A Two Cascaded Network Integrating Regional–based YOLO and 3D-CNN for Cerebral Microbleeds Detection

Mohammed A. Al-masni, Woo-Ram Kim, Eung Yeop Kim, Young Noh, and Dong-Hyun Kim*, *Member, IEEE*

***Abstract*— Cerebral Microbleeds (CMBs) are small chronic brain hemorrhages, which have been considered as diagnostic indicators for different cerebrovascular diseases including stroke, dysfunction, dementia, and cognitive impairment. In this paper, we propose a fully automated two-stage integrated deep learning approach for efficient CMBs detection, which combines a regional-based You Only Look Once (YOLO) stage for potential CMBs candidate detection and three-dimensional convolutional neural networks (3D-CNN) stage for false positives reduction. Both stages are conducted using the 3D contextual information of microbleeds from the MR susceptibility-weighted imaging (SWI) and phase images. However, we average the adjacent slices of SWI and complement the phase images independently and utilize them as a two-channel input for the regional-based YOLO method. The results in the first stage show that the proposed regional-based YOLO efficiently detected the CMBs with an overall sensitivity of 93.62% and an average number of false positives per subject (FP_{avg}) of 52.18 throughout the five-folds cross-validation. The 3D-CNN based second stage further improved the detection performance by reducing the FP_{avg} to 1.42. The outcomes of this work might provide useful guidelines towards applying deep learning algorithms for automatic CMBs detection.**

***Index Terms*— Cerebral Microbleeds; Computer-Aided Detection; Deep Learning**

I. Introduction

Cerebral Microbleeds (CMBs) are small foci of chronic brain hemorrhages that are generated by structural malformation of the small blood vessels and the deposits of blood products. CMBs have a high prevalence in several populations, including healthy elderlies [1]. The presence of microbleeds could increase the possible clinical implications of ischemic stroke, traumatic brain injury, and Alzheimer's diseases. Indeed, direct pathological observations have also revealed that CMBs bring about damage to the surrounding brain tissue, which cause dysfunction, dementia, and cognitive impairment [2]. Therefore, accurate differentiation of CMBs from different suspicious regions (i.e., CMB mimics) such as calcifications, irons, and veins is important for proper diagnosis and appropriate treatment.

In clinical practice, Magnetic Resonance (MR) imaging specifically with the modern advances of using gradient-echo (GRE) and susceptibility-weighted imaging (SWI) is usually preferred for screening the microbleeds in brain images [3]. Although SWI improves the identification of CMBs, visual inspection by neuroradiologists is still time-consuming, fault-prone, laborious, and subjective. Therefore, a need of automated computer-aided detection (CAD) system of CMBs is indispensable. However, automated detection of CMBs is a challenging task due to the large variation among the CMBs locations in the brain, their small sizes, and the presence of CMB mimics such as calcifications [2].

The past decade has witnessed numerous investigations to solving this challenging task. The two-dimensional fast radial symmetry transform (2D-FRST) algorithm has been widely utilized to detect putative CMBs [4]. Recently, convolutional neural networks (CNNs) have gained a lot of attention in various medical imaging applications, including brain, breast, lung, and skin cancers detection [5], classification [6, 7], segmentation [8, 9] as well as in the detection of CMBs. This is due to the capability of CNNs in designing and extracting robust and effective hierarchy features, which lead to significant improvement of detection performance. In fact, most of the recent works on automatic detection of CMBs were accomplished in a two-cascaded framework; detection stage for potential candidate revelation and discrimination stage for FPs reduction [10-12]. In 2016, Dou et al. proposed a two-stage CNNs for automatic detection of CMBs using SWI images [10]. They exploited the 3D fully convolutional network (3D-FCN) as screening stage followed by 3D-CNN as discrimination stage. Their cascaded network achieved a sensitivity of 93.16%, a precision of 44.31%, and an average number of FPs per subject (FP_{avg}) of 2.74. More recently, Liu et al. developed a 3D-FRST for candidate detection stage using SWI images and 3D deep learning residual network (3D-ResNet) for the FPs reduction stage using both SWI and phase images [11]. This framework obtained a sensitivity of 95.80%, a precision of 70.90%, and an FP_{avg} of 1.6. Similar work was proposed by Chen et al., which integrated 2D-FRST [4] and 3D-ResNet using 7.0 T SWI images [12]. The detection performance of CMBs was improved over using only the 2D-FRST and achieved a sensitivity, precision, and FP_{avg} of 94.69%, 71.98%, and 11.58, respectively.

The contributions of this paper are summarized as follows. We develop a completely integrated deep learning method for efficient CMBs identification throughout a combination of the regional-based CNN method based on You Only Look Once (YOLO) utilized for CMBs candidate detection and 3D-CNN used for FPs reduction. For the first stage, we average the adjacent slices of SWI and complement phase images

M. A. Al-masni and D.-H. Kim are with the Department of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul, Republic of Korea (e-mails: m.almasani@yonsei.ac.kr and donghyunkim@yonsei.ac.kr).

W.-R. Kim and Y. Noh are with the Department of Neurology, Gachon University Gil Medical Center, Incheon, Republic of Korea (e-mail: wrkim9201@gmail.com and ynoh@gachon.ac.kr).

E. Y. Kim is with the Department of Radiology, Gachon University Gil Medical Center, Incheon, Republic of Korea (e-mail: neuroradkim@gmail.com).

*Corresponding author.

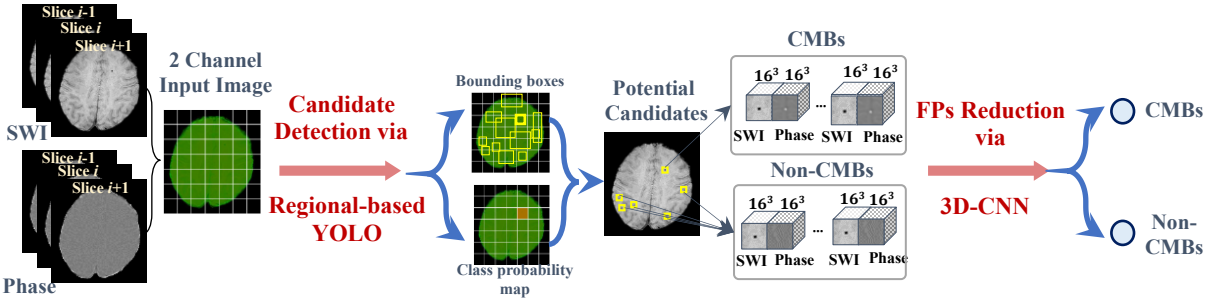


Figure 1. Overview diagram of the proposed two-stage deep learning approach.

independently and utilize them as a two-channel input for regional-based YOLO method. These settings enable YOLO to learn more reliable and representative hierarchal features and hence achieve better detection performance compared to the using of only one-channel image. The proposed method outperforms the commonly utilized FRST strategy and achieves superior detection performance in terms of sensitivity and FP_{avg} with lower computational cost.

II. Materials and Methods

The proposed deep learning two-stage approach for CMBs detection integrates a regional-based YOLO for potential candidate detection and a 3D-CNN for discrimination of actual CMBs from putative false positives. An overview diagram of the proposed work is illustrated in Fig.1.

A. 1st Stage: Regional-based YOLO for Candidates Detection

You Only Look Once (YOLO) is one of the recent regional deep learning CNN techniques that originated for object detection in the images [13]. It is able to simultaneously detect the locations of objects in the input images and classify them into different categories. Unlike the conventional R-CNN and Fast-RCNN, YOLO proceeds both detection and classification tasks utilizing a single convolutional network, which provides bounding boxes' coordinates along with their class predictions. Thus, YOLO is able to encode the global contextual information via looking into the entire input image only once. The input composite MR image is divided into $S \times S$ non-overlapped grid cells, where $S = 7$ in this work. For each microbleed presented on the input image, only one grid cell is responsible to detect it. In the end, each grid cell is represented by B bounding boxes known as anchors associated with five components: $x, y, width, height, and\ confidence$. The (x, y) represent the center location coordinates of CMB in the MR image, while the $confidence$ reflects the probability of the presence of microbleeds within a certain $width$ and $height$ from the center location. In other words, $confidence$ could be defined as an intersection over union (IOU) between the predicted boxes by YOLO and the ground-truth labels by neuroradiologists.

In this paper, we employed YOLOv2 for more accurate detection of CMBs from brain images [13]. This version of YOLO adds a variety of ideas to the initial YOLO for better detection results and faster performance, including batch normalization, fine-tuning with high-resolution network, using anchor boxes, multi-scale training, and pass-through layer concatenation. The configuration of the utilized regional-based YOLO consists of 23 convolutional layers with kernel sizes of 3×3 and various feature maps; and five max-pooling layers with sizes of 2×2 and strides of two. Finally, a softmax classifier is placed at the top of the network to produce a tensor of predictions, containing the detected bounding boxes information and the predicted class.

B. 2nd Stage: 3D-CNN for FPs Reduction

As an integration with the potential candidate detection stage, the second stage is designed to robustly reveal the positive samples (i.e., CMBs) from a large number of negative samples (i.e., CMB mimics) as a classification task. The input to this stage is 3D cropped regions around the center position of the detected bounding boxes associated with higher confidence scores via regional-based YOLO. As successfully applied in [11], we exploit training the proposed 3D-CNN using both SWI and phase images with sizes of $16 \times 16 \times 16$ voxels for each. Thus, the 3D-CNN stage is able to learn richer contextual spatial representations of CMBs from 3D training samples with sizes of $16 \times 32 \times 16$voxels. The input size is chosen to be large enough to include the microbleeds and limit the computational workload. The structure of the proposed 3D-CNN contains five convolutional layers with kernel sizes of $3 \times 3 \times 3$ and two max-pooling layers with sizes of $2 \times 2 \times 2$ and strides of two. To increase the stability of the network and speed up the learning process, each convolutional layer is appended by batch normalization and rectified linear unit (ReLU) activation function. Then, the extracted features from these convolutional layers are flattened and passed into three fully-connected neural networks (FC-NN). Additionally, one dropout layer is utilized after the first FC-NN layer to regularize and prevent networks from overfitting. This network is trained using learning rate, batch size, and number of epochs equals to 0.001, 50, and 200.

C. Dataset

To evaluate the proposed two-stage deep learning approach for CMBs detection, we have collected a set of MR brain volumes with a collaboration of Gachon University Gil Medical Center. Human data acquisition was performed in accordance with the relevant regulations and guidelines. Written informed consent was obtained from all participants and the study was approved by Institutional Review Board of Gachon University Gil Medical Center. Data was acquired using 3.0 T Verio Siemens MRI scanners (Siemens Healthineers, Germany). A total of 72 unhealthy subjects including 188 microbleeds were collected with the following imaging parameters: repetition time (TR) of 27 ms, echo time (TE) of 20 ms, flip angle (FA) of 15º, pixel bandwidth (BW) of 120 Hz/pixel, in-plane resolution of 0.50×0.50 mm^2, volume size of 512×448×72 voxels, slice thickness of 2 mm, field of view (FOV) of 256×224, and scan time of 4.45 minutes.

The labeling process of CMBs was performed by expert neuroradiologists using both the pre-processed SWI and phase

images. This process was accomplished using our in-house software developed by Matlab 2019a (MathWorks Inc., Natick, MA, USA). The neuroradiologists were able to simultaneously examine the axial plane of the SWI and phase brain images of the same subject.

D. Data Preparation

In this work, data normalization, brain region extraction, and SWI generation were applied. In order to increase the consistency among the input intensities, the input 3D volume of each subject was normalized in range 0 to 1. Moreover, we applied the fast, robust, and automated brain extraction tool (BET) on the MR magnitude images to segment the head into brain and non-brain [14]. The generated binary mask of brain tissue via BET is then applied to the MR SWI and phase images. This process enables computer-aided detection algorithms to avoid getting more incorrect candidates of CMBs located outer the brain tissue region. Furthermore, for the input of the regional-based YOLO stage, all images were resized into 448×448 pixels by eliminating 64 zero rows from the top and bottom of the data. Regarding the 3D-CNN stage, the input of 16×16×16 voxels from each SWI and phase images was captured from the original data. For appropriate training of CNNs and overcoming the issue of insufficient training data, we applied the augmentation process to enlarge the training samples in both proposed stages. This process included rotation operation and horizontal and vertical flipping.

E. Training and Testing

In this study, the proposed deep learning two-stage approach proceeded using a k-fold cross-validation (k=5) to increase the reliability and effectiveness of the proposed work and reduce the bias error. To this end, our data was randomly split into five subsets based on the subject-level. Then, we considered one subset as a testing set, while the remaining four subsets were utilized as training and validation sets. This process was repeated k times to involve all subsets as testing. For each k-fold, the proposed networks of the candidate detection stage and FPs reduction stage should be trained, validated, and tested using separate sets.

III. Results and Discussion

To quantitatively evaluate the capability of the proposed deep learning two-stage approach for CMBs detection, we computed the following evaluation indices. Sensitivity and precision. We also used the average number of false positives per subjects (FP_{avg}).

A. Experiments on Proper Composite Input Data

Several attempts and investigations for selecting proper input were tested. All the experiments were evaluated using the first fold test on the CMBs candidate detection stage with a confidence score equals to 0.01. We generated different input combinations from the original magnitude, high-pass filtered phase and SWI images. The use of different images was motivated by the fact that clinical radiologists also use these various images when diagnosing CMB.

We investigated five different combinations of inputs: single channel of SWI, two-channels of SWI and phase, three-channels of SWI, phase, and magnitude, two-channels of SWI and complement of phase, and two-channels of SWI and complement of phase with averaging of adjacent slices. We found that the information from both SWI and phase images facilitated the regional-based YOLO network to learn appropriate representations of microbleeds, leading to improve the detection of CMBs with a sensitivity of 97.22% and FP_{avg} of 275.50. However, the information from the magnitude images did not add further details. Also, using the complement of phase besides the original SWI images enable the detection network to obtain better detection of CMBs with a sensitivity of 94.44% and FP_{avg} of 56.44. Moreover, the detection stage via the regional-based YOLO achieved superior detection performance of CMBs with a sensitivity of 100 % and less number of FPs of 53.71 when it was trained using SWI and complement phase images with additional details of microbleeds from the adjacent slices as shown Table I.

TABLE I. Performance of Regional-based YOLO on Different Composite Inputs Throughout First Fold Test

Composite Input Image	Sensitivity (%)	FP_{avg}
Single Channel: SWI	91.67	23.57
Two-Channels: SWI and Phase	97.22	275.50
Three-Channels: SWI, Phase, and Magnitude	88.89	69.50
Two-Channels: SWI and Complement of Phase	94.44	56.44
Two-Channels: Averaging the Adjacent Slices of both SWI and Complement of Phase	100.00	53.71

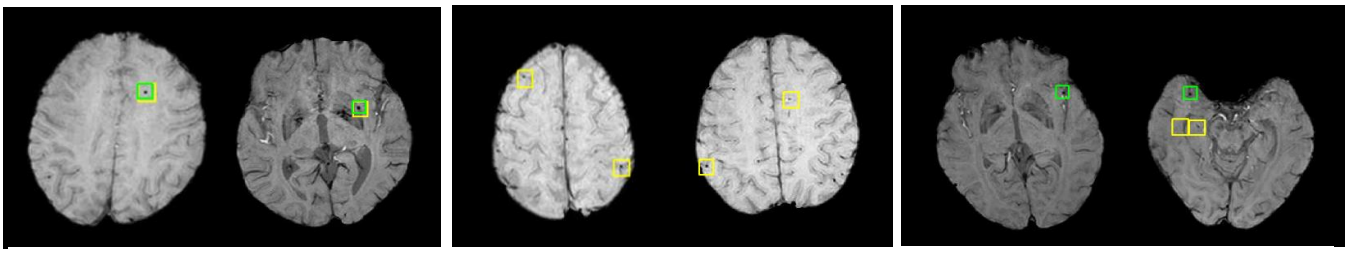
B. Performance of Regional-based YOLO (1st Stage)

For quantitative assessment, we present the overall candidate detection performance of the proposed regional-based YOLO technique throughout all five-folds test in Table II. The proposed candidate detection network achieves overall sensitivity and FP_{avg} of 93.62% and 52.18, respectively. It accurately detects 176 out of 188 microbleeds, while it only loses 12 microbleeds. Fig. 2 shows some examples of the true positives (TPs), false positives (FPs), and false negatives (FNs) that are generated by the regional-based YOLO.

C. Performance of 3D-CNN (2nd Stage)

The main aim of the 3D-CNN stage is to reduce the large number of false positive candidates that were generated through the candidate detection stage and further single out CMBs from the challenging ones. It is of note that the inputs to this stage represent all the generated potential candidates from the regional-based YOLO technique. Thus, the false negatives cases cannot be included in the FPs reduction stage. For the quantitative results of the proposed 3D-CNN, Table II presents the classification performance throughout all five-folds test. It is shown that the proposed network has the capability to differentiate between CMBs and non-CMBs candidates with high classification achievement. We obtain overall sensitivity of 94.32% and a precision of 61.94%. Finally, the proposed FPs reduction stage via the 3D-CNN was able to robustly reduce the FPs from 52.18 to 1.42.

We provide direct comparison through implementing the most commonly utilized 3D version of fast radial symmetry transform (FRST) and applied it to our dataset. As a comparison of the potential candidate detection stage, the proposed regional-based YOLO overcomes the 3D-FRST method in terms of sensitivity and FP_{avg}. The proposed regional-based YOLO generates a limited number of FP_{avg} of 52.18 compared to a large number of candidates generated by 3D-FRST with FP_{avg} of 497.4. Our proposed YOLO proved its effectiveness and feasibility compared to 3D-FRST with



(a) (b) (c)

Figure 2. Examples of the detected suspicious regions by the proposed regional-based YOLO (yellow boxes); (a), (b), and (c) indicate the TPs, FPs, and FNs cases, respectively. Green boxes refer to labeled ground-truth microbleeds by radiologists.

TABLE II. OVERALL PERFORMANCE OF THE CANDIDATE DETECTION STAGE VIA REGIONAL-BASED YOLO AND THE FPS REDUCTION STAGE VIA 3D-CNN THROUGHOUT ALL FIVE-FOLDS TEST

Fold Test	1st Stage: Candidate Detection via Regional-based YOLO				2nd Stage: FPs Reduction via 3D-CNN				
	TPs	FNs	Sensitivity (%)	FP_{avg}	TPs	FNs	Sensitivity (%)	Precision (%)	FP_{avg}
1st Fold	36	0	100.00	53.71	34	2	94.44	75.56	0.79
2nd Fold	27	1	96.43	41.14	27	0	100.00	52.94	1.71
3rd Fold	34	6	85.00	64.35	31	3	91.18	91.18	0.21
4th Fold	34	2	94.44	51.07	32	2	94.12	47.06	2.57
5th Fold	45	3	93.75	50.80	42	3	93.33	60.00	1.75
Total	**176**	**12**	**93.62**	**52.18**	**166**	**10**	**94.32**	**61.94**	**1.42**

lower computation cost. It only requires 0.69 seconds to process one subject and generate potential candidates. It is 47.9 times faster than the 3D-FRST.

IV. CONCLUSION

In this work, we presented a new two-stage integrated deep learning approach for efficient CMBs detection. The detection stage via the regional-based YOLO endeavors to deny the background regions and simultaneously retrieve potential microbleeds candidates. Subsequently, the 3D-CNN stage is developed to reduce the FPs and single out the accurate microbleeds. We found that training the networks using 3D contextual information by averaging the adjacent slices of the SWI and complement phase images could improve the detection performance.

ACKNOWLEDGMENT

This research was supported by the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2018M3C7A1056884) and (NRF-2019R1A2C1090635). This research was also supported by a grant of the Korea Healthcare Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI14C1135).

REFERENCES

[1] S. Martinez-Ramirez, S. M. Greenberg, and A. Viswanathan, "Cerebral microbleeds: overview and implications in cognitive impairment," *Alzheimers Research & Therapy,* vol. 6, no. 3, 2014.

[2] A. Charidimou, A. Krishnan, D. J. Werring, and H. R. Jager, "Cerebral microbleeds: a guide to detection and clinical relevance in different disease settings," *Neuroradiology,* vol. 55, no. 6, pp. 655-674, Jun, 2013.

[3] W. Chen, W. Zhu, I. Kovanlikaya, A. Kovanlikaya, T. Liu, S. Wang, C. Salustri, and Y. Wang, "Intracranial Calcifications and Hemorrhages: Characterization with Quantitative Susceptibility Mapping," *Radiology,* vol. 270, no. 2, pp. 496-505, Feb, 2014.

[4] W. Bian, C. P. Hess, S. M. Chang, S. J. Nelson, and J. M. Lupo, "Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images," *Neuroimage-Clinical,* vol. 2, pp. 282-290, 2013.

[5] M. A. Al-masni, M. A. Al-antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, E. Valarezo, M. T. Choi, S. M. Han, and T. S. Kim, "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," *Computer Methods and Programs in Biomedicine,* vol. 157, pp. 85-94, Apr, 2018.

[6] M. A. Al-antari, M. A. Al-masni, M. T. Choi, S. M. Han, and T. S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *International Journal of Medical Informatics,* vol. 117, pp. 44-54, Sep, 2018.

[7] A. Masood, B. Sheng, P. Li, X. H. Hou, X. E. Wei, J. Qin, and D. G. Feng, "Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images," *Journal of Biomedical Informatics,* vol. 79, pp. 117-128, Mar, 2018.

[8] M. A. Al-Masni, M. A. Al-antari, M. T. Choi, S. M. Han, and T. S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine,* vol. 162, pp. 221-231, Aug, 2018.

[9] X. M. Zhao, Y. H. Wu, G. D. Song, Z. Y. Li, Y. Z. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Medical Image Analysis,* vol. 43, pp. 98-111, Jan, 2018.

[10] Q. Dou, H. Chen, L. Q. Yu, L. Zhao, J. Qin, D. F. Wang, V. C. T. Mok, L. Shi, and P. A. Heng, "Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks," *Ieee Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1182-1195, May, 2016.

[11] S. F. Liu, D. Utriainen, C. Chai, Y. S. Chen, L. Wang, S. K. Sethi, S. Xia, and E. M. Haacke, "Cerebral microbleed detection using Susceptibility Weighted Imaging and deep learning," *Neuroimage,* vol. 198, pp. 271-282, Sep, 2019.

[12] Y. C. Chen, J. E. Villanueva-Meyer, M. A. Morrison, and J. M. Lupo, "Toward Automatic Detection of Radiation-Induced Cerebral Microbleeds Using a 3D Deep Residual Network (vol 32, pg 766, 2019)," *Journal of Digital Imaging,* vol. 32, no. 5, pp. 898-898, Oct, 2019.

[13] J. Redmon, and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6517 - 6525.

[14] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping,* vol. 17, no. 3, pp. 143-155, Nov, 2002.