Chapter 2
Review of Deep Learning Methods
in Mammography, Cardiovascular,
and Microscopy Image Analysis

Gustavo Carneiro, Yefeng Zheng, Fuyong Xing and Lin Yang

Abstract Computerized algorithms and solutions in processing and diagnosis mammography X-ray, cardiovascular CT/MRI scans, and microscopy image play an important role in disease detection and computer-aided decision-making. Machine learning techniques have powered many aspects in medical investigations and clinical practice. Recently, deep learning is emerging a leading machine learning tool in computer vision and begins attracting considerable attentions in medical imaging. In this chapter, we provide a snapshot of this fast growing field specifically for mammography, cardiovascular, and microscopy image analysis. We briefly explain the popular deep neural networks and summarize current deep learning achievements in various tasks such as detection, segmentation, and classification in these heterogeneous imaging modalities. In addition, we discuss the challenges and the potential future trends for ongoing work.

2.1 Introduction on Deep Learning Methods in Mammography

Breast cancer is one of the most common types of cancer affecting the lives of women worldwide. Recent statistical data published by the World Health Organisation (WHO) estimates that 23% of cancer-related cases and 14% of cancer-related
Deaths among women are due to breast cancer [1]. The most effective tool to reduce the burden associated with breast cancer consists of early detection in asymptomatic women via breast cancer screening programs [2], which commonly use mammography for breast imaging. Breast screening using mammography comprises several steps, which include the detection and analysis of lesions, such as masses and calcifications, that are used in order to estimate the risk that the patient is developing breast cancer. In clinical settings, this analysis is for the most part a manual process, which is susceptible to the subjective assessment of a radiologist, resulting in a potentially large variability in the final estimation. The effectiveness of this manual process can be assessed by recent studies that show that this manual analysis has a sensitivity of 84% and a specificity of 91% [3]. Other studies show evidence that a second reading of the same mammogram either from radiologists or from computer-aided diagnosis (CAD) systems can improve this performance [3]. Therefore, given the potential impact that second reading CAD systems can have in breast screening programs, there is a great deal of interest in the development of such systems.

2.2 Deep Learning Methods in Mammography

A CAD system that can analyze breast lesions from mammograms usually comprises three steps [3]: (1) lesion detection, (2) lesion segmentation, and (3) lesion classification. The main challenges involved in these steps are related to the low signal-to-noise ratio present in the imaging of the lesion, and the lack of a consistent location, shape, and appearance of lesions [4, 5]. Current methodologies for lesion detection involve the identification of a large number of candidate regions, usually based on the use of traditional filters, such as morphological operators or difference of Gaussians [6–13]. These candidates are then processed by a second stage that aims at removing false positives using machine learning approaches (e.g., region classifier) [6–13]. The main challenges faced by lesion detection methods are that they may generate a large number of false positives, while missing a good proportion of true positives [4]; in addition, another issue is the poor alignment of the detected lesion in terms of translation and scale within the candidate regions—this issue has negative consequences for the subsequent lesion segmentation that depends on a relatively precise alignment. Lesion segmentation is then addressed with global/local energy minimisation models on a continuous or discrete space [14–16]. The major roadblock faced by these methods is the limited availability of annotated datasets that can be used in the training of the segmentation models. This is a particularly important problem because, differently from the detection and classification of lesions, the segmentation of lesions is not a common task performed by radiologists, which imposes strong limitations in the annotation process and, as a consequence, in the availability of annotated datasets. In fact, the main reason behind the need for a lesion segmentation is the assumption that the lesion shape is an important feature in the final stage of the analysis: lesion classification. This final stage usually involves the extraction of manually or automatically designed features from the lesion image.
and shape and the use of those features with traditional machine learning classifiers [17–19]. In this last stage, the main limitation is with respect to the features being extracted for the classification because these features are usually hand-crafted, which cannot guarantee optimality for this classification stage.

The successful use and development of deep learning methods in computer vision problems (i.e., classification and segmentation) [20–24] have motivated the medical image analysis community to investigate the applicability of such methods in medical imaging segmentation and classification problems. Compared to the more traditional methods presented above (for the problem of mammogram analysis), deep learning methods offer the following clear advantages: automated learning of features estimated based on specific detection/segmentation/classification objective functions; opportunity to build complete “end-to-end” systems that take an image, detect, segment, and classify visual objects (e.g., breast lesion) using a single model and a unified training process. However, the main challenge faced by deep learning methods is the need for large annotated training sets given the scale of the parameter space, usually in the order of $10^6$ parameters. This problem is particularly important in medical image analysis applications, where annotated training sets rarely have more than a few thousand samples. Therefore, a great deal of research is focused on the adaptation of deep learning methods to medical image analysis applications that contain relatively small annotated training sets.

There has been an increasing interest in the development of mammogram analysis methodologies based on deep learning. For instance, the problem of breast mass segmentation has been addressed with the use of a structured output model, where several potential functions are based on deep learning models [25–27]. The assumption here is that deep learning models alone cannot produce results that are accurate enough due to the small training set size problem mentioned above, but if these models are combined with a structured output model that makes assumptions about the appearance and shape of masses, then it is possible to have a breast mass segmentation that produces accurate results—in fact this method holds the best results in the field in two publicly available datasets [19, 28]. Segmentation of breast tissue using deep learning alone has been successfully implemented [29], but it is possible that a similar structured output model could improve even more the accuracy obtained. Dhungel et al. [30] also worked on a breast mass detection methodology that consists of a cascade of classifiers based on the Region Convolutional Neural Network (R-CNN) [23] approach. The interesting part is that the candidate regions produced by the R-CNN contain too many false positives, so the authors had to include an additional stage based on a classifier to eliminate those false positives. Alternatively, Ertosun and Rubin [31] propose a deep learning-based mass detection method consisting of a cascade of deep learning models trained with DDSM [28]—the main reason that explains the successful use of deep learning models here is the size of DDSM, which contains thousands of annotated mammograms.

The classification of lesions using deep learning [32–34] has also been successfully implemented in its simplest form: as a simple lesion classifier. Carneiro et al. [35] have proposed a system that can classify the unregistered two views of a mammography exam (cranial–caudal and mediolateral–oblique) and their respective
segmented lesions and produce a classification of the whole exam. The importance of this work lies in its ability to process multi-modal inputs (images and segmentation maps) that are not registered, in its way of performing transfer learning from computer vision datasets to medical image analysis datasets, and also in its capability of producing high-level classification directly from mammograms. A similar high-level classification using deep learning estimates the risk of developing breast cancer by scoring breast density and texture [36, 37]. Another type of high-level classification is the method proposed by Qiu et al. [38] that assesses the short-term risk of developing breast cancer from a normal mammogram.

2.3 Summary on Deep Learning Methods in Mammography

Based on the recent results presented above, it is clear that the use of deep learning is allowing accuracy improvements in terms of mass detection, segmentation, and classification. All the studies above have been able to mitigate the training set size issue with the use of regularization techniques or the combination of different approaches that can compensate the relatively poor generalization of deep learning methods trained with small annotated training sets. More importantly, deep learning is also allowing the implementation of new applications that are more focused on high-level classifications that do not depend on lesion segmentation. The annotation for this higher level tasks is readily available from clinical datasets, which generally contain millions of cases that can be used to train deep learning models in a more robust manner. These new applications are introducing a paradigm shift in how the field analyzes mammograms: from the classical three-stage process (detection, segmentation, and classification of lesions) trained with small annotated datasets to a one-stage process consisting of lesion detection and classification trained with large annotated datasets.

2.4 Introduction on Deep Learning for Cardiological Image Analysis

Cardiovascular disease is the number one cause of death in the developed countries and it claims more lives each year than the next seven leading causes of death combined [39]. The costs for addressing cardiovascular disease in the USA will triple by 2030, from 273 billion to 818 billion (in 2008 dollars) [40]. With the capability of generating images of a patient’s inside body non-invasively, medical imaging is ubiquitously present in the current clinical practice. Various imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and nuclear imaging, are widely available in clinical practice to generate images
of the heart, and different imaging modalities meet different clinical requirements. For example, ultrasound is most widely used for cardiac function analysis (i.e., the pumping of a cardiac chamber) due to its low cost and free of radiation dose; nuclear imaging and MRI are used for myocardial perfusion imaging to measure viability of the myocardium; CT reveals the most detailed cardiac anatomical structures and is routinely used for coronary artery imaging; while fluoroscopy/angiography is the workhorse imaging modality for cardiac interventions.

Physicians review these images to determine the health of the heart and to diagnose disease. Due to the large amount of information captured by the images, it is time consuming for physicians to identify the target anatomy and to perform measurements and quantification. For example, many 3D measurements (such as the volume of a heart chamber, the heart ejection fraction, the thickness and the thickening of the myocardium, or the strain and torsion of the myocardium) are very tedious to calculate without help from an intelligent post-processing software system. Various automatic or semi-automatic cardiac image analysis systems have been developed and demonstrated to reduce the exam time (thereby increase the patient throughput), increase consistency and reproducibility of the exam, and boost diagnosis accuracy of physicians.

Cardiovascular structures are composed of the heart (e.g., cardiac chambers and valves) and vessels (e.g., arteries and veins). A typical cardiac image analysis pipeline is composed of the following tasks: detection, segmentation, motion tracking, quantification, and disease diagnosis. For an anatomical structure, detection means determining the center, orientation, and size of the anatomy; while, for a vessel, it often means extraction of the centerline since a vessel has a tubular shape [41]. Early work on cardiac image analysis usually used non-learning-based data-driven approaches, for example, from simple thresholding and region growing to more advanced methods (like active contours, level sets, graph cuts, and random walker) for image segmentation. In the past decade, machine learning has penetrated into almost all steps of the cardiac image analysis pipeline [42, 43]. The success of a machine learning-based approach is often determined by the effectiveness and efficiency of the image features.

The recent advance of deep learning demonstrates that a deep neural network can automatically learn hierarchical image representations, which often outperform the most effective hand-crafted features developed after years of feature engineering. Encouraged by the great success of deep learning on computer vision, researchers in the medical imaging community quickly started to adapt deep learning for their own tasks. The current applications of deep learning on cardiac image segmentation are mainly focused on two topics: left/right ventricle segmentation [44–52] and retinal vessel segmentation [53–60]. Most of them are working on 2D images as input; while 3D deep learning is still a challenging task. First, evaluating a deep network on a large volume may be too computationally expensive for a real clinical application. Second, a network with a 3D patch as input requires more training data since a 3D patch generates a much bigger input vector than a 2D patch. However, the medical imaging community is often struggling with limited training samples (often in hundreds or
thousands) due to the difficulty to generate and share patients’ images. Nevertheless, we started to see a few promising attempts \cite{61–63} to attack the challenging 3D deep learning tasks.

2.5 Deep Learning-Based Methods for Heart Segmentation

Carneiro et al. \cite{44} presented a method using a deep belief network (DBN) to detect an oriented bounding box of the left ventricle (LV) on 2D ultrasound images of the LV long-axis views. One advantage of the DBN is that it can be pre-trained layer by layer using unlabeled data; therefore, good generalization capability can be achieved with a small number of labeled training images. A 2D-oriented bounding box has five pose parameters (two for translation, one for rotation, and two for anisotropic scaling). Since an exhaustive searching in this five-dimensional pose parameter space is time consuming, they proposed an efficient search strategy based on the first- or second-order derivatives of the detection score, which accelerated the detection speed by ten times. Furthermore, the DBN has also been applied to train a boundary detector for segmentation refinement using an active shape model (ASM). The LV detection/segmentation module can also be integrated in a particle filtering framework to track the motion of the LV \cite{44}. This work was later extended to segment the right ventricle (RV) too \cite{46}. In follow-up work \cite{47}, the DBN was applied to segment the LV on short-axis cardiac MR images. Similarly, the LV bounding box is detected with a DBN. Furthermore, another DBN was trained to generate a pixel-wise probability map of the LV. Instead of using the ASM as \cite{44}, the level set method is applied on the probability map to generate the final segmentation.

Avendi et al. \cite{50} proposed a convolutional network (CNN)-based method to detect an LV bounding box on a short-axis cardiac MR image. Stacked autoencoder was then applied to generate an initial segmentation of the LV, which was used to initialize the level set function. Their level set function combines a length-based energy term, a region-based term, and the prior shape. Instead of running level set on the probability map as \cite{44}, it was applied on the initial image.

Different to \cite{44, 50}, Chen et al. proposed to use a fully convolutional network (FCN) to segment the LV on 2D long-axis ultrasound images \cite{52}. In \cite{44, 50}, deep learning was applied in one or two steps of the whole image analysis pipeline. Differently, the FCN can be trained end-to-end without any preprocessing or post-processing. It can generate a segmentation label for each pixel efficiently since the convolution operation is applied once on the whole image. Due to the limited training samples, a deep network often suffers from the over-fitting issue. There are multiple canonical LV long-axis views, namely apical two-chamber (A2C), three-chamber (A3C), four-chamber (A4C), and five-chamber (A5C) views. Instead of training an LV segmentation network for each task, the problem was formulated as a multi-task learning, where all tasks shared the low-level image representations. At the high level, each task had its own classification layers. The segmentation was refined iteratively by focusing on the LV region detected by the previous iteration. Experiments showed that
the iterative cross-domain deep learning approach outperformed alternative single-domain deep learning, especially for tasks with limited training samples.

Zhen et al. [49] presented an interesting method for direct estimation of a ventricular volume from images without performing segmentation at all. They proposed a new convolutional deep belief network. A DBN is composed of stacked restricted Boltzman machine (RBM), where each layer is fully connected to the previous layer. Due to the full connectivity, the network has more parameters than a CNN; therefore it is more prone to over-fit. In [49], the first RBM layer was replaced with a multi-scale convolutional layer. The convolutional DBN was trained without supervision on unlabeled data and the trained network was used as an image feature extractor. A random forest regressor was then trained on the DBN image features to directly output an estimate of the LV area on each MR slice. Summing LV areas from all images results in the final volume estimate.

Due to the difficulty of 3D deep learning, all the above-reviewed methods work on 2D images, even though the input may be 3D. A 3D volume contains much richer information than a 2D image. Therefore, an algorithm leveraging 3D image information may be more robust. For heart segmentation, we only found one example using 3D deep learning, namely marginal space deep learning (MSDL) [62]. MSDL is an extension of marginal space learning (MSL), which uses hand-crafted features (i.e., Haar-like features and steerable features) and a boosting classifier. Here, the hand-crafted features are replaced with automatically learned sparse features and a deep network is exploited as the classifier. In [62], Ghesu et al. demonstrated the efficiency and robustness of MSDL on aortic valve detection and segmentation in 3D ultrasound volumes. Without using GPU, the aortic valve can be successfully segmented in less than one second with higher accuracy than the original MSL. MSDL is a generic approach and it can be easily re-trained to detect/segment other anatomies in a 3D volume.

2.6 Deep Learning-Based Methods for Vessel Segmentation

Early work on vessel segmentation used various hand-crafted vesselness measurements to distinguish the tubular structure from background [64]. Recently, we saw more and more work to automatically learn the most effective application-specific vesselness measurement from an expert-annotated dataset [65, 66]. Deep learning has potential to replace those classifiers to achieve better segmentation accuracy. However, the current applications of deep learning on vessel segmentation are mainly focused on retinal vessels in fundus images [53–60]. We only found limited work on other vessels, e.g., the coronary artery [62, 63] and carotid artery [61]. We suspect that the main reason is that a fundus image is 2D; therefore, it is much easier to apply an off-the-shelf deep learning package on this application. Other vessels in a 3D volume (e.g., CT or MR) are tortuous and we have to take the 3D context for a reliable segmentation. With the recent development of 3D deep learning, we expect to see more applications of deep learning on other vessels too.
In most work, pixel-wise classification is performed by a trained deep network to
directly output the segmentation mask. For example, Wang et al. [53] applied a CNN
to retinal vessel segmentation. To further improve the accuracy, they also used the
CNN as a trainable feature extractor: activations of the network at different layers
are taken as features to train random forests (RF). State-of-the-art performance has
been achieved by an ensemble of RF classifiers on the public DRIVE and STARE
datasets. Li et al. [54] presented another method based on an FCN with three lay-
ers. They formulated the task as cross-modality data transformation from the input
image to vessel map. The first hidden layer was pre-trained using denoising autoen-
coder, while the other two hidden layers were randomly initialized. Different to [53]
(which generates a label of the central pixel of an input patch), Li et al. approach
outputs labels for all pixels in the patch. Since overlapping patches are extracted
during classification, a pixel appears on multiple patches. The final label of the pixel
is determined by majority voting to improve the classification accuracy. Fu et al. [60]
adapted a holistically nested edge detection (HED) method for retinal vessel seg-
mentation. HED is motivated by the FCN and deeply supervised network, where the
outputs of intermediate layers are also directly connected to the final classification
layer. After getting the the vessel probability map using HED, a conditional random
field is applied to further improve the segmentation accuracy.

Since pixel-wise classification is time consuming, Wu et al. [58] proposed to
combine pixel classification and vessel tracking to accelerate the segmentation speed.
Starting from a seed point, a vessel is traced in the generalized particle filtering frame-
work (which is a popular vessel tracing approach), while the weight of each particle is
set by the CNN classification score at the corresponding position. Since CNN classi-
fication is invoked only on a suspected vessel region during tracing, the segmentation
speed was accelerated by a factor of two. Besides retinal vessel segmentation, deep
learning has also been exploited to detect retinal vessel microaneurysms [56] and
diabetic retinopathy [57] from a fundus image.

Coronary artery analysis is the killer application of cardiac CT. Due to the tiny
size of a coronary artery, CT is currently the most widely used noninvasive imaging
modality for coronary artery disease diagnosis due to its superior image resolution
(around 0.2–0.3 mm for a state-of-the-art CT scanner). Even with a quite amount of
published work on coronary artery segmentation in the literature [64], we only found
one work using deep learning [62] for coronary artery centerline extraction. Coronary
centerline extraction is still challenging task. To achieve a high detection sensitivity,
false positives are unavoidable. The false positives mainly happen on coronary veins
or other tubular structures; therefore, traditional methods cannot reliably distinguish
false positives from true coronary arteries. In [41], a CNN is exploited to train a classi-
ifier which can distinguish leakages from good centerlines. Since the initial centerline
is given, the image information can be serialized as a 1D signal along the centerline.
Here, the input channels consist of various profiles sampled along the vessel such
as vessel scale, image intensity, centerline curvature, tubularity, intensity, and grad-
ient statistics (mean, standard deviation) along and inside a cross-sectional circular
boundary, and distance to the most proximal point in the branch. Deep learning-based
branch pruning increases the specificity from 50 to 90\% with negligible degradation of sensitivity.

Similar to heart segmentation reviewed in Sect. 2.5, almost all previous work on deep learning for vessel segmentation was focused on 2D. Recently, Zheng et al. [61] proposed an efficient 3D deep learning method for vascular landmark detection. A two-step approach is exploited for efficient detection. A shallow network (with one hidden layer) is used for the initial testing of all voxels to obtain a small number of promising candidates, followed by more accurate classification with a deep network. In addition, they proposed several techniques, i.e., separable filter decomposition and network sparsification, to speed up the evaluation of a network. To mitigate the over-fitting issue, thereby increasing detection robustness, small 3D patches from a multi-resolution image pyramid are extracted as network input. The deeply learned image features are further combined with Haar-like features to increase the detection accuracy. The proposed method has been quantitatively evaluated for carotid artery bifurcation detection on a head–neck CT dataset. Compared to the state-of-the-art, the mean error is reduced by more than half, from 5.97 to 2.64 mm, with a detection speed of less than 1 s/volume without using GPU.

Wolterink et al. [63] presented an interesting method using a 2.5D or 3D CNN for coronary calcium scoring in CT angiography. Normally, a standard cardiac CT protocol includes a non-contrasted CT scan for coronary calcium scoring [67] and a contracted scan (called CT angiography) for coronary artery analysis. If calcium scoring can be performed on a contrasted scan, the dedicated non-contrasted scan can be removed from the protocol to save radiation dose to a patient. However, calcium scoring on CT angiography is more challenging due to the reduced intensity gap between contrasted coronary lumen and calcium. In this work voxel-wise classification is performed to identify calcified coronary plaques. For each voxel, three orthogonal 2D patches (the 2.5D approach) or a full 3D patch are used as input. A CNN is trained to distinguish coronary calcium from other tissues.

2.7 Introduction to Microscopy Image Analysis

Microscopy image analysis can provide support for improved characterization of various diseases such as breast cancer, lung cancer, brain tumor, etc. Therefore, it plays a critical role in computer-aided diagnosis in clinical practice and pathology research. Due to the large amount of image data, which continue to increase nowadays, it is inefficient or even impossible to manually evaluate the data. Computerized methods can significantly improve the efficiency and the objectiveness, thereby attracting a great deal of attention. In particular, machine learning techniques have been widely and successfully applied to medical imaging and biology research [68, 69]. Compared with non-learning or knowledge based methods that might not precisely translate knowledge into rules, machine learning acquires their own knowledge from data representations. However, conventional machine learning techniques usually do not
directly deal with raw data but heavily rely on the data representations, which require considerable domain expertise and sophisticated engineering [70].

Deep learning is one type of representation learning methods that directly process raw data (e.g., RGB images) and automatically learns the representations, which can be applied to detection, segmentation, or classification tasks. Compared with hand-crafted features, learned representations require less human intervention and provide much better performance [71]. Nowadays, deep learning techniques have made great advantages in artificial intelligence, and successfully applied to computer vision, natural language processing, image understanding, medical imaging, computational biology, etc. [70, 72]. By automatically discovering hidden data structures, it has beaten records in several computer vision tasks such as image classification [73] and speech recognition [74], and won multiple competitions in medical image analysis such as brain image segmentation [75] and mitosis detection [76]. Meanwhile, it has provided very promising performance in other medical applications [77, 78].

Recently, deep learning is emerging as a powerful tool and will continue to attract considerable interests in microscopy image analysis including nucleus detection, cell segmentation, extraction of regions of interest (ROIs), image classification, etc. A very popular deep architecture is convolutional neural networks (CNNs) [70, 79], which have obtained great success in various tasks in both computer vision [73, 80–82] and medical image analysis [83]. Given images and corresponding annotations (or labels), a CNN model is learned to generate hierarchical data representations, which can be used for robust target classification [84]. On the other hand, unsupervised learning can also be applied to neural networks for representation learning [85–87]. Autoencoder is an unsupervised neural network commonly used in microscopy image analysis, which has provided encouraging performance. One of significant benefits of unsupervised feature learning is that it does not require expensive human annotations, which are not easy to achieve in medical computing.

There exist a number of books and reviews explaining deep learning principles, historical survey, and applications in various research areas. Schmidhuber [88] presents a historical overview of deep artificial neural networks by summarizing relevant work and tracing back the origins of deep learning ideas. LeCun et al. [70] mainly review supervised learning in deep neural networks, especially CNNs and recurrent neural networks, and their successful applications in object detection, recognition, and nature language processing. The book [71] explains several established deep learning algorithms and provides speculative ideas for future research, the monograph [87] surveys general deep learning techniques and their applications (mainly) in speech processing and computer vision, and the paper [83] reviews several recent deep learning applications in medical image computing (very few in microscopy imaging). Due to the emergence of deep learning and its impacts in a wide range of disciplines, there exist many other documents introducing deep learning or relevant concepts [74, 89–92].

In this chapter, we focus on deep learning in microscopy image analysis, which covers various topics such as nucleus/cell/neuron detection, segmentation, and classification. Compared with other imaging modalities (e.g., magnetic resonance imaging, computed tomography, and ultrasound), microscopy images exhibit unique com-
Fig. 2.1 Sample images of breast cancer, muscle, and pancreatic neuroendocrine tumor using different tissues and stain preparations. Hematoxylin and eosin (H&E) staining is used for the first two, while immunohistochemical staining is for the last. These image exhibit significant challenges such as background clutter, touching nuclei, and weak nucleus boundaries, for automated nucleus/cell detection and segmentation.

Deep learning methods involve multi-level representation learning, which starts from raw data input and gradually moves to more abstract levels via nonlinear transformations. With enough training data and sufficiently deep architectures, neural networks can learn very complex functions and
discover intricate structures in the data [70]. One significant advantage is that deep learning does not require much engineering work, which is not easy to achieve in some specific domains. Deep learning has been successfully applied to pattern recognition and prediction, and outperforms traditional machine learning methods in many domains including medical image computing [83]. More specifically, deep learning exhibits its great power in microscopy image analysis. To our knowledge, up to now there are mainly four commonly used deep networks in microscopy image analysis: CNNs, fully convolutional networks (FCNs), recurrent neural networks (RNNs), and stacked autoencoders (SAEs). More details related to optimization and algorithms can be found in [71, 89].

2.9 Microscopy Image Analysis Applications

In microscopy image analysis, deep neural networks are often used as classifiers or feature extractors to resolve various tasks in microscopy image analysis, such as target detection, segmentation, and classification. For the usage of a classifier, a deep neural network assigns a hard or soft label to each pixel of the input image in pixel-wise classification or a single label to the entire input image in image-level classification. CNNs are the most popular networks in this type of applications and their last layers are usually chosen as a multi-way softmax function corresponding to the number of target classes. For the usage of a feature extractor, a network generates a transformed representation of each input image, which can be applied to subsequent data analysis, such as feature selection or target classification. In supervised learning, usually the representation before the last layer of a CNN is extracted, but those from middle layers or even lower layers are also helpful to object recognition [111, 112]. To deal with limited data in medical imaging, it might be necessary to apply pretrain and fine-tune to the neural network. Tables 2.1 and 2.2 summarize the current deep learning achievements in microscopy image analysis.

2.10 Discussions and Conclusion on Deep Learning for Microscopy Image Analysis

Deep learning is a rapidly growing field and is emerging as a leading machine learning tool in computer vision and image analysis. It has exhibited great power in medical image computing with producing improved accuracy of detection, segmentation, or recognition tasks [83]. Most of works presented in this paper use CNNs or one type of the variants, FCNs, to solve problems in microscopy image analysis. Our conjecture is that CNNs provide consistent improved performance across a large variety of computer vision tasks and thus it might be straightforward to apply convolutional networks to microscopy image computing. More recently, FCNs have attracted a great
<table>
<thead>
<tr>
<th>Network</th>
<th>Usage</th>
<th>Topic</th>
<th>Data</th>
<th>Evaluation metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>[97]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Mitosis detection</td>
<td>Breast cancer images</td>
</tr>
<tr>
<td>[98]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Nucleus detection</td>
<td>Brain tumor, NET, breast cancer images</td>
</tr>
<tr>
<td>[99]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Cell detection</td>
<td>Breast cancer images</td>
</tr>
<tr>
<td>[100]</td>
<td>CNN</td>
<td>Pixel classification, feature extraction</td>
<td>Neutrophils identification</td>
<td>Human squamous cell carcinoma images</td>
</tr>
<tr>
<td>[102]</td>
<td>CNN</td>
<td>Patch classification</td>
<td>Mitosis detection</td>
<td>NIH3T3 scratch assay culture images</td>
</tr>
<tr>
<td>[103]</td>
<td>CNN</td>
<td>Patch scoring</td>
<td>Cell detection</td>
<td>NET, lung cancer images</td>
</tr>
<tr>
<td>[104]</td>
<td>CNN</td>
<td>Regression</td>
<td>Cell, nucleus detection</td>
<td>Breast cancer, NET, HeLa images</td>
</tr>
<tr>
<td>[105]</td>
<td>CNN</td>
<td>Regression</td>
<td>Nucleus detection, classification</td>
<td>Colon cancer images</td>
</tr>
<tr>
<td>[106]</td>
<td>FCN</td>
<td>Regression</td>
<td>Cell counting</td>
<td>Retinal pigment epithelial and precursor T Cell lymphoblastic lymphoma images</td>
</tr>
<tr>
<td>[107]</td>
<td>CNN</td>
<td>Voting</td>
<td>Nucleus detection</td>
<td>NET images</td>
</tr>
<tr>
<td>[108]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Mitosis detection</td>
<td>Breast cancer images</td>
</tr>
<tr>
<td>[109]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Hemorrhage detection</td>
<td>Color fundus images</td>
</tr>
<tr>
<td>[110]</td>
<td>SSAE</td>
<td>Feature extraction</td>
<td>Nucleus detection</td>
<td>Breast cancer images</td>
</tr>
</tbody>
</table>
Table 2.2 Summary of current deep learning achievements in microscopy image analysis. FCNN = fully connected neural network, DSC = dice similarity coefficient, PPV = positive predictive value, NPV = negative predictive value, IOU = intersection over union, MCA = mean class accuracy, ACA = average classification accuracy, and BAC = balanced accuracy

<table>
<thead>
<tr>
<th>Network</th>
<th>Usage</th>
<th>Topic</th>
<th>Data</th>
<th>Evaluation metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>[113]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Neuronal membrane segmentation</td>
<td>Ventral nerve cord images of a Drosophila larva</td>
</tr>
<tr>
<td>[114]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Neuronal membrane segmentation</td>
<td>Ventral nerve cord images of a Drosophila larva</td>
</tr>
<tr>
<td>[115]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Nucleus, cell segmentation</td>
<td>Developing C. elegans embryos images</td>
</tr>
<tr>
<td>[116]</td>
<td>CNN</td>
<td>Pixel classification</td>
<td>Nucleus, cytoplasm segmentation</td>
<td>Cervical images</td>
</tr>
<tr>
<td>[117]</td>
<td>FCN</td>
<td>Pixel classification</td>
<td>Neuronal membrane and cell segmentation</td>
<td>Ventral nerve cord images of a Drosophila larva, Glioblastoma-astrocytoma U373 cell and HeLa cell images</td>
</tr>
<tr>
<td>[118]</td>
<td>FCN</td>
<td>Pixel classification</td>
<td>Neuronal membrane segmentation</td>
<td>Ventral nerve cord images of a Drosophila larva</td>
</tr>
<tr>
<td>[119]</td>
<td>RNN</td>
<td>Pixel classification</td>
<td>Neuronal membrane segmentation</td>
<td>Ventral nerve cord images of a Drosophila larva</td>
</tr>
<tr>
<td>[120]</td>
<td>SDAE</td>
<td>Patch reconstruction</td>
<td>Nucleus segmentation</td>
<td>Brain tumor, lung cancer images</td>
</tr>
<tr>
<td>[121]</td>
<td>CNN</td>
<td>Image classification</td>
<td>Image classification</td>
<td>Human Epithelial-2 (HEp-2) cell images</td>
</tr>
<tr>
<td>[122]</td>
<td>FCNN</td>
<td>Cell classification</td>
<td>Cell classification</td>
<td>Optical phase and loss images</td>
</tr>
<tr>
<td>[123]</td>
<td>CNN</td>
<td>Feature extraction</td>
<td>Image classification</td>
<td>Glioblastoma multiforme and low grade glioma images</td>
</tr>
<tr>
<td>[124]</td>
<td>CNN</td>
<td>Feature extraction</td>
<td>Image classification</td>
<td>Colon cancer images</td>
</tr>
</tbody>
</table>

(continued)
Table 2.2 (continued)

<table>
<thead>
<tr>
<th>Network</th>
<th>Usage</th>
<th>Topic</th>
<th>Data</th>
<th>Evaluation metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>[125]</td>
<td>Autoencoder</td>
<td>Feature extraction</td>
<td>Image classification</td>
<td>Basal-cell carcinoma cancer images</td>
</tr>
<tr>
<td>[126]</td>
<td>SPSD</td>
<td>Feature extraction</td>
<td>Image classification</td>
<td>Glioblastoma multiforme and kidney clear cell carcinoma, tumorigenic breast cancer, and control cell line images</td>
</tr>
</tbody>
</table>
Another potential challenge of applying deep learning to microscopy image computing is to improve the network scalability, thereby adapting to high resolution images. In pathology imaging informatics, usually it is necessary to conduct quantitative analysis on whole-slide images (WSI) [134] instead of manually selected regions, since it can reduce biases of observers and provide complete information that is helpful to decision-making in diagnosis. The resolution of a WSI image is often over $50000 \times 50000$, and has tens of thousands or millions of object of interest (e.g., nuclei or cells). Currently, pixel-wise prediction with CNNs is mainly conducted in a sliding-window manner, and clearly this will be extremely computationally expensive when dealing with WSI images. FCNs are designed for efficient inference and might be a good choice for computation improvement.

This paper provides a survey of deep learning in microscopy image analysis, which is a fast evolving field. Specifically, it briefly introduces the popular deep neural networks in the domain, summarizes current research efforts, and explains the challenges as well as the potential future trends. Deep learning has benefitted the microscopy imaging domain and we expect that it will play a more important role in the future. New learning algorithms in artificial intelligence can accelerate the process of transferring deep learning techniques from natural toward medical images and enhance its achievements.

References

46. Ngo TA, Lu Z, Carneiro G (2016) Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med Image Anal 35:159–171
Deep Learning and Convolutional Neural Networks for Medical Image Computing Precision Medicine, High Performance and Large-Scale Datasets
Lu, L.; Zheng, Y.; Carneiro, G.; Yang, L. (Eds.)
2017, XIII, 326 p. 117 illus., 100 illus. in color., Hardcover
ISBN: 978-3-319-42998-4