Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks

Qi Dou, *Student Member, IEEE*, Hao Chen, *Student Member, IEEE*, Lequan Yu, Lei Zhao, Jing Qin, Defeng Wang, Vincent CT Mok, Lin Shi*, and Pheng-Ann Heng, *Senior Member, IEEE*

***Abstract*—Cerebral microbleeds (CMBs) are small haemorrhages nearby blood vessels. They have been recognized as important diagnostic biomarkers for many cerebrovascular diseases and cognitive dysfunctions. In current clinical routine, CMBs are manually labelled by radiologists but this procedure is laborious, time-consuming, and error prone. In this paper, we propose a novel automatic method to detect CMBs from magnetic resonance (MR) images by exploiting the 3D convolutional neural network (CNN). Compared with previous methods that employed either low-level hand-crafted descriptors or 2D CNNs, our method can take full advantage of spatial contextual information in MR volumes to extract more representative high-level features for CMBs, and hence achieve a much better detection accuracy. To further improve the detection performance while reducing the computational cost, we propose a cascaded framework under 3D CNNs for the task of CMB detection. We first exploit a 3D fully convolutional network (FCN) strategy to retrieve the candidates with high probabilities of being CMBs, and then apply a well-trained 3D CNN discrimination model to distinguish CMBs from hard mimics. Compared with traditional sliding window strategy, the proposed 3D FCN strategy can remove massive redundant computations and dramatically speed up the detection process. We constructed a large dataset with 320 volumetric MR scans and performed extensive experiments to validate the proposed method, which achieved a high sensitivity of 93.16% with an average number of 2.74 false positives per subject, outperforming previous methods using low-level descriptors or 2D CNNs by a significant margin. The proposed method, in principle, can be adapted to other biomarker detection tasks from volumetric medical data.**

***Index Terms*—3D convolutional neural networks, biomarker detection, cerebral microbleeds, deep learning, susceptibility-weighted imaging.**

Manuscript received December 30, 2015; accepted February 03, 2016. Date of publication February 11, 2016; date of current version April 29, 2016. This work was supported by a grant from the National Basic Research Program of China, 973 Program (Project 2015CB351706), grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project CUHK412412, GRF14203115 and CUHK14113214), and a grant from the Shenzhen-Hong Kong Innovation Circle Funding Program (Project SGLH20131010151755080 and GHP/002/13SZ). The first two authors contributed equally to this work. *Asterisk indicates corresponding author.*

Q. Dou, H. Chen, L. Yu, and P. A. Heng are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: qdou@cse.cuhk.edu.hk; hchen@cse.cuhk.edu.hk).

L. Zhao is with the Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong.

J. Qin is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Medicine, Shenzhen University, Shenzhen 518060, China.

D. Wang is with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Hong Kong.

V. Mok is with the Chow Yuk Ho Technology Center for Innovative Medicine, Therese Pei Fong Chow Research Center for Prevention of Dementia, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong.

*L. Shi is with the Chow Yuk Ho Technology Center for Innovative Medicine, Therese Pei Fong Chow Research Center for Prevention of Dementia, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong (e-mail: shilin@cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2016.2528129

I. Introduction

CEREBRAL microbleeds (CMBs) refer to small foci of chronic blood products in normal (or near normal) brain tissues. They are histopathologically considered to be composed of hemosiderin deposits that leak through pathological blood vessels [1]. CMBs are prevalent in patients with cerebrovascular and cognitive diseases (such as stroke and dementia), as well as present in healthy aging individuals. The existence of CMBs and their distribution patterns have been recognized as important diagnostic biomarkers of cerebrovascular diseases. For example, the lobar distribution of CMBs suggests probable cerebral amyloid angiopathy [2] and the deep hemispheric or infratentorial CMBs imply probable hypertensive vasculopathy [3]. More importantly, the presence of CMBs could dramatically increase the risk of symptomatic intracerebral hemorrhage and recurrent ischemic stroke [4]. Apart from indicating these vascular diseases, CMBs could also structurally damage their nearby brain tissues, and further cause neurologic dysfunction, cognitive impairment and dementia [5]. In this regard, reliable detection of the presence and number of CMBs is crucial for cerebral diagnosis and may guide physicians in determining which drugs to use for necessary treatment, such as stroke prevention [6].

Modern advances in magnetic resonance (MR) imaging technologies, e.g., susceptibility-weighted imaging (SWI) [7], make paramagnetic blood products sensitive to screening, and hence facilitate the recognition of CMBs [8]. As shown in Fig. 1, the CMB is radiologically visualized as rounded hypointensities of small size in the SWI scan [9] (see the yellow rectangle in Fig. 1 left). In general, the clinical routine to annotate CMB is based on visual inspection and manual localization [10], which suffers from limited reproducibility among different observers and could be laborious and time-consuming, especially within the context of large numbers of subjects. Alternatively, automatic detection methods can help relieve the workload on radiologists as well as improve the efficiency and reliability of

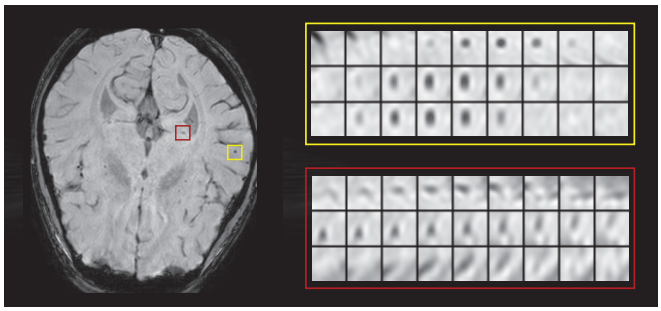


Fig. 1. Illustration of a CMB and a CMB mimic denoted with yellow and red rectangles, respectively. In each of the big rectangle, from top to down, the rows demonstrate adjacent slices in axial, sagittal and coronal planes. Best viewed in color.

the radiologic assessment. However, the automatic detection of CMBs faces several challenges. First, there is a large variation regarding the size of CMBs with a diameter ranging from 2 mm to 10 mm [1]. Second, the widespread distributed locations of CMBs make complete and accurate detection even harder [2], [3]. Third, there exist a lot of hard CMB mimics, e.g., flow voids, calcification and cavernous malformations, (see the red rectangle in Fig. 1 left) which would resemble the appearance of CMBs in SWI scans and heavily impede the detection process [9].

Previous automatic CMB detection methods mainly employed hand-crafted features based on shape, size and intensity information. For example, Fazlollahi *et al.* [11] utilized the radon transform to describe the shape information of CMBs, while Kuijf *et al.* [12] applied the radial symmetry transform (RST) to identify spherical regions as CMBs. To improve the capability of discrimination, Bian *et al.* [13] proposed to measure the geometric features after performing a 2D fast RST. Ghafaryasl *et al.* [14] further designed more comprehensive features that integrated the geometry, intensity, scale and local image structures. To improve the detection speed, some researchers proposed to first quickly remove the apparent non-CMB background regions and retrieve a small number of promising candidates for further classification based on these features [15], [16]. However, the design of these hand-crafted features heavily depends on the domain knowledge of CMBs. In addition, these low-level features are usually insufficient to capture the complicated characteristics of CMBs.

Recently, some investigations have been dedicated to learning features in a data driven way in order to more accurately detect CMBs [17], [18]. Among them, convolutional neural network (CNN) is one of the most promising solutions to meet the challenges of CMB detection by virtue of its high capability in extracting powerful high-level features. Actually, CNNs have achieved a great success with hierarchical feature representations in challenging natural image recognition tasks including object detection [19] semantic segmentation [20], image classification [21] and video action recognition [22]. Lately, CNNs have also presented outstanding effectiveness on 2D medical image computing problems such as standard plane localization from ultrasound images [23] and mitosis detection from histology images [24].

Our objective in this work is to detect biomarkers which are sparsely distributed in a 3D medical volume. However, how to effectively employ CNNs on 3D volumetric data still remains an open problem in medical image computing community. One straightforward way is to employ conventional 2D CNNs based on a single slice and process the slices sequentially [25]–[27]. Apparently, this solution disregards the contextual information along the third dimension, so its performance would be heavily degraded. Alternatively, some researches aggregate adjacent slices [18] or orthogonal planes (i.e., axial, coronal and sagittal) [28], [29] to enhance complementary spatial information. Nevertheless, this solution is still unable to make full use of the volumetric spatial information, because the input slices are independently treated and the convolution kernels are not shared along the third dimension. Note that the spatial information of all three dimensions is quite important for our CMB detection task. As shown in Fig. 1 right, the mimic can resemble the CMB in the view of one or two dimensions, but when taking the characteristics of all three dimensions into consideration, it is much easier to distinguish the CMB from the mimic. To the end, a 3D version of CNN, which utilizes 3D convolution kernels, is a more reliable solution to take full advantage of spatial contextual information in volumetric data for more accurate detection of CMBs.

To our best knowledge, the effectiveness of the 3D CNN on volumetric medical data has not been extensively explored. The main obstacles lie in the expensive computational cost, memory requirement and time consumption [30], [31]. Consider, if we detect lesions from a $512 \times 512 \times 150$ volume (as is the case in this work) using the traditional sliding window strategy [24], [32], [33], over 39 millions of 3D patches are sampled in a voxel-wise manner. Even with a larger sampling stride such as 4, we would still obtain over half a million of 3D patches. This brings about a large amount of computational workload, which is impractical in clinical practice. Fortunately, by taking a closer look at the sliding window strategy, we can find that convolutional operations are redundantly conducted due to overlapped sampling. In this case, if we can elegantly remove the massive redundant computations, the detection process can be dramatically speeded up.

In order to accurately and efficiently detect CMBs from volumetric brain SWI data, we propose a robust and efficient method by leveraging 3D CNNs. Specifically, our method consists of two stages that are designed in a cascaded manner. The first stage is the *screening* stage, in which a small number of candidates are retrieved using a novel 3D fully convolutional network (3D FCN) model. The screening strategy with the 3D FCN model can achieve significant acceleration compared with the conventional sliding window strategy under the same setting of the sampling stride. The second stage is the *discrimination* stage, where the candidates obtained from the screening stage are carefully distinguished with a 3D CNN discrimination model. This stage removes a large number of false positive candidates and yields the final detection results. To validate the effectiveness of the proposed method, we built a large dataset of cerebral SWI images including 126 stroke subjects and 194 normal aging subjects. Extensive experiments conducted on the large dataset corroborate that our method can achieve better re-

sults than the state-of-the-art methods in terms of sensitivity, precision and false positive rate. The main contributions of this work are summarized as follows:

1) We, for the first time, exploit the 3D CNN for automatic detection of CMBs from volumetric brain SWI images. The 3D CNN sufficiently encodes the spatial contextual information and hierarchically extracts high-level features in a data driven way. It demonstrates better performance than previous methods based on low-level 3D features or 2D CNNs. To our best knowledge, we are one of the pioneers to employ 3D CNN for automatic detection of key biomarkers from volumetric medical data.
2) To efficiently leverage 3D CNN, we propose a novel 3D FCN strategy to successfully avoid redundant computations in the traditional sliding window strategy. The 3D FCN is capable of inputting a whole volumetric data and directly outputting a 3D prediction score volume within a single forward propagation. In this way, the detection speed is dramatically accelerated.
3) We propose a two-stage cascaded framework to efficiently and accurately detect CMBs. The screening stage with the 3D FCN rapidly retrieves potential candidates, and the discrimination stage with the 3D CNN focuses on these candidates to further accurately single out the true CMBs from challenging mimics. In addition, this proposed framework is general and can be easily adapted to other biomarker detection tasks.

The remainder of this paper is organized as follows. We detail our method in Section II and report the experimental results in Section III. Section IV further analyzes some key issues of the proposed method and discusses future directions. The conclusions are drawn in Section V.

II. Methodology

Fig. 2 shows an overview of the proposed cascaded framework, which is composed of two stages: screening stage and discrimination stage. In the screening stage, the 3D FCN model takes a whole volumetric data as input and directly outputs a 3D score volume. Each value on the 3D score volume represents the probability of CMB at a corresponding voxel of the input volume. Subsequently, in the discrimination stage, we further remove false positive candidates by applying a 3D CNN discrimination model to distinguish true CMBs from challenging mimics with high-level feature representations.

A. 3D Convolutional Neural Network

Typically, a CNN alternatively stacks convolutional (C) and sub-sampling, e.g., max-pooling (M), layers. In a C layer, small feature extractors (kernels) sweep over the topology and transform the input into feature maps. In a M layer, activations within a neighborhood are abstracted to acquire invariance to local translations. After several C and M layers, feature maps are flattened into a feature vector, followed by fully-connected (FC) layers. Finally, a softmax classification layer yields the prediction probability. Readers can refer to [34] for more details about typical CNN constructions. Although CNNs have achieved remarkable successes in 2D medical image analysis [23], [24], they have been seldom extended to 3D volumetric image processing tasks.

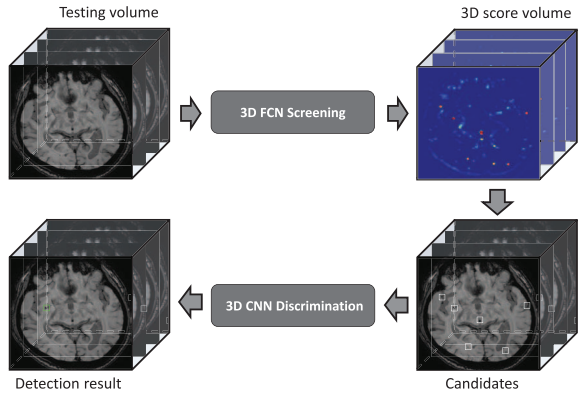


Fig. 2. An overview of the proposed cascaded framework for CMB detection.

1) 3D Convolutional Layers: In a typical C layer, a feature map is produced by convolving the input with convolution kernels, adding a bias term, and finally applying a non-linear activation function. By denoting the *i-th* feature map of the *l-th* layer as $\boldsymbol{h}_i^l$ and the *k-th* feature map of the previous layer as $\boldsymbol{h}_k^{l-1}$, a C layer is formulated as:

$$\boldsymbol{h}_i^l = \boldsymbol{\sigma}\left(\sum_k \boldsymbol{h}_k^{l-1} * \boldsymbol{W}_{ki}^l + \boldsymbol{b}_i^l\right) \tag{1}$$

where $\boldsymbol{W}_{ki}^l$ and $\boldsymbol{b}_i^l$ are the filter and bias term connecting the feature maps of adjacent layers, the $*$ denotes the convolution operation and the $\boldsymbol{\sigma}(\cdot)$ is the element-wise non-linear activation function.

In 2D natural image processing, the input of CNN usually consists of three color channels (i.e., RGB). Inspired by this, the most straightforward way to adapt 2D CNN to support volumetric data processing is replacing the color channels with slices of the volume. As shown in Fig. 3(a), given a volumetric image of size $X \times Y \times Z$, when we employ this scheme to generate a feature map, we first need to split the input volume along the third dimension into Z isolated slices, and then feed these Z isolated slices into the network. Correspondingly, Z 2D kernels are formed, with each single slice swept over by a unique kernel (see the red line). However, this scheme cannot sufficiently leverage the spatial information, since the Z 2D kernels are different from each other. In other words, due to the absence of kernel sharing across the third dimension, the encoded volumetric spatial information is inevitably deficient.

Learning feature representations from all three dimensions is vitally important for biomarker detection tasks from volumetric medical data, e.g., CMB detection from SWI images. In this regard, we propose to employ the 3D convolution kernel, in the pursuance of encoding richer spatial information of the volumetric data. In this case, the feature maps are 3D blocks instead of 2D patches (we call them *feature volumes* hereafter). As shown in Fig. 3(b), given the same volumetric image of size $X \times Y \times Z$, when we employ a 3D convolution kernel to generate

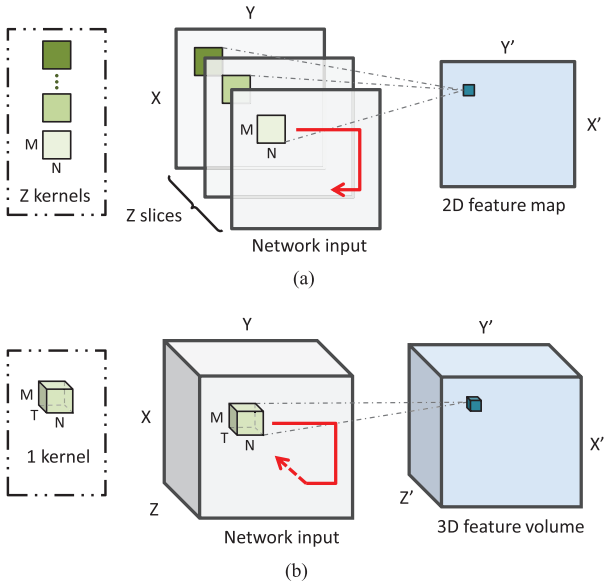


Fig. 3. Comparison of using 2D and 3D convolution kernels given volumetric image with size of $X \times Y \times Z$ in terms of network input, kernel behavior and generated feature map. Red lines represent the moving direction of kernels, i.e., sweeping over the 2D and 3D topologies, respectively. (a) With the 2D convolution (kernel size of $M \times N$), the volume is first split into Z isolated slices along the third direction and these slices are input to the network. Each generated feature map is a 2D patch. (b) With the 3D convolution (kernel size of $M \times N \times T$), the entire volume is input to the network. Each generated feature map is a 3D volume. (Note that kernel sizes M, N and T need not to be equal. Best viewed in color.).

a 3D feature volume, the input to the network is the entire volumetric data. Consequently, a 3D kernel is formed and it sweeps over the whole 3D topology (see the red line). By leveraging the kernel sharing across all three dimensions, the network can take full advantage of the volumetric contextual information.

Generally, the following equation formulates the exploited 3D convolution operation in an element-wise manner:

$$\boldsymbol{u}_{ki}^{l}(x,y,z) = \sum_{m,n,t} \boldsymbol{h}_{k}^{l-1}(x-m, y-n, z-t)\boldsymbol{W}_{ki}^{l}(m,n,t) \quad (2)$$

where $\boldsymbol{W}_{ki}^{l}$ is the 3D kernel in the *l-th* layer which convolves over the 3D feature volume $\boldsymbol{h}_{k}^{l-1}$, $\boldsymbol{W}_{ki}^{l}(m,n,t)$ is the element-wise weight in the 3D convolution kernel. Following (1) and (2), the 3D feature volume $\boldsymbol{h}_{i}^{l}$ is obtained by different 3D convolution kernels:

$$\boldsymbol{h}_{i}^{l} = \boldsymbol{\sigma}\left(\sum_{k} \boldsymbol{u}_{ki}^{l} + \boldsymbol{b}_{i}^{l}\right). \quad (3)$$

2) 3D CNN Hierarchical Architecture: After figuring out the 3D convolutional layers, we can hierarchically construct a deep 3D CNN model by stacking the C, M and FC layers, as shown in Fig. 4. Specifically, in the C layer, multiple 3D feature volumes are produced. In the M layer, the max-pooling operation is also performed in a 3D fashion, i.e., the feature volumes are sub-sampled based on a cubic neighborhood. In the following FC layer, 3D feature volumes are flattened into a feature vector as its input. The ultimate output layer employs the softmax activation to yield the prediction probabilities.

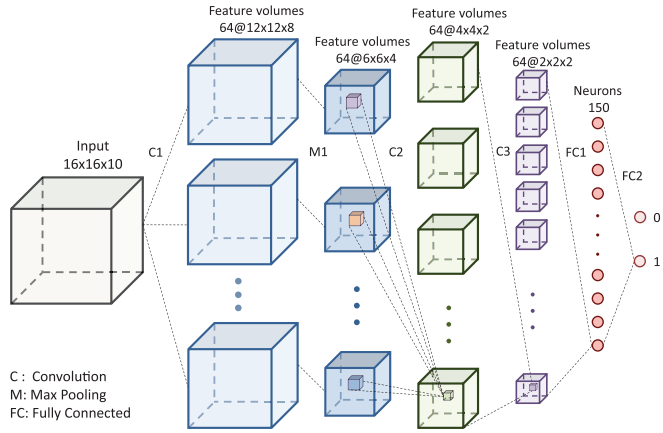


Fig. 4. The hierarchical architecture of the 3D CNN model.

In our 3D CNN implementation, the rectifier linear unit (ReLU) [35] is utilized for the non-linear activation function in the C and FC layers. The 3D convolution kernels are randomly initialized from the Gaussian distribution and trainable parameters in the network are tuned using the standard back-propagation with stochastic gradient descent by minimizing the cross entropy loss. Meanwhile, dropout strategy [36] is utilized to reduce the co-adaption of intermediate features and improve the generalization capability.

B. 3D Fully Convolutional Network

One of the main concerns about exploiting CNN in medical imaging domain lies in the time performance, as many medical applications require prompt responses for further diagnosis and treatment. The situation is more rigorous when processing volumetric medical data. Directly applying 3D CNNs to detect lesions using the traditional sliding window strategy is usually impracticable, especially when the input images are acquired with high resolutions, because thousands or even millions of 3D block samples need to be analyzed. In most biomarker detection applications, the targets are usually sparsely distributed throughout the volume, such as the CMBs in the 3D brain MR data. To this end, one promising solution is to first obtain the candidates with a high sensitivity and then perform fine-grained discrimination only on these candidates, so that the computational cost can be greatly reduced. Previous work proposed to retrieve CMB candidates (also called regions-of-interest (ROI) in some papers) in a MR volume by employing local statistical information, including size, intensity, shape and other geometric features [14], [15], [18]. However, due to the large variations of CMBs in different patients, only relying on these statistic values, it is difficult to precisely describe the characteristics of CMBs and detach them from the background regions. The results either neglect true CMBs or include a large number of false positives, which can complicate the following discrimination procedure.

We propose to use 3D CNN to robustly screen candidates by leveraging high-level spatial representations of CMBs learned from a large number of 3D training samples. However, we still face the challenge of time performance when employing 3D CNN to retrieve candidates with the traditional sliding window

strategy. To this end, inspired by the 2D fully convolutional networks [20], [37], we propose to extend the strategy into a 3D format for efficient retrieval of CMB candidates from MR volumetric data. The proposed 3D FCN can take an arbitrary-sized volume as input and produce a 3D score volume within a single forward propagation, and hence greatly speed up the candidate retrieval procedure without damaging the sensitivity.

1) Fully Convolutional Transformation: In the 3D CNN architecture, both the M and C layers can process arbitrary-sized input, where convolution or max-pooling kernels sweep over the input and generate the corresponding-sized output. However, the traditional FC layers flatten the feature volumes into vectors thus dismissing the spatial relationships. These FC layers then utilize vector-matrix multiplications to generate the output, as shown in the following:

$$\boldsymbol{h}^l = \boldsymbol{\sigma}(\boldsymbol{W}^l \boldsymbol{h}^{l-1} + \boldsymbol{b}^l) \tag{4}$$

where $\boldsymbol{h}^{l-1} \in \mathbb{R}^P$ and $\boldsymbol{h}^l \in \mathbb{R}^Q$ are the feature vectors in the *(l-1)-th* and the *l-th* FC layers, respectively, $\boldsymbol{W}^l \in \mathbb{R}^{Q \times P}$ is the weight matrix and $\boldsymbol{b}^l$ denotes the bias term.

In *traditional* CNN, once trained, the weight $\boldsymbol{W}^l$ is with a fixed shape, and hence the FC layer has fixed input/output sizes. As a result, a network with traditional FC layers requires that the initial inputs have a fixed size. For example, when the network is trained based on 3D samples of size $16 \times 16 \times 10$, errors will arise if we input a test sample of size $20 \times 16 \times 10$, due to the shape mismatch in the first dimension.

In this regard, we equivalently re-write the FC layers into the following convolutional format:

$$\boldsymbol{h}_q^l = \boldsymbol{\sigma}(\sum_p \boldsymbol{h}_p^{l-1} * \boldsymbol{W}_{pq}^l + \boldsymbol{b}_q^l) \tag{5}$$

where each neuron in the FC layer is regarded as a $1 \times 1 \times 1$ feature volume, $\boldsymbol{W}_{pq}^l \in \mathbb{R}^{1 \times 1 \times 1}$ is the 3D kernel and the $*$ is the 3D convolution operation described in (2). In this way, the vector-matrix multiplications are formulated as convolution operations with $1 \times 1 \times 1$ kernels. With the FC layers converted into convolutional layers, the network could therefore support arbitrary-sized input.

2) 3D Score Volume Generation: During the training phase, a *traditional* 3D CNN model is learned. Once training is done, to acquire the 3D FCN model, the FC layers in the *traditional* 3D CNN are transformed into the convolutional fashion. More specifically, the multiplication matrix $\boldsymbol{W}^l \in \mathbb{R}^{Q \times P}$ is reshaped into a 5D tensor $\boldsymbol{W}^l \in \mathbb{R}^{Q \times 1 \times P \times 1 \times 1}$ (the dimensions are ordered for the ease of implementation), and hence the weight matrix is converted into a series of convolution kernels. During the testing phase, the 3D FCN model directly inputs a volume (with size $512 \times 512 \times 150$ for our dataset) and outputs a 3D score volume (with reduced resolution compared with the original input size). The value at each location of score volume indicates the probability of CMB.

Some technical issues need to be handled when developing the 3D FCN model. Specifically, when converting the traditional FC layers into the convolutional fashion by casting the 2D multiplication matrix ($\mathbb{R}^{Q \times P}$) into the 5D tensor ($\mathbb{R}^{Q \times 1 \times P \times 1 \times 1}$), we should precisely maintain the spatial correlations. In addition, during the whole volume testing phase, we need to ensure the dimension consistency in the logistic regression layer, where the feature volumes are first flattened into vectors, then applied to the softmax function and finally reshaped back to form the 3D score volume.

TABLE I
THE ARCHITECTURE OF 3D FCN SCREENING MODEL

Layer	Kernel size	Stride	Output size	Feature volumes
Input	-	-	16× 16× 10	1
C1	5× 5× 3	1	12× 12× 8	64
M1	2× 2× 2	2	6× 6× 4	64
C2	3× 3× 3	1	4× 4× 2	64
C3	3× 3× 1	1	2× 2× 2	64
FC1	2× 2× 2	1	1× 1× 1	150
FC2	1× 1× 1	1	1× 1× 1	2

Note: FC layers are converted into the convolutional fashion.

Compared with the sliding window strategy which repeatedly crops overlapping samples, the 3D FCN strategy produces a 3D score volume within a single forward propagation. As a result, the 3D FCN successfully eliminates a large number of redundant convolutional computations, which dramatically speeds up the prediction process. For example, when employing the number of convolution operations to roughly estimate the computational cost of a testing volume with size $512 \times 512 \times 150$, the proposed 3D FCN strategy (with the architecture shown in Table I) is roughly 800 times faster than the voxel-wise sliding window strategy and 100 times faster than the sliding window strategy with a sampling stride of 2, which is the same stride as our 3D FCN architecture (i.e., generating the same resolution score volumes).

3) Score Volume Index Mapping: Due to successive layers of convolution and max-sampling operations, the size of the generated 3D score volume is reduced compared with the original input. Actually, the 3D score volume is a coarse version of the voxel-wise predictions which are produced by the sliding window strategy. Meanwhile, the locations on this coarse score volume can be traced back to the coordinates on the original input space.

Since all three dimensions follow the same index mapping mechanism, we demonstrate the mapping process with one dimension. In our formulation, indices are numbered from zero. Generally, for each C or M layer (supposing non-padding convolution and non-overlap pooling) in the model, the index mapping procedure with convolution or max-pooling operation can be calculated by:

$$x' = d \cdot x + \lfloor \frac{c-1}{2} \rfloor \tag{6}$$

where x' and x denote the coordinates before and after the convolution or max-pooling operation; d and c represent the stride and kernel size, respectively; the $\lfloor \cdot \rfloor$ represents the floor function.

When mapping the location x_s in the coarse score volume back through the architecture towards the location x_o in the original input volume, we successively deduce the index mapping procedures along all intermediate convolution and max-pooling layers until the initial input layer. For example, based on the

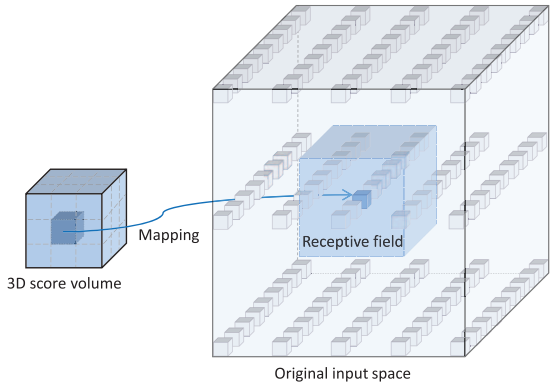


Fig. 5. The mapping from the 3D score volume onto the original input space.

network architecture shown in Table I, for each position index x_s in the coarse score volume, we can obtain its corresponding index x_o in the original input as follows:

$$x_o = \lfloor \frac{c_1 - 1}{2} \rfloor + \lfloor \frac{c_2 - 1}{2} \rfloor + d_2 \cdot (x_s + \lfloor \frac{c_3 - 1}{2} \rfloor + \lfloor \frac{c_4 - 1}{2} \rfloor + \lfloor \frac{c_5 - 1}{2} \rfloor + \lfloor \frac{c_6 - 1}{2} \rfloor) = D \cdot x_s + C \quad (7)$$

where, according to the network architecture, $c_1 = 5$, $c_2 = 2$, $d_2 = 2$, $c_3 = 3$, $c_4 = 3$, $c_5 = 2$, $c_6 = 1$, and we can calculate $D = 2$ and $C = 6$ for the $\boldsymbol{X}$ dimension.

As shown in Fig. 5, with this mechanism, each location in the 3D score volume can be mapped back to the centroid of the corresponding receptive field of the neuron. Equivalently, if the cubic patch centered on the traced position is input to the *traditional* 3D CNN, the prediction probability is indeed the value at the location on the coarse score volume. Consequently, the prediction scores are sparsely mapped back onto the input volume, and regions with high probabilities are retrieved as potential candidates.

C. Two-Stage Cascaded Framework

In order to detect CMBs from MR images, we employ 3D CNN based models to tap potentials of spatial information in all three dimensions and represent them as high-level features. We construct a 3D FCN model and a 3D CNN model tailored for two different stages and integrate them into an efficient and robust detection framework. In this cascaded framework for CMB detection, each stage serves its own mission. The screening stage with the 3D FCN aims to accurately reject the background regions and rapidly retrieve a small number of potential candidates. The discrimination stage with the 3D CNN focuses only on the screened set of candidates to further single out the true CMBs from challenging mimics.

1) Screening Stage: In this stage, the 3D FCN model with the architecture shown in Table I is exploited. Note that the FC layers are converted into the convolutional fashion, thus we present kernel sizes for them. We analyzed the bounding box sizes of CMB regions, which were measured in the number of voxels along the three axes. The sagittal and frontal axes take similar sizes and the longitudinal axis has a relatively low resolution due to the parameter settings of data acquisition, which are regular scanning settings in clinical practice. To

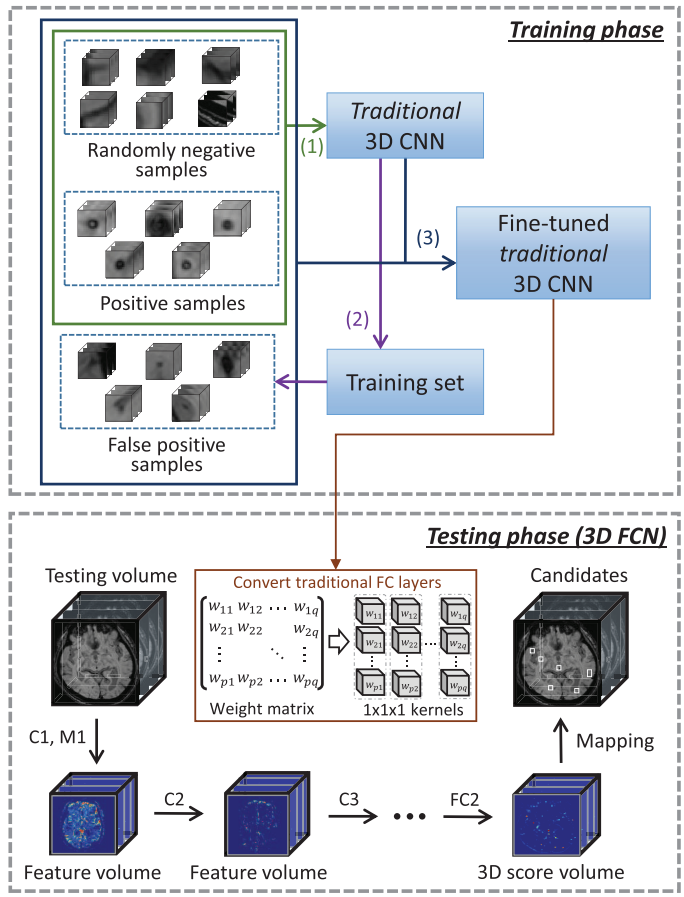


Fig. 6. Illustration of the workflow of the screening stage. The training phase is conducted in three sub-steps: (1) train an initial *traditional* 3D CNN with positive samples and randomly selected negative samples; (2) apply the initial model on training set and obtain false positive samples to enlarge the training database; (3) fine-tune the initial *traditional* 3D CNN model with the enlarged database to strengthen its discrimination capability. Once training is done, the traditional FC layers are converted into the convolutional fashion (as shown in the brown box). During the testing phase, the 3D FCN takes a whole volume as input, extracts representative feature volumes and finally produces a 3D score volume to retrieve candidates.

ensure the accuracy while limiting the computational workload, when training the *traditional* 3D CNN, we set the input size as $16 \times 16 \times 10$, because 99.39% of the CMB lesions are bounded by this sized box.

The workflow of the screening stage is illustrated in Fig. 6, including both training and testing phases. During the training phase, the positive samples are extracted from CMB regions and augmented by translation, rotation and mirroring to expand the training database. In practice, the network is trained with three sub-steps. We start from training an initial 3D CNN with randomly selected non-CMB regions throughout the brain as negative samples. Next, we add false positive samples acquired by applying the initial model on the training set. Finally, the initial model is fine-tuned with the enlarged training database which consists of 23.63% positives, 47.52% randomly selected negatives and 28.85% supplemental false positives. In this way, the discrimination capability of the network is further enhanced. Once training is done, the fine-tuned *traditional* 3D CNN is converted into the 3D FCN model by transforming the FC layers into the convolutional fashion. During the testing phase, the 3D

TABLE II
THE ARCHITECTURE OF 3D CNN DISCRIMINATION MODEL

Layer	Kernel size	Stride	Output size	Feature volumes
Input	-	-	20× 20× 16	1
C1	7× 7× 5	1	14× 14× 12	32
M1	2× 2× 2	2	7× 7× 6	32
C2	5× 5× 3	1	3× 3× 4	64
FC1	-	-	1× 1× 1	500
FC2	-	-	1× 1× 1	100
FC3	-	-	1× 1× 1	2

Note: FC layers remain the traditional construction.

FCN model takes the whole volume as input and generates the corresponding coarse 3D score volume.

Considering that the produced score volume could be noisy, we utilize the local non-max suppression in a 3D fashion as the post-processing. Locations in the 3D score volume are then sparsely traced back to coordinates in the original input space, according to the index mapping process presented in (7). Finally, regions with high prediction probabilities are selected as the potential candidates.

2) Discrimination Stage: In this stage, 3D small blocks are cropped centered on the screened candidate positions. The size of these blocks was carefully validated. We first found that a number of false positives were produced in the first stage with a training block size of $16 \times 16 \times 10$. By enlarging the block size, richer contextual information within larger surrounding neighborhood can provide additional clues to better distinguish CMBs from their mimics. However, due to the small size of CMB, the cropped block size can not be too large. Otherwise, redundant contextual information would be introduced and may degrade the performance. In this regard, we set the input size as $20 \times 20 \times 16$ in our experiments, in order to discriminate the challenging candidates with a suitable receptive field. The parameter setting of block size is detailed in Section III-D.

The extracted 3D candidate regions are classified by a newly constructed 3D CNN model. We notice that the randomly selected non-CMB samples are not strongly representative, especially when we aim to distinguish true CMBs from their mimics. To generate representative samples and improve the discrimination capability of the 3D CNN model, the obtained false positives (which take very similar appearance as CMBs) on the training set in the screening stage are taken as negative samples when training the 3D CNN in the second stage. The network architecture of the discrimination model is shown in Table II. Note that the FC layers remain as the traditional format without transformed into convolutional fashion, because this stage focuses on classification rather than overall screening and the matrix multiplications are more computationally efficient compared with convolution operations.

III. EXPERIMENTS

A. Dataset and Preprocessing

To validate the performance of the proposed method, we built a large dataset of SWI images for CMB detection, referred as *SWI-CMB*. The *SWI-CMB* includes 320 SWI images acquired from a 3.0T Philips Medical System with 3D spoiled gradient-echo sequence using venous blood oxygen level dependent series with the following parameters: repetition time 17 ms, echo time 24 ms, volume size $512 \times 512 \times 150$, in-plane resolution 0.45×0.45 mm, slice thickness 2 mm, slice spacing 1 mm and a 230×230 mm^2 field of view. The subjects came from two separated groups: 126 subjects with stroke (mean age $\pm$ standard deviation: 67.4 ± 11.3) and 194 subjects of normal aging (mean age $\pm$ standard deviation: 71.2 ± 5.0).

TABLE III
DETAILS OF DATASETS

Datasets	Stroke		Normal aging		Total	
	Subjects	CMBs	Subjects	CMBs	Subjects	CMBs
Training	91	701	139	223	230	924
Validation	15	81	25	27	40	108
Testing	20	78	30	39	50	117

The dataset was labeled by an experienced rater (Lei Zhao) and was verified by a neurologist (Dr. Zhaolu Wang) following the guidance of Microbleed Anatomical Rating Scale [10]. We employed the Pearson correlation coefficient (PCC) to assess the interobserver agreement between the two raters [38]. Due to the large dataset and expensive manual annotation efforts, we tested the interobserver agreement with a subset of 20 subjects (including 10 cases with stroke and 10 cases of normal aging). The PCC turned out to be 0.91 ($p < 0.01$), which indicates a high degree of agreement between the two raters. Overall, a total of 1149 CMBs were annotated from the whole dataset and regarded as the ground truth in our experiments. To our best knowledge, this is the largest benchmark dataset available for CMB detection.

In our experiments, we randomly divided the whole dataset into three sections for training, validation and testing, respectively. The details of these three sets are shown in Table III. In the preprocessing step, we normalized the volume intensities to the range of $[0, 1]$ with:

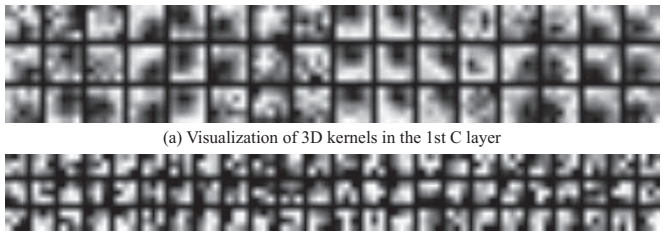
$$I' = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{8}$$

where I and I' denote the original and normalized intensity value, respectively. The $I_{\max}$ is the maximum intensity value after trimming the top 1% gray scale values and the $I_{\min}$ is the minimum intensity value of the volume. Since the MR image has sparse outliers with high intensity values, by trimming the extremes at the top end of the intensity range, more room could be made for the remaining intensities to be adjusted. This preprocessing has been widely performed in other related studies [11], [18].

B. Evaluation Metrics

We employed three commonly used metrics to quantitatively evaluate the performance of the proposed CMB detection method including sensitivity (S), precision (P) and the average number of false positives per subject ($\mathrm{FP_{avg}}$). They are defined as follows:

$$\mathrm{S} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \mathrm{FP_{avg}} = \frac{\mathrm{FP}}{\mathrm{N}}, \tag{9}$$



(a) Visualization of 3D kernels in the 1st C layer

(b) Visualization of 3D kernels in the 2nd C layer

Fig. 7. Visualization of typical learned filters in the screening 3D FCN model: (a) visualization of the C1 layer kernels, where each column represents a 3D kernel of size $5 \times 5 \times 3$, which is visualized as three 5×5 maps; (b) visualization of the C2 layer kernels, where each column represents a 3D kernel of size $3 \times 3 \times 3$, which is visualized as three 3×3 maps.

where TP, FP and FN denote the total number of true-positive, false-positive and false-negative detection results, respectively. The N represents the number of subjects in the testing dataset.

C. CMB Candidates Localization With 3D FCN

To provide a comprehensible insight into the learned kernels of the 3D FCN in the screening stage, typical 3D convolution kernels of the first two convolutional layers are visualized in Fig. 7. The sizes of kernels can be referred to the architecture in Table I. The Fig. 7(a) visualizes the C1 layer kernels (with size $5 \times 5 \times 3$), where each column represents a 3D kernel which is demonstrated as three 5×5 maps. Interestingly, the learned kernels attend to the spherical shapes of CMBs as well as the intensity difference between the CMBs and surrounding background. More importantly, the observed slight changes of the three maps within each column validate that the 3D kernels have effectively captured spatial information across the third dimension of the volumetric data, demonstrating 3D CNN can capture more contextual information than 2D CNN for more accurate detection in 3D medical data. The Fig. 7(b) visualizes the C2 layer kernels (with size $3 \times 3 \times 3$), where each column represents a 3D kernel which is visualized as three 3×3 maps. These kernels are difficult for straightforward interpretation since they try to construct some high-level concepts from the output features of the bottom layer. Nevertheless, we can observe that these kernels attain evidently organized patterns.

During the testing phase, the 3D FCN model inputs a whole volume and correspondingly produces a 3D score volume. Each location on the score volume is assigned a probability of belonging to a CMB region. Locations on the score volume are then sparsely mapped back to coordinates of the original input space according to (7). After post-processing, the score volume is thresholded and the regions of high probabilities are retrieved as candidates. Here, we set the threshold $T = 0.64$, which yields the best performance on the validation dataset.

We compared the screening performance of our proposed method with two state-of-the-art approaches which utilize low-level statistical features [15], [18]. We implemented these comparison approaches and employed them on our testing dataset. The results are listed in Table IV. The values of sensitivity mean the percentage of successfully retrieved CMBs while the values of $\mathrm{FP}_{\mathrm{avg}}$ describe the number of remaining false positives per subject. The fewer false positives produced, the more powerful discrimination capability a screening method has. The proposed 3D FCN model achieves the highest sensitivity with fewest average number of false positives, which highlights the efficacy of the proposed method. Note that our method outperforms the other two methods by a large margin, thanks to the 3D FCN model.

TABLE IV
COMPARISON OF DIFFERENT SCREENING METHODS

Methods	Sensitivity	$\mathrm{FP}_{\mathrm{avg}}$	Time per subject (s)
Barnes et al. [15]	85.47%	2548.2	81.46
Chen et al. [18]	90.48%	935.8	**12.00**
3D FCN model	**98.29%**	**282.8**	64.35

We have also recorded the average time for screening each subject and the results are listed in Table IV. From the clinical perspective, the time performance of our method is satisfactory; processing a whole volume with a size of $512 \times 512 \times 150$ takes around 1 minute in our experiments. The method of [15] is slower than ours because it calculates local thresholds using a voxel-wise sliding window strategy. In contrast, the method of [18] merely exploits global thresholding on intensity and size, hence it has a much faster screening speed.

For the candidate screening stage, the retrieval accuracy is vitally important, because we cannot re-find the CMBs that are missed by the screening stage in the following discrimination stage. Although [18] is faster, we achieved around 8% increase in sensitivity and reduced the number of $\mathrm{FP}_{\mathrm{avg}}$ from 935.8 to 282.8, when compared with this method. These results provide a much more reliable basis for further fine discrimination. By employing the 3D FCN, our method achieves a good balance between retrieval accuracy and speed.

Typical candidate screening results by the proposed 3D FCN are shown in Figs. 8 and 9. In our experiments, the local non-max suppression was performed in a 3D fashion. For the sake of clear illustration, we projected the 3D score volume together with its 3D suppression result onto the visualized slice planes, because after the 3D suppression, local-maximums may fall onto other adjacent planes instead of the visualized slice. To comprehensively present the results, we projected the volumetric results along two directions, i.e., the longitudinal and frontal axes. Examples of the obtained axial and sagittal plane projection results are shown in Figs. 8 and 9, respectively. It is observed that high values on the score volume mostly correspond to CMB lesions. In addition, most of the backgrounds have been successfully suppressed as zeros. After thresholding, only a small number of candidates are obtained (see those white rectangles), which dramatically reduces the computational workload in the following stage.

D. True CMB Discrimination With 3D CNN

Employing the CMB samples with augmentations and the false positives generated from the training set in the screening stage (overall 0.7 million training samples with around 13% positive samples), we train the 3D CNN discrimination model to remove false positive candidates and accurately identify true CMBs.

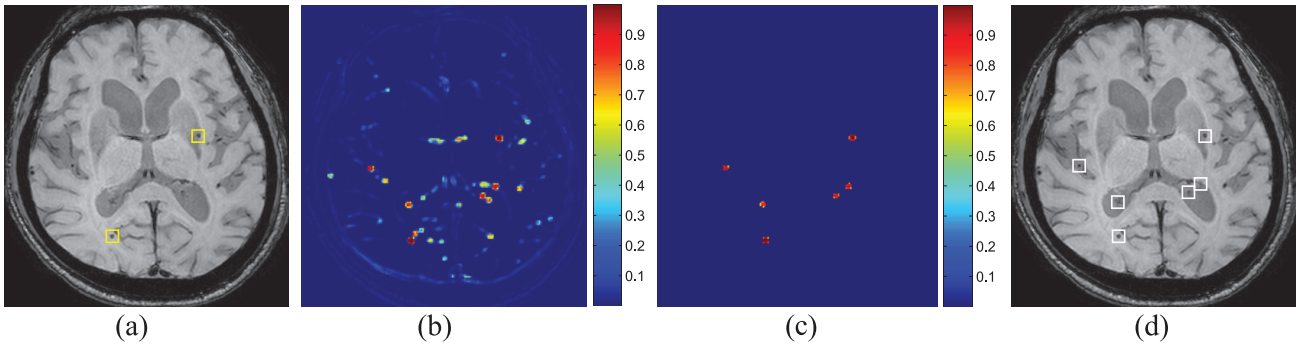


Fig. 8. Typical results of the 3D FCN screening model with score volume projection onto the axial plane. (a) Raw data with true CMBs (yellow rectangles). (b) 2D projection of the score volume generated with FCN. (c) 2D projection of the post-processed score volume. (d) Retrieved candidates (white rectangles). Best viewed in color.

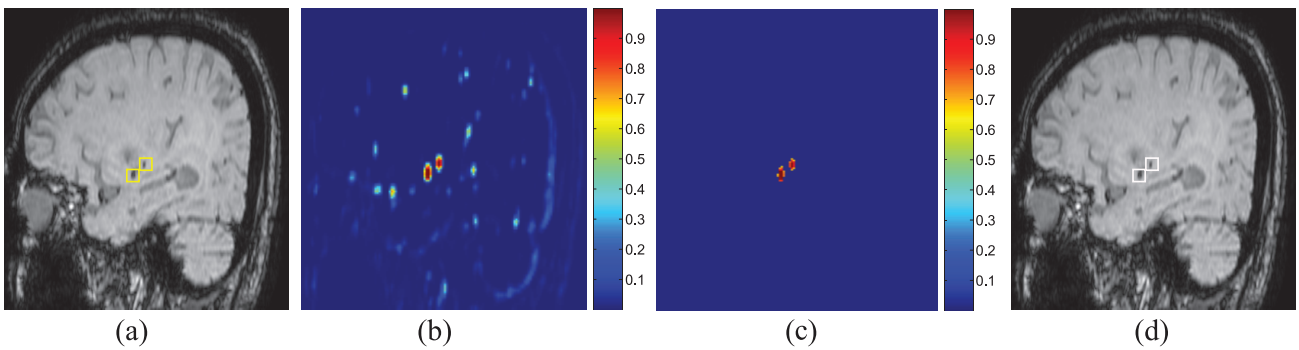


Fig. 9. Typical results of the 3D FCN screening model with score volume projection onto the sagittal plane. (a) Raw data with true CMBs (yellow rectangles). (b) 2D projection of the score volume generated with FCN. (c) 2D projection of the post-processed score volume. (d) Retrieved candidates (white rectangles). Best viewed in color.

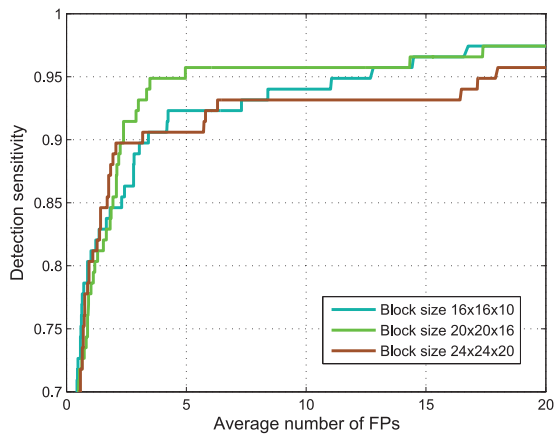


Fig. 10. FROC curves of 3D CNN with different block size configurations.

The 3D CNN demands a suitable receptive field (i.e., input size) to achieve fine discrimination. Specifically, we compared three different block size configurations, i.e., $16\times16\times10$ (which is the same size as the training block in the screening stage), $20 \times 20 \times 16$ and $24 \times 24 \times 20$. We tested larger sizes than the input block size of the screening model because we wanted to validate whether a larger input block with more contextual information can enhance the discrimination capability of the model. The results under these settings are shown in Table V. In addition, Fig. 10 presents the Free Response Operating Characteristic (FROC) curves of different sample size configurations. In

TABLE V
DETECTION RESULTS UNDER DIFFERENT BLOCK SIZE CONFIGURATIONS

3D block size	Sensitivity	Precision	FP_{avg}
$16 \times 16 \times 10$	91.45%	33.75%	4.20
$20 \times 20 \times 16$	**92.31%**	**42.69%**	**2.90**
$24 \times 24 \times 20$	91.45%	27.16%	5.74

the case of block size $16 \times 16 \times 10$, the detection sensitivity reached 91.45% with 4.20 false positives. The detection performance was improved to the sensitivity of 92.31% with 2.90 false positives under block size $20\times20\times16$, demonstrating that properly increasing contextual information can enhance the discrimination capability of 3D CNN. When block size $24\times24\times20$ was employed, the detection sensitivity decreased to 91.45% with average 5.74 false positives per subject. This may be because that too much redundant contextual information would disturb the actual CMB signals, and hence degrade the detection performance. Actually, regardless of the employed block sizes, our method achieved much better results than other methods using 2D CNNs or hand-crafted features. Derived from these experiments, by setting the block size as $20 \times 20 \times 16$, we can achieve an optimal detection performance.

We independently trained three models using the network architecture shown in Table II. The differences of the three models lie in the random weights initialization states and the number of training epochs. The neural network with a large number of

TABLE VI
EVALUATION OF DETECTION RESULTS

Methods	Sensitivity	Precision	FP_{avg}
Barnes et al. [15]	64.96%	5.13%	28.10
Random forest [40]	85.47%	17.24%	9.60
2D-CNN-SVM [18]	88.03%	22.69%	7.02
Ours (single)	92.31%	42.69%	2.90
Ours (average)	**93.16%**	**44.31%**	**2.74**

parameters is usually with a low bias and a large variance. By averaging multiple models with different weight initializations and early stopping conditions, we can reduce the model variance, and thus further boost the discrimination capability [39]. As shown in the last two rows in Table VI, the performance by averaging the three models was slightly better than that of a single 3D CNN model.

We compared the performance of our method with three typical approaches. These methods were implemented on our dataset for direct comparison. The first one employed hand-crafted features based on shape and intensity [15]. The second one constructed a random forest classifier based on low-level features, which is commonly used for 3D object detection tasks in medical applications [40]. The third one utilized a 2D CNN and concatenated 2D features as 3D representations [18].

For the method of [15], we employed its feature extraction procedure on our dataset and utilized a support vector machine (SVM) [41] classifier for prediction. Readers can refer to [15] for details.

For the random forest based method, we extracted intensity-based and geometry-based features. For intensity-based features, referring to [42], we employed the following four groups of features: 1) the local intensity; 2) the mean intensity of the local cuboid; 3) the difference of local intensity and random offset cuboid mean; 4) the difference of local and random offset cuboid means. In addition, to provide an overview of the surrounding tissues, we added the raw 3D intensities within the cubic block in a compact manner, i.e., we utilized PCA [43] to extract the leading 100 principal components as representations. For geometry-based features, we first performed the local thresholding to generate a binary mask. This process is similar to [15]. Specifically, for each voxel, we calculated the local threshold based on the mean and standard deviation of the surrounding tissue. If the voxel of interest was below the threshold, it was marked in the binary mask. Next, referring to [14], we extracted seven geometric features based on the marked regions in the binary mask, which are listed as follows: 1) the volume (number of voxels) of the region V; 2) the sorted sizes of the bounding-box containing the region denoted by $l_{\max}$, l_{med} and $l_{\min}$; 3) the ratios of the sizes $l_{\min}/l_{\max}$ and $l_{\max}/l_{med}$; 4) compactness $V/(l_{\max} \times l_{med} \times l_{\min})$. In total, we extracted 159-dimensional features for the random forest classifier. The classifier used 500 trees and the maximum depth of each tree is 10.

For the method of 2D CNN, we utilized the same network architecture as [18]. Specifically, this method extracted slices from cubic regions as input to the conventional 2D CNN and concatenated the learned features as 3D representations, following a SVM classifier to predict labels. In our experiments, the SVM parameters were optimized on the validation dataset. This method is hereinafter referred as 2D-CNN-SVM.

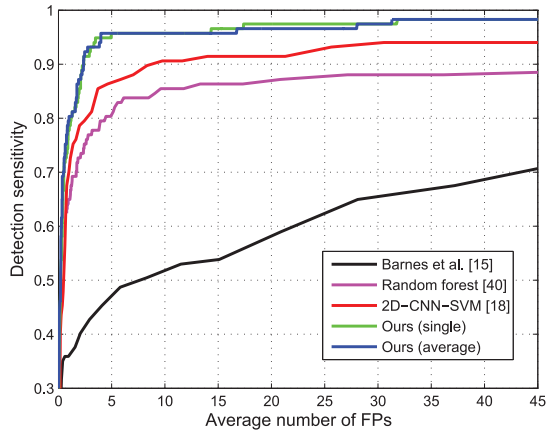


Fig. 11. Comparison of FROC curves of different methods. The top two lines are results produced by our 3D CNN based cascaded frameworks.

Table VI shows the comparison results of different methods and the FROC curves are presented in Fig. 11. Specifically, the method of *Ours (single)* denotes that a single 3D CNN model was used in the second stage, whereas the *Ours (average)* denotes that the model averaging was utilized in the second stage. It is clearly observed that our methods outperform the other three comparison methods by a large margin with the highest detection sensitivity and the fewest false positive predictions. Although the 2D-CNN-SVM method did not sufficiently leverage the 3D spatial characteristics of the CMBs, the high-level features even with limited spatial information obtained better performance than the other two methods employing low-level features. The comparison results between our methods and the 2D-CNN-SVM method demonstrate that our framework benefits from the high-level representations which can encode richer spatial information by leveraging the 3D convolutional architectures. Employing model averaging in the second stage can further improve the overall detection performance.

Figs. 12 and 13 present typical examples of successfully detected CMBs. In Fig. 12 left, there are a number of hard mimics (white rectangles) around the two true CMBs (green rectangles). Our method is able to precisely distinguish them. In Fig. 12 right, the two CMBs are sparsely distributed in the volume with one of them locating at almost the boundary of the volume. In this condition, our method can still accurately detect both of them. In Fig. 13, two CMBs with significantly different sizes and shapes appear in the same volume, while our method can successfully deal with the large variations and accurately figure out them. All these challenging examples demonstrate the effectiveness of the proposed method.

In order to illustrate the discrimination capability of intermediate representations, the features extracted by the 2D-CNN-SVM and 3D CNN discrimination model were embedded into the 2D plane using the t-SNE toolbox [44], as shown in Fig. 14. The CMB and non-CMB samples are

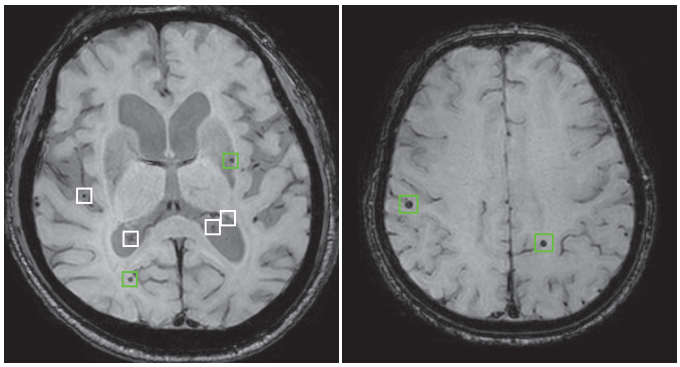


Fig. 12. Examples of CMB detection results (viewed in axial planes). Green rectangles denote the correctly detected CMBs and white rectangles denote the removed false positive candidates by our method.

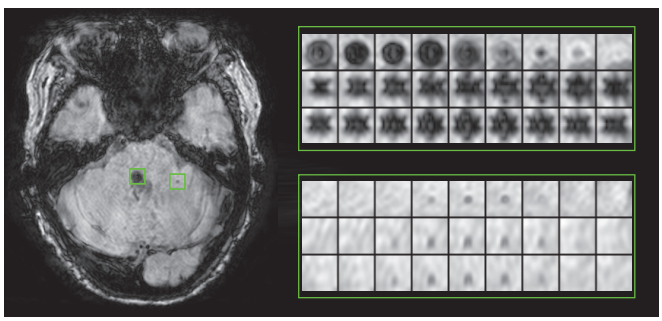


Fig. 13. Examples of correctly detected CMBs with various sizes and shapes. The right part shows the adjacent slices of the CMBs in three dimensions, i.e., axial, sagittal and coronal, from top to down.

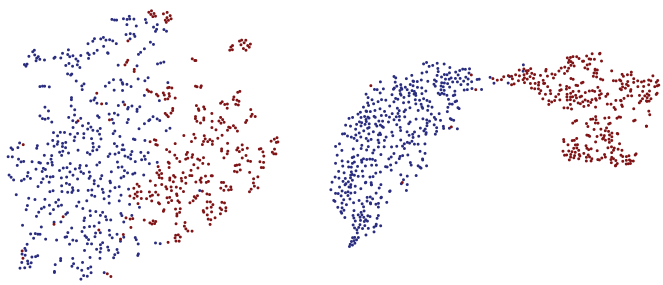


Fig. 14. Feature embedding from 2D-CNN-SVM [18] (left) and 3D CNN methods (right) with t-SNE toolbox. The red and blue colors correspond to the CMBs and non-CMBs, respectively. Best viewed in color.

distinctly separated based on the features extracted via our 3D CNN. In contrast, embedding of the 2D-CNN-SVM representations do not present such a clear partition boundary, highlighting the discrimination capability of the 3D CNN based features which can encode richer spatial information.

E. System Implementation

We implemented the proposed framework based on Theano[1] library using dual Intel Xeon(R) processors E5–2650 2.6 GHz and a GPU of NVIDIA GeForce GTX TITAN Z. The networks were trained with the following hyper-parameters: learning rate $= 0.03$, momentum $= 0.9$, dropout rate $= 0.3$, batch size $= 100$. The trainable weights were randomly initialized from the Gaussian distribution $(\mu = 0, \sigma = 0.01)$ and updated with standard back-propagation. The models converged in about 50 minutes. The 3D FCN inference would take around 1 minute to process a whole SWI volume with size of $512 \times 512 \times 150$, and the 3D CNN in the second stage was quite fast and could process a subject within 1 second. Readers can access our codes and data via our project webpage[2] for more implementation details.

IV. Discussion

The CMBs have been recognized as important biomarkers for cerebrovascular diseases diagnosis and neurologic dysfunction assessment. In current clinical routine, the manual annotation is laborious, time-consuming and error prone. In order to relieve the radiologists of their backbreaking labour and improve the diagnosis efficiency, we propose an efficient and robust two-stage framework for automatic detection of CMBs from SWI images. It is a full stack solution integrated with 3D CNNs for this real challenging medical image processing problem. The first stage efficiently screens the whole volume and retrieves a number of potential candidates with a high sensitivity. It can not only speed up the discrimination procedure but also assist the non-experienced radiologists by timely promoting the candidates for a closer inspection. The second stage robustly discriminates the true CMBs with only a few false positives generated, which can facilitate further segmentation as well as substantial quantification measurements of CMBs. Specifically, our method explores the 3D CNN with shared 3D convolution kernels, which can take full advantage of the spatial information of biomarkers in volumetric data. Extensive experimental results corroborate the efficacy and efficiency of our approach; its performance outperforms other state-of-the-art methods by a significant margin.

In medical image processing community, especially for 3D data computing tasks, 3D CNNs hold promising potentials but have not been well explored yet. Most previous approaches adapted 2D CNNs for processing 3D volumetric data [18], [45], [46], with difficulties being reported when attempting to employ 3D CNNs. To our best knowledge, few works [47], [48] ever utilized real 3D CNNs on medical images, and their architecture settings, convolution kernels and prediction score volumes were not comprehensively presented. One main concern for 3D CNNs is their high computational cost. In order to address this problem in the detection task, we propose a fast way to narrow down the search range to a limited number of candidates by employing the 3D FCN, which can eliminate the redundant convolutional computations during the forward propagation. Another concern of employing 3D CNNs is that the implementation of 3D CNN is effortful. In this paper, we detail the 3D CNN based solution for the challenging CMB detection task, and we shall release our codes with the hyper-parameter configurations to promote research on 3D CNNs. Researchers can avoid a lot of laborious development workload based on our codes.

Although we constructed a relatively larger dataset (i.e., including 320 annotated SWI volumes) than previous work, compared with the natural image domain which usually employs millions of training samples (e.g., ImageNet challenge provides 1.2 million images [21]), we still face the risk of

[1]http://deeplearning.net/software/theano/

[2]http://www.cse.cuhk.edu.hk/~qdou/cmb-3dcnn/cmb-3dcnn.html

over-fitting when training the 3D CNN models. Fortunately, different from the extremely challenging natural image processing tasks (e.g., ImageNet challenge classifies images into 1000 categories [21]), which require exceptionally large and deep models (e.g., AlexNet contains 60 million parameters [21]), most medical image processing tasks are not that complicated and therefore the CNN models employed for these tasks are not necessarily as complicated as those applied to natural image processing. Taking our application as an example, it is a binary classification task. Although the large variations of CMBs and the existence of many hard mimics make the detection task quite challenging, given the small size of CMBs, it is infeasible to construct a network with too many subsampling layers, because the lesions are too tiny to support many layers of feature abstraction. In this regard, we build the 3D FCN screening model consisting of 6 layers (3C, 1M and 2FC) with 0.08 million trainable parameters and the 3D CNN discrimination model also consisting of 6 layers (2C, 1M and 3FC) with 1.3 million parameters. The discrimination model is relatively large because it requires more feature volumes as well as FC layers to improve the representation capability and ensure the detection accuracy. By additionally leveraging the dropout regularization [36], the parameters of the two models are learned using 0.4 and 0.7 million training samples, respectively, without considerable over-fitting observed on the validation dataset.

With the design of two-stage cascaded framework, we keep two aims in mind: efficiency and accuracy. For an automatic lesion detection system targeting clinical practice, we believe that both of them are equally crucial. In the cascaded architecture, the first stage focuses on excluding massive background regions and screening potential candidates. In this stage, we develop the 3D FCN to reduce computational cost, thus meet the requirement of efficiency. The second stage focuses on the small number of candidates and remove the difficult false positives which are with similar appearance to CMBs. In this stage, we employ a discrimination 3D CNN to identify the true CMBs with a high sensitivity and low false positive rate, thus meet the requirement of accuracy. Quantitatively, with the first stage, we obtain around 280 false positives per subject. After the second stage, only less than 3 false positives remain. We can see that the second stage removes nearly 99% false positive candidates using the 3D CNN discrimination model.

The proposed automatic CMB detection framework has great significance in clinical practice. The CMB distribution patterns have been proven to be associated with many cerebrovascular diseases and cognitive dysfunction. For example, the lobar distribution of CMBs suggests probable cerebral amyloid angiopathy [2]. Another research shows that the topographic distribution of CMBs differs between typical and atypical presentations of Alzheimer's disease; the atypical presentations are with greater burden in the frontal lobes and deep brain regions, suggesting possible differences in underlying etiology [49]. In addition, in patients with Alzheimer's disease, the presence of non-lobar CMBs is associated with an increased risk for cardiovascular events and cardiovascular mortality while patients with lobar CMBs have an increased risk for stroke and stroke-related mortality, indicating that these patients should be treated with the utmost care [50]. According to the existing findings of association between CMBs and risk factors of some diseases, we can compare the distribution, location or density of the detected CMBs with that of normal controls and make suggestions for further clinical interventions for the patients.

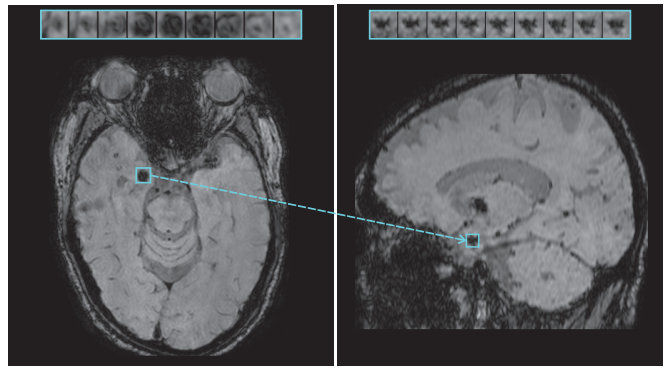


Fig. 15. Example of false negative detection result. Cyan rectangles denote the mis-detected CMB viewed in axial (left) and sagittal (right) planes. Best viewed in color.

Although the proposed method has achieved appealing performance with a high sensitivity of 93.16%, there are several limitations. First, the current detection scheme did not incorporate the SWI filtered phase images [51], which demonstrate the calcium as hyperintense, thus being easily able to differentiate the CMB from the mimic of calcification. In our future work, we shall take into account the phase information to exclude possible mimics of calcifications. Second, as shown in Fig. 15, the presented method may mis-detect some true CMBs with irregular sizes. The false negative shown in Fig. 15 is with quite large size of $12 \times 12 \times 10$ mm^3. However, it is observed that the number of CMBs with a diameter over 10 mm is scarce (less than 0.87%) in our dataset. To achieve a balance between the accuracy and speed, we set the input size as $20 \times 20 \times 16$ (about $10 \times 10 \times 8$ mm^3), which can fulfill the requirements of most cases in our dataset. In such a case, in terms of the mis-detected CMB in Fig. 15, almost no contextual information has been included to recognize it under the current model configurations. To address this issue without compromising detection speed, we plan to integrate the strategy of spatial pyramid pooling [52], which considers the multi-scale/size information during the feature representation phase, into our 3D CNN models in the future.

V. Conclusion

We present an efficient and robust method to automatically detect CMBs from MR volumes leveraging 3D CNNs. With the continuous accumulation of medical data, 3D CNN is a promising solution for many detection and segmentation tasks from 3D volumetric data, as it is capable of representing high-level features with rich spatial information of targets in a data driven way. However, the expensive computational cost of 3D CNN prohibits its use in clinical practice. We propose a two-stage framework to reduce its computational cost and improve the detection performance. The first stage retrieves a number of candidates with high probabilities of being CMBs by leveraging a novel 3D FCN strategy. Compared with traditional sliding window strategy, the 3D FCN strategy eliminates a large

number of redundant convolutional computations, and hence dramatically speeds up the detection procedure. In the second stage, a well-trained model is performed on the candidates to discriminate CMBs from hard mimics. Experimental results demonstrate that the proposed method outperforms previous methods by a large margin with a higher detection sensitivity and fewer false positives. The proposed method can be easily adapted to other detection and segmentation tasks and boost the application of 3D CNNs on volumetric medical data.

ACKNOWLEDGMENT

The authors sincerely thank Dr. Zhaolu Wang for his efforts to carefully label and verify the ground truth of the dataset.

REFERENCES

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

[2] A. Charidimou and D. J. Werring, "Cerebral microbleeds: Detection, mechanisms and clinical challenges," *Future Neurol.*, vol. 6, no. 5, pp. 587–611, 2011.

[3] M. Vernooij *et al.*, "Prevalence and risk factors of cerebral microbleeds the Rotterdam scan study," *Neurology*, vol. 70, no. 14, pp. 1208–1214, 2008.

[4] C. R. Cordonnier, A.-S. Salman, and J. Wardlaw, "Spontaneous brain microbleeds: Systematic review, subgroup analyses and standards for study design and reporting," *Brain*, vol. 130, no. 8, pp. 1988–2003, 2007.

[5] A. Charidimou and D. J. Werring, "Cerebral microbleeds and cognition in cerebrovascular disease: an update," *J. Neurolog. Sci.*, vol. 322, no. 1, pp. 50–55, 2012.

[6] Z. Wang, Y. O. Soo, and V. C. Mok, "Cerebral microbleeds is antithrombotic therapy safe to administer?," *Stroke*, vol. 45, no. 9, pp. 2811–2817, 2014.

[7] M. Akter *et al.*, "Detection of hemorrhagic hypointense foci in the brain on susceptibility-weighted imaging: clinical and phantom studies," *Acad. Radiol.*, vol. 14, no. 9, pp. 1011–1019, 2007.

[8] J. D. Goos *et al.*, "Clinical relevance of improved microbleed detection by susceptibility-weighted magnetic resonance imaging," *Stroke*, vol. 42, no. 7, pp. 1894–1900, 2011.

[9] S. M. Greenberg *et al.*, "Cerebral microbleeds: A guide to detection and interpretation," *Lancet Neurol.*, vol. 8, no. 2, pp. 165–174, 2009.

[10] S. Gregoire *et al.*, "The microbleed anatomical rating scale (MARS) reliability of a tool to map brain microbleeds," *Neurology*, vol. 73, no. 21, pp. 1759–1766, 2009.

[11] A. Fazlollahi *et al.*, "Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the radon transform," in *Proc. IEEE-ISBI Conf.*, 2014, pp. 113–116.

[12] H. J. Kuijf *et al.*, "Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform," *NeuroImage*, vol. 59, no. 3, pp. 2266–2273, 2012.

[13] 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, Clin.*, vol. 2, pp. 282–290, 2013.

[14] B. Ghafaryasl *et al.*, "A computer aided detection system for cerebral microbleeds in brain MRI," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, 2012, pp. 138–141.

[15] S. R. Barnes *et al.*, "Semiautomated detection of cerebral microbleeds in magnetic resonance images," *Magn. Resonance Imag.*, vol. 29, no. 6, pp. 844–852, 2011.

[16] T. van den Heuvel *et al.*, "Computer aided detection of brain microbleeds in traumatic brain injury," in *Proc. SPIE Med. Imag. Int. Soc. Opt. Photon.*, 2015, p. 94142F.

[17] Q. Dou *et al.*, "Automatic cerebral microbleeds detection from MR images via independent subspace analysis based hierarchical features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 7933–7936.

[18] H. Chen *et al.*, "Automatic detection of cerebral microbleeds via deep learning based 3d feature representation," in *Proc. IEEE-ISBI Conf.*, 2015, pp. 764–767.

[19] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5325–5334.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[22] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.

[23] H. Chen, Q. Dou, D. Ni, J.-Z. Cheng, J. Qin, S. Li, and P.-A. Heng, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2015*. New York: Springer, 2015, pp. 507–514.

[24] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2013*. New York: Springer, 2013, pp. 411–418.

[25] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *ArXiv Preprint ArXiv:1505.06448*, 2015.

[26] H. R. Roth *et al.*, "Anatomy-specific classification of medical images using deep convolutional nets," in *Proc. IEEE-ISBI Conf.*, 2015, pp. 101–104.

[27] H. R. Roth *et al.*, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," *ArXiv Preprint ArXiv:1506.06448*, 2015.

[28] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2013*. New York: Springer, 2013, pp. 246–253.

[29] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI*. New York: Springer, 2014, pp. 520–527.

[30] G. Urban, M. Bendszus, F. A. Hamprecht, and J. Kleesiek, "Multimodal brain tumor segmentation using deep convolutional neural networks," in *Proc. Winning Contribution MICCAI BraTS (Brain Tumor Segmentation) Challenge*, 2014, pp. 31–35.

[31] S. C. Turaga *et al.*, "Convolutional networks can learn to generate affinity graphs for image segmentation," *Neural Comput.*, vol. 22, no. 2, pp. 511–538, 2010.

[32] F. Ning *et al.*, "Toward automatic phenotyping of developing embryos from videos," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1360–1371, Sep. 2005.

[33] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

[34] Y. Bengio, I. J. Goodfellow, and A. Courville, Deep Learning. Cambridge, MA, MIT Press, book in preparation, 2015 [Online]. Available: http://www.iro.umontreal.ca/~bengioy/dlbook

[35] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2011, pp. 315–323.

[36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv Preprint ArXiv:1207.0580*, 2012.

[37] K. Kang and X. Wang, "Fully convolutional neural networks for crowd segmentation," *ArXiv Preprint ArXiv:1411.4464*, 2014.

[38] J. de Bresser *et al.*, "Visual cerebral microbleed detection on 7t MR imaging: reliability and effects of image processing," *Am. J. Neuroradiol.*, vol. 34, no. 6, pp. E61–E64, 2013.

[39] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, no. 1, pp. 1–58, 1992.

[40] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News* vol. 2, no. 3, pp. 18–22, 2002 [Online]. Available: http://CRAN.R-project.org/doc/Rnews/

[41] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[42] D. Zikic, B. Glocker, and A. Criminisi, "Atlas encoding by randomized forests for efficient label propagation," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2013*. New York: Springer, 2013, pp. 66–73.

[43] I. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer, 1986, vol. 1.
[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2579-2605, p. 85, 2008.
[45] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *ArXiv Preprint ArXiv:1505.03046*, 2015.
[46] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2013*. New York: Springer, 2013, pp. 246–253.
[47] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, "Multi-scale 3d convolutional neural networks for lesion segmentation in brain MRI," in *Proc. MICCAI Ischemic Stroke Lesion Segmentation Challenge*, 2015, p. 13.
[48] G. Urban, M. Bendszus, F. A. Hamprecht, and J. Kleesiek, "Multi-modal brain tumor segmentation using deep convolutional neural networks," in *Proc. Winning Contribution MICCAI BraTS (Brain Tumor Segmentation) Challenge*, 2014, pp. 31–35.
[49] J. L. Whitwell *et al.*, "A comparison of the regional distribution of microbleeds in atypical and typical presentations of Alzheimer's disease," *Alzheimer's Dementia*, vol. 10, no. 4, p. P20, 2014.
[50] M. R. Benedictus *et al.*, "Microbleeds, mortality, and stroke in Alzheimer disease: The mistral study," *JAMA Neurol.*, vol. 72, no. 5, pp. 539–545, 2015.
[51] Z. Wu *et al.*, "Identification of calcification with MRI using susceptibility-weighted imaging: A case study," *J. Magn. Reson. Imag.*, vol. 29, no. 1, pp. 177–182, 2009.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Comput. Vision-ECCV 2014*. New York: Springer, 2014, pp. 346–361.