PWML DETECTION IN 3D CRANIAL ULTRASOUND VOLUMES USING OVER-SEGMENTATION AND CLASSIFICATION WITH DEEP NEURAL NETWORKS

Flora Estermann [1], *Valérie Kaftandjian* [2], *Philippe Guy* [2], *Philippe Quetin* [3], *Philippe Delachartre* [1]

[1] Univ Lyon, INSA Lyon, Université Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France

[2] Univ Lyon, INSA Lyon, Laboratoire Vibrations Acoustique (LVA), F-69621 Villeurbanne, France

[3] CH Avignon, France

ABSTRACT

Punctate white matter lesions (*PWML*) are the most common white matter injuries observed in preterm neonates. Automatic detection of these lesions could better assist doctors in diagnosis. Recent advances in deep learning have resulted in optimistic results on many MR biomedical image benchmark datasets, but few methods seem to tackle the detection of very small lesions in ultrasound images. In this paper, we propose a two-phase strategy. Firstly, we highlight the foreground information by aggregating the lesions in the ground truth along the coronal projection of the brain, then we train a segmentation network to detect PWML with the resulting over-segmented masks. Secondly, a classification network is used to eliminate false alarms and improve the accuracy of the model. Experimental results demonstrate the effectiveness of our method to detect PWML in ultrasound images, improving the recall by 13% compared to the best published models, while limiting the number of false alarms efficiently.

***Index Terms*—** Deep Learning, Automatic Anomaly Detection, 3D Ultrasound Imaging, White Matter Injury, U-Net.

1. INTRODUCTION

Punctate white matter lesions (*PWML*) are responsible for neurodevelopmental sequelae in early childhood [1] [2]. Accurate detection and segmentation of these lesions by an automatic algorithm could better assist doctors in diagnosis. Currently, MRI is the gold standard for lesion detection, but this procedure is expensive and not easily to access. As ultrasound is routinely performed on newborns, this modality could actually be of real interest for the detection of lesions and would allow more children with PWML to benefit from medical care.

Many papers present unsupervised approaches using CNNs for breast tumors classification [3], and brain tumor segmentation [4], but few methods seem to tackle the detection of very small lesions. On the other hand, on the side of supervised methods, research on automatic detection of PWML in MR images was first tackled by Mukherjee et al. [5], that proposed a method that considers the correlation of pixels in 3D space. Besides, Liu et al. [6] proposed a spatiotemporal transformation structure in order to use the information between adjacent slices, which achieves higher performance than existing methods and consumes less computing resources than 3D CNNs. Despite the high contrast and low noise of MR images, the reported accuracy for the PWML detection task remains low with a Dice under 0.60 and a recall at 0.65. Finally, Erbacher et al. [7] started working on this task with cranial ultrasound (*cUS*) images and introduced a novel deep architecture based on the 2D U-Net backbone and a soft attention model focusing on the PWML localization, called Priority U-Net, with the recall and the precision in the PWML detection task reaching 0.5370 and 0.5043, respectively.

As shown above, the detection and segmentation of PWML on cUS is extremely challenging. First, despite a higher resolution than MRI, US images are difficult to analyse because of their low contrast, the presence of speckle and the high variability related to the data acquisition process. Second, PWML are very tiny (in our dataset, the median volume of the lesions is less than $1mm^3$), resulting in the lesion area being much smaller than the whole brain area, so the data imbalance problem between positive and negative pixels makes it a huge challenge for deep learning models to learn. Finally, the presence of numerous artifacts in the brain, very similar to true lesions in terms of gray levels, sometimes location or even shape, makes the distinction between true lesions and false alarms even more complicated. Recently, Dakak et al. [8] developed an automatic approach to analyze discontinuities on industrial CT volumes by reducing the number of false alarms using a U-Net for image segmentation followed by a classification step performed by a second neural network called CT-Casting Net.

In this work, we first train a Priority U-Net to perform over-segmentation on the expanded groundtruth in order to increase the number of detected lesions (higher recall). In a second step, a modified Casting Net is used to eliminate false alarms and improve the accuracy of our model (higher precision). Finally, when compared with other methods, our model achieves better results in the detection of PWML.

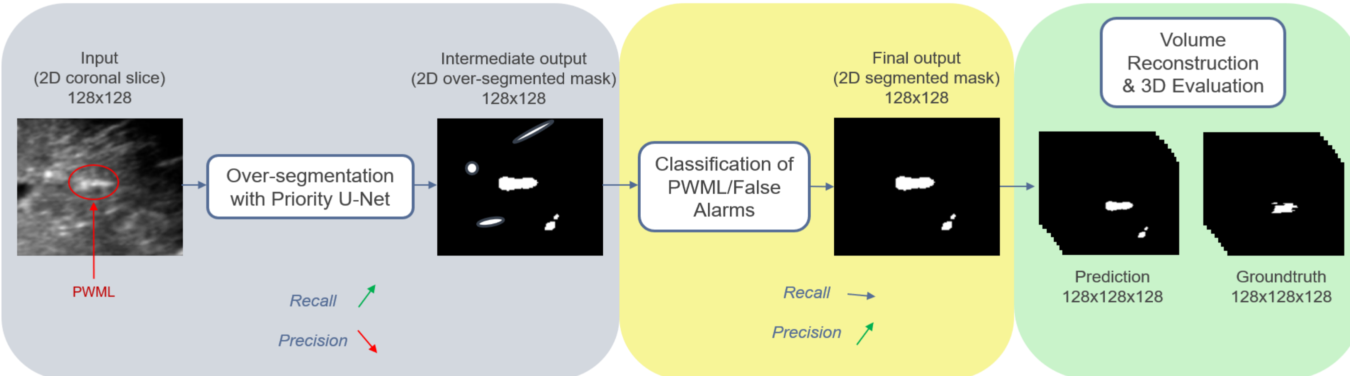

Fig. 1: PWML detection pipeline : The over-segmentation of lesions with the Priority U-Net improves the recall but produces many false positives. A second stage of classification allows us to differentiate true lesions from false alarms and improves the precision of the intermediate output. 2D predictions are then concatenated and compared to the 3D groundtruth for final assessment.

2. METHODOLOGY

2.1. Over-segmentation with Priority U-Net

Before training the model, the lesions in the ground truth are expanded by aggregating the foreground pixels within a 5-slice sliding block along the coronal projection of the brain volume. It results in a label image with a higher percentage of foreground than the original label image, causing additional losses and helping to make training more effective.

Segmentation training is then performed on the modified ground truth with the Priority U-Net model (Fig. 1), which includes layers relying on 3D probabilistic maps derived from a spatial prior knowledge of PWMLs location and computed on the training dataset.

The output of the network usually includes many false positives, that will be removed during the next classification step.

2.2. Classification of PWML & False alarms

Table 1: Architecture of the CT-Casting-Net. [8]

Layer	Layer Type	Kernel Size	#Filters
1	Conv2D + ELU	5x5	64
2	MaxPooling2D	2x2	64
3	Conv2D + ELU	5x5	128
4	MaxPooling2D	2x2	128
5	Conv2D + ELU	5x5	256
6	MaxPooling2D	2x2	256
7	Conv2D+ReLU+Dropout	5x5	512
8	MaxPooling2D	2x2	512
9	Flatten + FC + ReLU	-	256
10	FC + Softmax	-	2

In parallel, 3 CNN classifiers based on the Casting Net Architecture (Table 1) are trained on a different projection, to predict the class of smaller 2D patches (32x32) extracted from the volume and centered on the connected components of the predicted mask. The goal is to teach the network to differentiate true lesions from artifacts present in the brain, at a smaller scale.

During the testing phase, once the 3D intermediate mask is obtained after the first step of over-segmentation, 2D patches are extracted around the regions of interest (thumbnails from the image, centered around the connected components of the predicted mask) on the 3 orthogonal projections of the brain and sent to the corresponding classifier to identify the true PWML (Fig. 2). The intermediate mask is then corrected (lesions predicted as false alarms are removed).

2.3. Class imbalance issue

As mentionned in the introduction, because the lesions are very small, the data imbalance problem between lesionnal pixels and background makes it a huge challenge for deep learning models to learn correctly. Therefore, the use of specific loss functions is considered to overcome this issue.

Self-Balancing Focal Loss: For a better training of the Priority U-Net, the Self-Balancing Focal Loss (SBFL) introduced by Liu et al. [6] is used in addition to the Dice Loss. The SBFL divides the whole loss into foreground loss and background loss part. It can automatically balance these losses and ultimately boosts the performance of the model.

$$SBFL_0 = -(1 - y_{pred}) \times y_{pred} \times log(1 - y_{pred} + \epsilon)$$

$$SBFL_1 = -y_{pred} \times (1 - y_{pred}) \times log(y_{pred} + \epsilon)$$

$$\beta = \frac{0.4 \times sum(SBFL_0)}{sum(SBFL_0) + sum(SBFL_1)} + 0.5$$

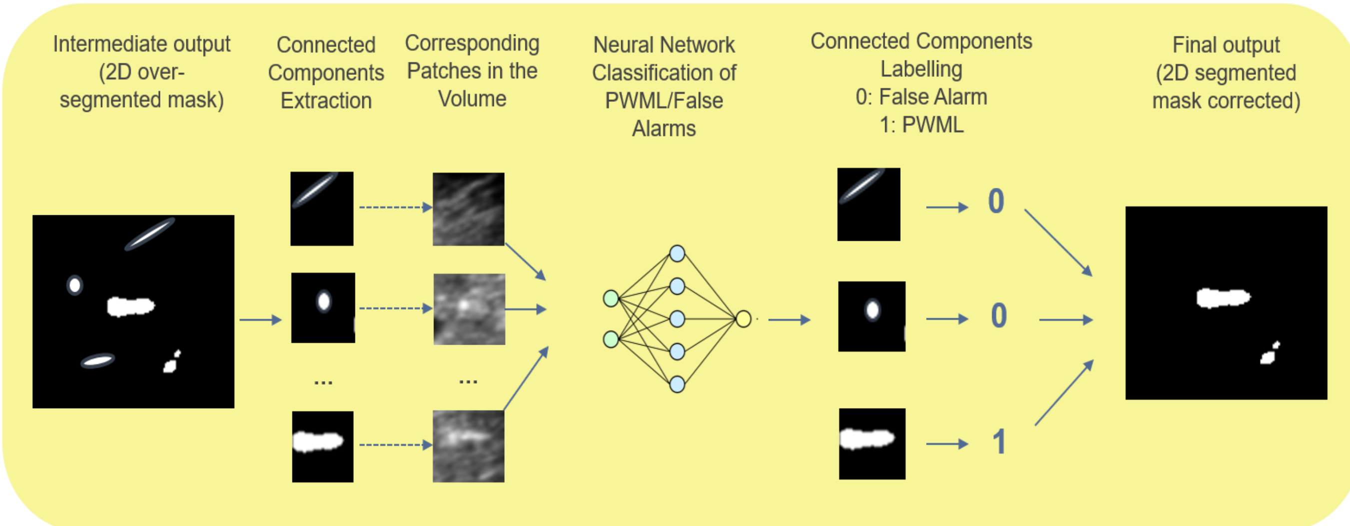

Fig. 2: PWML detection pipeline (2nd step): Classification of PWML/False Alarms. The classification network 1 helps to differentiate between actual lesions and other brain artifacts at a smaller scale. The previous mask obtained after the over-segmentation is corrected with the output of the classification.

$$SBFL = \beta \times SBFL_1 + (1 - \beta) \times SBFL_0$$

Where y_{pred} is the intermediate output of the Priority U-Net and $SBFL_0$ and $SBFL_1$ are the focal loss of background and foreground pixels respectively. In order to segment PWML, we will focus on the segmentation of the lesion areas when balancing the loss of positive and negative samples. That is, β should always be between 0.5 and 0.9 to ensure that the model does not only focus on the segmentation of positive areas. For that reason, we constrain β not to exceed 0.9 by applying a coefficient of 0.4 to the equation. Finally, the SBFL is composed of $SBFL_1$ weighted by β and $SBFL_0$ weighted by $1 - \beta$.

Weighted Cross-Entropy Loss: For the binary classification of the patches, the Casting Net was modified to include Weighted Cross-Entropy. Indeed, under class imbalance, the model is seeing much more zeros than ones, so it will also learn to predict more zeros than ones because the training loss can be minimized by doing so.

Weighted Cross Entropy applies a scaling parameter α to Binary Cross Entropy, allowing us to penalize false positives or false negatives more harshly. When α is greater than 1, the model penalizes more on false negatives, hence helping increase Recall. On the other hand, when α is less than 1, the model penalizes more on false positives, hence helping increase Precision.

$$L_{BCE} = -y_{true} \times log(y_{pred}) - (1 - y_{true}) \times log(1 - y_{pred})$$

$$L_{WBCE} = -\alpha y_{true} \times log(y_{pred}) - (1 - y_{true}) \times log(1 - y_{pred})$$

Because our training dataset contained more negative than positive examples, but as we wanted to limit the trade-off between Precision and Recall, we performed a grid-search over a 10-fold cross-validation and finally set α to 1.5.

3. EXPERIMENTS & RESULTS

3.1. Dataset

The 2D images are extracted from 54 reconstructed US brain volumes (including 29 with PWML) from 45 preterm babies whose mean age at birth was 31.6 ± 2.5 gestational weeks. In total, the dataset without preprocessing contains 414 lesions. The smallest lesion barely reaches $0.02mm^3$, while the largest is more than $41mm^3$. The median lesion size is $0.72mm^3$, which is extremely tiny. Besides, PWML have quite varied contrasts and do not really have specific shapes (punctate, ovoid or sometimes linear) 3. They are usually located in the center of hemispheres, near the lateral ventricles.

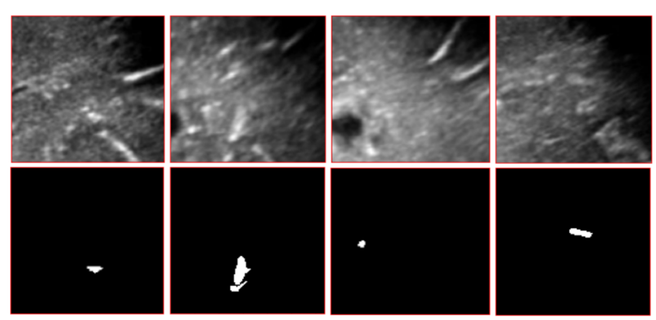

Fig. 3: PWML examples from the Brain US Dataset (images on top and label masks on bottom). PWML have varied contrasts and shapes, and are often difficult to distinguish from peripheral vessels or arteries in cross-section.

3.2. Data Preprocessing

As the acquisition process of the ultrasound images is performed manually by the pediatrician along the anterior/posterior axis of the brain, the brain scan does not always results in the same number of dynamic sequences (DICOM). In order to recover a complete volume, we completed this process by a reconstruction algorithm [9]. After that, a first preprocessing phase consists of extracting a sub-volume of size $128 \times 128 \times 128$ in the top-right hemisphere, periventricular region of the brain for each patient. Eventually, it is planned to process the entire brain by this process.

In order to compare with MRI, a first filtering is performed on the size of the lesions for each volume to limit the number of lesions that are too small and to make the problem less complex. As a result, only 90% of the lesional volume is kept for each patient, which allows us to get rid of the tiniest lesions, that are usually not even visible in the MRI.

For the first over-segmentation step, lesions in the ground truth are expanded as described in the section 2.1, and 2D images of size 128×128 containing PWML are extracted along the coronal projection of the brain to train the Priority U-Net. Horizontal flipping is randomly applied with a probability of 0.5 for data augmentation.

For the second classification step, patches of size 32×32 are extracted from the original volumes and used to train the 3 classifiers along each projection of the brain (axial, coronal and sagittal), including various examples with real lesions, artifacts/false alarms and normal examples. Random affine transforms (rotation, shearing, scaling, and translation) and flipping were employed for data augmentation.

3.3. Experimental Setup

The proposed pipeline was implemented in Python 3.9 with Keras and TensorFlow backend. All the models were trained and tested with GPU. For each model, we performed a 10-fold cross validation with 3 patients in the validation set and 27 patients in the training set.

The Priority U-Net was trained for 25 epochs with the Self-balancing Focal Loss and the Dice Loss, whereas the Casting-Net was trained for approximately 70 epochs using the Weighted-binary Cross-entropy Loss. The batch size is 4 for the segmentation network and 32 for the classification network. The initial learning rate was fixed at 10e-1 with the Adam optimizer and automatically decreased by a factor 0.1 when validation loss did not improved for 10 epochs.

3.4. Results

Table 2: Final results of the proposed approach (Priority U-Net + Classification) compared to the U-Net and the original Priority U-Net. All these results are the medians of 10-folds cross-validation.

Model	Recall	Precision	F1-Score
U-Net	53.41	54.12	50.86
Priority U-Net	60.22	51.40	55.46
Priority U-Net (HF)	**73.32**	47.72	55.77
Priority U-Net (HF) + Classification	**68.91**	**56.00**	56.23

To quantitatively assess the quality of the PWML detection produced by the target pipeline, we employed 3 criteria to evaluate each model : The recall, the precision and the F1-score. For each of these metric, the closer to 1 the better. The quantitative results are shown in Table 2.

The evolution of the results after training the Priority U-Net with the highlighted foreground (*HF*) seems to indicate that expanding the PWML in the label images and performing over-segmentation on them actually helps the network to detect smaller objects, thus significantly increasing the recall of 13% compared to the results obtained with the original Priority U-Net. After the classification, the recall slightly declines, while the precision improves from 47 to 56%, which compensates for the degradation observed in the previous step.

As a result, our model achieves better performance for PWML detection in US images, with a higher sensitivity to very small lesions. We believe that the precision could still be improved by providing more spatial information to the model.

4. DISCUSSION AND CONCLUSION

In this paper, we preprocessed the label images to get a higher percentage of foreground, before training a network to perform an over-segmentation of the original lesions. This first step was followed by a second step of classification to limit the number of missclassified artifacts in the brain. At the end of the pipeline, we reach a higher recall and precision (69% and 56% respectively) than those obtained with the Priority U-Net.

Many previous studies have focused on MRI-based lesion detection, and very few have addressed problems on a scale as small as PWML. While relatively few have conducted this task on US images, this work highlights once again the possibility of detecting brain lesions through ultrasound imaging.

Nevertheless, we anticipate that results might be improved by enriching the database with more patients. Improvements on the segmentation results and using multimodal fusion for classification are currently under study and will be the object of future works as well.

5. ACKNOWLEDGMENTS

This work was supported by the LABEX CELYA (ANR-10-LABX-0060) operated by the French National Research Agency (ANR) and the University of Lyon, within INSA Lyon.

6. COMPLIANCE WITH ETHICAL STANDARDS INFORMATION

The data from human subjects used in this work were obtained and treated in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committees of the institutions involved in creating the PWML database, from which these data were accessed. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

[1] Annie L.A. Nguyen, Yang Ding, Sabrina Suffren, Irène Londono, David Luck, and Gregory A. Lodygensky, “The brain’s kryptonite: Overview of punctate white matter lesions in neonates,” *International Journal of Developmental Neuroscience*, vol. 77, no. 1, pp. 77–88, October 2019.

[2] Clara Adriana Maria de Bruijn, Stefano Di Michele, Maria Luisa Tataranno, Luca Antonio Ramenghi, Andrea Rossi, Mariya Malova, Manon Benders, Agnes van den Hoogen, and Jeroen Dudink, “Neurodevelopmental consequences of preterm punctate white matter lesions: a systematic review,” *Pediatric Research*, September 2022.

[3] Jiao Z., Gao X., Wang Y., and Li J., “A parasitic metric learning net for breast mass classification based on mammography,” *Pattern Recognition*, vol. 75, pp. 292–301, 2018.

[4] Liu Y., Li J., Wang Y., Wang M., Li X., Jiao Z.and Yang J., and Gao X., “Refined-segmentation r-cnn: A two-stage convolutional neural network for punctate white matter lesion segmentation in preterm infants,” *arXiv e-prints*, vol. abs/1906.09684, pp. 1265–1274, June 2019.

[5] Subhayan Mukherjee, Irene Cheng, Steven Miller, Ting Guo, Vann Chau, and Anup Basu, “A fast segmentation-free fully automated approach to white matter injury detection in preterm infants,” *Medical and Biological Engineering and Computing*, vol. 57, no. 1, pp. 71–87, January 2019.

[6] Yalong Liu, Jie Li, Miaomiao Wang, Zhicheng Jiao, Jian Yang, and Xianjun Li, “Trident segmentation cnn: A spatiotemporal transformation cnn for punctuate white matter lesions segmentation in preterm neonates,” *Med Biol Eng Comput*, 2019.

[7] Pierre Erbacher, Carole Lartizien, Matthieu Martin, Pedro Foletto Pimenta, Philippe Quetin, and Philippe Delachartre, “Priority u-net: Detection of punctuate white matter lesions in preterm neonate in 3d cranial ultrasonography,” *submitted to MIDL*, 2020.

[8] Abdel Dakak, Valérie Kaftandjian, Philippe Duvauchelle, and Patrick Bouvet, “Establishment of acceptance criteria for tomography inspection of aluminium alloy castings,” *International Symposium on Digital Industrial Radiology and Computed Tomography*, October 2019.

[9] M. Martin, B. Sciolla, M. Sdika, X. Wang, P. Quetin, and P. Delachartre, “Automatic segmentation of the cerebral ventricle in neonates using deep learning with 3d reconstructed freehand ultrasound imaging,” *IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2018.