NON-UNIFORM PATCH SAMPLING WITH DEEP CONVOLUTIONAL NEURAL NETWORKS FOR WHITE MATTER HYPERINTENSITY SEGMENTATION

M. Ghafoorian[⋆†], *N. Karssemeijer*[⋆], *T. Heskes*[†], *I.W.M. van Uden*[◇], *F.E. de Leeuw*[◇], *E. Marchiori*[†], *B. van Ginneken*[⋆] and *B. Platel*[⋆]

[⋆] Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands

[†] Institute for Computing and Information Sciences, Radboud University, Nijmegen, the Netherlands

[◇] Donders Institute for Brain, Cognition and Behaviour, Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands

ABSTRACT

Convolutional neural networks (CNN) have been widely used for visual recognition tasks including semantic segmentation of images. While the existing methods consider uniformly sampled single- or multi-scale patches from the neighborhood of each voxel, this approach might be sub-optimal as it captures and processes unnecessary details far away from the center of the patch. We instead propose to train CNNs with non-uniformly sampled patches that allow a wider extent for the sampled patches. This results in more captured contextual information, which is in particular of interest for biomedical image analysis, where the anatomical location of imaging features are often crucial. We evaluate and compare this strategy for white matter hyperintensity segmentation on a test set of 46 MRI scans. We show that the proposed method not only outperforms identical CNNs with uniform patches of the same size (0.780 Dice coefficient compared to 0.736), but also gets very close to the performance of an independent human expert (0.796 Dice coefficient).

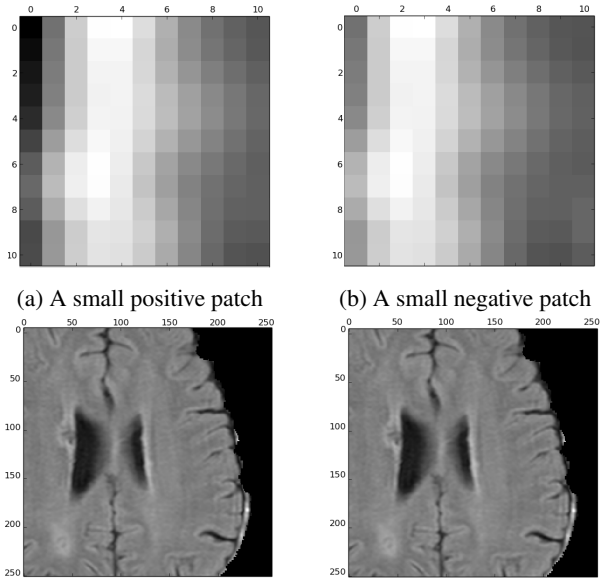
***Index Terms*—** non-uniform patch, convolutional neural network, deep learning, white matter hyperintensity

1. INTRODUCTION

White matter hyperintensities (WMH) are a common finding on brain MR images of patients diagnosed with small vessel disease (SVD) and several other neurological disorders. WMHs often represent areas of demyelination found in the white matter of the brain and are best observable in fluid-attenuated inversion recovery (FLAIR) MR images, as high value signals [1].

As manual segmentation of WMHs is laborious and subject to inter- and intra-rater variability, in the past decade a multitude of algorithms have been proposed to automate this process. Most of these methods use either an unsupervised clustering of WMHs as outliers or a supervised learning approach with hand-crafted features. None of these methods has been accurate enough to be considered as a stand-alone system [2].

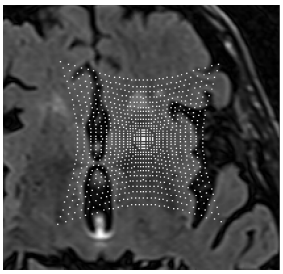
Since the past few years convolutional neural networks (CNN) have been reported to be the state of the art in most of visual recognition tasks and in particular in image classification and object detection. A popular way to extend image classifying CNNs for a segmentation problem is to train them to predict the label for each voxel given a small patch representing a local neighborhood of that voxel [3]. Nonetheless the chosen patch size might impose natural limitations hindering success of such segmentation systems in many medical image analysis applications, where the anatomical location of the imaging features is of crucial importance; small patches lack enough contextual information, while larger patch sizes, apart from higher computational costs, decrease the localization accuracy [4]. Figure 1 illustrates this.



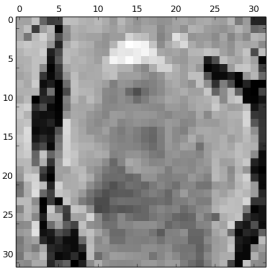
(a) A small positive patch (b) A small negative patch

(c) A large positive patch (d) A large negative patch

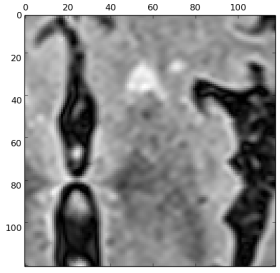
Fig. 1: A comparison of visual differences between two adjacent positive and negative patches in a small (11×11) and a large patch size (256×256). Evidently it is much easier to differentiate (a) from (b) rather than (c) from (d).



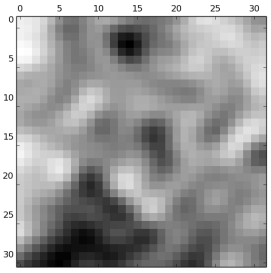
(a) Non-uniformly sampled voxels



(b) Resulted non-uniformly sampled patch



(c) Uniformly sampled patch with a similar extent



(d) Uniformly sampled patch with the same size

Fig. 2: An illustration of the patch sampling process from a FLAIR slice ($\alpha = 0.04$).

A way to address this problem is to break the unnecessary assumption of uniform patch sampling. The human visual system also non-uniformly perceives the world, with a lot of details at the focal point but a compact contextual representation from the surroundings. Inspired by the way our natural visual system performs, we propose to take non-uniformly sampled patches to train deep CNNs and we apply such a system for segmentation of WMHs, where a comprehensive inclusion of contextual information matters for a decent segmentation [5]. We show that this sampling approach outperforms similar networks with uniform sampling.

2. METHODS

2.1. Non-uniform patch sampling

Suppose $P_{i,j,k}$ is a $n \times n$ patch that we want to non-uniformly sample to represent a local neighborhood of voxel coordinate (i, j, k) from an image I. Then we have:

$$P_{ijk}(a + \lfloor \frac{n}{2} \rfloor, b + \lfloor \frac{n}{2} \rfloor) = I(i + l, j + m, k) \qquad (1)$$

where a and b, integers belonging to the interval $[-\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor)$, are offsets from the center of the patch being sampled and l and m, offsets of the corresponding voxel from the image I, are computed as:

$$l = \lfloor a.e^{\alpha\sqrt{a^2+b^2}} + \frac{1}{2} \rfloor \qquad (2)$$

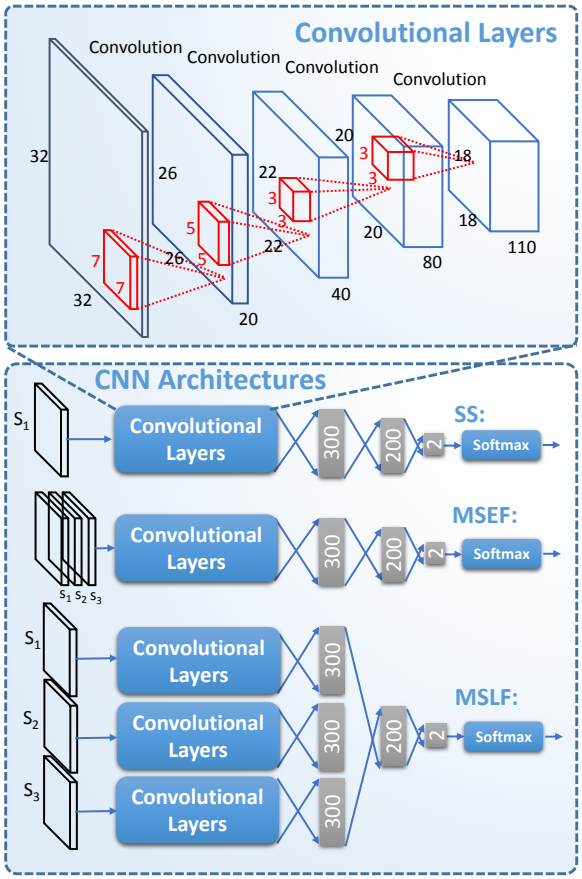


Fig. 3: CNN architectures used in this study. From top to bottom: single-scale, multi-scale early fusion and multi-scale late fusion with patches from three scales (S_1, S_2, S_3).

$$m = \lfloor b.e^{\alpha\sqrt{a^2+b^2}} + \frac{1}{2} \rfloor \qquad (3)$$

where α is a controlling factor indicating the extent of the patch, and $\alpha = 0$ will result in uniformly sampled patches. An intuitive way to see these equations is as we get further away from the center of the patch (larger absolute values for a and b) the x- and y-axis offsets of the voxels to be sampled from the image (l and m) grow exponentially. This implies a dense sampling on the center and less dense sampling from the sides. Figure 2 visualizes the sampled voxels for the mentioned non-uniform patch creation (2a) and the resulted non-uniformly sampled patch (2b) and compares it with uniformly sampled patches with a similar patch extent (2c) and the same patch size (2d).

2.2. CNN architecture and the training procedure

We create input patches with $n = 32$ and three different values for the α parameter ($\alpha = 0.01, 0.02, 0.04$). We use an eight layers CNN as depicted in the top architecture in Figure 3. This network consists of four convolutional layers that have

20, 40, 80 and 110 filters of size 7×7, 5×5, 3×3, 3×3 respectively. Then we apply three layers of fully connected neurons of size 300, 200 and 2. Finally the resulting responses are turned into probability values using a softmax classifier. The type of non-linearity that we apply to each neuron is a rectified linear unit, which is known to prevent the vanishing gradient problem in deep networks. We do not use pooling as it results in a shift-invariance property [6], which is not desired in segmentation tasks.

We train our network with the stochastic gradient descent algorithm with a mini-batch size of 128 and the cross-entropy cost function. We also use RMSPROP algorithm to speedup the learning process by adaptively changing the learning rate for each parameter. Random initialization of the weights is crucial in order to break the symmetry among the units the same layer. Thus we randomly sample the initial weights from a $(0, \frac{1}{\sqrt{m}})$ Gaussian distribution. CNNs are complex architectures that are likely to easily overfit training-set-specific patterns, thus to add a form of regularization and also to prevent co-adaptation of feature detectors, we use drop-out with a ratio of 0.3 on all of the layers in the network. We train our network for 10 epochs, which we found was sufficient for convergence, and we pick the set of weights with the best A_z on a validation set. We utilize the Theano library [7] for the implementation.

3. EXPERIMENTAL SETUP

3.1. Data

The data used for training, validation and evaluation of the proposed methods, is provided by the RUN DMC [8], which is a cohort study including T1 and FLAIR images of SVD patients. The part of the dataset that we use for this study consists of 466 cases that were annotated by either one (420) or two trained readers (46). We use the 46 subjects with two annotations for testing purposes and separate the rest into two sets of 378 and 42 for training and validation respectively. There are several preprocessing steps that have to be taken before the images are ready for patch extraction: We first perform a rigid registration of T1 to FLAIR images. Then we extract the brains from the T1 images and transfer and apply the resulting masks to the FLAIR images. A bias field correction is then performed. We use the FSL package [9] for the three mentioned steps. Finally we normalized the image intensities to be within the range of $[0, 1]$. Extracting patches from the training and validation sets of subjects results in a balanced dataset of 3.88M and 430K patches respectively.

3.2. Evaluation and comparison

In order to evaluate the effectiveness of our non-uniform patch sampling method, we compare it to three alternative approaches trained and validated with uniformly sampled patches on the same datasets as illustrated in Figure 3:

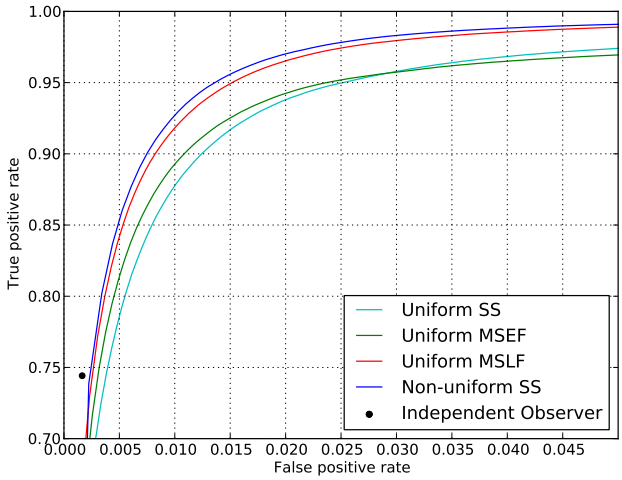


Fig. 4: An ROC comparison of different methods and the independent human observer.

- *Single-scale (Uniform SS)*: Similar to most of the methods, we train a similar single-scale network on uniformly sampled patches with the same size (32×32).
- *Multi-scale early fusion (Uniform MSEF)*: As smaller patches are better for an accurate localization, while larger patches capture more contextual information, a CNN given multiple patches with varying sizes would benefit from both. One possible architecture to fuse the information from multi-scale patches is to fuse them in the input layer. For each candidate voxel we extract 128×128, 64×64 and 32×32 patches. Then we downsample the two larger patches to 32×32 and feed them to a single CNN as different input channels.
- *Multi-scale late fusion (Uniform MSLF)*: Another fusion possibility to leverage the information in the multi-scale patches, is to input each scale separately into several convolutional layers. Then we can fuse the representation features from each scale and pass it forward to more fully connected layer. We use the same three scales as mentioned for MSEF.

The metrics that we use to compare these methods are Dice similarity coefficient and area under the receiver operating characteristic (ROC) curves (A_z). We also provide p-values for a statistical significance test with bootstrapping and measuring the Dice coefficient.

4. RESULTS

Table 1 demonstrates the performance of the proposed algorithm given three different α values. Figure 4 and Table 2

Table 1: A comparison of the non-uniform sampling method with different α values.

Method	Validation A_z	Test Dice	Test A_z
$\alpha = 0.01$	0.9958	0.756	0.9943
$\alpha = 0.02$	0.9963	0.780	0.9955
$\alpha = 0.04$	0.9955	0.779	0.9954

Table 2: A comparison of the test set Dice and A_z of the non-uniform sampling method ($\alpha = 0.02$) to different methods.

Method	Dice	A_z
Uniform SS	0.736	0.9895
Uniform MSEF	0.759	0.9867
Uniform MSLF	0.776	0.9937
Non-uniform SS	0.780	0.9955
Independent observer	0.796	-

compare the best performing non-uniform sampling method ($\alpha = 0.02$) to uniform sampling methods and an independent human observer with ROC curves, Dice and A_z on the test set. Table 3 shows statistical significance test p-values for pairwise comparison of different methods.

5. DISCUSSION AND CONCLUSIONS

As shown by the experiments, a CNN with non-uniform patch sampling can significantly outperform an identical network with the same amount of uniformly sampled data from the input image (uniform SS). This happens as non-uniform sampling enlarges the extent of the patch and thus provides more contextual information to the CNN. Multi-scale approaches also aim to capture more contextual information with larger scales and improve over the uniformly sampled single-scale patches. However, the experimental results suggest an advantage of single-scale non-uniform sampling over uniform multi-scale approaches although their sample size is larger by a factor of 3. This seems to be a consequence of the fact that a single non-uniformly sampled patch not only contains both details on the focal part and large context information, but also demands a simpler model with less weights for training. As an inherent limitation for this method, we do not know yet if it is possible to benefit from a practical speed-up by turning it into a fully convolutional network.

Table 3: Statistical significance tests for comparison of different methods. p_{ij} represents p-vlaue for a one-sided test checking whether method in row i is better than method in column j.

Method	UMSEF	UMSLF	NUSS	Ind. Obs.
USS	<0.01	<0.01	<0.01	<0.01
UMSEF	-	<0.01	<0.01	<0.01
UMSLF	-	-	0.23	0.05
NUSS	-	-	-	0.03

6. REFERENCES

[1] J. M Wardlaw, E. E Smith, G. J Biessels, et al., "Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration," *The Lancet Neurology*, vol. 12, no. 8, pp. 822–838, 2013.

[2] M. E. Caligiuri, P. Perrotta, A. Augimeri, F. Rocca, A. Quattrone, and A. Cherubini, "Automatic detection of white matter hyperintensities in healthy aging and pathology using magnetic resonance imaging: A review," *Neuroinformatics*, vol. 13, no. 3, pp. 1–16, 2015.

[3] D. Ciresan, A. Giusti, L. M Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, 2012, pp. 2843–2851.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer International Publishing, 2015.

[5] M. Ghafoorian, N. Karssemeijer, I. van Uden, F.E. de Leeuw, T. Heskes, E. Marchiori, and B. Platel, "Small white matter lesion detection in cerebral small vessel disease," in *SPIE Medical Imaging*, 2015, vol. 9414, pp. 941411–941411.

[6] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks–ICANN 2010*, vol. 6354 of *Lecture Notes in Computer Science*, pp. 92–101. 2010.

[7] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. J. Goodfellow, A. Bergeron, N. Bouchard, and Y. Bengio, "Theano: new features and speed improvements," Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.

[8] A. G. van Norden, K. F. de Laat, R. A. Gons, et al., "Causes and consequences of cerebral small vessel disease. The RUN DMC study: a prospective cohort study. Study rationale and protocol," *BMC Neurology*, vol. 11, pp. 29, 2011.

[9] M. Jenkinson, C. F. Beckmann, T. EJ Behrens, M. W Woolrich, and S. M Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.