CHAPTER

6 Deep Cascaded Networks for Sparsely Distributed Object Detection from Medical Images

Hao Chen*, Qi Dou*, Lequan Yu*, Jing Qin†, Lei Zhao*, Vincent C.T. Mok*, Defeng Wang*, Lin Shi*, Pheng-Ann Heng*

*The Chinese University of Hong Kong, Hong Kong, China**

The Hong Kong Polytechnic University, Hong Kong, China†

CHAPTER OUTLINE

Deep Learning for Medical Image Analysis. DOI: 10.1016/B978-0-12-810408-8.00008-0
Copyright © 2017 Elsevier Inc. All rights reserved.

CHAPTER POINTS

- A cascaded deep learning framework, composed of a coarse retrieval model and a fine discrimination model, is proposed to efficiently and accurately detect sparsely distributed objects from medical images
- We demonstrate the importance of volumetric feature representations by exploiting 3D deep learning on volumetric imaging modalities
- Extensive experimental results on two challenging applications including both 2D and 3D imaging modalities corroborated the outstanding performance and generalization capability of our method in terms of speed and accuracy

6.1 INTRODUCTION

Computer-aided diagnosis (CAD) has become one of the major research subjects in medical image computing and clinical diagnosis [1]. In the pipeline of a CAD system, clinically related object detection from medical images plays an important role, which has a wide range of applications, such as mitosis detection [2], nuclei detection [3], cerebral microbleed detection [4], lung nodule detection [5], cell tracking, lesion localization [6], etc. In general, the clinical routine to annotate targets of interest is based on visual inspection and manual localization [1], which suffers from a low reproducibility among different observers and could be laborious and time-consuming, especially within the context of large numbers of subjects. Robust object detection from medical images can release doctors from considerable workload, provide reliable quantitative assessments, and accelerate the diagnosis procedure.

However, how to efficiently and effectively detect the sparsely distributed objects from large-scale data remains a challenging problem. Although the situation may vary depending on the imaging modalities and confronting detection tasks, there are several common challenges: (i) There is a high intra-class variation of object appearance in medical images. Take the mitosis for example, it is characterized by a large variety of shape configurations, which are related to the high variation of biological structures, as shown in the green rectangle in Fig. 6.1A. Furthermore, the different conditions of histology image acquisition process, including sampling, cutting, staining, and digitalizing, increase the variabilities of mitosis appearance [2,7]. This is common when the tissue samples are acquired from different patients or at different time slots. (ii) The sparsely distributed objects require the algorithm to be efficient and robust when applied to large-scale data. In the example of cerebral microbleeds (CMBs), the widespread distributed locations of CMBs make complete and accurate detection very challenging [8,9]. (iii) There are strong mimics that hinder the detection process, therefore the automatic methods should be robust and discriminative to

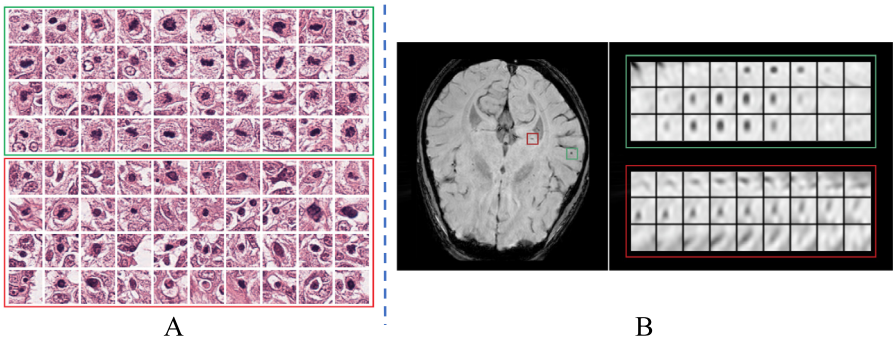

FIGURE 6.1

Examples of target objects and mimics. (A) Mitosis detection (green rectangle encloses the true mitoses and red rectangle encloses the mimics that carry similar appearance); (B) CMB and mimics are denoted with green and red rectangles, respectively (in each of the rectangle, from top to down, the rows show adjacent slices in axial, sagittal and coronal planes).

single out the true objects from hard mimics. In the case of mitosis detection, some cell types (e.g., apoptotic cells) carry similar morphological appearance with mitosis, as shown in the red rectangle in Fig. 6.1A, resulting in lots of false positives in the detection process. Similarly, in the task of CMB detection, there exist many sources of hard CMB mimics, e.g., flow voids, calcification and cavernous malformations, see the red rectangle in Fig. 6.1B, which would resemble the appearance of CMBs in magnetic resonance (MR) scans and heavily impede the detection process [10].

Recently, deep neural networks have made breakthroughs in image recognition related tasks including object detection [11], semantic segmentation [12], and image classification [13,14]. Inspired by this and aiming to tackle the aforementioned challenges, we propose a fast and accurate method to detect sparsely distributed objects from medical images by designing deep cascaded neural networks (CasNN). Extensive experimental results on two typical applications including mitosis detection from histopathological images and CMB detection from MR volumes demonstrated promising performance and high generalization capability of our framework.

The remainder of this chapter is organized as follows. In Section 6.2, we describe the cascaded framework based on deep convolutional neural networks. Extensive experimental results on 2D histopathological images in Section 6.3, and volumetric data in Section 6.4 demonstrate the robustness and efficiency of the proposed method. Meanwhile, Section 6.4 elaborates the 3D CNN and verifies the effectiveness of volumetric feature representations for 3D detection tasks. This chapter concludes with a discussion in Section 6.5. An early version of this work was published in [4] and [15].

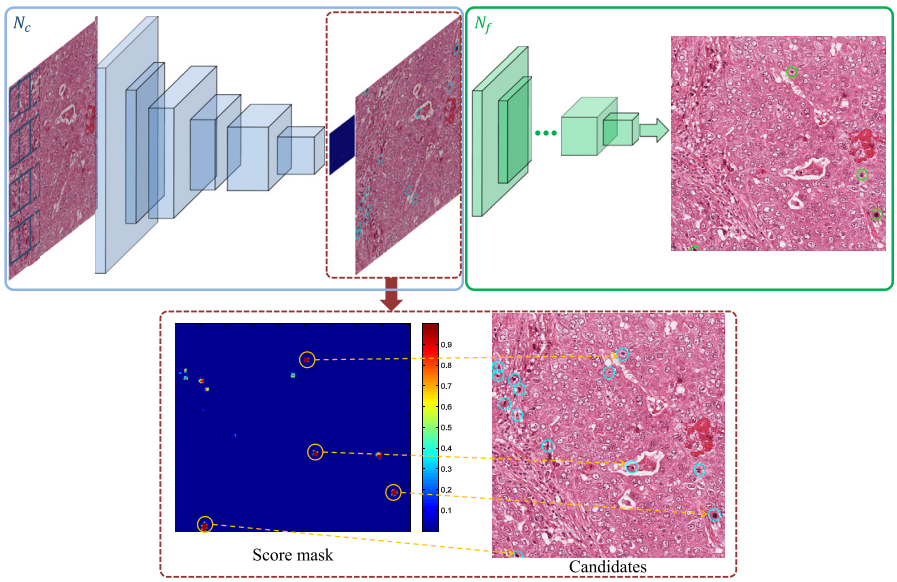

FIGURE 6.2

An overview of deep cascaded networks for sparsely distributed object detection. The cascaded framework consists of a coarse retrieval model N_c and a fine discrimination model N_f (we take the task of mitosis detection from histopathological images for illustration).

6.2 METHOD

Fig. 6.2 shows the architecture of the proposed method (we take the example of mitosis detection for illustration), which consists of two models of convolutional neural networks integrated in a cascaded manner. The first model quickly retrieves the object candidates while preserving a high sensitivity by taking the advantage of the fully convolutional network. We call it the *coarse retrieval model* N_c, which outputs a score map indicating the probability of candidates. The second model is effectively constructed by a deeper convolutional neural network. We call it the *fine discrimination model* N_f, which has a higher capability of feature representation, hence it accurately discriminates the true targets and mimics. Note that as the N_f performs only on the candidates generated by N_c instead of the whole image, hence the detection process can be significantly accelerated.

6.2.1 COARSE RETRIEVAL MODEL

Considering that objects are sparsely distributed in the whole image, a step of retrieving the regions of interest (ROI), e.g., mitosis candidates, could reduce the detection time dramatically, as the subsequent detection process could focus only on

the candidates. Previous studies obtained the candidates relying on the pre-defined measurements on domain-specific morphological textures, color ratios, or histogram distribution [16–19]. However, these methods were prone to losing targets, as these handcrafted features could not accurately describe the complicated characteristics of objects. Recently, there have been studies that explored the fast scanning approach with deep max-pooling convolutional neural networks for the mitosis detection [20, 21]. Although these methods are more accurate than previous studies based on low-level features, they are computation-intensive and time-consuming in a pixel-wise classification way.

Different from previous methods, we utilize a fully convolutional network (FCN) for fast retrieving of the mitosis candidates. A traditional convolutional neural network (CNN) contains the convolutional (C), sub-sampling, e.g., max-pooling (M), and fully connected (FC) layers. Both C and M layers are translation-invariant and can be operated on input of arbitrary size. However, the introduction of FC layers requires the input with a fixed size,

$$\boldsymbol{h}^l_j = \sigma(\boldsymbol{W}^l_j \boldsymbol{h}^{l-1} + b^l_j), \tag{6.1}$$

where $\boldsymbol{W}^l_j$ is the weight matrix connecting the neurons $\boldsymbol{h}^{l-1}$ in $(l-1)$th FC layer and jth index neuron $\boldsymbol{h}^l_j$ in the lth FC layer, b^l_j is the bias and $\sigma(\cdot)$ is the element-wise nonlinear activation function. In fact, the fully connected layers are equivalent to the convolutional layers with kernel size 1×1,

$$\boldsymbol{h}^l_j = \sigma(\sum_k \boldsymbol{W}^l_{jk} \otimes \boldsymbol{h}^{l-1}_k + \boldsymbol{b}^l_j), \tag{6.2}$$

where $\otimes$ denotes the 2D spatial convolution, $\boldsymbol{W}^l_{jk}$ is the convolution kernel connected to jth feature map $\boldsymbol{h}^l_j$ and the kth feature map in the previous layer $\boldsymbol{h}^{l-1}$. By employing Eq. (6.2), we can convert the fully connected layers into a fully convolutional fashion. Once the filters have been trained, the fully convolutional network can be applied to an input image of arbitrary size.

The proposed coarse retrieval model has several advantages. First, since the FCN can take the whole image as input and generate the score mask with only one pass of forward propagation, it is capable of retrieving candidates efficiently while preserving a high sensitivity. Second, it can also help to build a representative training database for the fine discrimination model N_f. Specifically, the targeting objects are sparsely distributed and rarely appeared in the whole image, the mimics can be well represented by putting the false positives from N_c into the training samples of N_f. In this way, the capability of the model N_f in distinguishing the true targets from the hard mimics can be greatly enhanced.

6.2.1.1 Score Mask Generation

The proposed coarse retrieval model is trained on the training samples with a fixed size input ($94 \times 94 \times 3$, architecture of N_c for mitosis detection is shown in Table 6.1)

Table 6.1 The architecture of coarse retrieval model N_c

Layer	Kernel size	Stride	Output size	Feature maps
Input	—	—	94×94	3
C1	5×5	1	90×90	32
M1	3×3	3	30×30	32
C2	3×3	1	28×28	32
M2	2×2	2	14×14	32
C3	3×3	1	12×12	32
M3	2×2	2	6×6	32
FC4	—	—	100	—
FC5	—	—	2	—

by minimizing the cross entropy loss. Once the training is done, the N_c can be converted into a FCN model by Eq. (6.2). Then, the trained filters can be applied to scan the whole image instead of employing the traditional patch-by-patch manner, which speeds up the detection process dramatically. Hence, the score mask indicating probabilities of candidates can be obtained with only a single forward propagation of the converted coarse retrieval model. Each position of the output score mask corresponds to a specific region (size 94×94) in the original image. Actually, it is equivalent to scanning the whole image with a fixed stride, which is mainly determined by the stride of max-pooling layers. We will detail this in the following section.

6.2.1.2 Object Candidate Localization

Derived from the proposed model, the candidates can be located by mapping the index with higher scores on the score mask into the original coordinates of input image. Assuming non-overlapping region pooling, index mapping with convolution and max-pooling operations is formulated as

$$\hat{x}_i = \frac{x_i - c_i}{s_i} + 1, \tag{6.3}$$

where c_i denotes the kernel size of convolutional or max-pooling layer, $\hat{x}_i$ is the position index after C or M operation on x_i, and s_i denotes the stride of convolutional or max-pooling layer. The original position index can be obtained by inverting above operations. For example, based on the network architecture shown in Table 6.1, for each position $\hat{p}_s$ in the score mask, we can get the index p_o in the original image as

$$\begin{aligned} p_o = {} & c_1 - 1 + s_{m_1}(c_2 - 1) + s_{m_1}s_{m_2}(c_3 - 1) \\ & + s_{m_1}s_{m_2}s_{m_3}\hat{p}_s = p_0 + s\,\hat{p}_s, \end{aligned} \tag{6.4}$$

where $p_0 = 22$ and $s = 12$ according to the architecture in Table 6.1. Thus we can retrieve the candidates with a sparse distribution based on the above index mapping. This is quite efficient when the target objects are sparsely distributed in the whole image. Despite with max-pooling layers, the probability maps can give quite dense

predictions considering the equivalent stride 12 compared with the image size 2048 × 2048. Therefore, this approach can efficiently retrieve the candidates with a high sensitivity while reducing the computational workload. Subsequently, candidates are input into the model N_f for fine discrimination after local smoothing and non-max suppression.

6.2.2 FINE DISCRIMINATION MODEL

In order to increase the representation capability for singling out true targets from mimics, the fine discrimination model is constructed by a deeper convolutional neural network (e.g., we used CaffeNet [22] with 8 layers in mitosis detection). In addition, the training samples are augmented with different transformations to enlarge the training dataset. Finally, the fine discrimination model N_f is optimized by minimizing the following cross-entropy function with standard back-propagation:

$$\arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \sum_{k=1}^{K} -t_k \log p(y_k = 1|\boldsymbol{I}_n) + \lambda ||\boldsymbol{W}||_2^2, \tag{6.5}$$

where $\boldsymbol{\theta} = \{\boldsymbol{W}, \boldsymbol{b}\}$ denotes the parameters of N_f, λ is the parameter for controlling the balance between the data loss term and the regularization term, $p(y_k = 1|\boldsymbol{I}_n)$ is the output probability for kth class given the input sub-window patch $\boldsymbol{I}_n$, t_k is the corresponding ground truth, K and N are the total number of classes and training samples, respectively. In the training process, dropout method [23] was utilized to reduce the co-adaption of intermediate features. In order to improve the robustness, we trained multiple models of N_f for reducing the variance and improving the robustness capability [24]. For example, three architectures with different number of neurons in three fully connected layers (i.e., FC6–FC8), 1024–256–2, 1024–512–2, 512–256–2, were trained in the mitosis detection. The sub-window sample was categorized as a true target when its averaged output posterior probability was above the threshold T (determined with cross-validation in our experiments), otherwise, categorized as non-target.

6.3 MITOSIS DETECTION FROM HISTOLOGY IMAGES

6.3.1 BACKGROUND

Breast cancer is the most common cancer for women and a major cause of cancer death worldwide [25]. According to the Nottingham Grading System, three morphological features in histology sections, including tubule formation, nuclear pleomorphism and number of mitotic figures, are critical for the diagnosis of breast cancer [26]. Among them, the number of mitoses gives an important aggressiveness indication of the invasive breast carcinoma.

Previous studies employed domain-specific handcrafted features to describe the morphological, statistical, or textural characteristics of mitosis [16–19]. However,

these handcrafted features usually require considerable efforts to design and validate. Furthermore, they cannot sufficiently represent the characteristics of mitoses with a large variation of shapes and textures, therefore resulting in a low detection accuracy. Compared with methods based on handcrafted features, the deep CNN with powerful feature representation learning has achieved state-of-the-art performance in object recognition related problems [13,14,27]. Regarding the mitosis detection, Ciresan et al. utilized a deep CNN as a pixel-wise classifier to detect mitosis and achieved the best performance at 2012 ICPR MITOSIS challenge [21] and 2013 MICCAI challenge [2], respectively. However, the pixel-wise classifier of deep CNN is computation-intensive and time-consuming. Considering a single whole slide that consists of thousands of high-power fields (HPFs), it takes a long time to run across all sub-windows for detection, which may prohibit its potential in clinical practice.

6.3.2 TRANSFER LEARNING FROM CROSS-DOMAIN

The deep CNN with powerful feature representation achieved remarkable performance on recognition related tasks with large scale training data available. However, limited data in the field of medical applications increases the difficulties for training a powerful model to discriminate objectives from their mimics. The situation is further deteriorated with existence lots of false positives carrying similar appearance. Although various transformation augmentations could be used to enlarge the training database, the training samples may be still insufficient to train a powerful model. Further improvement can be obtained by transferring knowledge learned from related auxiliary tasks, where the training data can be easily acquired. Previous studies demonstrated that the filters trained on large scale images of ImageNet [28] could be transferred to different applications in other domains empirically [22,29–31]. Therefore, we optimized the new medical task by employing an off-the-shelf model CaffeNet [22]. The parameters of the previous layers (C1–C5) in the N_f were initialized by the pre-trained filters of CaffeNet model, which was trained on large scale images of ImageNet. Then the whole network was jointly optimized on the specific task.

6.3.3 DATASET AND PREPROCESSING

The datasets were obtained from the 2012[1] and 2014 ICPR MITOSIS contests.[2] In our experiments, we evaluated our method on the HPF images acquired by the widely-used Apero-XT scanner. The 2012 ICPR MITOSIS dataset consisted of 50 images (train/test = 35/15) with 326 annotated mitoses, while the 2014 ICPR MITOSIS dataset was extensively enlarged (train/test = 1200/496). The centroids of mitoses were annotated by experienced pathologists and the ground truths of 2014 ICPR testing data were held out by the organizers for evaluation. For each dataset, we split training data with ground truth into two sets for training and validation (about 1/7 of total training data), respectively. Patches extracted from mitotic regions were augmented by different transformations, including translation, rotation and flipping, for enlarging the training dataset.

Table 6.2 Results of 2012 ICPR MITOSIS dataset

Method	Precision	Recall	F_1 score
SUTECH	0.699	0.720	0.709
IPAL [16]	0.698	0.740	0.718
DNN [21]	**0.886**	0.700	0.782
Retrieval model N_c	0.211	**0.891**	0.342
RCasNN	0.720	0.713	0.716
CasNN (single)	0.738	0.753	0.745
CasNN (average)	0.804	**0.772**	**0.788**

6.3.4 QUANTITATIVE EVALUATION AND COMPARISON

According to the challenge evaluation, a detection would be regarded as a true positive if its distance to a ground truth mitosis is less than 8 μm. All the detections that are not fallen within 8 μm of a ground truth are counted as false positives. All the ground truths that do not have detections within 8 μm are counted as false negatives. The evaluation measurements include recall $R = N_{tp}/(N_{tp} + N_{fn})$, precision $P = N_{tp}/(N_{tp} + N_{fp})$, and F_1-score $F_1 = 2RP/(R + P)$, where N_{tp}, N_{fp}, and N_{fn} are the number of true positives, false positives, and false negatives, respectively.

In order to build a representative dataset for training the fine discrimination model, false positives from the coarse retrieval model were employed. Note that the size of training data (total 394,275 training samples including 7.4% mitoses, 67.8% random selected negative samples, and 24.8% false positives from N_c) is smaller compared to the training dataset used in the state-of-the-art method [21] (1 million training instances with 6.6% as mitoses). Although with less aggressive data augmentation, our method achieved a competitive performance on testing data with a much faster speed. The detailed results are reported in Table 6.2. Compared to the method with the best performance in 2012 ICPR contest [21], our method (CasNN) with model averaging achieved a comparable F_1-score of 0.788 and a higher recall of 0.772. The single CasNN model outperformed the randomly initialized model (RCasNN) with respect to all the evaluation measurements, demonstrating that the knowledge transferred from deep and rich hierarchies can help to improve the performance. Four typical detection examples on the testing data of 2012 ICPR MITOSIS contest are shown in Fig. 6.3. It is observed that the tissue appearance has large variations, which increases the difficulties for automated mitosis detection. Nevertheless, our method can successfully detect most of mitoses from histology images with a low false negative rate and only a few false positives, which verifies the efficacy of our method qualitatively.

As the organizers from 2012 ICPR MITOSIS contest indicated that the dataset in this contest is too small for a good assessment of reliability and robustness of different algorithms [32], we further evaluated our method on the 2014 ICPR MITOSIS dataset, which was extensively expanded (1200 training images and 496 testing images). One of the most difficult challenges in this dataset is the variability of tissue appearance, mostly resulted from the different conditions during the tissue acquisi-

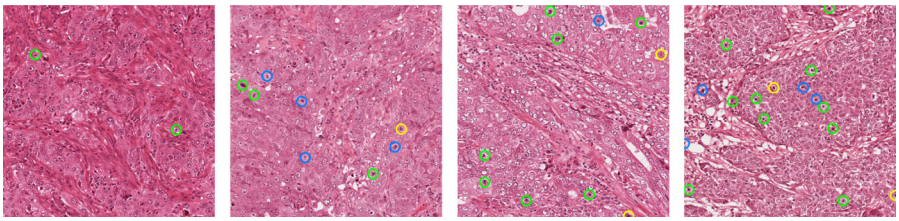

FIGURE 6.3

Mitosis detection results of our method on the testing data of 2012 ICPR MITOSIS contest. Yellow, blue, and green circles denote the false negatives, false positives, and true positives, respectively.

Table 6.3 Results of 2014 ICPR MITOSIS dataset

Method	Precision	Recall	F_1 score
STRASBOURG	—	—	0.024
YILDIZ	—	—	0.167
MINES-CURIE-INSERM	—	—	0.235
CUHK	—	—	0.356
RCasNN	0.360	0.424	0.389
CasNN (single)	0.411	0.478	0.442
CasNN (average)	**0.460**	**0.507**	**0.482**

tion process. As a result, the dataset is much more challenging than that in 2012 ICPR. The results of different methods are reported in Table 6.3. Our approach achieved the best performance with F_1-score of 0.482, outperforming other methods by a large margin. The performance of the single CasNN model outperformed the RCasNN model, demonstrating the efficacy of transfer learning strategy consistently.

6.3.5 COMPUTATION COST

In histopathological examination of breast cancer diagnosis, a single whole slide usually consists of thousands of HPFs. Thus, the processing time of one HPF should be taken into account in clinical applications [2]. The superior advantage of the proposed cascaded framework is that it can reduce detection time dramatically while achieving a satisfactory accuracy. The coarse retrieval model took about 0.45 s to process per 4 megapixel HPF and the fine discrimination model with 10 input variations cost about 0.49 s using a workstation with a 2.50 GHz Intel(R) Xeon(R) E5-2609 CPU and a NVIDIA GeForce GTX TITAN GPU. Totally, it took about 0.5 s for each input variation and was roughly 60 times faster than the state-of-the-art method [21], which took about 31 s with an optimized GPU implementation. Meanwhile, our approach achieved comparable detection accuracy to [21]. This makes our approach possible for real-world clinical applications.

6.4 CEREBRAL MICROBLEED DETECTION FROM MR VOLUMES

6.4.1 BACKGROUND

Cerebral microbleeds refer to small foci of chronic blood products in (near) normal brain tissues, as shown in Fig. 6.1B, which are considered to be composed of hemosiderin deposits that leak through pathological blood vessels [33]. The existence of CMBs and their distribution patterns have been recognized as important diagnostic biomarkers of cerebrovascular diseases. In addition, CMBs could also structurally damage their nearby brain tissues, and further cause neurologic dysfunction, cognitive impairment, and dementia [34]. In this regard, reliable detection of the presence and number of CMBs is crucial for cerebral diagnosis and guiding physicians in determining necessary treatment.

Previous computer-aided CMB detection methods mainly employed handcrafted features based on shape, size and intensity information. For example, Fazlollahi et al. [35] utilized the radon transform to describe the shape information of CMBs, and Kuijf et al. [36] have applied the radial symmetry transform (RST) to identify spherical regions. To improve the capability of discrimination, Bian et al. [37] proposed to measure the geometric features after performing a 2D fast RST. Ghafaryasl et al. [38] further designed more comprehensive features that integrated the geometry, intensity, scale and local image structures. However, the design of these handcrafted features heavily depends on domain knowledge of the lesion. Furthermore, 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 [39,31]. Among them, convolutional neural network 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.

Our objective in this task is to detect CMBs which are sparsely distributed in a 3D brain volume. However, how to effectively employ CNNs on volumetric data still remains an open problem in medical image computing community, even though CNNs have presented outstanding effectiveness on 2D medical imaging applications [40–42]. One straightforward solution is to employ conventional 2D CNNs based on a single slice and process the slices sequentially [43–45]. Apparently, this solution disregards the contextual information along the third dimension, thus its performance would be heavily degraded. Alternatively, some researches aggregate adjacent slices [31] or orthogonal planes (i.e., axial, coronal and sagittal) [46,47] to enhance complementary spatial information. Nevertheless, this solution is still unable to make full use of the volumetric spatial information. Note that the spatial information of all three dimensions is quite important for our CMB detection task. As shown in Fig. 6.1B, 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 is a more promising and reliable solution to take full advantage of spatial contextual information in volumetric data for accurate detection of CMBs.

6.4.2 3D CASCADED NETWORKS

We extend the detection framework of 2D cascaded networks into a volumetric version. In the first screening stage, the 3D fully convolutional network 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. In the second discrimination stage, we further remove a large number of false positive candidates by applying a 3D CNN discrimination model to distinguish the true CMBs from challenging mimics.

6.4.2.1 3D Convolutional Neural Network

Learning feature representations from volumetric contextual information is vitally important for biomarker detection tasks from 3D medical images. In this regard, we propose to employ the 3D convolution kernel, in the pursuit 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). Given a volumetric image, when we employ a 3D convolution kernel to generate a 3D feature volume, the input to the network is the entire volumetric data. 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 fashion:

$$\boldsymbol{u}^{l}_{jk}(x, y, z) = \sum_{m,n,t} \boldsymbol{h}^{l-1}_{k}(x-m, y-n, z-t)\ \boldsymbol{W}^{l}_{jk}(m, n, t), \tag{6.6}$$

where $\boldsymbol{W}^{l}_{jk}$ is the 3D kernel in the lth layer which convolves over the 3D feature volume $\boldsymbol{h}^{l-1}_{k}$ in the $(l-1)$th layer, $\boldsymbol{W}^{l}_{jk}(m, n, t)$ is the element-wise value of the 3D convolution kernel. The 3D feature volume $\boldsymbol{h}^{l}_{j}$ is obtained by different 3D convolution kernels as

$$\boldsymbol{h}^{l}_{j} = \boldsymbol{\sigma}(\sum_{k} \boldsymbol{u}^{l}_{jk} + \boldsymbol{b}^{l}_{j}). \tag{6.7}$$

Then a 3D CNN model is constructed by stacking the C, M and FC layers hierarchically, as shown in Fig. 6.4. 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 rectifier linear unit (ReLU) [48] is utilized for the nonlinear activation function in the C and FC layers. Finally, the output layer employs the softmax regression to yield the prediction probabilities.

6.4.2.2 Screening Stage

Previous works proposed to screen CMB candidates in a MR volume by employing local statistical information, including size, intensity, shape and other geometric features [49,31,38,39]. However, due to the large variations of CMBs, only relying on

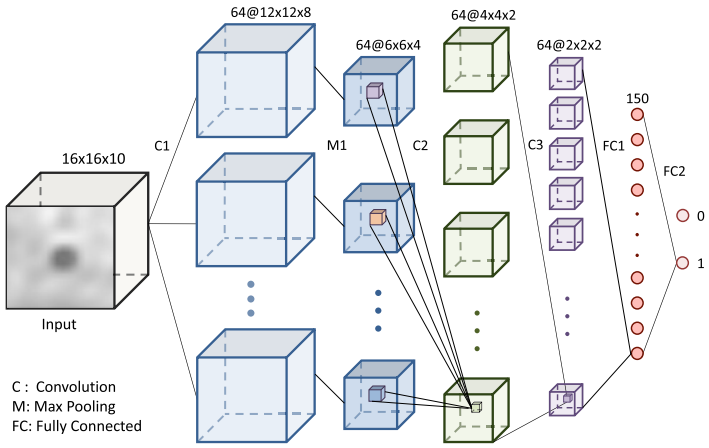

FIGURE 6.4

The hierarchical architecture of the 3D CNN model.

these statistic values is difficult to precisely describe the characteristics of CMBs and detach them from the background regions. The results would easily 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 discriminative representations of CMBs learned from a large number of 3D samples. Inspired by the 2D FCN, we extend the strategy into a 3D format, i.e., 3D FCN, for efficient and effective retrieval of CMB candidates from MR volumetric data. During the training phase, the positive samples are extracted from CMB regions and augmented by translation, rotation, and mirroring to expand the training database. During the testing phase, the 3D 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 and architecture in Table 6.4. Finally, regions with high prediction probabilities are selected as the potential candidates.

6.4.2.3 *Discrimination Stage*

The screened 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 therefore improve the discrimination capability of the 3D CNN model, the obtained false positives (which take very similar appearance to true CMBs) on the training set in the screening stage are taken as negative

Table 6.4 The architecture of 3D FCN screening model

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

Table 6.5 The architecture of 3D CNN discrimination model

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

samples when training the 3D CNN in the second stage. The network architecture of the discrimination model is shown in Table 6.5.

6.4.3 DATASET AND PREPROCESSING

A large dataset of susceptibility-weighted imaging (SWI) images, referred to as *SWI-CMB*, was built to validate the framework for CMB detection in volumetric data. 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). The dataset was labeled by an experienced rater and was verified by a neurologist following the guidance of Microbleed Anatomical Rating Scale [50]. Overall, a total of 1149 CMBs were annotated from the whole dataset and regarded as the ground truth in our experiments. We randomly divided the whole dataset into three sections for training (230 subjects), validation (40 subjects), and testing (50 subjects), respectively. In the preprocessing step, we normalized the volume intensities to the range of [0, 1] after trimming the top 1% gray-scale values [31].

Table 6.6 Comparison of different screening methods

Methods	Recall	FP_{avg}	Time/subject (s)
Barnes et al. [49]	85.47%	2548.2	81.46
Chen et al. [31]	90.48%	935.8	**12.00**
3D FCN model	**98.29%**	**282.8**	64.35

6.4.4 QUANTITATIVE EVALUATION AND COMPARISON

6.4.4.1 *CMB Candidate Localization*

Three commonly used metrics were employed to quantitatively evaluate the performance of the proposed CMB detection method including recall, precision, and the average number of false positives per subject (FP_{avg}).

We compared the screening performance of our proposed method with two state-of-the-art approaches which utilize low-level statistical features [49,31]. The results are listed in Table 6.6. The values of sensitivity mean the percentage of successfully retrieved CMBs while the values of FP_{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 recall 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.

We have also recorded the average time for screening each subject and the results are shown in Table 6.6. From the clinical perspective, the time performance of our method is satisfactory; processing a whole volume with a size of $512\times512\times150$ takes around 1 min in our experiments. The method of [49] is slower than ours because it calculates local thresholds using a voxel-wise sliding window strategy. In contrast, the method of [31] 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 [31] is faster, we achieved around 8% increase in recall and reduced the number of FP_{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 Fig. 6.5. 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, which dramatically reduces the computational workload in the following stage.

6.4.4.2 *True CMB Discrimination*

Employing the CMB samples with augmentations and the false positives generated from the training set in the screening stage, we trained the 3D CNN discrimination

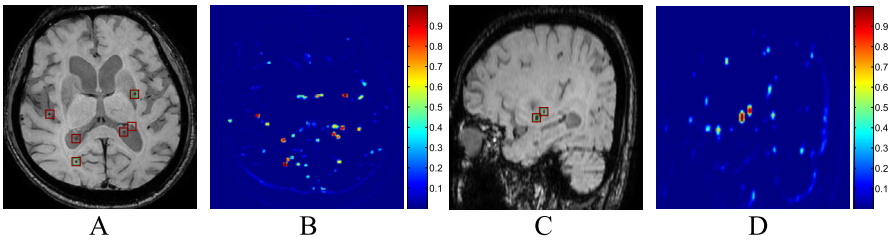

FIGURE 6.5

Example results of the 3D FCN screening model. (A), (C) Raw data with true CMBs (green rectangles) and screened candidates (red rectangles) in the axial and sagittal plane, respectively. (B), (D) Corresponding 2D projection of the score volume generated with 3D FCN.

Table 6.7 Evaluation of detection results on SWI-CMB dataset

Methods	Recall	Precision	FP_{avg}
Barnes et al. [49]	64.96%	5.13%	28.10
Random forest [51]	85.47%	17.24%	9.60
2D-CNN-SVM [31]	88.03%	22.69%	7.02
CasNN (single)	92.31%	42.69%	2.90
CasNN (average)	**93.16%**	**44.31%**	**2.74**

model to remove false positive candidates and accurately identify true CMBs. We independently trained three models using the network architecture shown in Table 6.5. The differences of the three models lie in the random weights initialization states and the number of training epochs. This is because that the neural network with a large number of parameters is usually with a low bias and a high variance. By averaging multiple models with different training conditions, we can reduce the model variance, and therefore further boost the discrimination capability [24]. As shown in the last two rows of results in Table 6.7, the performance by averaging the three models was 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 handcrafted features based on shape and intensity [49]. We employed its feature extraction procedure on our dataset and utilized a support vector machine (SVM) [52] classifier for prediction. The second one constructed a random forest (RF) classifier based on low-level features, which is commonly used for 3D object detection tasks in medical applications [51]. Specifically, we extracted a total of 159 intensity and geometry-based features. The RF classifier used 500 trees and the maximum depth of each tree is 10. The third one 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 [31]. We followed the same network architecture as [31], which is hereinafter referred as 2D-CNN-SVM.

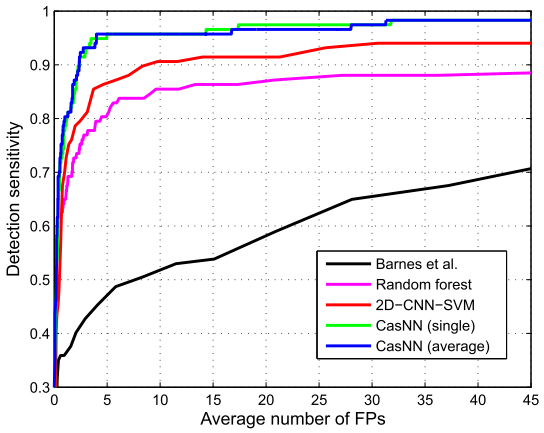

FIGURE 6.6

Comparison of FROC curves of different methods.

Table 6.7 shows the comparison results of different methods and the FROC curves are presented in Fig. 6.6. It is clearly observed that our methods outperform the other three comparison methods by a large margin with the highest detection recall and the lowest false positive rates. 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.

6.4.5 SYSTEM IMPLEMENTATION

The proposed framework was implemented based on Theano library [53] 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 weights of network were randomly initialized from the Gaussian distribution ($\mu = 0, \sigma = 0.01$) and trained with standard back-propagation.

6.5 DISCUSSION AND CONCLUSION

Detecting objects from medical images that provide clinical significance has been an indispensable component in computer-aided diagnosis. In order to save doctors from backbreaking labor and improve diagnosis reliability as well as efficiency, we pro-

pose an efficient and robust framework for automatic detection of sparsely distributed objects from high-resolution images/volumes. With the design of deep cascaded networks, we keep two aims in mind, namely, efficiency and accuracy. For an automatic object detection system targeting clinical practice, we believe that both of them are equally crucial. In our framework, the first stage efficiently screens the whole image and retrieves a number of potential candidates with a high recall. It can not only speed up the discrimination procedure but also assist the non-experienced trainees by timely promoting the candidates for a closer inspection. The second stage robustly discriminates the true detection targets with only few false positives generated, which can facilitate further quantification measurements. Extensive experimental results on both 2D/3D typical applications corroborated the efficacy and efficiency of our approach, outperforming other state-of-the-art methods by a significant margin.

In medical image computing tasks, especially for volumetric image processing, 3D CNNs hold promising potentials but have not been fully explored yet. Most previous approaches adapted 2D CNN variants for processing 3D volumetric data [54], with difficulties being reported when attempting to employ 3D CNNs. One main concern for employing 3D CNNs is their high computational cost in a sliding window way. 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.

The proposed method is inherently generic and can be easily adapted to different detection tasks in 2D and 3D medical data. Future investigations include assessing the proposed method on more detection tasks and accelerating the algorithm with GPU optimization.

ACKNOWLEDGEMENTS

This work is supported by Hong Kong RGC General Research Fund (No. CUHK412513) and Shenzhen–Hong Kong Innovation Circle Funding Program (No. SGLH20131010151755080 and GHP/002/13SZ). The authors gratefully thank Dr. Ludovic Roux for providing mitosis datasets and helping with the evaluation.

REFERENCES

1. Kunio Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Comput. Med. Imaging Graph. 31 (4) (2007) 198–211.
2. Mitko Veta, Paul J. van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B.L. Larsen, Jacob S. Vestergaard, Anders B. Dahl, et al., Assessment of algorithms for mitosis detection in breast cancer histopathology images, Med. Image Anal. 20 (1) (2015) 237–248.

3. Humayun Irshad, Antoine Veillard, Ludovic Roux, Daniel Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: a review – current status and future potential, IEEE Rev. Biomed. Eng. 7 (2014) 97–114.
4. Qi Dou, Hao Chen, Lequan Yu, Lei Zhao, Jing Qin, Defeng Wang, Vincent C.T. Mok, Lin Shi, Pheng-Ann Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1182–1195.
5. Bram van Ginneken, Samuel G. Armato, Bartjan de Hoop, Saskia van Amelsvoort-van de Vorst, Thomas Duindam, Meindert Niemeijer, Keelin Murphy, Arnold Schilham, Alessandra Retico, Maria Evelina Fantacci, et al., Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study, Med. Image Anal. 14 (6) (2010) 707–722.
6. Tao Chan, Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain, Comput. Med. Imaging Graph. 31 (4) (2007) 285–298.
7. Angel Cruz-Roa, Gloria Díaz, Eduardo Romero, Fabio A. González, Automatic annotation of histopathological images using a latent topic model based on non-negative matrix factorization, J. Pathol. Inform. 2 (2011).
8. Andreas Charidimou, David J. Werring, Cerebral microbleeds: detection, mechanisms and clinical challenges, Future Neurol. 6 (5) (2011) 587–611.
9. M.W. Vernooij, Aad van der Lugt, Mohammad Arfan Ikram, P.A. Wielopolski, W.J. Niessen, Albert Hofman, G.P. Krestin, M.M.B. Breteler, Prevalence and risk factors of cerebral microbleeds: the Rotterdam scan study, Neurology 70 (14) (2008) 1208–1214.
10. Steven M. Greenberg, Meike W. Vernooij, Charlotte Cordonnier, Anand Viswanathan, Rustam Al-Shahi Salman, Steven Warach, Lenore J. Launer, Mark A. Van Buchem, Monique Breteler, Cerebral microbleeds: a guide to detection and interpretation, Lancet Neurol. 8 (2) (2009) 165–174.
11. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
12. Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
13. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
14. Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556, 2014.
15. Hao Chen, Qi Dou, Xi Wang, Jing Qin, Pheng Ann Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016.
16. Humayun Irshad, Automated mitosis detection in histopathology using morphological and multi-channel statistics features, J. Pathol. Inform. 4 (2013).
17. Christopher D. Malon, Eric Cosatto, Classification of mitotic figures with convolutional neural networks and seeded blob features, J. Pathol. Inform. 4 (2013).
18. Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, Anant Madabhushi, Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2014, 90410B.

19. F. Boray Tek, Mitosis detection using generic features and an ensemble of cascade AdaBoosts, J. Pathol. Inform. 4 (2013).
20. Alessandro Giusti, Dan C. Cireşan, Jonathan Masci, Luca M. Gambardella, Jürgen Schmidhuber, Fast image scanning with deep max-pooling convolutional neural networks, arXiv:1302.1700, 2013.
21. Dan C. Cireşan, Alessandro Giusti, Luca M. Gambardella, Jürgen Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Springer, 2013, pp. 411–418.
22. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, Caffe: convolutional architecture for fast feature embedding, arXiv:1408.5093, 2014.
23. Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv:1207.0580, 2012.
24. Stuart Geman, Elie Bienenstock, René Doursat, Neural networks and the bias/variance dilemma, Neural Comput. 4 (1) (1992) 1–58.
25. Peter Boyle, Bernard Levin, et al., World Cancer Report 2008, IARC Press, International Agency for Research on Cancer, 2008.
26. C.W. Elston, I.O. Ellis, Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up, Histopathology 19 (5) (1991) 403–410.
27. Hao Chen, Chiyao Shen, Jing Qin, Dong Ni, Lin Shi, Jack C.Y. Cheng, Pheng-Ann Heng, Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, 2015, pp. 515–522.
28. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. (2014) 1–42.
29. Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.
30. Hao Chen, Dong Ni, Jing Qin, Shengli Li, Xin Yang, Tianfu Wang, Pheng Ann Heng, Standard plane localization in fetal ultrasound via domain transferred deep neural networks, IEEE J. Biomed. Health Inform. 19 (5) (2015) 1627–1636.
31. Hao Chen, Lequan Yu, Qi Dou, Lin Shi, Vincent C.T. Mok, Pheng Ann Heng, Automatic detection of cerebral microbleeds via deep learning based 3D feature representation, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 764–767.
32. Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, Gilles Le Naour, Metin N. Gurcan, Mitosis detection in breast cancer histological images: an ICPR 2012 contest, J. Pathol. Inform. 4 (2013).
33. Andreas Charidimou, Anant Krishnan, David J. Werring, H. Rolf Jäger, Cerebral microbleeds: a guide to detection and clinical relevance in different disease settings, Neuroradiology 55 (6) (2013) 655–674.
34. Andreas Charidimou, David J. Werring, Cerebral microbleeds and cognition in cerebrovascular disease: an update, J. Neurol. Sci. 322 (1) (2012) 50–55.

35. Amir Fazlollahi, Fabrice Meriaudeau, Victor L. Villemagne, Christopher C. Rowe, Paul Yates, Olivier Salvado, Pierrick Bourgeat, Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the radon transform, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, 2014, pp. 113–116.
36. Hugo J. Kuijf, Jeroen de Bresser, Mirjam I. Geerlings, Mandy Conijn, Max A. Viergever, Geert Jan Biessels, Koen L. Vincken, Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform, NeuroImage 59 (3) (2012) 2266–2273.
37. Wei Bian, Christopher P. Hess, Susan M. Chang, Sarah J. Nelson, Janine M. Lupo, Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images, NeuroImage Clin. 2 (2013) 282–290.
38. Babak Ghafaryasl, Fedde van der Lijn, Marielle Poels, Henri Vrooman, M. Arfan Ikram, Wiro J. Niessen, Aad van der Lugt, Meike Vernooij, Marleen de Bruijne, A computer aided detection system for cerebral microbleeds in brain MRI, in: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, 2012, pp. 138–141.
39. Qi Dou, Hao Chen, Lequan Yu, Lin Shi, Defeng Wang, Vincent C.T. Mok, Pheng Ann Heng, Automatic cerebral microbleeds detection from MR images via independent subspace analysis based hierarchical features, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 7933–7936.
40. Hao Chen, Qi Dou, Dong Ni, Jie-Zhi Cheng, Jing Qin, Shengli Li, Pheng-Ann Heng, Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, 2015, pp. 507–514.
41. Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, Jianming Liang, Convolutional neural networks for medical image analysis: full training or fine tuning?, IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312.
42. Hao Chen, Xiao Juan Qi, Jie Zhi Cheng, Pheng Ann Heng, Deep contextual networks for neuronal structure segmentation, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016.
43. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle, Brain tumor segmentation with deep neural networks, arXiv:1505.03540, 2015.
44. Holger R. Roth, Christopher T. Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, Ronald M. Summers, Anatomy-specific classification of medical images using deep convolutional nets, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 101–104.
45. Holger R. Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B. Turkbey, Ronald M. Summers, DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, 2015, pp. 556–564.
46. Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, Mads Nielsen, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Springer, 2013, pp. 246–253.
47. Holger R. Roth, Le Lu, Ari Seff, Kevin M. Cherry, Joanne Hoffman, Shijun Wang, Jiamin Liu, Evrim Turkbey, Ronald M. Summers, A new 2.5 D representation for lymph node de-

tection using random sets of deep convolutional neural network observations, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, Springer, 2014, pp. 520–527.
48. Xavier Glorot, Antoine Bordes, Yoshua Bengio, Deep sparse rectifier neural networks, in: International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.
49. Samuel R.S. Barnes, E. Mark Haacke, Muhammad Ayaz, Alexander S. Boikov, Wolff Kirsch, Dan Kido, Semiautomated detection of cerebral microbleeds in magnetic resonance images, Magn. Reson. Imaging 29 (6) (2011) 844–852.
50. S.M. Gregoire, U.J. Chaudhary, M.M. Brown, T.A. Yousry, C. Kallis, H.R. Jäger, D.J. Werring, The microbleed anatomical rating scale (MARS) reliability of a tool to map brain microbleeds, Neurology 73 (21) (2009) 1759–1766.
51. Andy Liaw, Matthew Wiener, Classification and regression by randomForest, R News 2 (3) (2002) 18–22, http://CRAN.R-project.org/doc/Rnews/.
52. Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1–27:27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
53. Theano Development Team, Theano: a Python framework for fast computation of mathematical expressions, arXiv:1605.02688, 2016.
54. Hayit Greenspan, Bram van Ginneken, Ronald M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, IEEE Trans. Med. Imaging 35 (5) (2016) 1153–1159.

NOTES

1. http://ipal.cnrs.fr/ICPR2012/.
2. http://mitos-atypia-14.grand-challenge.org/.