Apport du Transfer Learning pour la segmentation automatique de lésions cérébrales radio-induites chez des patients atteints de glioblastome à partir d'un nombre restreint d'IRMs annotées

Thomas Coudert[1], Sophie Ancelet[2], Nadya Pyatigorskaya[3], Lucia Nichelli[3], Damien Ricard[4], Dimitri Psimaras[3], Marie-Odile Bernier[2], Michel Dojat[1], Florence Forbes[5], Alan Tucholka[6]

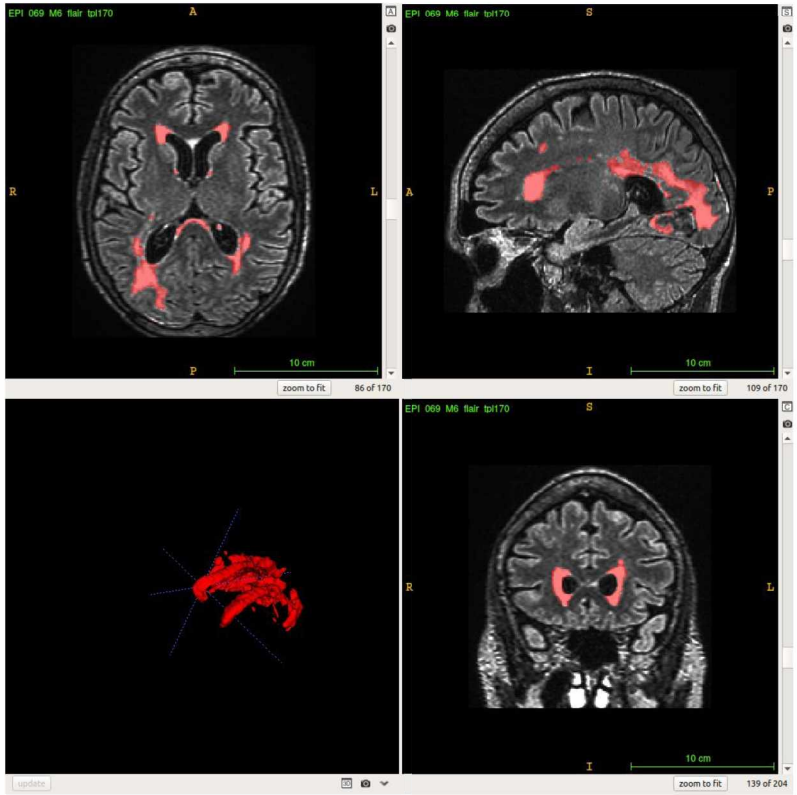
[1]INSERM U1216, GIN ;[2]IRSN, PSE-SANTE/SESANE/LEPID ;[3]AP-HP Pitié-Salpêtrière ;[4]HIA Percy & CentreGiovanni Borelli ;[5]INRIA Grenoble Rhône-Alpes, Statify ;[6]Pixyl

Journées de biostatistique 2021

Introduction

What is an MRI?

- Non invasive technique
- 3D acquisition
- Visualization of soft tissues (perfect for brain imaging)
- Each voxel (3D pixel) = an intensity value
- Different contrasts T1, T2, FLAIR...

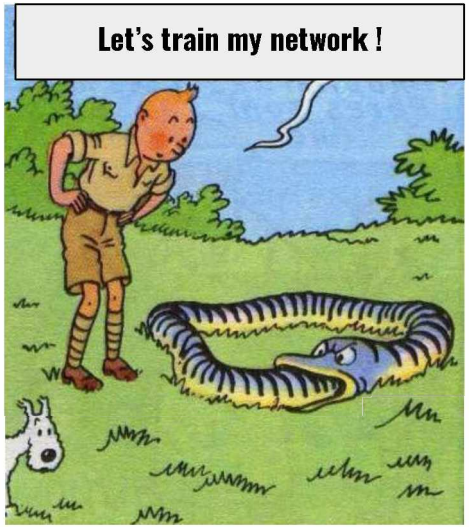


What is a segmentation?

- 3D labellisation of all voxel
- Binary values
- Red mask on the figure
- Volumetry, lesion detection, following of a pathology...

Introduction

- ➔ Importance of brain MRI pathologies **segmentation** during diagnosis
- ➔ Doing it **manually** and **from scratch** is very time consuming and shows bad repeatability
- ➔ Need of artificial intelligence (deep learning algorithms) to **predict this segmentation** ⇒ save several hours of work, could highlight something a specialist has missed

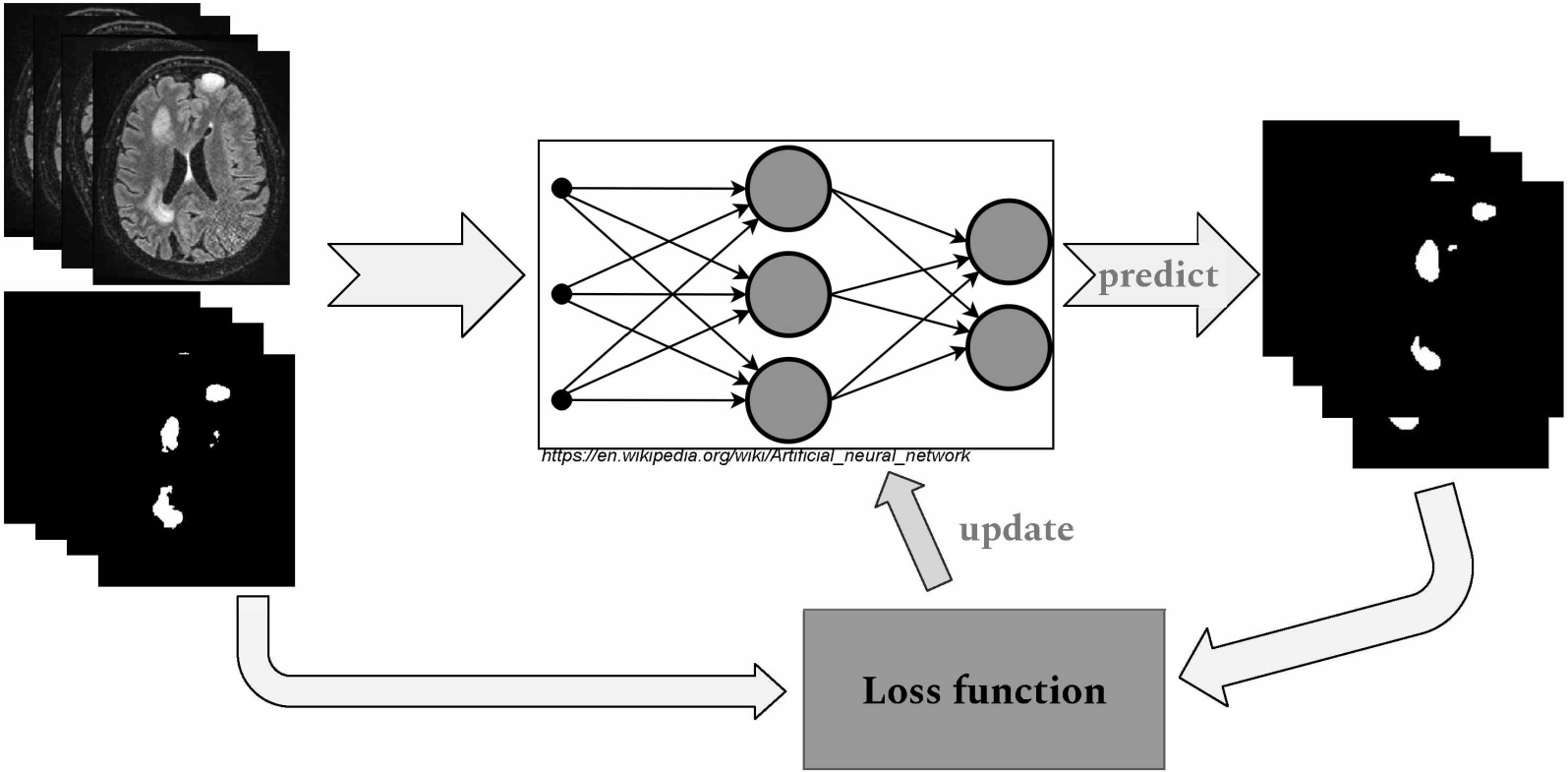


Problem : train supervised DL algorithms necessitates **already segmented** data to learn segmentation task ⇒ vicious circle.

⇒ Unsupervised learning : not enough efficient for now

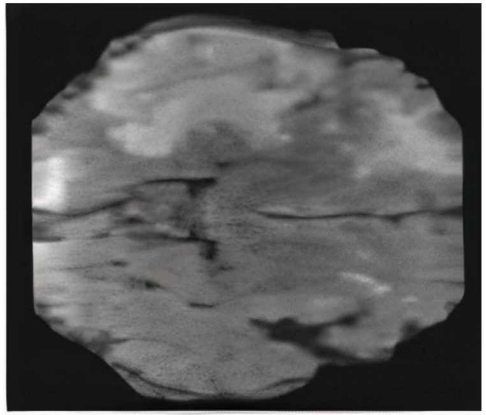
Introduction

Neural network reminders



Introduction *Prediction example*

Flair image

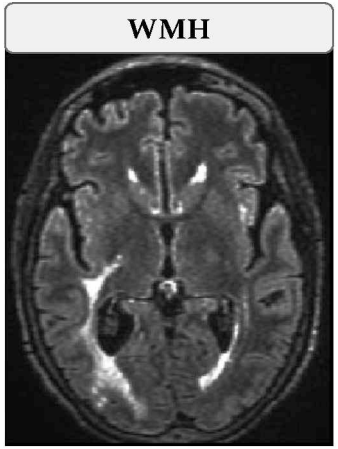


Ground Truth

Prediction

Introduction *RADIO-AIDE project*

High Grade Glioma (HGG) diagnostic → **Radiotherapy** treatment (RT) → Neurotoxicity, **radiation-induced leukoencephalopathy (RIL)** →



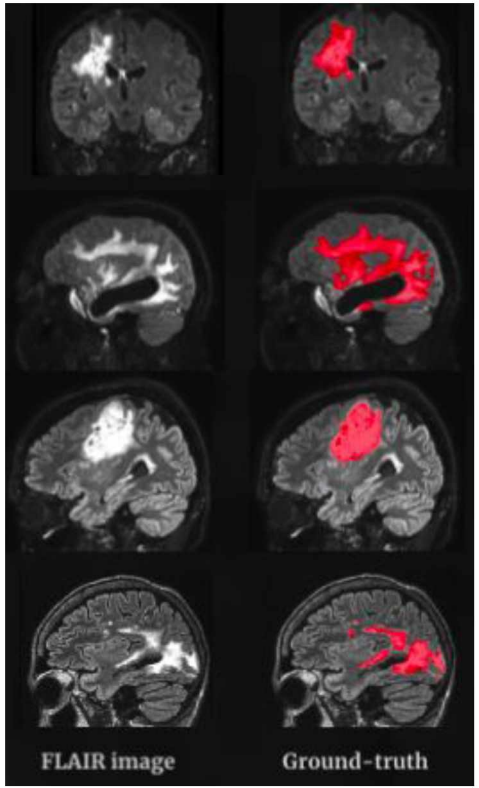
➔ following of this lesions is essential to evaluate the consequence of RT in the short and long term ⇒ segmentation by Artificial Intelligence in 3D FLAIR images where WMH are hyperintenses ⇒ evaluate and minimize the impact of RT

Introduction *EpibrainRad cohort*

EpiBrainRad[1] cohort : 224 patients newly-diagnosed for a HGG since 2015 and treated by RT (post-op data)

Multicentric data

Problem : actually only 20 image **3D FLAIR** segmented available : how to deal with such a small amount of data ?

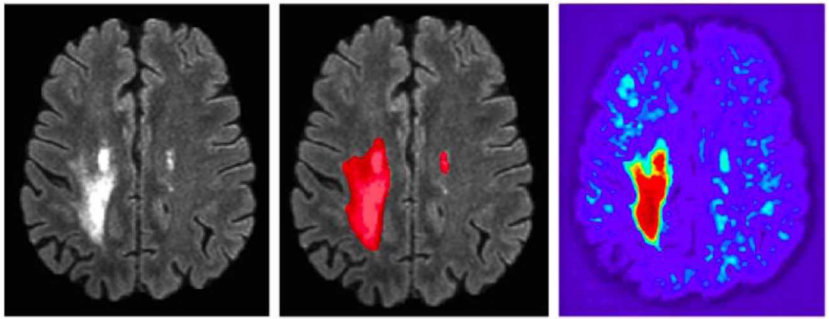


[1] T. Durand, S. Jacob, L. Lebouil, H. Douzane, P. Lestaevel, A. Rahimian, D. Psimaras, L. Feuvret, D. Leclercq, Bruno Brochet, R. Tamarat, F. Milliat, M. Benderitter, N. Vayatis, G. Noël, K. Hoang-Xuan, J.-Y. Delattre, D. Ricard, and M.-O. Bernier. **EpiBrainRad An epidemiologic study of the neurotoxicity induced by radiotherapy in high grade glioma patients**. BMC Neurology , 15(1):261, December 2015.

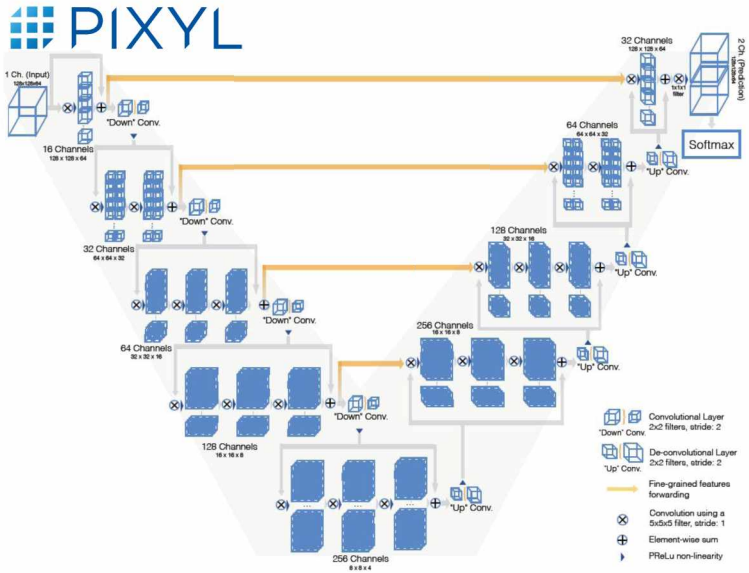
I. Material and methods

State of the art material :

- ➔ **3D V-Net network**[1] with attention gates[2]
- ➔ **Cross-Entropy** (CE) **Loss** and **Dice Loss** functions
- ➔ **Pre-processing** of the data : **skull-stripping**...



FLAIR image | Prediction | Attention map

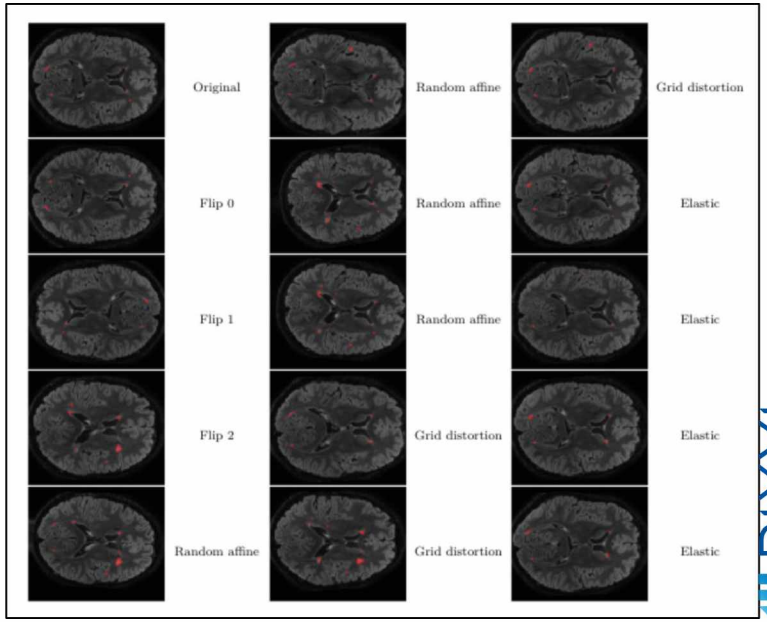


[1] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. **V-net: Fully convolutional neural networks for volumetric medical image segmentation.** CoRR, abs/1606.04797, 2016.

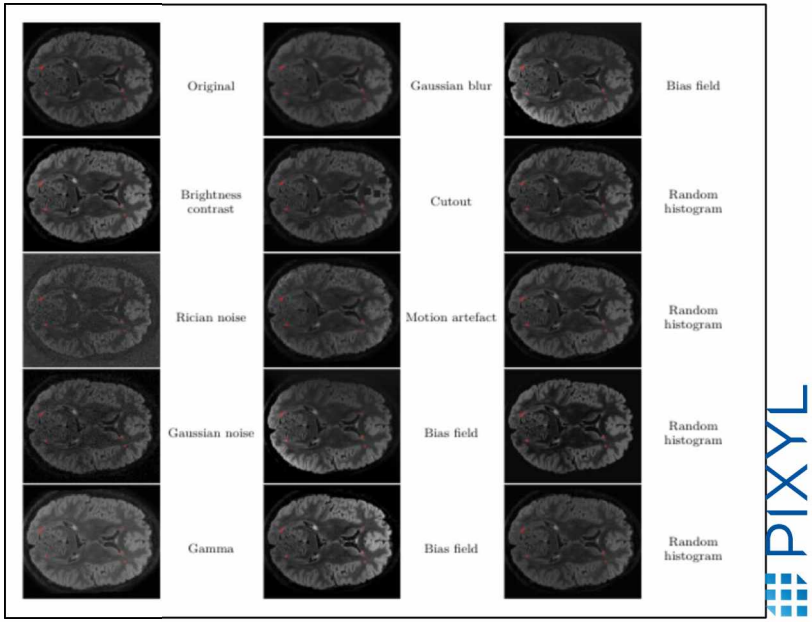
[2] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. **Attention u-net: Learning where to look for the pancreas**. arXiv preprint arXiv:1804.03999, 2018.

I. Material and methods

➔ random **data augmentation** during training (at each step of the training, the image is randomly transformed to bring variability and robustness)



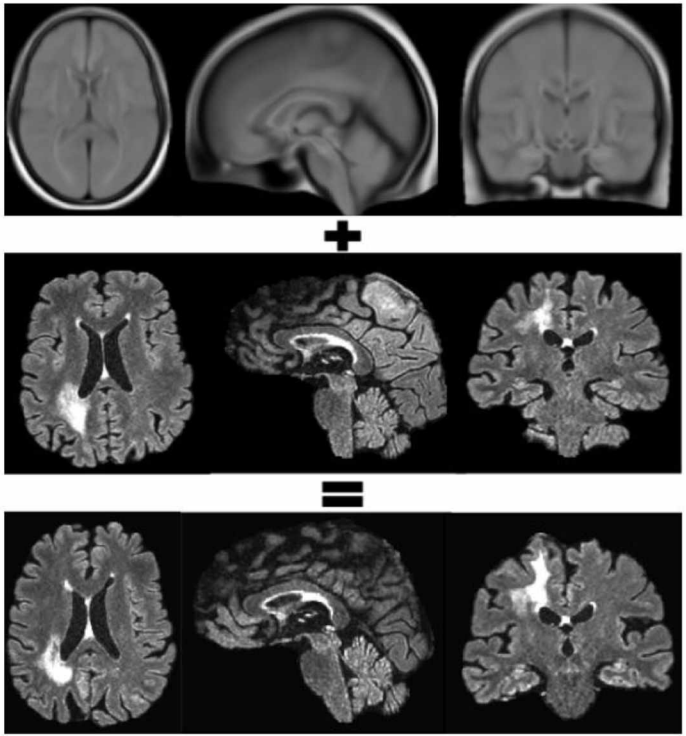
Spatial transformations



Intensity transformations

I. Material and methods

➔ **registration** on a MNI template



One image :
170 * 170 * 204 = almost 6 millions of voxel of size 1mm

I. Material and methods

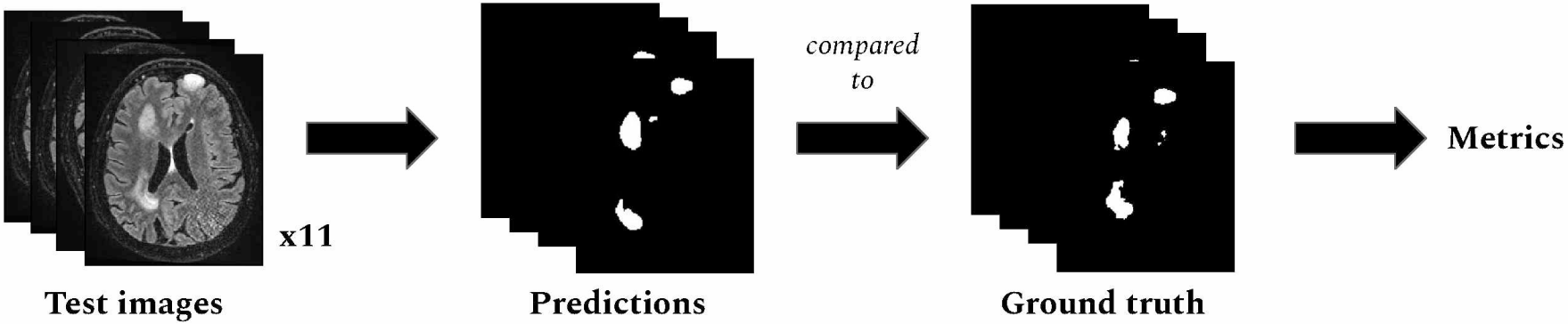
How to evaluate a model ?

STEP 1 : Use 9 3D EpiBrainrad images as training

STEP 2 : Generate 11 predictions for 11 other de-correlated 3D EpiBrainRad images

STEP 3 : Compare this 11 predictions with the 11 specialist segmentations associated

STEP 4 : Obtains metrics



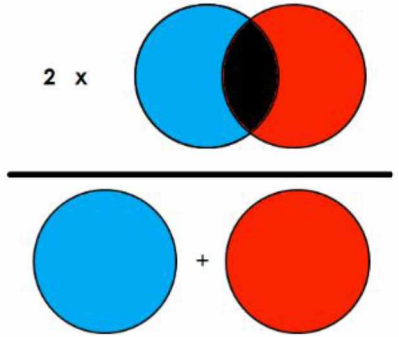
I. Material and methods

Metrics :

$$DSC = 2TP/(2TP + FP + FN)$$

$$Sensitivity = TP/(TP + FN)$$

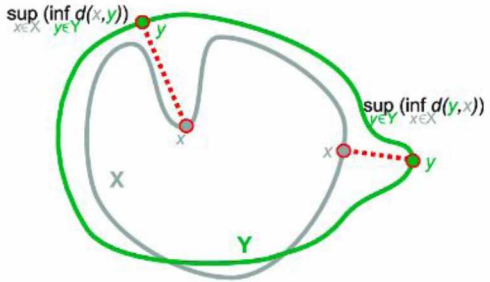
$$Precision = TP/(TP + FP)$$



https://towardsdatascience.com/

Hausdorff distance :

$$d_h(X,Y) = max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\}$$



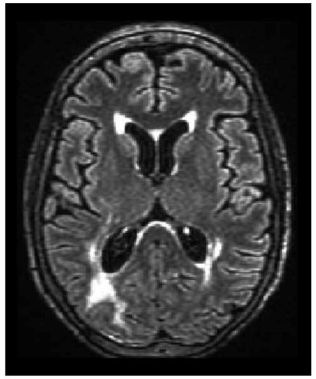
source: Jeanne Pellerin thesis available on research-gate

BUT : sometimes some segmentations look good qualitatively but has a few stray voxels
This few voxels totally disturb the computation of HD and increases it value
⇒ 95th percentile Hausdorff distance (**95HD**)

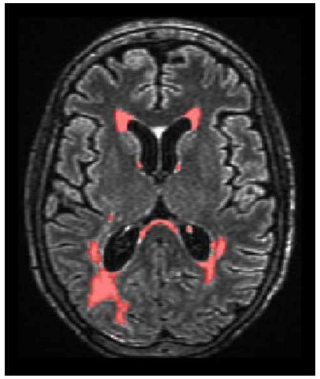
II. Preliminary results

Results when taking only 9 3D EpiBrainRad data as input of a neural network :

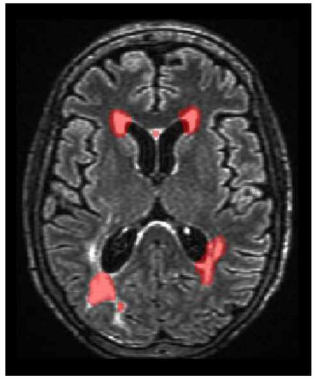
Dataset	WMH DSC	Sensitivity	Precision	95HD (mm)
9 EpiBrainRad LR=5e-3	0.557 ± 0.203	0.507 ± 0.229	0.703 ± 0.228	21.089 ± 15.469
9 EpiBrainRad LR=1e-5	0.705 ± 0.088	**0.869 ± 0.140**	0.624 ± 0.144	**6.838 ± 4.468**
9 EpiBrainRad dropout=0.3	**0.740 ± 0.127**	0.719 ± 0.182	**0.799 ± 0.112**	13.105 ± 17.237
9 EpiBrainRad max epochs set to 150	0.578 ± 0.257	0.578 ± 0.275	0.623 ± 0.258	19.848 ± 24.362



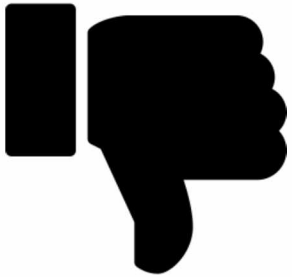
FLAIR



Ground truth



Prediction



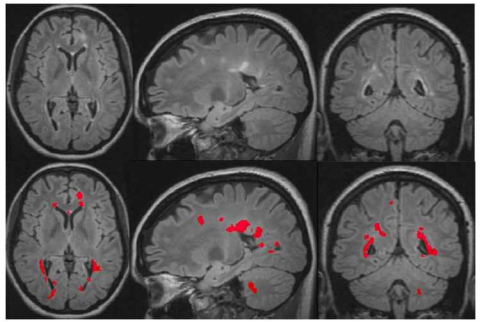
Next step ?

III. Use of pre-existent

First idea : create a mixing dataset containing some EpiBrainRad data put aside for training mixed with other data anatomically similar to WMH (e.g. hyperintense in FLAIR scan) as :

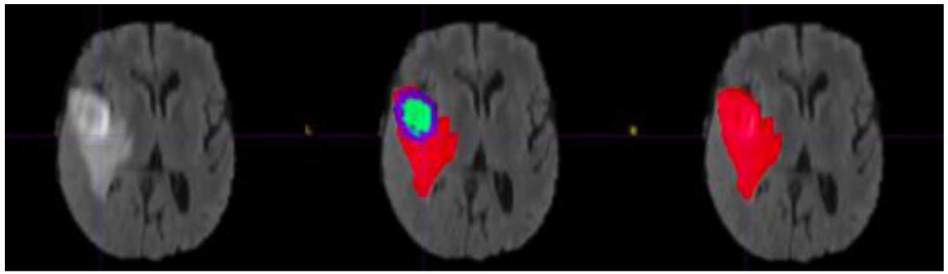
Multiple Sclerosis (**MS**) data from Pixyl database

- 200 3D FLAIR images
- Already binarized segmentation



Gliomas data from **BraTS Challenge**[1] dataset 2016 and 2017

- 484 3D FLAIR images
- Binarized for the purposes of the study

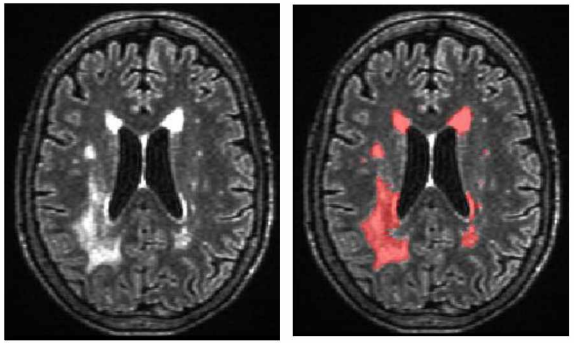


[1] B. H. Menze et al. **The multimodal brain tumor image segmentation benchmark** (**brats**). IEEE Transactions on Medical Imaging, 34(10):1993–2024, 2015.

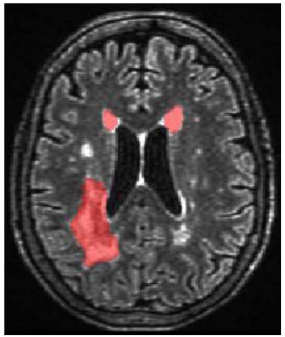
III. Use of pre-existent datasets

Inference is made on 11 EpiBrainRad data

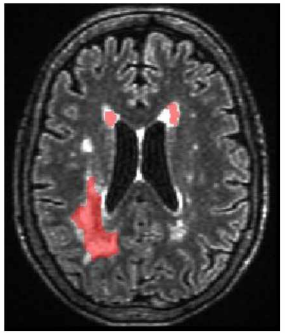
Dataset	WMH DSC	Sensitivity	Precision	95HD (mm)
200 BraTS only	**0.841 ± 0.073**	0.872 ± 0.101	0.817 ± 0.068	8.939 ± 10.250
200 BraTS + 5 EpiBrainRad	0.838 ± 0.071	0.866 ± 0.107	0.825 ± 0.089	**5.532 ± 4.703**
200 BraTS + 9 EpiBrainRad	0.836 ± 0.075	**0.879 ± 0.106**	0.803 ± 0.073	10.246 ± 12.439
200 MS only	0.662 ± 0.199	0.574 ± 0.246	**0.871 ± 0.108**	25.700 ± 20.958
200 MS + 5 EpiBrainRad	0.654 ± 0.098	0.593 ± 0.154	0.854 ± 0.098	16.362 ± 13.394
Preliminary results	0.740 ± 0.127	0.869 ± 0.140	0.799 ± 0.112	6.838 ± 4.468



Ground truth



BraTS + EpiBrainRad



MS + EpiBrainRad

- ➔ **High 95HD** = missed lesion
- ➔ **Low dice** for MS = under-segmentation so less TP

Be careful with metrics and always take a look

III. Use of pre-existent datasets

EpiBrainRad data are particular : **multiple small and big lesions**

➔ BraTS = unique big tumor (hyper-intensity)

➔ MS = multiple very small lesions

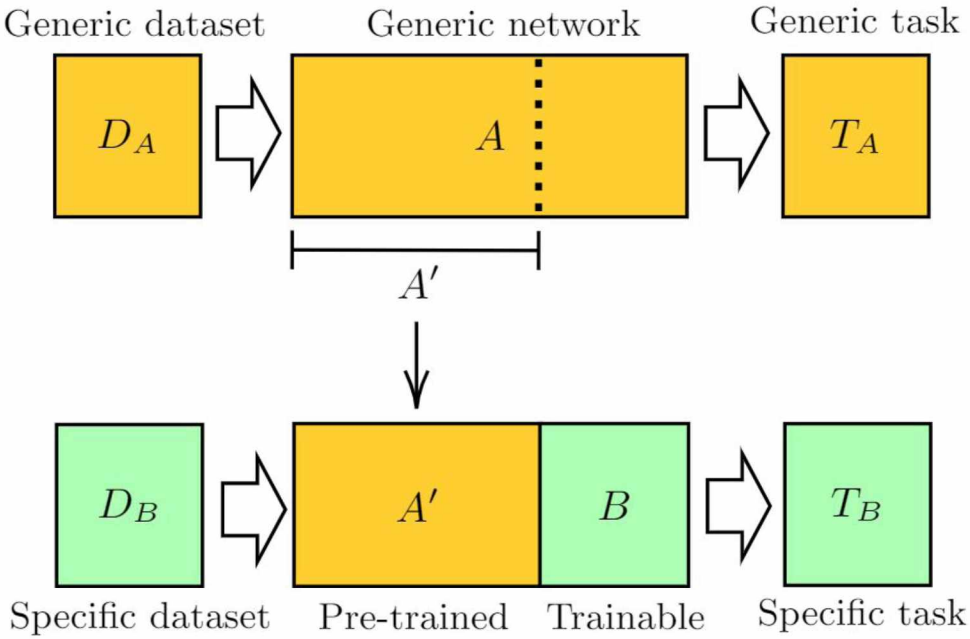
Only 5% of cohort of interest in the input dataset doesn't seem sufficient to transmit enough info to the network ...

Lack of data issue is still there.

<u>New idea : use of transfer learning</u>[1]

[1] Davood Karimi, Simon K. Warfield, and Ali Gholipour. **Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations.** Artificial Intelligence in Medicine, 116:102078, 2021.

IV. Transfer Learning

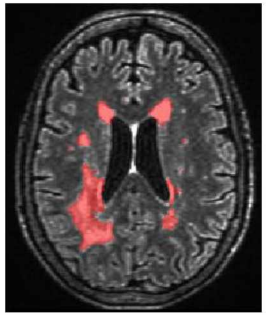


source : https://pennylane.ai/qml/demos/tutorial_quantum_transfer_learning.html

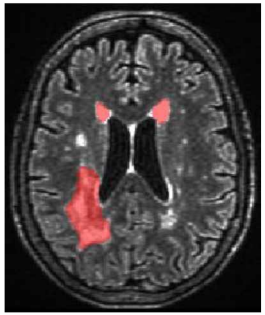
In our case : take weights of a pre-trained network and use them as initialization for weights of new network.

IV. Transfer Learning

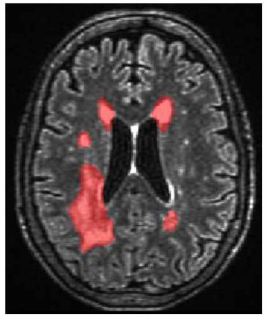
	Dataset distribution	WMH DSC	Sensitivity	Precision	95HD (mm)
without	200 BraTS + 9 EpiBrainRad into one training	0.836 ± 0.075	0.879 ± 0.106	0.803 ± 0.073	10.246 ± 12.439
with	200 BraTS pre-trained model, transferred on 9 EpiBrainRad	**0.858 ± 0.064**	**0.890 ± 0.104**	**0.835 ± 0.057**	**2.806 ± 3.523**



Ground Truth



Without TL



With TL

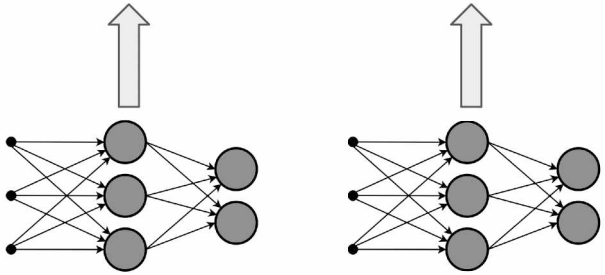
⇒ The increase of DSC isn't so high because of small over-segmentation, equilibrium between more FP and less FN. Little increase of TP number.
⇒ Hig decrease of Hausdorff Distance (95HD) ⇒ less missed lesions

⇒ Clear benefit of transfer learning, **best results with exactly the same amount of input images**.

IV. Transfer Learning

Some interesting models

Dataset (Pre-model/Trained model)		WMH DSC	Sensitivity	Precision	95HD (mm)
100 MS + 100 BraTS	9 EpiBrainRad	0.852 ± 0.048	0.915 ± 0.061	0.802 ± 0.067	2.517 ± 2.117
200 MS	9 EpiBrainRad	0.709 ± 0.217	0.679 ± 0.262	**0.855 ± 0.097**	7.372 ± 6.333
200 BraTS	9 EpiBrainRad	**0.865 ± 0.059**	0.895 ± 0.093	0.844 ± 0.065	2.504 ± 3.046
400 BraTS	9 EpiBrainRad	0.855 ± 0.056	**0.936 ± 0.065**	0.794 ± 0.085	**2.494 ± 2.268**

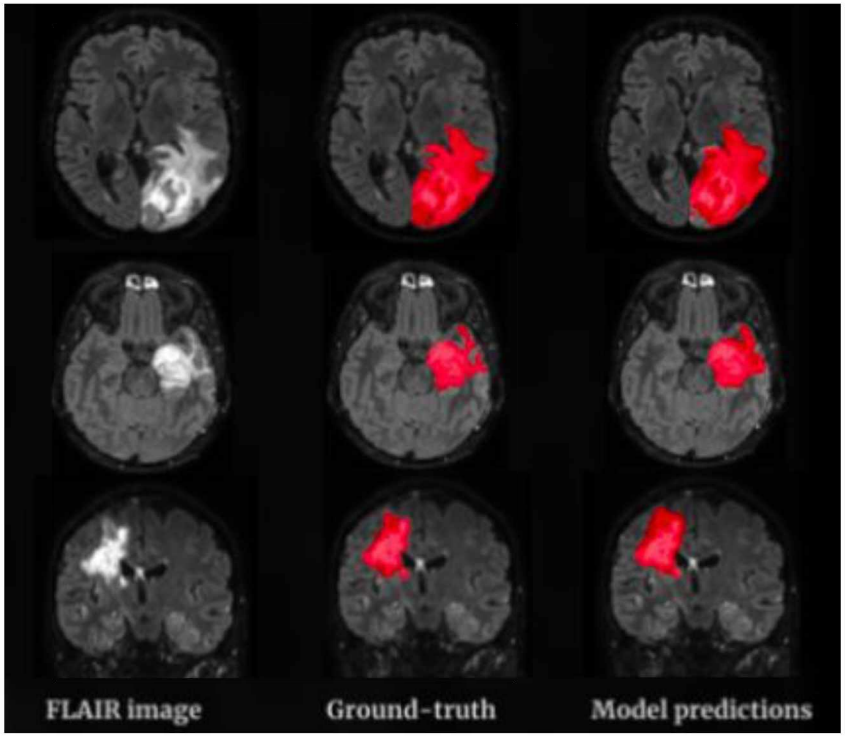


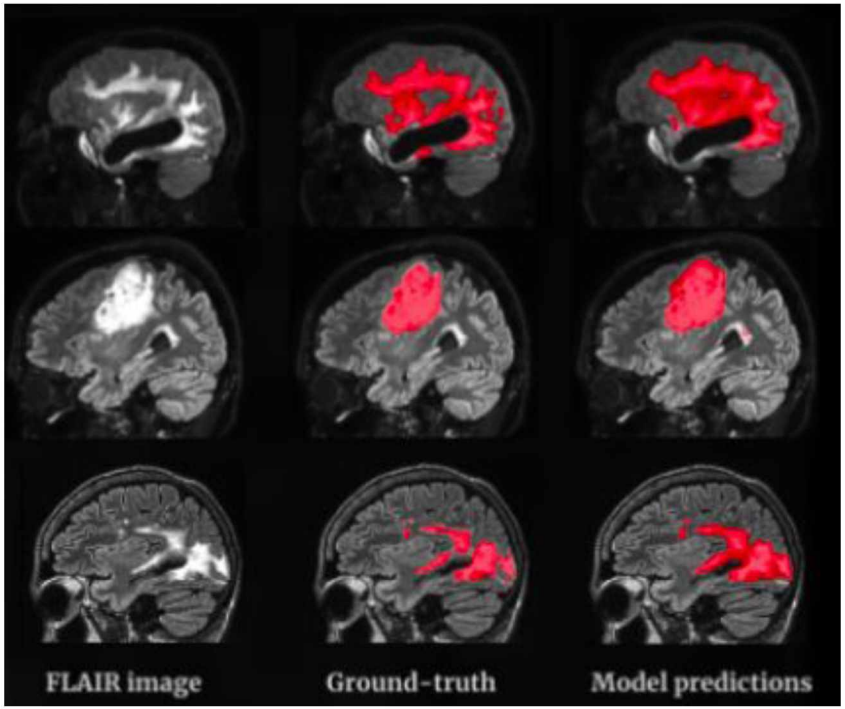
Pre-trained network with optimized weights

New network initialized with pre-trained weights

IV. Transfer Learning

Some interesting results





Conclusion

- ➔ Proved benefit of transfer learning methods
- ➔ Able to learn well with small amounts of data

Trained with **9 datas**, our best model is able to predict segmentations on a test set of 11 images with **13% of variability with a neurologist segmentation** on this 11 images.

⇒ could be interessant to evaluate the inter-clinician (or inter-rater) variability on this 11 images in order to quantify the "human like" performance VS our model

⇒ for example in BraTS challenge, the **inter-rater variability** is comprise between **10 and 20%**[1].

[1] B. H. Menze et al. **The multimodal brain tumor image segmentation benchmark (brats)**. IEEE Transactions on Medical Imaging, 34(10):1993–2024, 2015.

Prospects

➔ **Ensembling models** : several neural networks using different input datasets and presenting different predictions.
The predictions are then aggregated by a simple vote, where each model's say is ponderated by the uncertainty of this model (more the model is sure, more the given weight to his prediction is high).
It could allows to balance between benefits of MS and benefits of BraTS by ensembling the different obtained models.
Already tested on BraTS and showed good results[1].

➔ **Generating more data :** a transfer learning alternative
Use of generative network such as GAN to generate false data and false ground truth to expanded the input dataset[2]
<u>But</u> : it also needs a bit more data to train GAN efficiently

[1] Richard McKinley, Micheal Rebsamen, Katrin Daetwyler, Raphael Meier, Piotr Radojewski, and Roland Wiest. **Uncertainty-driven refinement of tumor-core segmentation using 3d-to-2d networks with label uncertainty**, 2020

[2] T. Neff, Christian Payer, D. Štern, and M. Urschler. **Generative adversarial network based synthesis for supervised medical image segmentation**. 2017.

Thank you for your attention !

Thomas Coudert, PhD student INSERM U1216, Grenoble Institut Neurosciences
thomas.coudert10@gmail.com
thomas.coudert1@grenoble-inp.org

Special thanks to : M.Dojat, F.Forbes A.Tucholka and S.Ancelet for their involvement