlin-Bicêtre, France
[2] LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
[3] Huazhong University of Science and Technology, Wuhan, China

1 Context

2 Method

3 Experiments and results

4 Conclusion

1 Context

2 Method

3 Experiments and results

4 Conclusion

White matter hyperintensities (WMH)

WMH:

- are a manifestation of small vessel diseases,
- can be everywhere in white matter,
- play a key role in stroke, demantia and ageing.

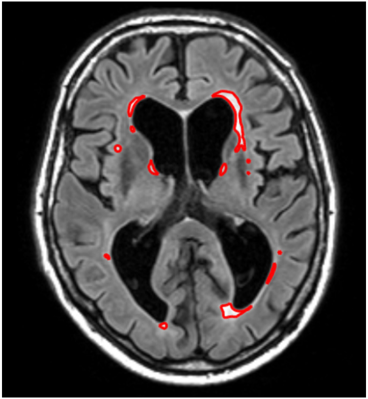
Importance of WMH study:

- analysis (shape, volume, location) is needed for clinical research studies,
- associated with clinical symptoms, can help prognosis, diagnosis, treatment monitoring etc.

Problem: manual segmentation is time-consuming and observer-dependent.

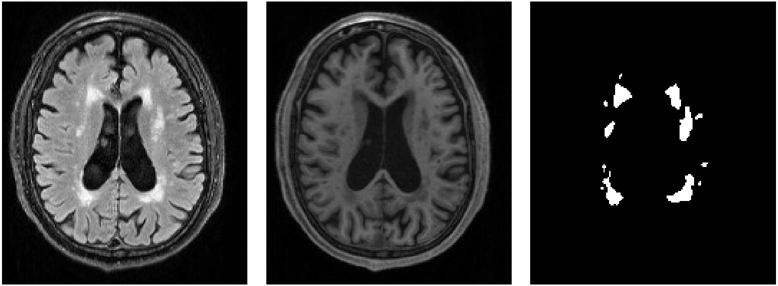
Challenge MICCAI

- Segmentation of WMH
- Part of Brain Lesion (BrainLes) MICCAI 2017 Workshop
- Method submitted in a Docker container
- Test data was not released



http://wmh.isi.uu.nl/

Data



(a) FLAIR image (b) T1 image (c) Ground Truth

Origin of the datasets

Institute	Scanner	Train	Test
UMC Utrecht	3T Philips Achieva	20	30
NUHS Singapore	3T Siemens TrioTim	20	30
VU Amsterdam (AMS)	3T GE Signa HDxt	20	30
	1.5T GE Signa HDxt	0	10
	3T Philips Ingenuity	0	10

Each volume is composed of about 45 slices used to generate the 2D input images for the network.

1 Context

2 Method

3 Experiments and results

4 Conclusion

Previous work

Our WMH segmentation is inspired by our previous work[1] on brain segmentation.

Reminder: VGG is a network:

- pretrained on ImageNet
  (database of hundreds of color natural images),
- dedicated to visual object detection in **2D color images**,
- including a *base* network.

[1] Y. Xu *et al.* From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning. In Proc. of IEEE Intl. Conf. on Image Processing (**ICIP**), pp. 4417–4421, Beijing, China, Sep **2017**.

Previous work

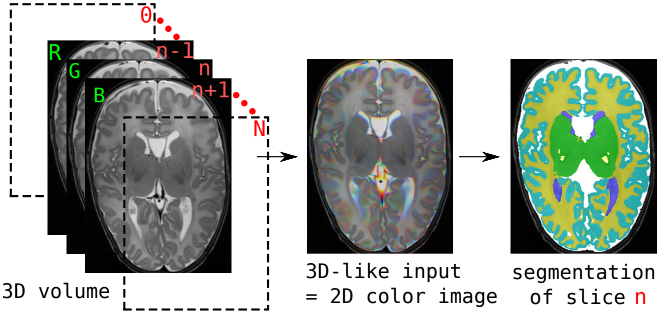
A segmentation method based on VGG-16:

1. **Preprocessing:**
   preparation of a set of 2D RGB images from a 3D volume
   $\rightsquigarrow$ pseudo-3D approach
2. **Learning:**
   transfer learning and modification of VGG-16 network
3. **Results:**
   inference on 2D color images, and reconstruction of 3D images

Key idea: A 2D color image encodes also 3D information.

Previous work

For each slice n do



Preprocessing: the idea

First experiment:

Use of the pseudo-3D approach with VGG-16 on WMH data.

Observation:

The network fails to detect small lesions.

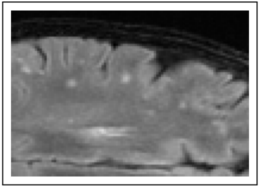
Idea:

Help the network by enhancing these small lesions in the input data.

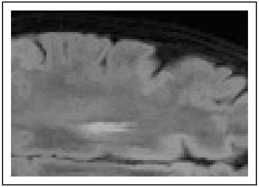
A morphological preprocessing

$$\textit{top-hat}(I) = I - \gamma(I),$$

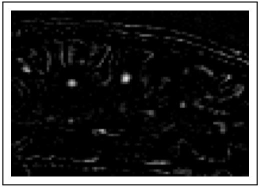
where $\gamma(I)$ is the morphological area opening of I.



FLAIR



opening of FLAIR



top-hat

We apply this procedure for each slice of a FLAIR volume.

A morphological preprocessing

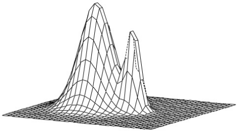
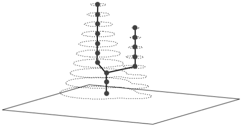


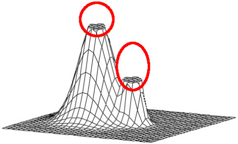
image I

$\equiv$



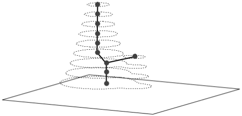
max-tree $\mathcal{T}_{\max}(I)$

$\downarrow$



opening $\gamma(I)$

$\equiv$

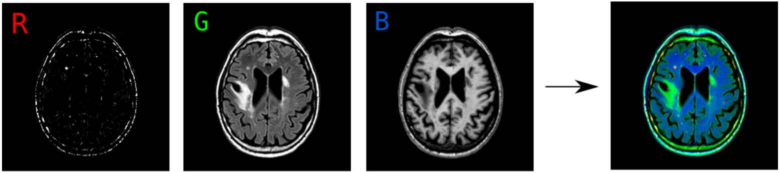


pruning of $\mathcal{T}_{\max}(I)$
based on component area

The top-hat is the **residue** (difference) between I and $\gamma(I)$.

Preprocessing: from 3D volumes to 2D images

For each slice n, do:

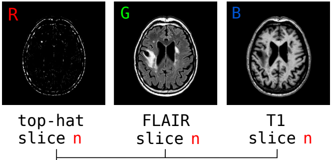


top-hat slice n | FLAIR slice n | T1 slice n | Input

2 gray 3D volumes (FLAIR + T1) + 1 generated (top-hat)
→ a set of RGB 2D images

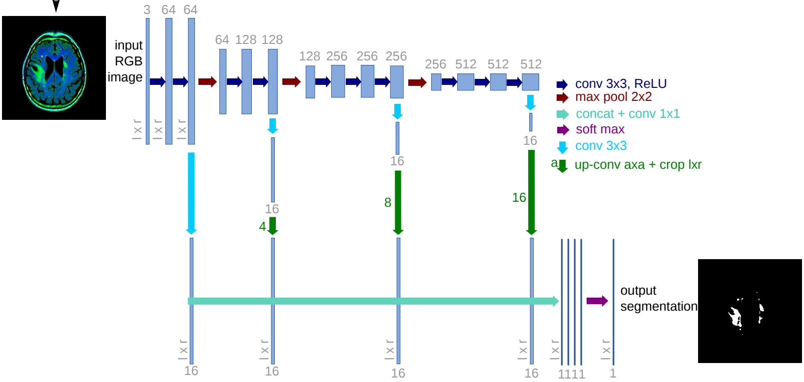
(It is not pseudo-3D anymore: each input 2D color image comes only from one slice.)

Network



← use of 2 modalities

+ small lesion enhancement



Parameters

- Total number of iterations: 150k
- Learning rate:
  10^{-8} for the first 50k iterations
  10^{-10} for the last 100k
- Momentum:
  0.99 for the first 50k iterations
  0.999 for the next 100k
- Weight decay: 0.0005
- 4 stages only

1 Context

2 Method

3 Experiments and results

4 Conclusion

Training phase: development

- 30 patients for training/30 patients for testing
  (10 from each hospital).
- Augmentation of training data
  (with scale variations and rotations).
- Input images:
  a series (3D volume) of 2D color images.

Training phase: for the challenge

- Model trained on all the 60 "expanded" patients.
- For each patient in the test dataset:
  pre-processing, centering, inference and reconstruction
  are fully automated
- Runtime on a 3D volume is less than **10 seconds** on average.

Evaluation

Dice: Dice coefficient

H95: Hausdorff distance (modified, 95th percentile)

AVD: Average volume difference

Recall: Sensitivity for individual lesions

F1: $\text{F1} = 2\,PR\,/\,(P + R)$,

where P and R are respectively the precision and recall for individual lesions:

P = true positives / (trues positives + false positives)

R = true positives / (trues positives + false negatives)

Validation of top-hat influence

Type	Dice ↑	AVD ↓	Recall ↑	F1 ↑
pseudo-3D	0.72	23.90	0.38	0.46
2D without top-hat	0.72	28.24	0.39	0.48
2D with top-hat	0.75	22.63	0.61	0.63

↑ means the higher the better / ↓ means the lower the better

Two conclusions:

- Pseudo-3D is useless here.
- Adding a morphological pre-processing gives much better results.

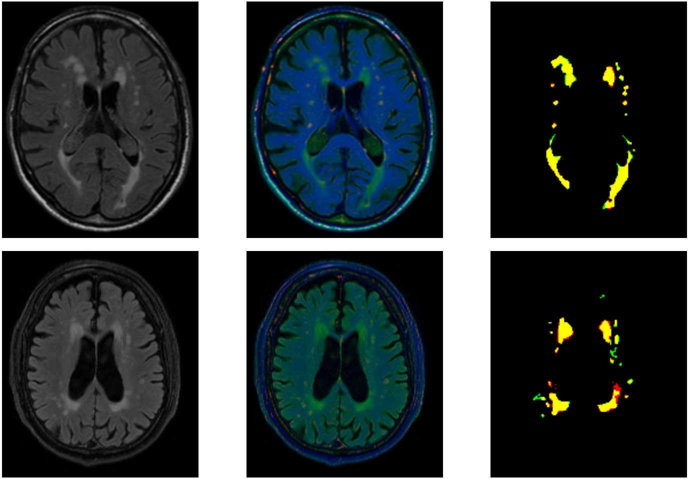
Quantitative results

Results of our method on the **challenge** dataset:

Origin	Dice ↑	H95 ↓	AVD ↓	Recall ↑	F1 ↑
UMC Utrecht	0.74	11.22	19.07	0.70	0.66
NUHS Singapore	0.77	8.28	17.64	0.61	0.68
AMS GE 3T	0.75	6.75	21.91	0.62	0.71
AMS GE 1.5T	0.73	10.94	16.66	0.60	0.71
AMS Philips 3T	0.50	70.27	46.33	0.57	0.53
Weighted average	**0.73**	**14.54**	**21.71**	**0.63**	**0.67**

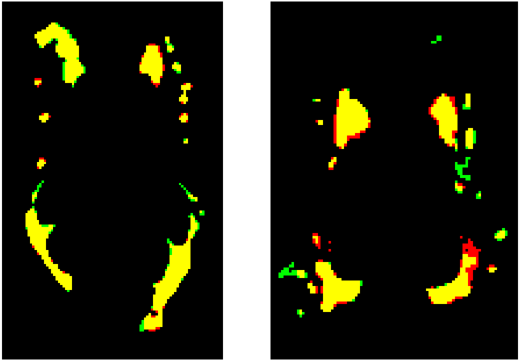
Rank: 6^{th} place of the challenge (among 21 competitors).

Some qualitative results



(d) FLAIR image (e) RGB input (f) Our result vs GT

Some qualitative results: zoomed in



Yellow: true positives; Red: false positives; Green: false negatives

1 Context

2 Method

3 Experiments and results

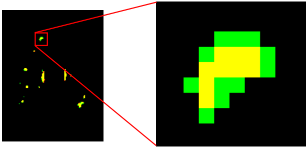
4 Conclusion

Conclusions

- Fast, robust and automated method for WMH segmentation:
  - inspired from our *pseudo-3D* approach (ICIP'17)
  - *transfer learning* works for some med. image segmentation tasks
- Segmentation of a 3D volume in less than 10 seconds:
  - Benefits from merging modalities in a *color* image...
  - ...and using a simple 2D network
- Effective benefits of morphological preprocessing:
  - highly non-linear
  - helps the network to identify objects of interest
- `Docker` container downloadable on our website:
  - *reproducible research* is important...

Perspectives

- Improvement possible thanks to post-processing?



- Application to other segmentations, pathological or not

- Going further with predictions? (prediction of tumor proliferation score for breast cancer, prediction of patient overall survival from the study of brain lesions, etc.)

.

The end

Supplementary materials and Docker file:
https://www.lrde.epita.fr/wiki/NeoBrainSeg

Thanks for your attention! Any questions ?

Results of the methods of the challenge

Sorted by increasing AVD:

Team	Dice ↑	H95 ↓	AVD ↓	Rec ↑	F1 ↑
nlp_logix	0.77	7.16	18.37	0.73	0.78
k2	0.77	9.79	19.08	0.59	0.70
ipmi-bern	0.69	9.72	19.92	0.44	0.57
misp	0.72	14.88	21.36	0.63	0.68
LRDE	**0.73**	**14.54**	**21.71**	**0.63**	**0.67**
...					
median	0.68	14.55	34.34	0.58	0.52
...					