



White Matter Hyperintensities Segmentation with Prototype Learning

Trabajo fin de master

Máster en I.A. avanzada: fundamentos, métodos y aplicaciones



Autor: Óscar Alarcón Palomar

Directora: Olga C. Santos Martín

Curso académico 2019-2020 – Septiembre de 2020

ETSI Informática UNED

Departamento de Inteligencia Artificial

White Matter Hyperintensities Segmentation with Prototype Learning

Abstract

This paper proposes a new method -based on meta-learning- for the WMH Segmentation Challenge, organized by UMC Utrecht, VU Amsterdam, and NUHS Singapore hospitals. The purpose of this challenge is to compare methods for the semantic segmentation of white matter hyperintensities (WMH), which are brain white matter lesions, of presumably vascular origin in brain imaging obtained with magnetic resonance. White matter hyperintensities are found in patients with brain diseases like Parkinson, Alzheimer or stroke. Semantic segmentation refers to the process of linking each pixel in an image to a class label. The semantic segmentation of images has had a great advance with convolution neural networks, but they require a large number of images to be able to obtain good results. Convolutional neural networks are a type of neural networks specialized on images which architecture is similar to neurons' pattern in human brain and they were inspired by the organization of the visual cortex. With the aim to reduce the number of images required in training, in this work, we propose the use of meta-learning algorithms, in particular prototype learning, to do this semantic segmentation. In addition, this approach also allows the network to be used in a different task for which it was not trained, which can improve its potential use. Results obtained suggest that it could be possible to use the network trained to a specific task (i.e., detect WMH in the brain), to another task (i.e. detect any kind of tumors in the brain).

Terms

white matter hyperintensities, white matter lesions, meta learning, few-shot learning, prototype learning, semantic segmentation, convolutional neural network

I. Introduction

White matter hyperintensities (WMH) are brain white matter lesions (also known as leukoaraiosis). Usually found in a larger number of brain images from patients diagnosed with the small vessel disease (SVD) [1], multiple sclerosis [2], Parkinson [3], stroke [4], Alzheimer [5] and dementia [6]. WMH are associated with cognitive decline, tripling the risk of stroke and doubling the risk of dementia [10]. WMH do become more common with advancing age [11], but with a prevalence highly variable. Although, a large number of studies shows that they are implied in important risk factors: In a meta-analysis of 22 studies, WMH were associated with progressive cognitive impairment, increasing the risk of stroke and dementia [12]. WMH also generate abnormal gait [13] and disturbed balance [14] (related with physical function); and also increase the risk of late onset depression [15]. WMH were also highly heritable [16] and are less frequent in subjects with long-lived parents [17]. They are also inversely associated with intelligence in youth [18] and with the educational attainment [19]. If these associations reflect brain resilience to damage [20] are yet unknown.

WMH are seen on brain imaging. Magnetic Resonance Image (MRI) can be used to obtain them. MRI is a non-invasive diagnostic method that provides anatomic and functional images with high spatial resolution and contrast [21]. As a result, it is highly accurate in both tissue characterization and pathology detection, as stated in the paper referenced before [21].

Different types of MR images can be obtained by controlling the radio frequency pulse and gradient waves [22].

- T1-weighted: It provides complete anatomical information.
- T2 (T2-weighted): Useful in identifying pathological lesions that are often characterized by increased water content.
- Enhanced in proton intensity: It is used to characterize lesions of the white substance. It has been replaced for FLAIR type images.
- Fluid attenuation inversion recovery (FLAIR): It is a special inversion recovery sequence with a long inversion time. This removes signal from the cerebrospinal fluid in the resulting images. Brain tissue on FLAIR images appears similar to T2 weighted images with grey matter brighter than white matter but CSF is dark instead of bright.

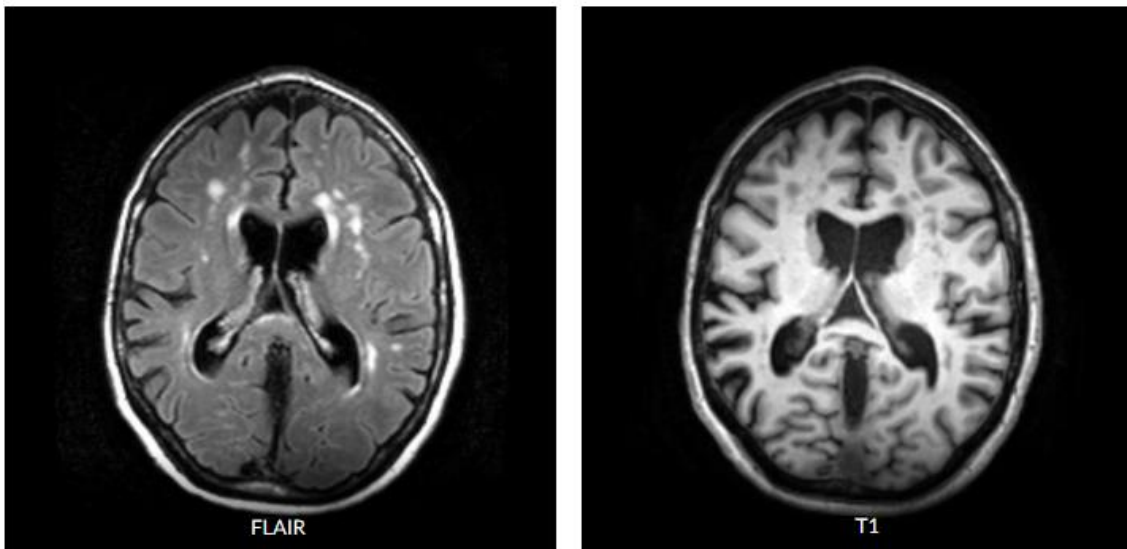


Figure 1. Example of FLAIR and T1 images taken from the WMH Challenge website.

The two basic types of MRI images are T1-weighted and T2-weighted images, often referred to as T1 and T2 images. Nevertheless, WMH can be better observed in FLAIR as high value signals [7]. WMH have a signal intensity brighter and with light boundaries compared to the surrounding white matter [8]. Unfortunately, in FLAIR images, grey matter can also be presented with high signal intensity. This leads to numerous false positives when trying to isolate WMHs. However, grey matter structures tend to be narrow; therefore, they have predominantly high spatial frequency components [8]. Cerebrospinal fluid is also bright in all images, but using FLAIR we can overcome it and maintain a good contrast between white matter and WMH lesions. There are also other challenges to overcome, including less contrast between grey matter and white matter; spatial variations in MRI caused by variable radiofrequency response; and background noise [8]. Thus, there is a need for a method to mark out of a WMH in the brain image.

Semantic segmentation does it: link each pixel in an image to a class label (WMH, brain, etc.). With semantic segmentation we could know which pixels of the image are the crane, or the brain, or a tumor or a white matter hyperintensity. The problem is that current segmentation methods using CNN needs a lot of images, but MR images are expensive because the MRI equipment are very expensive, and also there are no many free datasets.

Thus, our research goal is to find a method to let us do semantic segmentation using few images. Current methods in meta learning do image classification using few images (i.e. Siamese networks and Prototypical networks). In this paper, we research how to adapt these methods to do this classification classifying image's pixels.

This paper is divided in the following sections. First, we make a summary of the methods presented, and the current state of the art. In particular we, present the previous works in the segmentation of the magnetic resonance brain images, focusing on those carried out in the traditional methods with convolutional networks, to go on to comment on the newest ones carried out with few-shot learning. Then, we present the images and method used, making a summary of it. Finally, we present the results obtained and our conclusions about them, providing an overview of the improvement of our method with respect to the previous ones.

II. Related Work

We have looked up about 150 papers; several books on computer vision, looking for filters to extract features from images; online courses on machine learning, deep learning and convolutional neural networks in Coursera and Udacity, including the Stanford University course "CS231n: Convolutional Neural Networks for Visual Recognition".¹

This section summarizes the most important papers that attempt to solve the semantic segmentation problem, which is the most important subject in this work.

II.1. Semantic segmentation

Semantic segmentation is the process of dividing a digital image into several parts or objects. The objective of segmentation is to simplify and/or change the representation of an image into one that is more meaningful and easier to analyze.

The task of semantic segmentation is classifying each pixel of an image into a set of predefined semantic classes. These classes can represent cars, animals, people, road signs, tumors, etc. Most of the current methods are based on convolutional neural networks (CNN) [1], [2], [10], [13], [23]. There have been many improvements using CNN. For example, using Fully Convolutional Network (FNC) as an improvement of deep CNN, like Long et al. [13] did. Or dilated convolutions [2], [24] used to increase the receptive field without losing spatial resolution.

Most of the papers from WMH Challenge participants and in the literature, or practically all, do semantic segmentation with CNN. We think we can do the semantic segmentation with other approaches with different kind of networks, or with less image requirements.

II.2. Search for alternatives

As we have seen, CNNs are the most widely used networks for semantic segmentation: at this moment there is no better method to perform it [25].

However, there are clear limitations [26]. The main one is overfitting, when there are not enough images to train these networks, which makes the trained model not generalizable to new images [27]. For example, in the case of the WHM Challenge, the small number of images provided by the contest organizers [28] has make that quite a few of them used data augmentation techniques to avoid this.

¹ <http://cs231n.stanford.edu/index.html>

In recent years an alternative has emerged that makes learning faster and with fewer images. This is meta learning [29]. Thus, in this research we wonder if it can be an alternative for the problem at hand.

11.3. Meta learning

Meta-learning is difficult to define. It can be understood as the process of learning to learn, or the process of improving the learning algorithm through multiple learning episodes. There are two parts in the learning process: during the learning base, an internal algorithm solves a task like image classification [30]. In the meta-learning phase, an external algorithm updates the internal algorithm, so that the learned model improves an external objective [29].

Meta-learning was first mentioned in 1987 in two separate papers, one by J. Schmidhuber [31], and the other by G. Hinton [32]. The first one mentions a theoretical framework for methods that can learn how to learn. The second proposed that each neural network should have two sets of weights, a slow one that acquires knowledge slowly, and a fast one that acquires it quickly. Later, Thrun et al. coined the term learning to learn [29].

Depending on which aspects of the learning strategy are to be learned, the learning methods try to improve different aspect of the algorithm. Some of them are [29].

Parameter initialization. Here the goal is to learn the initial parameters of the neural network. In these cases, Model Agnostic Meta Learning (MAML) is used, as reported in [33], [34], [35].

Optimizer. Here the goal is to learn the optimal learning rate and the direction of update of the descent gradient with Meta-SGD (Stochastic Gradient Descent) [36].

Embedding Functions (Metric Learning). Here the meta-optimization learns an embedded network that transforms the raw input into a representation that allows its learning [37], [38], [39], [40] (like the Euclidean distance that is used in the prototype learning [41]).

Among the applications of meta learning is the few-shot learning. Its objective is to train Deep networks, like CNN, successfully with small data sets [29]. One of the applications is the multi-class classification of images as shown in [40].

A model of few-shot learning must learn with two small sets: 1) a small training set, referred as a support set, with a few labelled examples, with n examples for each class; and 2) a small validation set, referred as query set, with an unseen instances of the same classes contained in the support set [29].

Few-shot learning is also called k -shot learning, where k denotes the number of examples in each of the classes in the support set. If we have only one example for each class, then it is called one-shot learning. And, if we have no example for each class, then it is called zero-shot learning. In this case, the algorithm will learn from the meta information about each of the classes [42].

Meta-learning is also one of the most promising and trend-setting research areas in the field of Artificial Intelligence at the moment. It is believed to be a springboard for achieving General Artificial Intelligence (GAI) [43]. GAI is the hypothetical Artificial Intelligence capable of performing any intellectual task like humans [43]. It is hypothetical because it has not yet been obtained.

II.5. Few-shot learning

Few-shot learning methods learn with only a few examples and what it has learned is transferable knowledge to different tasks for which it has not been trained. Some of these methods are: metric learning [39], [37] and learning the optimization process [44], [33]. Vinyals et al [39] used this method into unsupervised learning. Snell et al. [37] proposed to represent each class into one feature vector (prototype), defining a Prototypical Network. And Sung et al. [45] proposed to use a separate module to directly learn the relation between support features and query features.

II.6. Few-shot segmentation

As introduced before, our research problem is how to perform semantic segmentation with few images. CNN needs a lot of images to avoid overfitting. As it has been seen in the last section, few-shot learning methods learn with few examples, and they have been used to classify images [40]. The next step is use few-shot with semantic segmentation (which is classifying each pixel of an image into a set of predefined semantic classes).

Semantic segmentation with few-shot learning algorithms has increased recently. Zhang et al. used an average pooling to differentiate foreground and background from the support set. Rakelly et al. [46] use a decoder with the concatenated features of the support and query set to get segmentation results. Shaban et al. [47] proposed a model with a branch to generate a set of parameters for the support set. These parameters are used to tune segmentation process of the query set. Dong et al. [41] has the idea of prototypical networks as a few-shot segmentation with metrical learning.

II.7. Summary

The vast majority of papers we have read use convolutional neural networks to perform semantic segmentation. Moreover, the results obtained do not vary much either, with success percentages that can be differentiated in tenths (Figure 2). For example, since 2014 it is very difficult to improve the results obtained in ImageNet and it is also difficult to attribute those very small increases to a better or more sophisticated architecture [25].

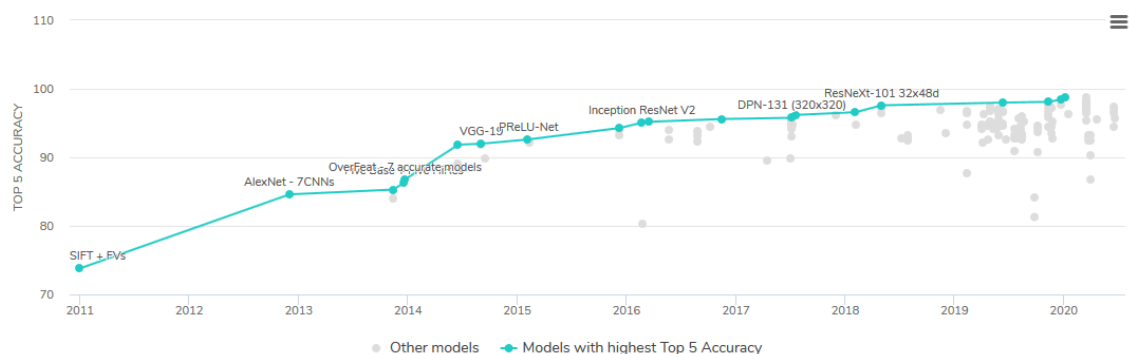


Figure 2. ImageNet Benchmark from Papers With Code website²

In the challenge where we have developed our solution, one of the main obstacles was the lack of images: participants are asked for results with little information. This lack of images means that the performance of convolution networks is not optimal (these networks need a lot of data to avoid overfitting). Participants are also allowed to use other datasets but free

² <https://paperswithcode.com/sota/image-classification-on-imagenet>

MRI datasets are very scarce. These facts made us to look for new approaches that can overcome this problem.

One of them was Meta learning, which solves the problem of having a small number of images. In addition, this approach works well on problems for which it has not been trained. This additional advantage means that our research work could be considered to recognize other types of brain injuries.

Among the papers read to inform our research background, we have chosen the paper "Few-Shot Semantic Segmentation with Prototype Learning" [41] as a basis for the work because they have developed a solution to do semantic segmentation with few-shot learning, but applied to do semantic segmentation in color images showing where is a car, a horse, a person, etc. We have used their architecture as a starting point, but added some modifications to it so it can be applied to our research problem: semantic segmentation of brain images. Thus, starting from the "Prototypical networks", whose objective is to classify images, we have made some adaptations so that these networks classify pixels (which is what semantic segmentation is all about). Classify an image is classify all the pixels in the image, and classify a pixel, or a group of pixels, is classify a piece of the image as if this piece of the image was a complete image.

III. Materials and methods

In this section we introduce the WMH segmentation challenge, the datasets we have used in our experiments and our proposed method.

A. WMH Segmentation Challenge, MICCAI 2017³

As the organizers explain in challenge's website: "The purpose of this challenge is to directly compare methods for the automatic segmentation of White Matter Hyperintensities (WMH) of presumed vascular origin." Manual delimitation of WMH is a time-consuming task and observer-dependent procedure [9]. Due to they are looking for an automatic segmentation algorithm. In this section we summarize it.

A total of twenty teams took part into the challenge, with the sysu_media team [48] obtaining the best results using a U-Net convolutional neural network [49].

To evaluate the results, the organizers have used the following metrics.

- Dice similarity coefficient.
It is a statistical validation metric to evaluate the reproducibility performance of manual segmentations and the precision of spatial overlap of automated probabilistic fractional segmentation of MR images, illustrated in two clinical examples.
The DICE similarity coefficient has been adopted to validate the segmentation of white matter lesions on MRI.

$$DSC = \frac{2 * (A \cap B)}{(A + B)} \quad (1)$$

³ <https://wmh.isi.uu.nl/>

- Hausdorff distance (modified, 95th percentile, mm)
The Hausdorff distance is commonly used in computer vision. In that field, a typical problem is that you are given a picture and a model of what you want to match. The goal is to find all the locations in the image that match the model. In other words, it would be about looking for a shape within an image.

- Average volume difference (in percentage)
It is the average of the difference in volume between two lesions. Let V_G y V_P be the volume of the lesion in the G and P regions respectively. So, the average volume difference (AVD) in percent is.

$$AVD = \frac{ABS(A - B)}{A} \quad (2)$$

- Sensitivity for individual lesions (recall)
It is the sensitivity for individual lesions.

$$recall = \frac{\text{number of detected WMH}}{\text{number of real WMH}} \quad (3)$$

- F1-score for individual lesions
In the statistical analysis of the binary classification, the F1 score (also F score or F measure) is a measure of the precision of a test. Consider both the precision p and the recovery r of the test to calculate the score: p is the number of correct positive results divided by the number of all positive results returned by the classifier, and r is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive). The F1 score is the harmonic mean of accuracy and recovery, where an F1 score reaches its best value at 1 (perfect precision and recovery) and its worst at 0.

$$F1 = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}} \quad (4)$$

The ranking of the WMH Segmentation Challenge - MICCAI 2017 was obtained averaging over all test scans. For each metric, the participants were sorted from best to worst. The best team receives the rank 0 and the worst, rank 1. The other teams are ranked (0,1). Finally, all team's ranks are averaged into an overall rank.

Thirteen teams used convolutional neural network architectures (such as U-Net [48], VGG-16 [50], DenseNet [51], HighResNet [52] or DeepMedic [53]), four used random forest [54], and the remaining three used deep neural networks [55] to perform the semantic segmentation of brain images by magnetic resonance.

Two types of brain imaging were offered by the organizers: T1 and FLAIR. Eighteen of the participants trained with both [56], while two of them trained with FLAIR imaging only [52].

Because of the limited number of images provided by the competition organizers: MRI from 60 different patients; for each patient we have T1 and FLAIR images, each image has another image with manual annotation of the WMH (a binary masks). Ten teams used data

enhancement techniques [57], and another two used additional data (MR images from other datasets) than the provided by the organizers to train their models [58].

Since our research work aims to improve semantic segmentation, the next section summarizes the most relevant papers about it.

#	Team	Rank	DSC	H95 (mm)	AVD (%)	Recall	F1
1	pgs	0.0185	0.81	5.63	18.58	0.82	0.79
2	sysu_media_2	0.0187	0.80	5.76	28.73	0.87	0.76
3	sysu_media	0.0288	0.80	6.30	21.88	0.84	0.76
4	buckeye_ai	0.0314	0.79	6.17	22.99	0.83	0.77
5	coroflo	0.0493	0.79	5.46	22.53	0.76	0.77
6	neuro.ml_2	0.0511	0.78	6.33	30.63	0.82	0.73
7	cian	0.0571	0.78	6.82	21.72	0.83	0.70
8	bioengineering_espil_team	0.0596	0.78	6.24	28.26	0.78	0.74
9	nlp_logix	0.0678	0.77	7.16	18.37	0.73	0.78
10	nih_cidi_2	0.0803	0.75	7.35	27.26	0.81	0.69
11	bigrbrain_2	0.0847	0.77	9.46	28.04	0.78	0.71
12	bigrbrain	0.0910	0.78	6.75	23.24	0.70	0.73
13	nic-vicorob	0.0927	0.77	8.28	28.54	0.75	0.71
14	rasha_improved	0.0964	0.77	7.42	24.97	0.76	0.67
15	wta	0.1012	0.78	6.78	16.20	0.66	0.73
16	dice	0.1046	0.77	7.63	19.77	0.69	0.71
17	fmrib	0.1114	0.75	8.91	27.93	0.72	0.70
18	wta_2	0.1132	0.76	8.21	21.31	0.67	0.72
19	misp_2	0.1178	0.78	11.10	19.71	0.68	0.71
20	uned_2	0.1237	0.76	8.90	28.83	0.73	0.63
21	uned_contrast	0.1355	0.75	9.90	44.33	0.77	0.60
22	k2	0.1531	0.77	9.79	19.08	0.59	0.70
23	uned	0.1552	0.73	11.04	55.84	0.81	0.54
24	rasha_simple	0.1562	0.74	9.05	29.73	0.65	0.64
25	acunet	0.1735	0.65	9.22	29.35	0.67	0.67
26	lrde	0.1797	0.73	14.54	21.71	0.63	0.67
27	misp	0.1821	0.72	14.88	21.36	0.63	0.68
28	ipmi-bern	0.2625	0.69	9.72	19.92	0.44	0.57
29	nih_cidi	0.2843	0.68	12.82	196.38	0.59	0.54
30	scan	0.2898	0.63	14.34	34.67	0.55	0.51
31	tig_corr	0.3039	0.68	17.48	39.07	0.54	0.46
32	livia	0.3044	0.61	22.70	38.39	0.54	0.61
33	achilles	0.3081	0.63	11.82	24.41	0.45	0.52
34	skkumedneuro	0.3604	0.58	19.02	58.54	0.47	0.51
35	tignet	0.3909	0.59	21.58	86.22	0.46	0.45
36	tig	0.3955	0.60	17.86	34.34	0.38	0.42
37	knight	0.4239	0.70	17.03	39.99	0.25	0.35
38	upc_dlmi	0.4449	0.53	27.01	208.49	0.57	0.42
39	himinn	0.4500	0.62	24.49	44.19	0.33	0.36
40	nist	0.4830	0.53	15.91	109.98	0.37	0.25

41	text_class	0.5781	0.50	28.23	146.64	0.27	0.29
42	neuro.ml	0.6074	0.51	37.36	614.05	0.71	0.21
43	hadi	0.8940	0.23	52.02	828.61	0.58	0.11

Table 1. Current results of the WMH Segmentation Challenge. In bold the best results on each metric. In this work we have followed the paper [9], but for the time of its publication there were more submits and now the best team is “pqs” and not “sysu_media”, as we mention in this paper. Table taken from WMH Challenge website⁴

B. Datasets

We have used two datasets, one of them is the WMH Challenge presented in the previous section. All of them have MR images from brain with injuries (tumors or WMH) and its corresponding label images with injuries segmented manually by expert board-certified neuroradiologists. We have chosen these datasets because all are related, all have the same type of images, and because of that, we can use them to pre-train our network to test if there is an improvement with pre-trained networks doing semantic segmentation.

B.1. WMH Challenge Dataset

The data with which we are going to test our solution are those provided in the “WMH Segmentation Challenge”. These data come from three hospitals in the Netherlands and Singapore. To obtain them, five different scanners from three different commercial brands have been used. Each hospital has provided images of 20 different patients. For each patient a 3D T1 image and a 2D multi-layer FLAIR image have been provided. In addition, for each patient a standard manual reference (a mask) is provided indicating where the hyper intensities are. FLAIR type images have been used to draw this reference.

B.1.1. Data detail

The data (images) have the following distribution.

Institute	Scanner	# training	Size of FLAIR scans	# test
UMC Utrecht	3 T Philips Achieva	20	240 x 240 x 48	30
NUHS Singapore	3 T Siemens Trio Tim	20	252 x 232 x 48	30
VU Amsterdam	3 T GE Signa HDxt	20	132 x 256 x 83	30
	3 T Philips Ingenuity	0	Unknown	10
	1.5 T GE Signa HDxt	0	Unknown	10

Table 2. WMH dataset data distribution.

The participants have been provided with 60 sets of brain images. The jury has, in turn, 110 images to test the presented solution. Those 110 images are not available to participants.

The images were provided in two formats.

- Original: Original images, anonymized and having the face removed
- Preprocessed: All images were pre-processed with SPM12 r6685 to correct bias field inhomogeneities.

B.1.2. Standard Reference Manual (or WMH masks)

One mask has been provided for each image of each patient. These masks are images with three possible pixel values.

- 0: background.
- 1: WMH.

⁴ <https://wmh.isi.uu.nl/>

- 2: other pathologies.

B.1.3. Image pre-processing

The images of each hospital have different sizes, so it has been necessary to resize them.

As for the WMH masks, pixels with a value of 2, other pathologies, have been eliminated since they are not the scope of this challenge.

B.2. Multimodal Brain Tumor Segmentation Challenge 2019 dataset

The MICCAI 2019 BraTS Challenge [59][60][61] focuses on evaluation of state-of-the-art methods for segmentation brain tumors in multimodal magnetic resonance imaging (MRI) scans. It focuses on the segmentation of brain tumors.

The challenge focuses on the segmentation of intrinsically heterogeneous brain tumors (gliomas). It also focuses on the prediction of patient overall survival, via integrative analyses of radiomic features and machine learning algorithms. Finally, BraTS'19 intends to experimentally evaluate the uncertainty in tumor segmentations.

B.2.1. Data detail

All scans are available as NIfTI files (.nii.gz), like WMH Challenge images; and are in four formats: a) native (**T1**), b) post-contrast weighted (**T1Gd**), c) T2-weighted (**T2**), and d) T2 Fluid Attenuated Inversion Recovery (**T2-FLAIR**).

The dataset has been segmented manually by clinical experts with three labels [61]: GD-enhancing tumor (**label 4**), the peritumoral edema (**label 2**), and the necrotic and non-enhancing tumor core (**NCR/NET — label 1**).

We have also converted all the labels into 1 because in [61] points that to segment the whole tumor extent with have to unite all the labels. We need to do it because we are going to segment images with two segmentation classes: brain healthy tissue and tumors.

B.2.2. Image pre-processing

With the mask image, we have converted every pixel value to one if its value is greater than one. In other words, we have joined all the labels.

We did because we want to train the network to identify the whole tumor, and originally, these mask images are labeled with three labels as we have explained in the previous section.

C. The proposed method

As anticipated in section II.7, we are going to use the same method described in [41]. Their method is a framework for N-way k-shot semantic segmentation based on prototype learning. Prototype means that some elements are more representative than others with the same category [62]. Prototypical networks classify data (images) computing their distances with the prototype representation of each class [37]. The novelty for their method is that they have used to do semantic segmentation (to classify each pixel in the image).

We have adopted their method, partially, to do semantic segmentation of one channel images (MR images are grey scale) unlike them who have used three channels images (colour images from real life and from ImageNet).

C.1. Architecture

We use mostly the same architecture than [41] with some differences.

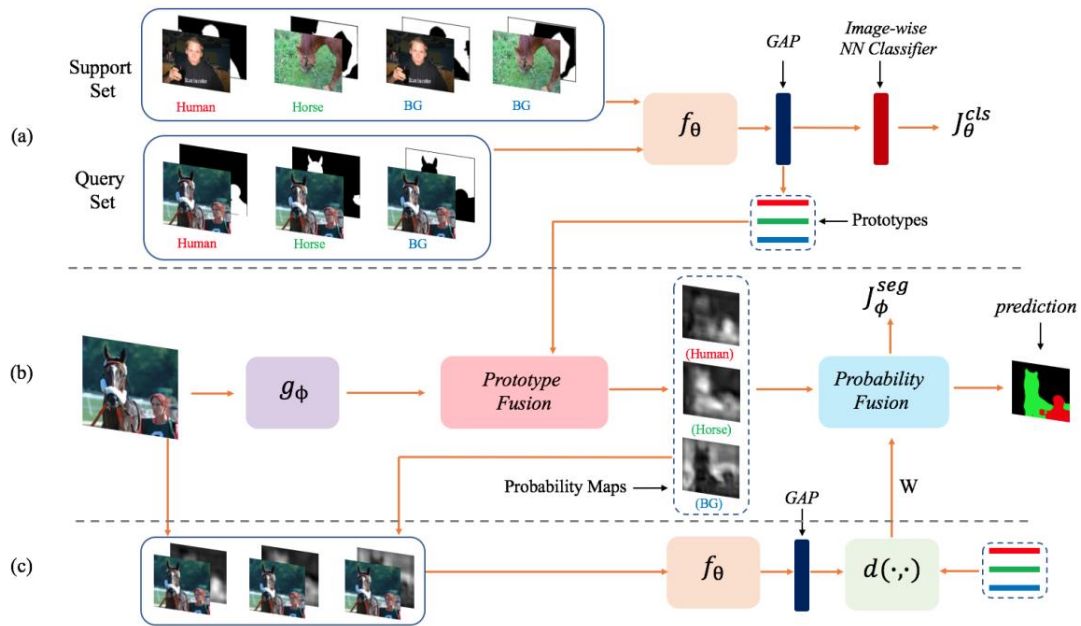


Figure 3. Architecture and data flow of [41] taken from their paper.

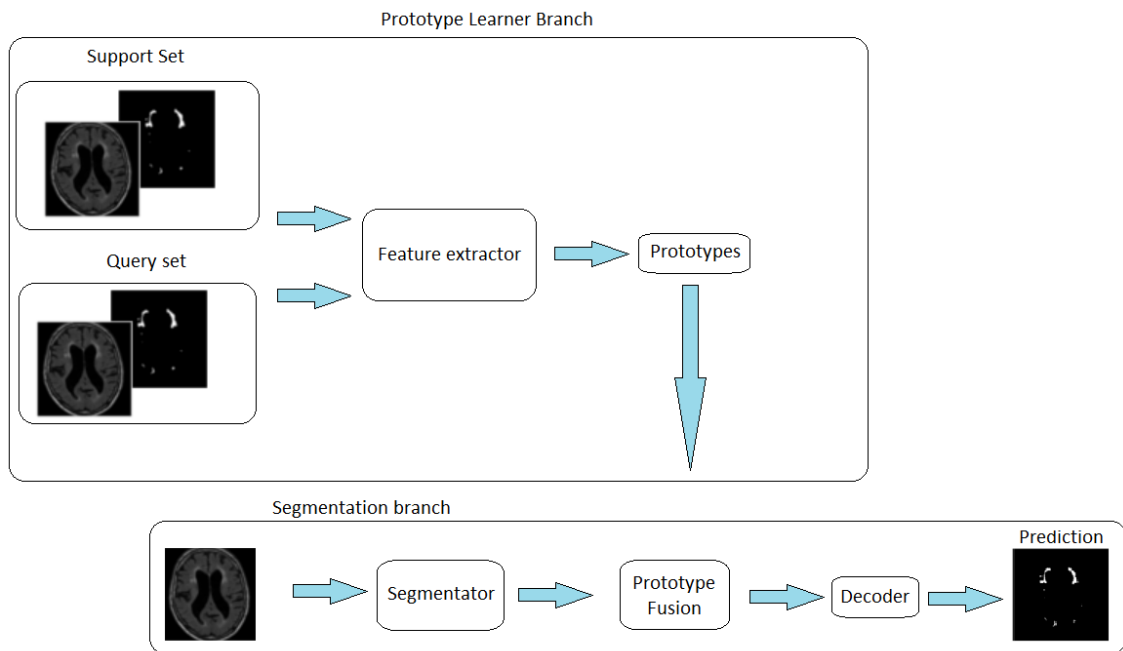


Figure 4. Our architecture and data flow.

The network has two branches: one to extract the prototypes (prototype learner branch), and another branch, segmentation branch, to do the semantic segmentation of the query images (the input images for this second branch).

The input for the learner branch is the brain image masked by the WMH mask, and its output is a feature maps with 1 channel. Instead of using a GAP layer, we use the mean of all of the feature vectors get in the training.

For the segmentator we use the same network architecture as the feature extractor. We fuse the output of this branch with the output of the segmentator in the prototype fusion, and we pass the result to a decoder to get the prediction.

Instead of using VGG16 network only, we use VGG16 and U-Net implementation from [48], to compare their results. We use VGG16 because is the same network used in [41], and we have used also U-Net because it has been designed for tasks with biomedical images [63] and it has been used by [48] in their winning solution of the challenge and you can check its architecture in Figure 5.

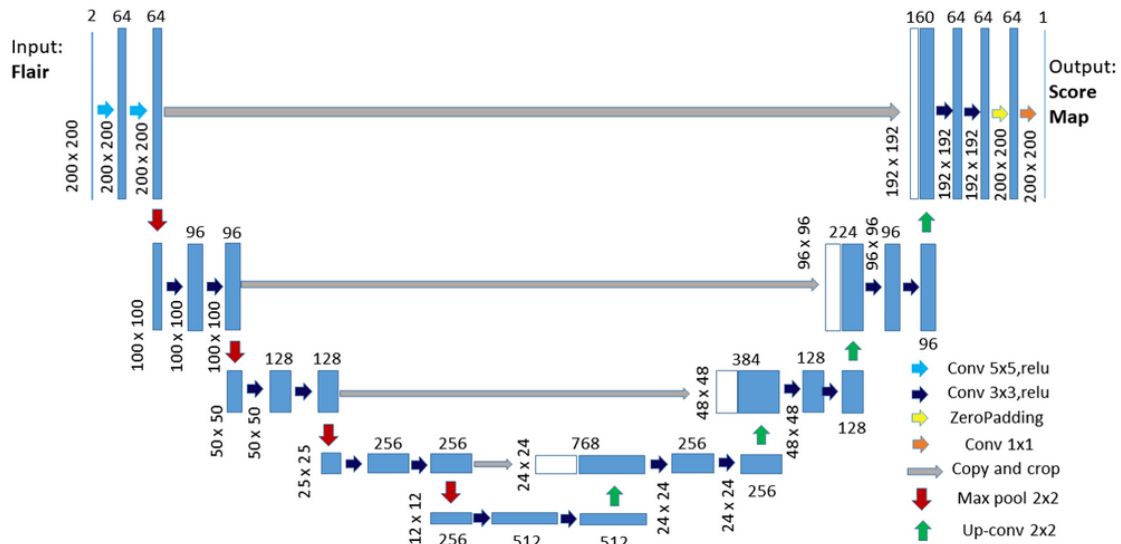


Figure 5. U-Net network architecture [48] solution.

VGG-16 network is an encoder, that encodes an image and its output is used to classify it. In our method we are using two VGG-16 networks in an encoder-decoder architecture. The encoder generates a high-dimensional feature vector. The decoder takes this high-dimensional feature vector and generates a semantic segmentation mask. Figure 6 shows this architecture.

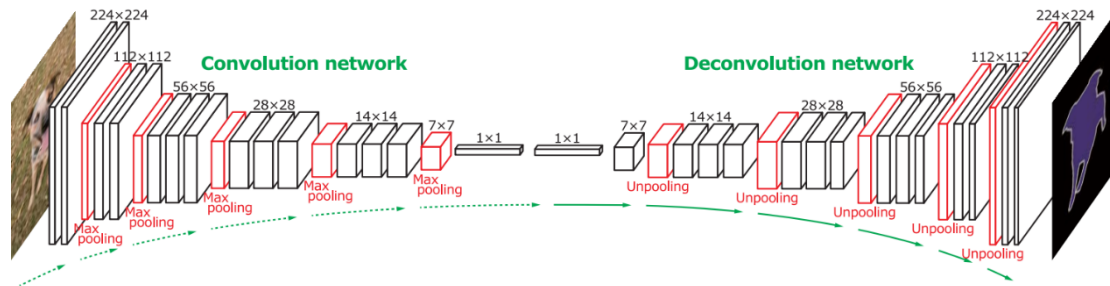


Figure 6. VGG-16 encoder-decoder architecture. Picture taken from [64].

C.2. The proposed algorithm

The algorithm to train the feature branch is.

- Initialize the feature extractor network's weights with the best initialization method (explained in the next section).
- With the whole support set, we train the feature extractor branch in episodic fashion.
- While training, we are getting the features arrays for the WMH class.
- At the end of the training, we do the mean of these features' arrays.

To get a prediction with the segmentation branch, we do.

- Initialize the segmentation network's weights with the same initialization method as feature extractor.
- To make a prediction, we take one image that it has not been used in the training, and we use it as the input for this branch to generate an output.
- We use this output as the input for a decoder to get the predicted WMH segmentation.

We have changed the architecture from paper [41] because we want to simplify it and take advantage of the location information contained in the class features' array obtained in the Feature branch.

C.3. Weight Initialization

They initialize network's weights training with the ImageNet dataset. Instead, we have trained several methods to initialize our network. You can check a table with a comparison in the following section "Experiments setup". We cannot train our network with ImageNet because it is a dataset with three channels images and we work with one channel images. Brain images, as mentioned before, are one channel images.

The methods used are.

1. Gaussian distribution.
Layers' weights are initialized with the values of a normal distribution with a mean of 0 and a standard deviation of 1.
2. Xavier normal initializer.
Sets layers' weights to values chosen from a random uniform distribution centered in.

$$\pm \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}} \quad (5)$$

where n_i is the number of incoming connections for current layer, and n_{i+1} is the number of outgoing connections.

3. Weights from the winning team sysu_media [48].
This team has shared their solution's weights and we have loaded into our U-Net network. Tensorflow lets train a model (in this case U-Net network) and save their weights into a binary file. This let research to not train the network every time we have to test our algorithm. And also, this let us to share our training between other researches.
4. The MICCAI 2019 BraTS Challenge dataset [59][60][61]
We have trained our networks with images from this dataset and then save the weights to use them later into our experiments.

C.4. Optimizer and loss function

For the optimizer we use ADAM optimizer [65] in both branches and dice coefficient function [66] to measure the similarity. The initial learning rate is 10^{-4} on both branches.

For the loss function we have used the metric "Dice similarity coefficient", which is one of the metrics for the WMH Challenge. We have used this and no other because this function is differentiable and we can use in backpropagation during training.

IV. Experiments

In this section we report the outcomes of the experiments carried out with the datasets and architecture described in the previous section.

1. Experiments setup

Following Figure 4, the experiment was divided in two steps.

1. Train the prototype learner with support set and query set using the four weights initialization methods explained in section E. Support set and query set have images from the WMH Challenge dataset.
2. Then we use the class prototypes obtained in the previous step to get the segmentation mask for new unseen brain images from the WMH Challenge dataset in the experiments.

We have considered two classes: brain and white matter hyperintensity. And we have trained the prototype learner with five data points as they did in [37]. We are going to do the same methods than [37], 2-way 5-shot learning and 2-way 1-shot learning, to compare results. The number of episodes in the prototype learner train was 200 because there are 960 brain images from the three hospitals. If we are using 5-shot, it means that we are using 5 images on each episode; so $\frac{960}{5} = 192$ episodes. We choose these 5 images randomly on each episode; In order to be more or less sure that we use all the images in the set, we decided to use 200 episodes.

To compute the mean for the class prototype, instead of getting one value as the mean of all the values in the features array (as they did in the referential paper [41]), we have computed the mean as an array, where each value in this array is the mean of the elements at the same row and column for all the features array compute it on the training. We did it this way due to the configuration of the WMH, since they appear surrounded by a characteristic brain tissue and we have found more interesting to have the average of each characteristic “pixel” that makes up a WMH, instead of a single number.

We have preprocessed all the images from the WMH Challenge, 960 images. A part from the preprocessing done by the WMH challenge organizers, we have done the same pre-process than the winning challenge solution [28] to: guarantee a uniform size of all data and normalize voxel intensity to reduce variation across subjects. We enforce these: cropping or padding each image to a uniform size (200 x 200) and Gaussian normalization on the brain voxel intensity.

The experiments were implemented by Tensorflow 2.2.0 [67] on several environments: i) NVIDIA GeForce GTX 970, ii) Google Collaboratory using a GPU notebook and iii) NVIDIA JETSON TX2. We have used these three environments because the first one is my personal computer and we have not been able to run the experiments due to the lack of memory. We decided to move the environment to Google Collaboratory because it is free and its hardware resources, like GPUs, are far from enough to let test our algorithm [68]. Finally, we have also tried the algorithm on NVIDIA JETSON TX2 which has the best performance.

Execution times in each of the tasks.

VGG-16				
Environment	Load images	Training	Predict	Total
NVIDIA GTX 970	32,426854848861694	412,695604801178	1,1420652866363525	446,264524936676
Google Collaboratory	72,16633415222168	81,3402853012085	1,6061344146728516	155,112753868103
NVIDIA JETSON TX2	67,26481236587416	75,6698214536974	0,9552147378934212	143,889848557465

Table 3. VGG-16 execution time comparison (seconds).

U-NET				
Environment	Load images	Training	Predict	Total
NVIDIA GTX 970	38,1671826839447	301,79126167297363	1,1170639991760254	341,07550835609436
Google Collaboratory	50,01747369766235	56,29819703102112	1,5159823894500732	107,84133577346802
NVIDIA JETSON TX2	45,32658412536972	52,36574921556958	0,8511365974521312	98,5434699383913

Table 4. U-NET execution time comparison (seconds).

As expected, artificial intelligence algorithms run better with GPU. NVIDIA Jetson TX2 has a better architecture than free services from Google and from my modest NVIDIA GTX 970 GPU.

Weights initialization

In the experiments, both branches, the prototype learner and the segmentor, use the encoder part of the VGG16 network and U-Net network. The weights of the network were initialized in several forms.

1. Gaussian distribution.
2. Xavier normal initializer
3. Weights from the winning team sysu_media [48].
4. The MICCAI 2019 BraTS Challenge dataset [59][60][61].

Environment	Network	1	2	3	4
Google Colab	VGG-16	1,0	0,99605644	0,99788135	0,99788135
Google Colab	U-Net	0,9757263	0,97694886	0,99658155	0,9995242

Table 5. Dice coef. loss comparison.

We have implemented the dice similarity coefficient used in the WMH Challenge. Dice coef. Is a statistical validation metric to evaluate the reproducibility performance of manual segmentations and the accuracy of spatial overlap of automated fractional probabilistic segmentation of MR images. The dice coef. Loss function is $1 - dice_coef$. The closer the value is to 1 the worse the network prediction is.

In table 6 we show the network prediction for each of the weight initialization methods.

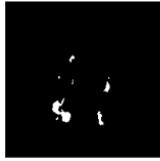


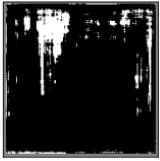

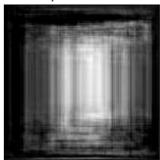


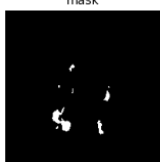
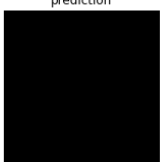
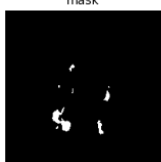
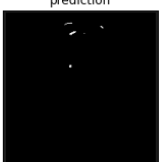
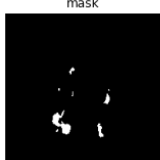
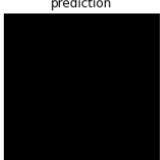
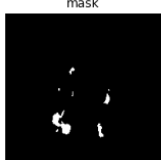
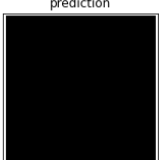
Weights	VGG-16		U-Net	
1	mask 	prediction 	mask 	prediction 
2	mask 	prediction 	mask 	prediction 
3	mask 	prediction 	mask 	prediction 
4	mask 	prediction 	mask 	prediction 

Table 6. Output comparison for the different weight initialization methods and networks.

The two best initialization methods were sysu_media (3) and BraTS19 (4) because they use the same kind of images that we are going to use in our experiment. Obviously, we cannot use sysu_media initialization because they use the same images that we use in the experiment and we can get wrong results due to overfitting.

2. Experiments execution

We have trained our model in an episodic fashion—that is, in each episode, we sample a few data points (images) from our dataset. On each episode we take a few images to train it and another few images to test it. So, over series of episodes, our model learns how to learn from a smaller dataset. This follows the idea behind few-shot learning, which is learn with few data.

We did quite a few experiments changing the network, VGG-16 or U-Net, and changing the step one (the way we initialize network's weights). So, with VGG-16 we did four experiments, one for each weight initialization method. And with U-Net we did another four.

One of the most interesting features in Meta Learning is that we can use our trained network in a task which it was not trained. To test it, we have initialized the net with the weights of the WMH Challenge winning solution, and use it to predict tumors with BraTS19 images dataset.

V. Experimental results

1. Results

We have found that using a pre-trained network gets better results than if we initialize the network with another initialization method (Gaussian distribution or Xavier normal initializer). In table 5 and table 6, we found that the best results are for the network pre-trained with similar images than the WMH Challenge images. So, we have pretrained our networks with

BraTS19 images dataset. We have not used the sysu media dataset [48] because they use the same images as us and we can falsify the results by obtaining overfitting. In other words, we have pretrained our networks with BraTS19 images dataset.

The results do not differ much if we use VGG-16 or U-Net. In all the cases analyzed the loss function scores over 0.997, a little better with U-Net, and no difference between 2-way 5-shot and 2-way 1 shot. It is better for U-Net because the network was created for Biomedical Image Segmentation [49] and there is no different between 5-shot and 1-shot because the prediction problem probably does not depend on the number of images you train with. Maybe the problem is in the network architecture.

Network	2-way 5-shot	2-way 1-shot
VGG-16	0,99788135	0,99788135
U-Net	0,9995242	0,9995242

Table 7. Loss for 2-way 5-shot and 2-way 1-shot.

The ground truth compared with the network predict are in Table 5.


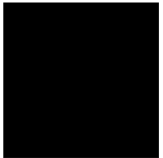

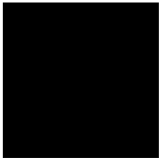

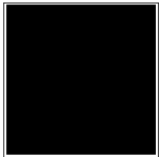


Network	2-way 5-shot		2-way 1-shot	
	mask	prediction	mask	prediction
VGG-16				
U-Net				

Table 8. Comparison between mask and network prediction.

In Table 8 we seen that the predictions are the same for both experiments (5-shot and 1-shot).

Comparing the results from the paper [41], in Figure 7, and our experiments we noticed that predictions from [41] are quite goods. The predictions in Figure 7 are the color mask in top of a person, or the horse, the dog or in top of the motorbike. In the Horse image (c) we see that the color mask for the horse is bigger than the horse but it is quite good, not perfect, to know where is horse in the image. But our experiments results show that maybe WMH are too small to be segmented or even the pixels for all of the WMH are not grouped like the horse: there is one horse in picture (c) and in the last mask, bottom right corner in table 9, there are more than four WMH.

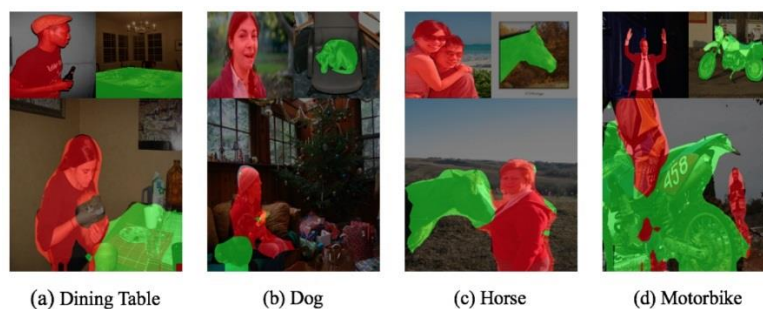


Figure 7. Predictions from paper [41].


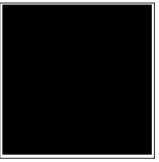

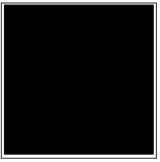

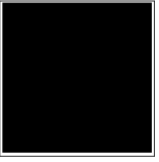

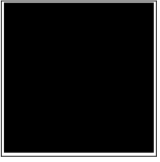
Network	2-way 5-shot		2-way 1-shot	
VGG-16	mask 	prediction 	mask 	prediction 
U-Net	mask 	prediction 	mask 	prediction 

Table 9. Predictions from our algorithm.

2. Discussion

Comparing our method with method in paper [41], we have the following differences.

We have pre-trained both branches with one-channel images (from “The MICCAI 2019 BraTS Challenge” dataset) instead of using ImageNet dataset. We did this way because we work with one-channel images and ImageNet images are three channels and we can use a network pre-trained with three-channels images with one-channel images.

We have computed the class prototype as an array instead of a single value. We did it this way due to the configuration of the WMH, since they appear surrounded by a characteristic brain tissue and we have found more interesting to have the average of each characteristic “pixel” that makes up a WMH, instead of a single number.

Instead of using a Prototype Fusion we have used a decoder to decode the fusion between the features array from the segmentor branch with the class prototype into a WMH mask.

With our changes in the architecture from paper [41] we wanted create a simpler architecture but semantic segmentation is a difficult task in which we have to take into account the spatial location, within the image, of what you are segmenting. Maybe we have lost this information and this is why appears a white border into U-Net predictions.

Finally, we have also initialized both networks with sysu media weights, and try to predict brain tumors in the BraTS19 dataset with these results.

Network	2-way 5-shot	2-way 1-shot
VGG-16	0,9978261	0,9978261
U-Net	0,79084265	0,6719284

Table 10. Loss predicting BraTS19 images with sysu media weights.









Network	2-way 5-shot		2-way 1-shot	
	mask	prediction	mask	prediction
VGG-16				
U-Net				

Table 11. Networks' predictions images.

U-Net network has a better performance with this configuration. Looks like pretraining the network with brain MR images with WMH and try to predict brain tumors from BraTS19 images has a better performance. It seems that the improvement is due to that we are trying to predict tumors instead of WMH, and this is easier for the network than predict WMH. If we get different results with the same network and changing the images, then, in future works, we have to improve the preprocess step with the images.

VI. Conclusion and future works

The field of semantic segmentation done with prototype learning is in early stages of research. The problem we have deal with is to find the way to identify the WMH in brain images. In paper [41] they cropped the relevant area in an image, i.e. an image of a landscape with a horse, the cropped the horse getting an image with a black background and a horse. We think cropping the images, getting only the brain to get the WMH and its surrounding tissue is a good idea, because the networks focuses on the relevant tissue. As we commented in the introduction, "WHM have a signal intensity brighter and with light boundaries compared to the surrounding white matter [8]".

The research presented in this paper is framed in the project "Intelligent systems for active aging and rehabilitation" offered by the phyUM research line on physical user movement in the Doctorate Program at UNED, where this research is expected to be continued. Specifically, the contribution of this work to this line of research is focused on the search for computer vision techniques for the analysis of brain images that allow the identification of brain damage, for example, those that occur after a stroke and can affect a person's ability to move. In this way, teaching/learning strategies could be designed to adapt to the motor rehabilitation needs associated with stroke. In addition, this work can also provide some insights to the INT2AFF project, which studies -in an intrasubject approach- how emotions evolve when users are learning tasks. These tasks can be psychomotor actions such as learning again to walk after a stroke. In particular, functional MRI, which collects MRI along time, to measure the evolution of brain activity, has proven useful to study human emotions [69].

Future works can lead to find alternative ways of identifying the relevant features instead of convolutional filters (a method to identify relevant features of an image and where in the image they are). Or even trying to find another way to teach a program to learn meta information containing a high-level description of each class. In this work, the application of an architecture with a decoder at the end of the second branch has suggested that we need a method to get also the location information for what the network has found.

Another future works that can be addressed are.

- Find a way to create a sharper image of the image obtained from the network as a prediction. For example, by treating the image as if it were out of focus, apply some algorithm to bring it into focus.
- Have another kind of layer to extract the features of an image and where in the image have been located.
- Use the output of the decoder to get the cost function and use it to train the network.

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the NVIDIA Jetson TX2 used for this research.

And we also gratefully of Josep M. Tormos, director of research at the Institut Guttmann.

This work is partially framed in the project INT2AFF funded under Grant PGC2018-102279-B-I00 (MCIU/AEI/FEDER, UE) by the Spanish Ministry of Science, Innovation and Universities, the Spanish Agency of Research and the European Regional Development Fund (ERDF).

References

- [1] van Norden AG, de Laat KF, Gons RA, et al. "**Causes and consequences of cerebral small vessel disease. The RUN DMC study: a prospective cohort study. Study rationale and protocol.**" *BMC Neurol.* 2011;11:29. Published 2011 Feb 28. doi:10.1186/1471-2377-11-29
- [2] Schoonheim MM, Vigeveno RM, Rueda Lopes FC, et al. "**Sex-specific extent and severity of white matter damage in multiple sclerosis: implications for cognitive decline**". *Hum Brain Mapp.* 2014;35(5):2348-2358. doi:10.1002/hbm.22332
- [3] Marshall GA, Shchelchkov E, Kaufer DI, Ivanco LS, Bohnen NI. "**White matter hyperintensities and cortical acetylcholinesterase activity in parkinsonian dementia**". *Acta Neurol Scand.* 2006;113(2):87-91. doi:10.1111/j.1600-0404.2005.00553.x
- [4] Weinstein G, Beiser AS, Decarli C, Au R, Wolf PA, Seshadri S. "**Brain imaging and cognitive predictors of stroke and Alzheimer disease in the Framingham Heart Study.**" *Stroke.* 2013;44(10):2787-2794. doi:10.1161/STROKEAHA.113.000947
- [5] Hirono N, Kitagaki H, Kazui H, Hashimoto M, Mori E. "**Impact of white matter changes on clinical manifestation of Alzheimer's disease: A quantitative study.**" *Stroke.* 2000;31(9):2182-2188. doi:10.1161/01.str.31.9.2182
- [6] Smith CD, Snowdon DA, Wang H, Markesbery WR. "**White matter volumes and periventricular white matter hyperintensities in aging and dementia.**" *Neurology.* 2000;54(4):838-842. doi:10.1212/wnl.54.4.838
- [7] Wardlaw JM, Smith EE, Biessels GJ, et al. "**Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration.**" *Lancet Neurol.* 2013;12(8):822-838. doi:10.1016/S1474-4422(13)70124-8

- [8] Zhong Y, Utraiainen D, Wang Y, Kang Y, Haacke EM. "**Automated White Matter Hyperintensity Detection in Multiple Sclerosis Using 3D T2 FLAIR.**" *Int J Biomed Imaging*. 2014;2014:239123. doi:10.1155/2014/239123
- [9] Kuijf HJ, Biesbroek JM, De Bresser J, et al. "**Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation Challenge.**" *IEEE Trans Med Imaging*. 2019;38(11):2556-2568. doi:10.1109/TMI.2019.2905770
- [10] Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. "**What are white matter hyperintensities made of? Relevance to vascular cognitive impairment**" [published correction appears in *J Am Heart Assoc*. 2016 Jan 13;5(1):e002006]. *J Am Heart Assoc*. 2015;4(6):001140. Published 2015 Jun 23. doi:10.1161/JAHA.114.001140
- [11] Morris Z, Whiteley WN, Longstreth WT, Weber F, Lee YC, Tsushima Y, Alphas H, Ladd SC, Warlow C, Wardlaw JM, Al-Shahi Salman R. "**Incidental findings on brain magnetic resonance imaging: systematic review and meta-analysis**". *BMJ*. 2009; 339:b3016.
- [12] Debette S, Markus HS. "**The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis.**" *BMJ*. 2010;341:c3666. Published 2010 Jul 26. doi:10.1136/bmj.c3666
- [13] de Laat KF, Tuladhar AM, van Norden AG, Norris DG, Zwiars MP, de Leeuw FE. "**Loss of white matter integrity is associated with gait disorders in cerebral small vessel disease.**" *Brain*. 2011;134(Pt 1):73-83. doi:10.1093/brain/awq343
- [14] Baezner H, Blahak C, Poggesi A, et al. "**Association of gait and balance disorders with age-related white matter changes: the LADIS study.**" *Neurology*. 2008;70(12):935-942. doi:10.1212/01.wnl.0000305959.46197.e6
- [15] Herrmann LL, Le Masurier M, Ebmeier KP. "**White matter hyperintensities in late life depression: a systematic review.**" *J Neurol Neurosurg Psychiatry*. 2008;79(6):619-624. doi:10.1136/jnnp.2007.124651
- [16] Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. "**Heritability of leukoaraiosis in hypertensive sibships.**" *Hypertension*. 2004;43(2):483-487. doi:10.1161/01.HYP.0000112303.26158.92
- [17] Altmann-Schneider I, van der Grond J, Slagboom PE, et al. "**Lower susceptibility to cerebral small vessel disease in human familial longevity: the Leiden Longevity Study.**" *Stroke*. 2013;44(1):9-14. doi:10.1161/STROKEAHA.112.671438
- [18] Valdés Hernández Mdel C, Booth T, Murray C, et al. "**Brain white matter damage in aging and cognitive ability in youth and older age.**" *Neurobiol Aging*. 2013;34(12):2740-2747. doi:10.1016/j.neurobiolaging.2013.05.032
- [19] Dufouil C, Alperovitch A, Tzourio C. "**Influence of education on the relationship between white matter lesions and cognition.**" *Neurology*. 2003;60(5):831-836. doi:10.1212/01.wnl.0000049456.33231.96
- [20] Deary IJ, Bastin ME, Pattie A, et al. "**White matter integrity and cognition in childhood and old age.**" *Neurology*. 2006;66(4):505-512. doi:10.1212/01.wnl.0000199954.81900.e2
- [21] G. Wright, "**Magnetic resonance imaging**", *IEEE Signal Process. Mag.*, 14 (1997), pp. 56-66
- [22] S. A. Trip, David H. Miller, "**Imaging in multiple sclerosis**", *Neurology in Practice*, 76, 3, 9 2005
- [23] Q. Mahmood and A. Basit, "**Automated Segmentation of White Matter Hyperintensities in Multi-modal MRI Images Using Random Forests**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/hadi/>
- [24] —, "**Multi-dimensional Gated Recurrent Units for the Segmentation of White Matter Hyperintensities**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/cian/>

- [25] Litjens G, Kooi T, Bejnordi BE, et al. "**A survey on deep learning in medical image analysis.**" *Med Image Anal.* 2017;42:60-88. doi:10.1016/j.media.2017.07.005
- [26] Gary Marcus. "**Deep Learning: A Critical Appraisal**", 1 2018
- [27] Yamashita R, Nishio M, Do RKG, Togashi K. "**Convolutional neural networks: an overview and application in radiology**". *Insights Imaging.* 2018;9(4):611-629. doi:10.1007/s13244-018-0639-9
- [28] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze, "**Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images,**" *NeuroImage*, vol. 183, pp. 650–665, 2018.
- [29] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. 2020. "**Meta-learning in neural networks: A survey**". arXiv preprint arXiv:2004.05439.
- [30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "**Imagenet Classification With Deep Convolutional Neural Networks,**" in *NeurIPS*, 2012.
- [31] J. Schmidhuber, "**Evolutionary Principles In Self-referential Learning,**" *On learning how to learn: The meta-meta-... hook*, 1987
- [32] G. E. Hinton and D. C. Plaut, "**Using Fast Weights To Deblur Old Memories,**" in *Conference Of The Cognitive Science Society*, 1987
- [33] C. Finn, P. Abbeel, and S. Levine. "**Model-agnostic meta-learning for fast adaptation of deep networks**". In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1126–1135. 2017.
- [34] C. Finn, K. Xu, and S. Levine, "**Probabilistic Model-agnostic Meta-learning,**" in *NeurIPS*, 2018.
- [35] C. Finn, A. Rajeswaran, S. Kakade, and S. Levine, "**Online Metalearning,**" *ICML*, 2019
- [36] Li, Zhenguo & Zhou, Fengwei & Chen, Fei & Li, Hang. "**Meta-SGD: Learning to Learn Quickly for Few Shot Learning**". 2017.
- [37] Jake Snell, Kevin Swersky, and Richard Zemel. "**Prototypical networks for few-shot learning**". In *Advances in Neural Information Processing Systems*, 2017
- [38] S. Qiao, C. Liu, W. Shen, and A. L. Yuille, "**Few-Shot Image Recognition By Predicting Parameters From Activations,**" *CVPR*, 2018
- [39] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra et al., "**Matching Networks For One Shot Learning,**" in *NeurIPS*, 2016
- [40] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. Wang, and J.-B. Huang, "**A Closer Look At Few-Shot Classification,**" in *ICLR*, 2019
- [41] N. Dong and E. P. Xing. "**Few-shot semantic segmentation with prototype learning**". In *BMVC*, volume 3, page 4, 2018
- [42] H. Larochelle, D. Erhan, and Y. Bengio, "**Zero-data Learning Of New Tasks.**" in *AAAI*, 2008
- [43] Goertzel, B. (2014). "**Artificial General Intelligence: Concept, State of the Art, and Future Prospects**", *Journal of Artificial General Intelligence*, 5(1), 1-48.
- [44] S. Ravi and H. Larochelle (2016) "**Optimization as a model for few-shot learning**".
- [45] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales (2018) "**Learning to compare: relation network for few-shot learning**". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208.
- [46] K. Rakelly, E. Shelhamer, T. Darrell, A. Efros, and S. Levine. "**Conditional networks for few-shot semantic segmentation**". 2018.
- [47] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots. "**One-shot learning for semantic segmentation**". 2017. arXiv preprint arXiv:1709.03410.

- [48] H. Li, G. Jiang, L. Zhao, R. Wang, J. Zhang, and W.-S. Zheng, "**Automatic White Matter Hyperintensity Segmentation via Two-channel U-Net**," 2017. [Online]. Available: http://wmh.isi.uu.nl/results/sysuf_gmedia/
- [49] O. Ronneberger, Philipp Fischer, and T. Brox, "**U-Net: Convolutional Networks for Biomedical Image Segmentation**," in Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. Lecture Notes in Computer Science. Springer, Cham, 2015, vol. 9351, pp. 234–241.
- [50] —, "**White Matter Hyperintensities Segmentation in a Few Seconds Using Fully Convolutional Network and Transfer Learning**," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017. Lecture Notes in Computer Science, A. Crimi, S. Bakas, H. Kuijff, B. Menze, and M. Reyes, Eds. Springer, Cham, 2018, pp.501–514.
- [51] R. McKinley, A. Jungo, R. Wiest, and M. Reyes, "**Pooling-free fully convolutional networks with dense skip connections for semantic segmentation, with application to segmentation of white matter lesions**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/scan/>
- [52] A. Georgiou, "**WMH segmentation challenge MICCAI 2017: Team name - Achilles**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/achilles/>
- [53] A. Safiullin, "**NeuroML team: Brief description of the solution**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/neuro-ml/>
- [54] M. Bento, R. de Souza, R. Lotufo, R. Frayne, and L. Rittner, "**WMH Segmentation Challenge: A Texture-Based Classification Approach**," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017. Lecture Notes in Computer Science, A. Crimi, S. Bakas, H. Kuijff, B. Menze, and M. Reyes, Eds. Springer, Cham, 2018, pp. 489–500.
- [55] S. Andermatt, S. Pezold, and P. C. Cattin, "**Automated Segmentation of Multiple Sclerosis Lesions using Multi-Dimensional Gated Recurrent Units**," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, A. Crimi, S. Bakas, H. Kuijff, B. Menze, and M. Reyes, Eds. Springer, Cham, 2018, pp. 31–42
- [56] M. Luna and S. H. Park, "**3D Convolutional Neural Network with Skip Connections for WMH Segmentation**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/misp/>
- [57] S. Valverde, M. Cabezas, J. Bernal, K. Kushibar, S. González-Villà, M. Salem, J. Salvi, A. Oliver, and X. Lladò, "**White matter hyperintensities segmentation using a cascade of three convolutional neural networks**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/nic-vicorob/>
- [58] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladò, "**Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach**," *NeuroImage*, vol. 155, pp. 159–168, 2017.
- [59] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "**The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)**", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
- [60] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "**Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features**", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
- [61] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "**Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge**", arXiv preprint arXiv:1811.02629 (2018)

- [62] Eleanor H Rosch. "**Natural categories. Cognitive Psychology**", 4(3):328–350, 1973
- [63] G. Zeng and G. Zheng, "**Deeply Supervised Multi-Scale Fully Convolutional Networks for Segmentation of White Matter Hyperintensities**," 2017. [Online]. Available: <http://wmh.isi.uu.nl/results/ipmi-bern/>
- [64] H. Noh, S. Hong and B. Han, "**Learning deconvolution network for semantic segmentation**", Proc. IEEE Int. Conf. Comput. Vis., pp. 1520-1528, 2015.
- [65] Diederik P Kingma and Jimmy Ba. "**Adam: A method for stochastic optimization**". In International Conference on Learning Representations, 2015
- [66] Milletari, Fausto & Navab, Nassir & Ahmadi, Seyed-Ahmad. (2016). "**V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation**". 565-571. 10.1109/3DV.2016.79.
- [67] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. "**Tensorflow: a system for large-scale machine learning**". In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, pages 265–283. USENIX Association, 2016.
- [68] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De Albuquerque and P. P. R. Filho, "**Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications**", IEEE Access, 2018.
- [69] Gu S., Wang F., Cao C., Wu E., Tang Y. Y., Huang J. H. (2019). "**An integrative way for studying neural basis of basic emotions with fMRI**". Front. Neurosci. 13:628. 10.3389/fnins.2019.00628