Deep Learning in Microscopy Image Analysis: A Survey

Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang, *Member, IEEE*

***Abstract*— Computerized microscopy image analysis plays an important role in computer aided diagnosis and prognosis. Machine learning techniques have powered many aspects of medical investigation and clinical practice. Recently, deep learning is emerging as a leading machine learning tool in computer vision and has attracted considerable attention in biomedical image analysis. In this paper, we provide a snapshot of this fast-growing field, specifically for microscopy image analysis. We briefly introduce the popular deep neural networks and summarize current deep learning achievements in various tasks, such as detection, segmentation, and classification in microscopy image analysis. In particular, we explain the architectures and the principles of convolutional neural networks, fully convolutional networks, recurrent neural networks, stacked autoencoders, and deep belief networks, and interpret their formulations or modelings for specific tasks on various microscopy images. In addition, we discuss the open challenges and the potential trends of future research in microscopy image analysis using deep learning.**

***Index Terms*— Classification, deep learning, detection, microscopy image analysis, segmentation.**

I. Introduction

MICROSCOPY image analysis provides quantitative support for improving characterizations of various diseases, such as breast cancer, lung cancer, brain tumor, and so on. Therefore, it plays a critical role in computer-aided diagnosis and prognosis. Due to the large amount of image data, which continue to increase nowadays, it is inefficient or even impossible to manually process the image data. Computerized methods significantly improve the efficiency and objectiveness, thereby attracting considerable attention in recent literature. In particular, machine learning techniques have been widely and successfully applied to medical and biology research [1]–[4]. Compared with the nonlearning-based methods that might not precisely translate domain knowledge into rules, machine learning acquires its 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 [5].

Deep learning is a representation learning method that directly processes raw data (e.g., RGB images) and automatically learns the representations, which can be applied to object detection, image segmentation, or target classification. Compared with hand-crafted features, learned representations require less human interventions and provide much better performance [6]. Nowadays, deep learning techniques have made great advances in artificial intelligence, and they have been successfully applied to computer vision [5], [7], [8], natural language processing [9], speech recognition [5], [10], medical imaging [11], computational biology [12], and so on. By automatically discovering hidden data structures, it has achieved the best performance in several contests, such as image classification [13] and speech recognition [14], and won multiple competitions in biomedical image analysis, such as brain image segmentation [15] and mitosis detection [16]. Meanwhile, it has provided very promising performance in other biomedical applications as well [17].

Recently, deep learning is emerging as a powerful tool that attracts considerable interest in microscopy image analysis, including nuclei detection, cell segmentation, tissue segmentation, image classification, and so on. One popular deep architecture is convolutional neural networks (CNNs) [5], [18], which have obtained great success in various tasks of computer vision [13], [19]–[21] and biomedical image analysis [22]. 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 [23]. On the other hand, unsupervised learning can also be applied to neural networks for representation learning [9], [24], [25]. Autoencoder is an unsupervised neural network that has been commonly used in microscopy image analysis with promising accuracy. One significant benefit of unsupervised feature learning is that it does not require expensive human annotations.

There exist several books and reviews explaining deep learning principles, historical surveys, and applications in various research areas. Schmidhuber [26] has presented a historical

Manuscript received June 8, 2016; revised December 3, 2016 and October 6, 2017; accepted October 16, 2017. Date of publication November 22, 2017; date of current version September 17, 2018. *(Corresponding author: Fuyong Xing.)*

F. Xing is with the Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Denver, CO 80045 USA, and also with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: fuyong.xing@ucdenver.edu).

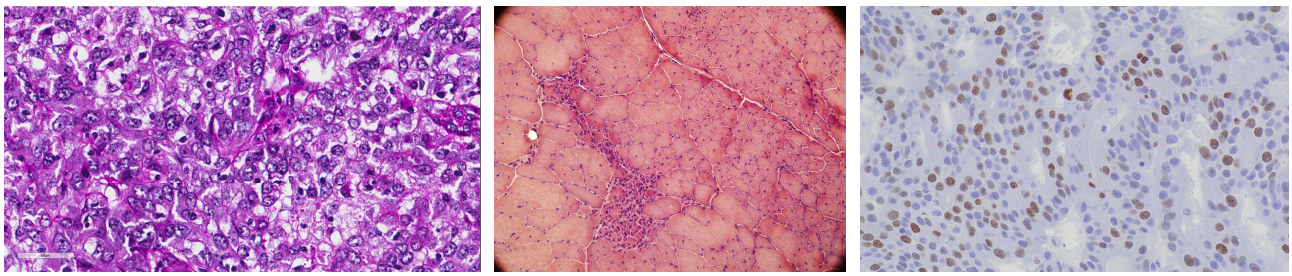
Y. Xie and H. Su are with the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611 USA.

F. Liu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA.

L. Yang is with the Department of Electrical and Computer Engineering, the Department of Computer and Information Science and Engineering, and the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611 USA.

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

Digital Object Identifier 10.1109/TNNLS.2017.2766168



(a) (b) (c)

Fig. 1. Sample images of (a) breast cancer, (b) muscle, and (c) pancreatic neuroendocrine tumor (NET). Haemotoxylin and eosin (H&E) staining is used for the first two, while immunohistochemical (IHC) staining is used for the last. These images exhibit significant challenges for automated nucleus/cell detection and segmentation, including but not limited to background clutter, touching nuclei, and weak nucleus/cell boundaries.

overview of deep artificial neural networks by summarizing relevant works and tracing back the origins of deep learning ideas. LeCun *et al.* [5] have mainly reviewed supervised learning in deep neural networks (DNNs), especially CNNs and recurrent neural networks (RNNs), and their successful applications in object detection, recognition, and natural language processing. The book [6] explains several established deep learning algorithms and provides speculative ideas for future research; the monograph [9] surveys general deep learning techniques and their applications (mainly) in speech processing and computer vision. Several recent deep learning applications in medical image computing are reviewed in [22]. The applications of DNNs in processing biomedical data, including omics, images, and signals, are reviewed in [12]. Due to the emergence of deep learning and its impacts in a wide range of disciplines, there exist many other reports introducing deep learning or relevant concepts [14], [27]–[31]. However, to the best of our knowledge, these surveys cover only a few, if any, publications related to microscopy image analysis.

In digital pathology and cell biology, a large amount of microscopic image data have been collected for image analysis assessment and the data acquisition rate is continuously and rapidly increasing. The high-dimensional microscopic data contain complex patterns and dependences in images such that they exhibit very complicated relationships with disease expressions or image labels. Additionally, there are extensive variations in the images due to different patients' specimen acquisition or data preparations. All of these situations significantly challenge many generic image analysis methods and conventional machine learning algorithms in the tasks of nucleus/cell detection, segmentation, classification, and so on. Deep learning, which learns abstract feature representations in a hierarchical way, can take advantage of large-scale and high-dimensional image data and discover hidden data structures for better microscopy image analysis. Meanwhile, deep learning can significantly reduce or eliminate the burden of feature engineering in conventional machine learning techniques. Nowadays, deep learning is the major method among the best solutions for many tasks in microscopy image analysis, and thus it holds great promise for the field.

In this paper, 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 such as magnetic resonance (MR) imaging, computed tomography, and ultrasound, microscopy imaging exhibits unique complex characteristics. In digital pathology, image data are usually generated with a certain staining and present significant challenges [11], [32]–[35], including background clutter, inhomogeneous intensity, touching or overlapping nuclei/cells, and so on. Fig. 1 shows several sample microscopy images of different tissue preparations. The motivation of this survey is to provide a comprehensive overview of deep learning in microscopy image analysis as well as a dedicated discussion on open challenges, unsolved problems, and potential future trends. Specifically, we introduce the deep learning's contributions to different microscopy image analysis tasks, with providing overview tables and figures for easy assessment. Then, we discuss the issues of applying deep learning to image understanding, including microscopy image computing, and present the current effort devoted to addressing these issues. Thereafter, we explicitly highlight the unique challenges and unsolve problems that deep learning needs to tackle in microscopy image analysis. In addition, we point out several potential future research trends of deep learning in microscopy image computing. Finally, we present an overview of deep learning techniques in the Supplementary Material for reference. This survey aims to help other researchers to catch a glimpse of the state-of-the-art methods in the field of microscopy image analysis and facilitate the applications of these technologies in their own research tasks.

II. Microscopy Image Analysis Applications

In microscopy image analysis, DNNs are often used as classifiers or feature extractors for various tasks, including object detection, segmentation, and classification. For the usage as a classifier, a DNN assigns a hard or soft label to each pixel of the input image for pixelwise classification, or a single label to the entire input image for image-level classification. CNNs are the most popular networks in these classification applications and their last layers are often chosen as a multiway softmax function corresponding to the number of target classes. For the usage as a feature extractor, a network generates a transformed representation of each input image, which can be applied to subsequent data analysis, such as

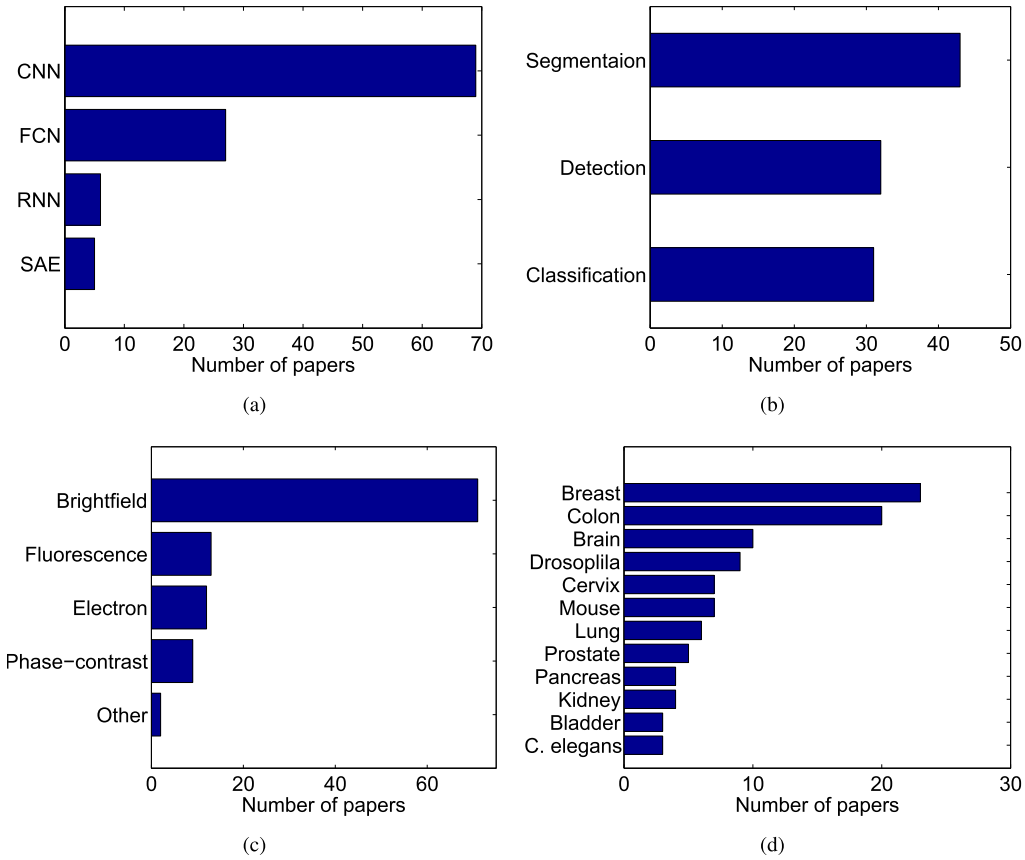


Fig. 2. Applications of deep learning in microscopy image analysis in terms of (a) network architectures, (b) tasks, (c) microscope types, and (d) image data (organs/specimens). For image data, the number of papers is listed in the Supplementary Material if it is not greater than 2. CNN, FCN, RNN and SAE represent convolutional neural network, fully convolutional network, recurrent neural network, and stacked autoencoder, respectively.

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 can also be helpful to object recognition [46], [47]. To deal with limited data in medical imaging, it is often necessary to apply pretraining and fine-tuning to the neural network. Deep belief networks (DBNs) have been used in natural image understanding [24] as well as medical image computing (e.g., MR image analysis) and biomedical signal processing [12]; however, to the best of our knowledge, there exist a very few publications reporting the applications of DBNs in microscopy image computing. The overview of deep learning techniques is provided in the Supplementary Material.

Fig. 2 shows the applications of deep learning in microscopy image analysis in terms of network architectures, tasks, microscope types, and image data (organs/specimens). Table I summarizes some key deep learning achievements of object detection in microscopy image analysis, and the tables for other tasks (i.e., segmentation and classification) are provided in the Supplementary Material. In each table, we list the types of networks, image data, tasks, experimental results, computational complexity, and pros/cons. Following [13], we represent the complexity with the number of weights and layers; we also calculate the computational workload [48], [49] in the testing phase as follows. Suppose there are C input feature maps with size $M \times N$ and D output feature maps with size $P \times Q$. For a convolutional layer with a $K \times K$ kernel, its computational workload is $\mathcal{O}(K^2PQCD)$ and it has K^2CD weights (not including the bias); if a stride of S and a padding of T are applied, then $P = (M - K + 2T)/S + 1$ and $Q = (N - K + 2T)/S + 1$. Although cross-channel pooling [50] is proposed to facilitate optimization, for the applications of deep learning in biomedical image computing, especially microscopy image analysis, all CNNs and others involving max (or average) pooling conduct the operation only over the spatial dimensions. This means the pooling operation is performed in each feature map (channel) separately, and thus the number of feature maps does not change after the pooling, thereby $C = D$. For a max-pooling layer with an $L \times L$ kernel, a stride of U, and a padding of V, the computational workload is $\mathcal{O}(L^2PQD)$ and no weights are learned, where $P = (M - L + 2V)/U + 1$ and $Q = (N - L + 2V)/U + 1$. Note that in microscopy image analysis, the number of feature maps does not change for pooling but not convolutional layers. A fully connected layer has a computational workload of $\mathcal{O}(MNPQCD)$ and the number of weights is $MNPQCD$. For upsampling, it can be implemented with standard bilinear interpolation. These types of layers are commonly used in CNNs, fully convolutional networks (FCNs), and stacked autoencoders (SAEs), and the overall computational workload of one network is the sum of those in each layer. Unlike a conventional neural

TABLE I

SUMMARY OF SOME KEY DEEP LEARNING ACHIEVEMENTS FOR OBJECT DETECTION IN MICROSCOPY IMAGE ANALYSIS. P = PRECISION, R = RECALL, F_1 = F_1-SCORE, AND AUC = AREA UNDER CURVE

Network	Data and Tasks	Experimental Results	Complexity ($1K = 1000$)	Pros and Cons
[36] CNN	breast cancer nucleus detection and area measurement	Bland-Altman plot	weights: 512K layers: 15	Pros: model area measurement with classification to avoid nucleus segmentation; use fully convolutional inference to improve computational efficiency. Cons: unable to handle large scale variations.
[37] CNN	lung cancer cell detection	P=0.83, R=0.84, F_1=0.83	weights: 709K layers: 8	Pros: accelerate forward networks with sparse kernels; diminish disk I/O time with asynchronous prefetching. Cons: do not consider spatial topology.
[38] CNN	breast cancer mitosis detection	P=0.88, R=0.70, F_1=0.78	weights: 14K layers: 13	Pros: effective hard negative mining; simple postprocessing. Cons: inefficient sliding-window prediction; fixed-size image inputs.
[39] CNN	nucleus detection in brain tumor, NET, and breast cancer images	brain: P=0.72, R=0.88, F_1=0.77 NET: P=0.84, R=0.93, F_1=0.88 breast: P=0.71, R=0.88, F_1=0.78	weights: 7040K layers: 8	Pros: train with patches to avoid extensive data annotations; easy to parallelize model training. Cons: sophisticated postprocessing to handle touching objects; sliding-window prediction.
[40] CNN	cell detection in NET, lung cancer images	NET: P=0.90, R=0.94, F_1=0.92 lung: P=0.88, R=0.92, F_1=0.90	weights: 22K layers: 7	Pros: predict only on candidates; generate training data with foveation. Cons: need good candidate generation; unable to handle scale variations.
[41] CNN	nucleus/cell detection in breast cancer, NET, cervix images	breast: P=0.92, R=0.91, F_1=0.91 NET: P=0.86, R=0.96, F_1=0.91 cervix: P=0.94, R=0.97, F_1=0.96	weights: 3464K layers: 8	Pros: regression modeling with considering topological information; fast inference with strided prediction. Cons: need proper selection of input image size.
[42] FCN	cell counting in retinal pigment epithelial and precursor T-Cell lymphoblastic lymphoma images	cell count difference: 2.9$\pm$0.2 (model A) 3.2$\pm$0.2 (model B)	weights: 1.3 million(A), 3.6 million (B) layers: 8(A), 8(B)	Pros: able to process arbitrary-size test images; efficient to train and inference. Cons: fixed-size network receptive fields; loss of high resolution information.
[43] CNN	NET nucleus detection	P=0.85, R=0.79, F_1=0.82	weights: 3523K layers: 9	Pros: fuse neighboring information for localization; well handle touching objects. Cons: fixed input size; need to define neighboring region.
[44] CNN	breast cancer mitosis detection	F_1=0.61, AUC=0.87	weights: 14K layers: 6	Pros: data augmentation via crowdsourcing; robust to noisy data. Cons: slow training process; sliding-window prediction.
[45] SAE	breast cancer nucleus detection	P=0.89, R=0.83, F_1=0.85	weights: 90K layers: 3	Pros: unsupervised feature learning. Cons: not end-to-end training; no mechanisms to handle touching objects.

network (e.g., feedforward neural networks), an RNN has the same parameters across all steps. In order to handle static images that do not have a sequence format, it can split one image into a set of nonoverlapping patches, which are then reorganized into an acyclic region sequence [51]–[53]. For a single time step, assuming the patch size is $H \times W \times C$,

where H, W, and C denote the height, width, and depth, respectively, and the dimensionality of the hidden state is D, the computational workload is $\mathcal{O}(HWCD)$ [54].

A. Detection

Detection of object of interests (e.g., nuclei and cells) on digitized specimens is very important in microscopy image analysis. In particular, nucleus or cell detection can provide significant support for object counting, segmentation, and tracking. Nowadays, CNNs, FCNs, and SAEs have all been successfully applied to object detection in microscopy images, and the locations of objects are often marked with detected single points near object centroids, which are referred to as seeds or markers. Intuitively, object detection can be formulated as a pixelwise classification problem. For a testing input image, the network outputs a probability map, where each pixel value indicates the probability of one pixel to be a seed. Therefore, the target objects can be located, in principle, by seeking local maxima in the generated probability map. In practice, non-maximum suppression is often used to improve the accuracy. We review the recent literature using CNN-, FCN-, and SAE-based methods for object detection as follows.

1) CNN Classification: Ciresan *et al.* [38] have applied CNNs to mitosis detection in hematoxylin and eosin (H&E) stained breast cancer histology images. To improve rotational invariance, data augmentation is conducted by applying arbitrary rotations and/or mirroring to training images. In the testing stage, the outputs from CNNs processing rotated and mirrored versions of input images are averaged to produce final probability maps, and kernel smoothing is used to further reduce noise such that mitosis centroids can be easily detected by seeking local maxima (with non-maximum suppression). This approach significantly outperforms all other competing techniques in the ICPR 2012 mitosis detection contest [55]. Similarly, CNNs are applied to pixelwise prediction for nucleus detection in pancreatic neuroendocrine tumor (NET), brain tumor, and breast cancer images [39], [56], [57] and circulating tumor cell in phase contrast microscopy images [58]. Wang *et al.* [59] have applied an eight-layer CNN to inflammatory bowel disease images to generate neutrophils candidates and then model cell context with a Voronoi diagram of clusters to detect true neutrophils.

It is not always necessary to conduct CNN prediction on each image pixel for target object detection in microscopy images. For cell detection in wide-field microscopy zebrafish images [60], an SVM classifier is first used to detect cell region candidates, which are further classified as cells or image background by a trained CNN. It demonstrates that this strategy provides significantly better detection accuracy than CNN prediction on each pixel. For mitosis detection in [61], a CNN classifier is applied to mitosis candidate patches (binary values), which are generated by using simple image processing techniques such as enhancement, thresholding, and filtering. Liu and Yang [40] have employed a CNN model to assign scores to cell candidates, which are generated by using various other algorithms, and selected a subset of candidates as final detection results by solving a maximum-weight-independent set problem. This approach is a general cell detection framework that can be applied to various data sets with different staining preparations, including H&E and Ki-67 immunohistochemical (IHC) staining.

2) CNN Regression: In order to enforce pixels close to object centers to have higher probability values than those elsewhere, spatial topology can be incorporated in deep learning models. Xie *et al.* [41] have replaced the last classification layer in a conventional CNN with a structured regression layer. Given an input Image I, the corresponding proximity mask $\mathcal{M}$ can be computed as follows:

$$\mathcal{M}_{ij} = \begin{cases} \dfrac{1}{1+\alpha D(i,j)} & \text{if } D(i,j) \le r \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\mathcal{M}_{ij}$ denotes the ij-th entry of $\mathcal{M}$. $D(i,j)$ is the Euclidean distance from (i,j) to the human annotated cell center closest to it. r represents a distance threshold and α is a predefined decay ratio.

To train the network, the paper [41] proposes to extract a set of local image patches and corresponding local proximity patches as the training data. Since the value of $\mathcal{M}_{ij}$ is within the interval $\mathcal{V} = [0,1]$, the proximity patch for patch $\boldsymbol{x} \in \mathcal{R}^{d\times d}$ can be defined as $\boldsymbol{y} \in \mathcal{V}^{d'\times d'}$, where $d \times d$ and $d' \times d'$ are the sizes of the local image patch and the proximity patch, respectively. The CNN model can be represented as a complex function ψ with parameter $\boldsymbol{\Theta}$. Given a set of training samples $\{(\boldsymbol{x}^i, \boldsymbol{y}^i)\}_{i=1}^N$, and the loss function l defined for one training sample, the model parameters can be learned by solving the following optimization problem:

$$\arg\min_{\boldsymbol{\Theta}} \frac{1}{N}\sum_{i=1}^{N} l(\psi(\boldsymbol{x}^i;\boldsymbol{\Theta}), \boldsymbol{y}^i). \tag{2}$$

For one pair of training sample $(\boldsymbol{x}^i, \boldsymbol{y}^i)$, we can obtain the model's output as $\boldsymbol{o}^i = \psi(\boldsymbol{x}^i;\boldsymbol{\Theta})$. Let o_j^i, a_j^i, and y_j^i represent the jth element of $\boldsymbol{o}^i$, $\boldsymbol{a}^i$, and $\boldsymbol{y}^i$, respectively. The loss function l can be defined as

$$l(\boldsymbol{o}^i, \boldsymbol{y}^i) = \frac{1}{2}\sum_{j=1}^{p} \left(y_j^i + \lambda\right)\left(y_j^i - o_j^i\right)^2 \tag{3}$$

where λ is a parameter used to tune the weights of losses coming from different parts of the proximity patch. A small λ value indicates that the model is forced to pay less attention to the area that has a lower proximity value.

Given a new testing image, all the local image patches are evaluated and the obtained local proximity masks are then fused to get the final proximity mask. This approach has been extensively tested for cell detection in pancreatic NET, breast cancer, and HeLa cell images, and outperforms other popular traditional methods.

Similarly, spatial regression is introduced into a CNN for nucleus detection and classification in colon cancer histology images [62]. Different from [41], the method in [62] explicitly constrains the output from the neural network. In addition, it further classifies nuclei into four categories with a weighted neighboring patch fusion technique. Another regression with

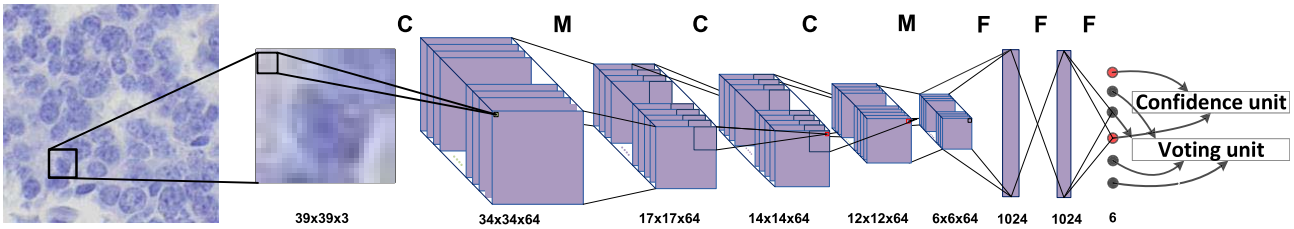


Fig. 3. Architecture of the deep voting model [43]. The C, M, and F represent the convolutional layer, the max-pooling layer, and the fully connected layer, respectively. In the last layer, the voting offset and confidence units are marked with different colors. In this case, the number of voting positions for each local patch k is 2.

CNNs (actually implemented with FCNs) for microscopy image analysis is reported in [42]. Specifically, it casts cell counting as a supervised learning-based density estimation problem [63]. The goal is to learn a mapping from a raw image to a density map, where the number of cells in a certain region is achieved by integrating the density map over that region. In this way, it does not require explicit cell detection, which usually is a nontrivial task. Another significant benefit is that the regression is developed with FCNs, which allows end-to-end training and can handle input images with arbitrary sizes. The proposed networks are evaluated with both synthetic and real data (retinal pigment cell images and precursor T-cell lymphoblastic lymphoma images), and provide very promising cell counting results.

Without conducting individual pixelwise hard classification, Xie *et al.* [43] have proposed a CNN-based voting approach, namely, deep voting, for nucleus detection in NET images. Specifically, it learns a CNN model as an implicit codebook and maps the current local image patch to a set of voting offsets which specify possible target (nucleus) positions and the corresponding voting confidences. The voting confidence is used to weight each corresponding vote. Note that the number of votes is a predefined value. The architecture is shown in Fig. 3.

For each training image patch, it collects the coordinates of k nucleus centers from a group of human annotations (gold standards) that are closest to the patch centers. Let $\mathcal{D}^i = (\mathcal{P}^i, \mathcal{T}^i)$ represent the ith training data, where $\mathcal{P}^i \in \mathcal{P}$ is the ith input image patch and $\mathcal{T}^i \in \mathcal{T}$ is the corresponding target information. $\mathcal{T}^i$ can be further defined as $\{\alpha_j^i, \boldsymbol{v}_j^i\}_{j=1}^k$, where $\boldsymbol{v}_j^i$ is the 2-D offset vector specifying the displacement from the center of patch $\mathcal{P}^i$ to the jth nearest gold standard nucleus. For each $\boldsymbol{v}_j^i$, the voting confidence α_j^i can be defined as

$$\alpha_j^i = \begin{cases} 1, & \text{if } \left|\boldsymbol{v}_j^i\right| \le r_1 \\ \dfrac{1}{1+\beta\left|\boldsymbol{v}_j^i\right|}, & \text{if } r_1 < \left|\boldsymbol{v}_j^i\right| \le r_2 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where r_1 and r_2 are used to specify the effective voting range, and β is the confidence decay ratio with respect to the length of each voting offset. Note that (4) is merely one of the possible definitions of α_j^i, and it can also be defined based on other properties of the local image patches.

The CNN model can be represented as ψ with $\boldsymbol{\Theta}$ as the model's parameters. Given a set of training data $\{\mathcal{D}^i = (\mathcal{P}^i, \mathcal{T}^i)\}_{i=1}^N$, where N is the number of training samples, the model's parameters can be obtained by solving the optimization problem

$$\arg\min_{\boldsymbol{\Theta}} \frac{1}{N}\sum_{i=1}^{N} l(\psi(\mathcal{P}^i;\boldsymbol{\Theta}), T^i) \tag{5}$$

where l is the loss function defined in the following.

Let $\psi(\mathcal{P}^i;\boldsymbol{\Theta}) = \{w_j^i, \boldsymbol{d}_j^i\}_{j=1}^k$ denote the model's output given the input $\mathcal{P}^i$, where w_j^i and $\boldsymbol{d}_j^i$ correspond to the voting confidences and offsets, respectively. Recall that the target information of $\mathcal{P}^i$ is $\mathcal{T}^i = \{\alpha_j^i, \boldsymbol{v}_j^i\}_{j=1}^k$. The loss function on one training sample can be given as

$$l(\psi(\mathcal{P}^i;\boldsymbol{\Theta}), \mathcal{T}^i) = \sum_{j=1}^{k} \frac{1}{2}\left(\alpha_j^i w_j^i \left\|\boldsymbol{d}_j^i - \boldsymbol{v}_j^i\right\|^2 + \lambda\left(w_j^i - \alpha_j^i\right)^2\right) \tag{6}$$

where λ is a constant used to tune the weights of losses coming from voting confidences and offsets. This loss function has the following useful properties. First, $\alpha_j^i w_j^i$ in the first term punishes uninformative votes, which have low voting confidences in the target information or predicted voting confidences. This property alleviates the potential problem caused by the fixed number of votes for each image patches. For uninformative votes, the loss coming from voting offsets vanishes, and this loss function degenerates into a squared error loss defined on the voting confidences. Second, not only does the second term penalize the error coming from voting confidences, but it also acts as a regularization term to prevent the network from producing trivial solutions by predicting all voting confidences as zero.

Since the proposed loss function is end-to-end differentiable, it can be trained using the efficient back-propagation algorithm [18]. Meanwhile, since the voting confidence and offset have different value ranges, the authors also propose a hybrid nonlinear transformation as the new activation function in the last layer. For voting units, they use a simple linear activation function; for confidence units, the sigmoid activation function is utilized to squash the activation.

In the testing stage, each local image patch is evaluated separately and votes to several possible positions. The votes are then accumulated and smoothed by using a Gaussian filter to obtain the final voting density map. The nuclei can be localized by finding all the local maxima. In practice, they also use a large testing stride (means they do not evaluate every single local image patch) to speed up the testing with a light

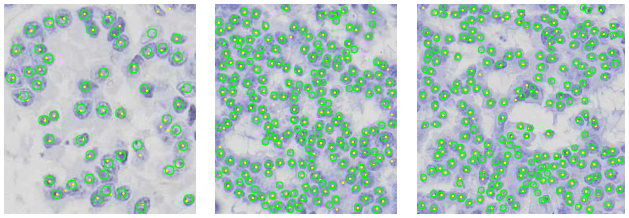


Fig. 4. Nucleus detection using deep voting [43] on several example NET images. Yellow dots: detected cell centers. Small green circles: gold standards.

sacrifice of the localization accuracy. The nucleus detection results using the deep voting method on several sample NET images are shown in Fig. 4.

3) Data Consideration: Albarqouni *et al.* [44] have presented a multiscale CNN framework, namely, AggNet, to learn from crowds for mitosis detection in H&E stained breast cancer images. By introducing an additional crowdsourcing layer, AggNet can directly deal with data aggregation in the learning process such that image annotation from nonexperts can be achieved for biomedical applications. Specifically, the proposed framework mainly consists of the following steps: train multiple CNNs with gold standard annotations, conduct mitosis detection with CNNs on new data, send detected candidates to the crowds for annotation via a web-platform, and collect the annotations to refine the CNN models and generate ground-truth labels. One benefit of AggNet is that it is robust to noisy annotations.

van Grinsven *et al.* [64] have proposed a training data sampling heuristic to improve and accelerate CNN training, and it has been successfully applied to hemorrhage detection in color fundus images. Since normal training samples are overrepresented and most of them are highly correlated with each other, uniformly treating the data will lead to slow convergence. The proposed method assigns different weights to training data during the training procedure such that informative samples are more likely to be selected for parameter update in next iteration. In this way, it can greatly improve the efficiency of CNN training.

4) SAEs: A stacked sparse autoencoder (SSAE) for nuclei detection in breast cancer images is reported in [45]. The network is learned with unsupervised pretraining followed by supervised fine-tuning. Specifically, an SSAE is first trained on raw image data with a sparsity constraint and a reconstruction loss. A softmax layer is then attached on the top of the SSAE and trained via fine-tuning in which the image patches aligned to nuclei are treated as positive samples and those misaligned to nuclei are treated as negative samples. In the testing stage, a sliding window method is employed for model inference. One advantage of this approach is that the parameters of the model are further refined through the supervised training to obtain more discriminative power such that better classification performance can be achieved.

B. Segmentation

Image segmentation, especially nucleus or cell segmentation, is a critical prerequisite in image-based computer-aided diagnosis [65], serving a basis of many image analyses, such as cellular morphology computation, characteristic quantification, cell recognition, and so on. In addition, accurate segmentation of neural structures is an essential step of dense neural circuit reconstruction. Therefore, automatic and effective segmentation is in an urgent need for microscopy image analysis. Recently, DNNs have been used to segment microcopy images and provided very promising performance. Typically, CNNs formulate it as a pixelwise classification problem: use a sliding window on input images to generate probability maps and then achieve image segmentation with thresholding; alternatively, FCNs trained in an end-to-end manner can directly output probability maps that have the same dimensions as input images, and thus computational efficiency can be significantly improved.

1) CNNs: Ciresan *et al.* [66] have used CNNs as binary pixel classifiers to label each pixel for neural membrane segmentation in serial-section transmitted electron microscopy (ssTEM) images. In order to improve network performance, foveation and nonuniform sampling are used to manipulate the data, and a polynomial function post-processing is used to calibrate network outputs followed by thresholding. Averaging predictions of multiple similar networks is also applied to robustness enhancement. This method outperforms the other competitors by a large margin and have won the ISBI 2012 EM segmentation challenge [15]. Later, CNNs with different architectures [67] are used on the same data set and a watershed merge tree technique [68] is applied to postprocessing for segmentation improvement.

Ning *et al.* [69] have designed an automatic CNN-based framework for analysis of developing *C. elegans* embryos. More specifically, a CNN is trained as a pixel classifier to conduct coarse image segmentation into five categories: nucleus, nucleus membrane, cytoplasm, cell wall, and outside medium. Next, the segmented image is further processed with an energy-based model and a set of elastic models for phenotyping analysis. In [70], a multiscale CNN is applied to feature learning for cervical image segmentation, which first classifies each pixel as background, cytoplasm, or nucleus, and then refines this coarse segmentation with graph partition [71]. In order to segment individual objects of interest, these methods require data-specific postprocessing to obtain desired performance. For breast cancer region segmentation in histopathology images [72] and muscle perimysium segmentation in skeletal images [73], CNNs provide pixelwise prediction followed by thresholding to obtain final results.

2) FCNs: Inspired by [21], Ronneberger *et al.* [75] have proposed a convolutional network, namely, U-net, that can be trained end-to-end for microscopy image segmentation. Not only does U-net unsample feature maps in the expansion path, but it also copies and crops corresponding feature maps from the contraction path such that the network can take advantage of context information. To deal with touching objects, it introduces a weighted loss such that the separation borders between touching objects are encouraged to learn in the end-to-end training. Additionally, it exploits elastic deformations to process the annotated training images for data augmentation [76], and this can effectively tackle the issue of

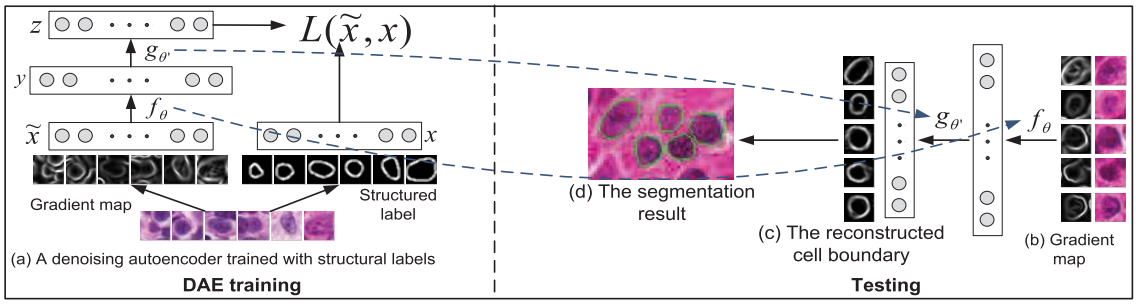


Fig. 5. Overview of the nuclei segmentation based on a denoising autoencoder in [74]. (a) Denoising autoencoder trained with structural labels. (b) Gradient map. (c) Reconstructed cell boundary. (d) Segmentation result.

limited training data, which occurs very often in medical image computing. U-net has been used to segment neural membranes in ssTEM images [15], glioblastoma-astrocytoma U373 cells in phase-contrast microscopy images, and HeLa cells in differential interference contrast microscopy images [77].

Chen *et al.* [78] have modified and extended the FCN [21] by incorporating multilevel contextual information and auxiliary supervised classifiers into more deeper networks for neuronal structure segmentation. Similar to [21], the contraction path consists of convolution and pooling operations, and this aims at classifying semantic information. On the other hand, the expansion path contains multiple convolutional and deconvolutional layers from different levels, and the hierarchical context information is fused and fed to a softmax layer. In order to alleviate the vanishing gradient problem and improve discriminative power of features in intermediate layers, auxiliary classifiers [79], [80] are introduced into the network for end-to-end training. In the testing stage, an overlap-tile strategy is applied to robustness improvement of image segmentation.

3) RNNs: RNNs have been widely used in traffic scene image understanding [81], financial signal processing [82], and biomedical applications with sequential data such as gene expression analysis [83], protein classification [84], EEG signal processing [85], and so on. Meanwhile, they have also been applied to biomedical image segmentation.

Recently, Stollenga *et al.* [86] have proposed a variant of multidimensional LSTM called PyraMiD-LSTM, which is easy to parallelize for segmentation of neuronal structures [15] and MR brain images. The basic idea is to rotate the topology structure of traditional LSTM by 45°, making the current time step only rely on units in either the row or column direction and thus renders it possible to calculate a row or column at once rather than one unit at a time. In addition, a traditional LSTM or RNN usually needs to compute many different directional sweepings starting from each corner to cover the whole context information; PyraMiD-LSTM borrows the same idea, but only needs to sweep the entire image or volume from each edge or plane rather than vertex. This can further reduce the necessary computation for 3-D volume data. The experimental results show that the PyraMiD-LSTM [86] achieves very competitive results on EM images and the best results on MR brain images.

Inspired by multidimensional RNNs [52], [86] and the Clockwork RNN (CW-RNN) [87], Xie *et al.* [54] have proposed a spatial CW-RNN for efficient and accurate muscle perimysium segmentation on H&E stained images. The central idea shares the similar spirit with [52], but they propose a 2-D version of CW-RNN, which contains much fewer parameters and requires a lower computational cost than LSTM [88]. Different from [86], which uses a time step for every single pixel, the approach in [54] splits the entire image into nonoverlapping regions and their dependences with each other are modeled with the spatial CW-RNN. This means that context information of the entire image is used for predicting the semantic label of each single local image region. Experimental results demonstrate that their proposed spatial CW-RNN is very efficient and can provide the state-of-the-art segmentation accuracy.

4) SAEs: Su *et al.* [74] have proposed a nucleus segmentation approach based on stacked denoising autoencoders (SDAEs). Traditionally, SDAEs are trained by minimizing the reconstruction error with respect to the original image. However, in [74], after the nuclei detection, the SDAE takes the noisy gradient map of an image patch centered on a nucleus as the input, and aims to reconstruct the structured label that is the true nuclear boundary annotated by human. Intuitively, the noisy gradient map can be considered as the corrupted version of the human annotation. Therefore, the model is trained in compliance with the *denoising criterion* [89]. The learned model is able to remove fake edges and correct broken or weak edges. Specifically, let $\tilde{\boldsymbol{x}}$ denote the noisy gradient map of the original image patch centered on a nucleus and $\boldsymbol{x}$ represent the human annotated contour. One layer of the denoising autoencoder is defined by

$$f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}) = s(\mathbf{W}\tilde{\boldsymbol{x}} + \boldsymbol{b}) \tag{7}$$

$$g_{\boldsymbol{\theta}'}(\boldsymbol{y}) = s(\mathbf{W}'\boldsymbol{y} + \boldsymbol{b}'). \tag{8}$$

The model can be learned by minimizing the squared loss over the parameters $\{\mathbf{W}, \boldsymbol{b}, \mathbf{W}', \boldsymbol{b}', \}$

$$L(\tilde{\boldsymbol{x}}, \boldsymbol{x}) = \|\boldsymbol{x} - g_{\boldsymbol{\theta}'}(f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}))\|^2. \tag{9}$$

To form a deep architecture, different layers are trained in a sequential way. First, the first layer is trained on the gradient maps and the structured labels, as shown in Fig. 5. Then, the second layer is trained using the hidden representations computed from the gradient maps and the hidden representations computed from the corresponding structured labels by the first layer. The trained layers are stacked together to form an SDAE,

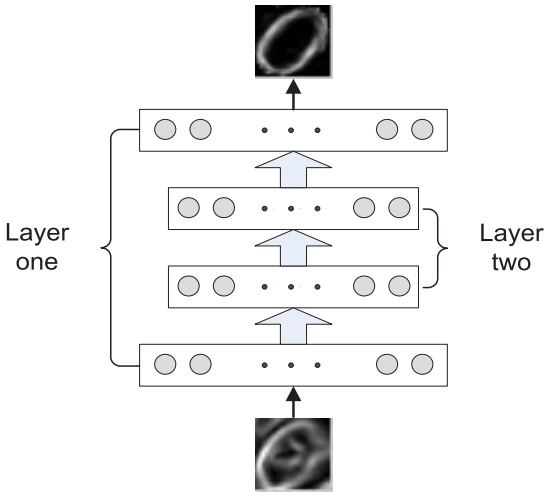


Fig. 6. SDAE in [74].

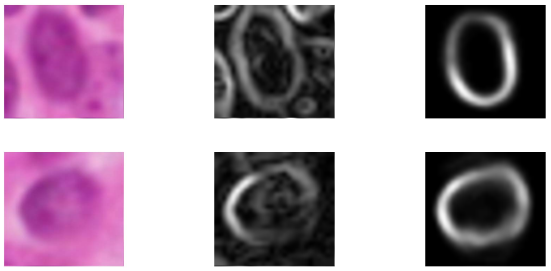


Fig. 7. Results on sample images [74]. (a) Original image patches. (b) Gradient maps. (c) Outputs of the SDAE.

as shown in Fig. 6. Fig. 7 shows that fake edges are removed and broken/weak edges are corrected.

C. Classification

Instead of performing pixelwise prediction, image-level classification assigns a single label to each input image. Supervised learning based networks, especially CNNs, are powerful tools for microscopy image classification, but unsupervised feature learning has been used to solve many computer vision tasks [24], [90] and it could also be applicable to medical imaging. One straightforward way for image classification is to use neural networks as classifiers which directly output an individual prediction for one image. Alternatively, the networks, which are trained with large-scale data sets, can be used as feature extractors to generate data representations, which are fed to other target classifiers.

Gao *et al.* [91] have used deep CNNs to classify human epithelial-2 cell images into six categories. Specifically, they experimentally investigate the effects of CNN hyperparameters, data augmentation, and image foreground masks on the classification performance. Meanwhile, the experiments also demonstrate that CNNs, which are pretrained on a significantly larger data set and then fine-tuned on a smaller and related data set, can provide higher accuracy than those trained from scratch on the smaller data set. Actually, this strategy might be used to handle limited training data and has been experimentally verified in other tasks [47]. Chen *et al.* [92] have applied a fully connected neural network to label-free cell classification. Different from the widely used CNNs trained with cross-entropy and backpropagation, this network is globally trained using a heuristic genetic algorithm with the area under the curve of the receiver operating characteristics. This learning algorithm provides higher classification accuracy than traditional classifiers and CNNs on their data sets, but it is computationally expensive. In addition, the network is trained with a set of biophysical features instead of raw images such that domain expertise might be required for input generation. Carneiro *et al.* [93] have employed a deep CNN model to estimate the number of and the proportion of classes of microcirculatory supply units (MCSU) in human squamous cell carcinoma. For one pair of input images, it first uses four different classifiers to produce multiple output maps, which are fed into a CNN to determine MCSU classes.

A pretrained CNN can be used as a fixed feature extractor for image categorization. In [94], the CNN is trained with a sufficiently large data set [95], such that it can provide very powerful generic descriptors for brain tumor histopathology image classification. The CNN is applied to overlapping patches, and features are extracted from the penultimate layer of the network; then, feature pooling [96] is exploited to generate a single representation vector for the entire image and feature selection is employed to eliminate redundancy; finally, a linear SVM classifier with the final representation is used to perform image classification of glioblastoma multiforme (GBM) and low grade glioma. This approach is also extended to segment necrosis regions in GBM images, and this is achieved by simply conducting patchwise instead of image-level classification. Using CNNs as feature extractors can also be found in [97], where an SVM and a weakly supervised learning model, multiple instance relearning [98], [99], are applied to colon cancer image classification.

A pioneer work of using autoencoder in microscopy image classification is reported in [100], in which a sparse autoencoder is exploited to learn visual features from the raw image data. Different from the traditional autoencoder, a sparsity constraint is introduced to the model

$$\mathcal{L}_{\text{sparse}}(\mathbf{W}) = \mathcal{L}(\mathbf{W}) + \beta \Sigma_{j=1}^{k} KL(\rho||\hat{\rho}_j) \tag{10}$$

where $\mathbf{W}$ is the model parameter and $\mathcal{L}(\mathbf{W})$ represents a typical cost function used for network training. β controls the sparsity, $KL(\cdot)$ denotes the Kullback–Leibler divergence, ρ is the specified sparsity parameter, $\hat{\rho}_j$ is the average activation of the hidden unit j, and k represents the number of the hidden units. The image representation is obtained by conducting a convolution on the image and then an average pooling on the resulting feature map. A softmax classifier is trained to classify the image regions into cancerous regions and noncancerous regions.

The encoder function can also be learned via predictive sparse decomposition (PSD) that is an efficient variant of sparse coding [101], [102]. In PSD, the expensive optimization used in the original sparse coding is replaced with elementwise nonlinearity and matrix multiplication in the encoder function [103]. This makes the algorithm computationally efficient for large scale feature extraction. Specifically, assuming

$\mathbf{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \in \mathbb{R}^{m \times N}$ denotes a set of N training samples, PSD aims to learn a set of parameters $\{\mathbf{B}, \mathbf{Z}, \mathbf{G}, \mathbf{W}\}$ by solving

$$\min_{\mathbf{B},\mathbf{Z},\mathbf{G},\mathbf{W}} \|\mathbf{X} - \mathbf{B}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\| + \|\mathbf{Z} - \mathbf{G}\sigma(\mathbf{W}\mathbf{X})\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{b}_i\|_2^2 = 1 \quad \forall i = 1, \ldots, h \tag{11}$$

where $\mathbf{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_h]$ is a set of basis vectors (dictionary), $\mathbf{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N]$ represents the sparse code feature for each sample, $\mathbf{W}$ is the autoencoder matrix representing weights, $\mathbf{G}$ denotes a diagonal scaling matrix, and $\sigma(\cdot)$ is the sigmoid function. The model is trained in an alternative way. $\mathbf{Z}$ is learned using the orthogonal matching pursuit [104] with other parameters fixed, $\mathbf{G}$ is calculated analytically as the solution to a simple least-square problem when others are fixed, and $\{\mathbf{B}, \mathbf{W}\}$ are learned by stochastic gradient descent with $\mathbf{Z}$ and $\mathbf{G}$ fixed. The advantage of this method is that the image feature can be approximately and efficiently computed using the encoder function $\mathbf{Z} = \mathbf{G}\sigma(\mathbf{W}\mathbf{X})$. In [103], two layers of the PSD are stacked together such that the features computed by the first layer are used as the input to the second layer. The output of the second layer is used to train an SVM for differentiating distinct tumor regions in histopathology images.

III. Discussion

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 [22]. Despite this success, there are several challenges or issues to be addressed. First, how DNNs achieve excellent performance is not completely and theoretically understood, and this would be an issue when it is necessary to interpret the results in the medical domain. Second, the large amount of medical image data, including microscopy images, needs a high processing rate, thereby requiring computation acceleration in DNNs. Third, currently DNNs have not comprehensively understood visual scenes provided by the images yet, and this would affect the interpretation of medical image data which provide very rich information for disease characterization.

A. Behind DNNs

Although a mathematical justification for the success of DNNs is currently not completely clear, there exists some research from theoretical perspectives. Nitta [105] has pointed out that critical points appear due to a hierarchical structure in DNNs and derived a sufficient condition for a DNN without critical points such that pretraining might be not necessary for DNN learning. In [106], a novel topological concept-based measurement is proposed to evaluate the complexity of functions implemented in neural networks, and it is observed that DNNs can realize more complex functions than shallow architectures using the same number of resources; in [107], DNNs show higher efficiency than shallow networks in encoding functions that exhibit repeated patterns, thereby giving rise to significant reduction in complexity. A comparative complexity analysis of deep and shallow networks is conducted with generative models such as restricted Boltzmann machines (RBMs) and DBNs, and it positively resolves the approximation properties of the networks [108], [109]. Chang [110] has shown that, theoretically, the depth of a multilayer neural network is encoded by the topological conjugacy of its hidden layers, and provided an approach to determine the existence of topologically conjugated hidden spaces.

One potential interpretation of DNNs is from the manifold learning perspective, which hypothesizes that manifold-shaped data can be flattened in higher layers. Brahma *et al.* [111] have quantified this flattening in DBNs and justified the unfolding and separation of low-dimensional manifolds in the input data. Moreover, Shao *et al.* [112] have presented a multispectral embedding algorithm to map hierarchical discriminative manifolds into a compact representation and thus enhance the model robustness and reduce the error rate. Some other effort has been devoted to sparseness analysis in the pretraining of DNNs [113]. It has shown that pretraining generates more sparseness in sigmoid-activated RBMs and denoising autoencoders; it has also provided several sufficient conditions for pretraining to converge to a sparse activation. On the other hand, Gong *et al.* [114] have proposed a multiobjective sparse feature learning model based on autoencoders, which can automatically determine the sparsity of hidden units.

One popular way to understand the networks is visualizing the learned weights or activation/heat maps. The first layer usually detects primitive image features such as edges and color blobs, and higher layers learn more abstract representations that are specific to target tasks [13], [24], [115]. There exist several attempts to understand and visualize the networks by retrieving input images that maximally activate some neurons [19], [116] or directly analyzing visual information contained in the learned representations [117]. In addition, a deconvolution method is presented to identify and link patterns in input images to specific CNN predictions [118], a sensitivity analysis approach is leveraged to visualize input sensitivities in images [119], and a layerwise relevance propagation (LRP) algorithm is proposed to quantify pixelwise relevances for overall classification scores [120]. A comparative study of these three visualization methods is presented in [121], where LRP exhibits superior performance to the other two. It presents a general framework to evaluate the visualization and could be helpful in understanding the DNNs.

B. Network Acceleration

A large number of optimization algorithms have been proposed to improve computational efficiency in deep model training [6], [122], and many of them have been implemented in those popular deep learning frameworks, such as Caffe [123], Theano [124], TensorFlow [125], Torch [126], and so on. Furthermore, parallel and distributed computing techniques have been employed to accelerate DNN learning [127]–[130]. Parallel programming with graphics processing units (GPUs) is widely used in various deep learning platforms, and this is a great advantage for both model

training and testing; distributed computing allows model optimization and runtime inference on multiple machines. Compared with standalone applications, distributed computing is more suitable for large-scale learning tasks. For instance, we can divide an image generated by whole slide imaging (WSI) techniques into a set of image tiles, then send them to multiple standard PC machines for data processing via a distributed scheme, and finally collect and stitch individual outputs from different machines to obtain desired results. In this way, it can address the memory and computational limitations of one single machine. It is worth noting that compared with training acceleration, the improvement of runtime speed is always more significant, because that is directly related to the waiting time of the end users.

Currently, hardware acceleration for deep learning is dominated by GPU-based programming [131], especially the NVIDIA CUDA programming [132], [133]. GPUs exhibit a great ability for parallel computing and provide much better computational efficiency than standard central processing units. However, GPUs consume high power such that they are not suitable to resource-limited embedded systems. Custom architectures that have less power consumption with more efficient performance have attracted research attention for accelerating DNNs and other algorithms [134]–[140]. Field-programmable gate arrays (FPGAs) that exhibit more flexible hardware configuration and lower power requirement than GPUs provide an attractive alternative for DNN acceleration [141], [142]. The early effort devoted to FPGA-based acceleration for neural networks can be found in [143] and [144], where the applications are limited by the FPGA size constraints. Recently, DNN implementations on FPGAs have achieved the state-of-the-art performance [145]–[147].

Application specific integrated circuit (ASIC) accelerators are another promising alternative for runtime reduction of DNNs. Compared with FPGAs, ASICs can provide high-throughput performance with lower power consumption, thereby potentially handling more complex tasks [148]. Reagen *et al.* [149] have presented an automated codesign method across the algorithm, architecture, and circuit levels to optimize the ASIC-based acceleration, which enables DNN applications in power-constrained devices. Chen *et al.* [150] have implemented and fabricated a reconfigurable accelerator, which can adapt to various CNN shapes for energy efficiency optimization. Except CNN accelerators, research effort has also been devoted to ASIC design for other neural networks, such as DBNs [151] and RNNs [152]. Recently, a custom ASIC called tensor processing unit [153] has been specifically designed for machine learning applications, which aims to drive DNNs to be tailored for TensorFlow [125]. Generally, ASICs are less flexible than GPUs and might be expensive to produce at the current stage. It is obvious that the aforementioned endeavors will be transferred to the biomedical domain in the near future. Recently, this is a promising and hot research area as well.

C. Automated Image Interpretation

Although exciting progress has been made in image understanding, computer vision and microscopy image analysis have not yet reached the stage of the comprehensive understanding of visual scenes, which is a holy grail of computer vision [154]. Compared with the tasks of object detection, segmentation, and recognition, it would be more interesting to perform reasoning about the visual world from images. A more detailed understanding of biomedical images could significantly benefit both physicians and patients by providing stronger support for diagnosis. To this end, usually natural language processing techniques are combined with image understanding algorithms to interpret image data. Recently, Shin *et al.* [155] have presented an interleaved text/image mining system, which uses DNNs to automatically generate descriptive attributes of patient images from a large-scale radiology database and detect frequent disease types for more specific interpretation. This might pave a new path for automated understanding of large-scale medical image data sets.

Visual question answering (VQA) [156], [157] is a method toward cognitive scene understanding, which can be used to evaluate the image interpretation ability of an image understanding system. VQA extends the idea of describing the visual content given an image/video to question answering and produces a natural language answer based on the provided image/video and the corresponding question. By taking advantage of two channels, vision and language, of the same scene, VQA provides a more accurate interpretation [158], [159]. Nowadays, DNN-based VQA has applied on natural image data sets, producing very impressive performance [159], [160], and it might draw considerable attention in the community of microscopy image analysis. However, how to effectively conduct VQA on microscopy images that exhibit unique characteristics is still an open challenge.

IV. Conclusion

A. Unique Challenges in Microscopy Image Analysis

1) High Image Dimension: In microscopy image analysis, in particular pathology imaging informatics, it is critical to conduct quantitative analysis on WSI images instead of manually selected regions, such that quantitative analysis is allowed for the entire morphologic landscape [161]. In addition, WSI image quantification can effectively eliminate the sampling errors, which frequently occur in selected region analysis. However, a WSI image often has a dimension of over $50\,000 \times 50\,000$ pixels and contains tens of thousands or millions of objects of interest (e.g., nuclei or cells). This is very different from general/natural image data in computer vision or other medical imaging modalities such as MR imaging, computed tomography, and ultrasound, which usually have much smaller sizes.

The large WSI images exhibit several significant and unique challenges. First, it might take a long time for a DNN, which needs to conduct extensive computation, to process one single entire image (in the testing stage). For instance, current pixelwise prediction of CNNs is mainly based on the sliding-window approach, which processes one image patch at a time. Clearly, this will be extremely computationally expensive for processing WSI images. FCNs are designed for fast inference, and from an algorithmic standpoint, they would be good

candidates to improve the speed, but it would be difficult to conduct large-scale matrix computation in GPUs due to limited memory. Second, it is not sensible to directly resize the whole WSI image into a smaller one (e.g., a 256×256 or 512×512 image), which can be fed into those commonly used pretrained DNNs (e.g., AlexNet [13]). Resizing will result in the loss of massive information, which is very important for cellular morphology characterization, and thus it will degrade the performance of target detection, segmentation, or classification in medical diagnosis. Finally, it is common to divide the entire WSI image into a set of smaller patches and then stitch computational results for all the patches to achieve the assessment of whole WSI image analysis. However, it is not easy to design an effective and efficient patch stitching method. A WSI image usually consists of thousands of patches and probably only a few of them exhibit tumor characteristics (e.g., at the early stage of breast cancer), with all the others being normal tissue regions or image background. This configuration might be prone to false negative in image-level classification. In addition, each patch is analyzed independently such that the correlation of topological information between different or adjacent patches is not taken into consideration, and thus it would lower the accuracy of object localization. Therefore, it is valuable to develop DNN-based models that are able to not only conduct efficient WSI image analysis but also effectively detect rare events with ignoring nonrelated information.

2) Image Artifacts and Batch Effects: Another unique challenge in microscopy image analysis is image artifacts and batch effects. Microscopy image acquisition is very different from that in other medical imaging modalities, where the anatomy is roughly in the same position with an identical scale. In pathology and histology, microscopic images are acquired with a series of processing steps, such as tissue collection, embedding, sectioning, and imaging [33]. Errors in the preparation would lead to image artifacts such as tissue folds, shadows, blurred regions, and so on [162]. These undesired outcomes significantly affect the image quality and challenge automatic image computing including deep learning. Although there are some reports on the correction of image artifacts in microscopy images [163], generally the literature on this topic is scarce.

Compared with image artifacts, batch effects present more important challenges, especially for collaborative research where the images are generated from multiple institutions. Due to different data preparations and microscope devices, color or scale variations will occur between two batches of specimens, even though they are from the same patient. Color variations can lead to significantly distinct and inconsistent color intensity in pathology images. On the other hand, scale batch effects are more difficult to detect and address, since object scale might naturally change due to disease progression and this is not easy to differentiate from that caused by disparate imaging procedures. These batch effects can cause bias in performance evaluation of predictive models [162]. Therefore, algorithms developed for microscopy image computing need to be robust to these variations. There is a large amount of literature for eliminating color or scale batch effects [164]–[166]; however, these methods require manual proper parameter selection or hand-crafted feature engineering. Recently, deep learning-based methods have been proposed for stain normalization to remove color variations [167], [168], and we expect the number of literature using DNNs for this task to increase in the future.

3) Object Crowding and Overlapping: It is very challenging to achieve robust object detection and segmentation due to crowding and partial overlapping of nuclei or cells in microscopy images, especially for digital histopathology specimens [see Fig. 1(a) and (c)]. Unlike most radiology images that usually contain only one or a few objects of interest per image, a pathology image often has tens of thousands of nuclei or cells and many of them are clustered into clumps such that they touch or partially overlap each other. In this scenario, the target objects might exhibit weak or misleading boundaries, thereby forming an obstacle for automatic individual object detection and segmentation.

Nowadays, many DNN-based approaches have been applied to nucleus or cell detection in different kinds of pathology images, which are acquired using bright-field or phase-contrast microscopes with H&E or IHC staining, leading to promising performance [41], [62]. This success might be partially attributed to the design of task-specific objective functions for model learning and lack of requirement of accurate delineation of object boundaries. However, this problem has not been completely solved and there is still room for improvement in object detection. In addition, it remains challenging to apply DNNs to direct nucleus or cell segmentation. Currently, it usually requires nontrivial postprocessing to achieve final object segmentation after DNN prediction on microscopy images. Although FCNs with end-to-end training are widely used for image segmentation in radiology, it is difficult for FCNs to effectively handle a large number of partially overlapped objects in histopathology images, which often exhibit severe background clutter. Furthermore, in microscopy image analysis, it is critical to preserve the shapes and sizes of nuclei and cells when conducting object segmentation such that it can facilitate subsequent cellular morphology computation for disease expression [169]. A potential solution is to incorporate shape prior modeling into DNNs, in particular CNNS and FCNs, but how to achieve this goal is still an open question in microscopy image analysis.

4) Insufficient, Imbalanced, and Inconsistent Data Annotations: Deep learning applications in general/natural image analysis usually assume that sufficient annotated training data are available, such as ImageNet [95]; however, this assumption does not hold in microscopy image analysis. Although The Cancer Genome Atlas [170] and Genotype-Tissue Expression [171] provide a number of WSI images for some diseases, only case-level or image-level labels are available. This is not helpful for individual nucleus/cell detection and segmentation, which, respectively, require patchwise and pixelwise data annotations. Meanwhile, it is prohibitively expensive to achieve extensive data annotations due to the shortage of medical expertise and privacy constraints [22]. One way to tackle this limited data problem in deep learning is to conduct data augmentation. Typically, the data sets can be enlarged with label-preserving

transforms [13], [172], such as translation, rotation, reflection, random crops, and color jittering. These methods are simple and computationally efficient, but the augmented data might be highly correlated. Alternatively, automated or semiautomated image annotation methods can be exploited to generate additional annotated data; however, designing efficient approaches to generate correct annotations remains a challenging problem in microscopy image analysis [173]. Crowdsourcing [174], [175] is one type of online activity in which a group of individuals are involved to undertake a task in order to generate large-scale data annotations from the online community. However, it is difficult to control the annotation consistency and quality, because many contributors are nonprofessional individuals, and this problem is more severe in digital pathology. There are some attempts to conduct deep learning from crowds for mitosis detection in histopathological images [44], but it is still necessary to put more effort into effective learning from crowds for microscopy image analysis applications.

In microscopy image analysis, we might have limited data annotation but with sufficient nonannotated training images present. In this scenario, DNNs can be learned with unsupervised pretraining followed by supervised fine-tuning. Unsupervised pretraining is usually conducted in a greedy layer-wise manner, aiming at learning latent representations of the data; supervised fine-tuning often involves learning a target classifier with previously learned representations [6], [29]. This scheme has been successfully applied to training various deep networks, including DBNs and SAEs [79], [176], [177], and demonstrates its effectiveness in natural and medical image analysis [178]–[181] but with very sparse literature in microscopy image computing. For supervised networks like CNNs, one way to handle limited data is to view a pretrained model that is learned with other data sets, either natural or medical images, as a fixed feature extractor, and use this model to generate features to train a target classifier for pixelwise or image-level prediction. Additionally, if the target data size is sufficiently large, it might be beneficial to initialize the network with a pretrained model and then fine-tune it toward the target task. The initialization can be conducted in the first several or all layers depending on the data size and properties. It is worth noting that compared with training from scratch, the initialization with the transferred features followed by fine-tuning to the target task can often significantly boost the generalization performance [182]. Currently, there is no public pretrained deep models with microscopy or pathology image data; however, it is expected that this kind of model would provide better performance compared with those pretrained models with the natural image data (e.g., ImageNet) due to a higher degree of similarity. In particular, a generic deep model, which is trained with digital pathology images acquired with distinct types of microscopes (e.g., bright-field, phase-contrast, fluorescence, and so on) and staining preparations (e.g., H&E and IHC), would be in urgent need, since it is not modality specific and would be able to deal with wide variations in pathology images generated with different imaging protocols.

Another data issue is that microscopy data sets are often imbalanced, especially for the applications of case-level or image-level classification. The number of healthy cases or images is significantly larger than that of diseased instances. Imbalanced training data might greatly degrade the performance of DNNs in classification tasks or slow down the convergence process. One straightforward method to address this issue is to leverage different sampling rates to generate training data with a balanced class distribution [183]. Dynamic sampling approaches [184], [185], which dynamically update the training data set during the model learning procedure, can also be used for imbalanced learning. By emphasizing informative samples, dynamic sampling methods can accelerate the training process of neural networks [64], [185]. Alternatively, cost-sensitive learning [186] assigns different costs to misclassified samples, and it has been applied to neural network training to deal with imbalanced data [187]–[189]. Other algorithms for imbalanced learning can be found in [190] and [191]. However, none of these algorithms have been applied to DNN learning in microscopy image analysis, and how to effectively incorporate imbalance learning into DNNs is still an open question.

In digital pathology, it is not unusual that data annotations or labeling are inconsistent. For instance, in diagnosis of the subtypes of nonsmall cell lung cancer (NSCLC), it is difficult to differentiate adenocarcinoma from squamous cell carcinoma based on examination of histopathologic tissue images. Therefore, data annotations for these two subtypes of NSCLC can be significantly inconsistent [192]. There would be large interobserver variations in the annotations due to the complex nature of histopathologic images and different background knowledge of observers. This is very different from labeling of general/natural images in computer vision, where data labeling does not require medical expertise and the label noise is much less. Thus, it is necessary to take label noise into consideration when designing a DNN model for microscopy image analysis, and this still remains an open challenge.

B. Future Trends of Deep Learning in Microscopy Image Analysis

Although deep learning has achieved state-of-the-art performance in several microscopy image analysis applications, it still needs considerable effort to tackle the aforementioned challenges and unsolved problems. In particular, it would be very valuable to develop efficient and effective deep learning-based WSI image analysis methods. To the best of our knowledge, there are no efficient algorithms for individual nuclease/cell detection or segmentation in WSI images, although the cell-level information can help differentiate many diseases' grades or stages and individual object quantification is an essential prerequisite. For image-level or patientwise classification, the current trend is to divide the WSI image into a set of overlapped or nonoverlapped small patches and then fuse the DNN predictions of a few selected patches for final outcomes [193], [194]. In this scenario, it can avoid the potential expensive computation of individual object localization. In addition to image classification, this patch-based strategy has been also applied to WSI lung and brain cancer survival

analysis [195]. We expect that it will become a potential future trend in high-dimensional microscopy image analysis.

Combing different types of data inputs in deep learning is a highly promising avenue for microscopy image analysis research. In the medical domain, physicians or doctors usually make decisions and deliver treatments based on a combination of various data such as radiographies, pathology images, clinical information (e.g., patient gender, age, disease history, and so on) and others. Several conventional machine learning-based data fusion algorithms have been proposed to combine pathologic, radiologic, and/or proteomic measurements for better disease characterization [196], [197], and one can expect that deep learning would be applied on this topic in the near future. Recently, the integration of medical images and diagnostic reports using DNNs has attracted research interest in medical image commuting, leading to improved image classification [198] or automatic text keyword prediction [155]. In particular, some effort has been devoted to diagnostic report generation on pathology image data sets [199], [200], where a CNN and an RNN are jointly trained with an attention mechanism such that the overall network is semantically and visually interpretable. This integration of image and text using a joint CNN-RNN network opens a new research area and will continually draw attention in microscopy image analysis.

Data-driven feature representations from DNNs have achieved great success in some microscopy image computing applications; however, it would be more interesting and convincing to design DNN architectures with taking domain expertise into consideration. Unlike general/natural image understanding, the analysis of digital pathology images involving domain-specific knowledge is more interpretable and helpful to pathologists and patients. In order to address the issue of lack of intuition in DNN-based feature learning, a straightforward solution is to concatenate the data-driven deep learning representations and domain inspired, hand-crafted features for specific tasks [201]; however, it would be desirable to learn a feature and/or decision fusion model for better performance. On the other hand, task-specific DNN architecture design has been witnessed a potential future direction. Although one can adopt the well-trained and widely used architectures such as AlexNet [13], VGGNet [202], and ResNet [203] for microscopy image computing applications, it might not be the optimal solution for detection or segmentation tasks because all these general networks are primarily designed for classification. By taking advantage of task-specific objective functions and the characteristics of corresponding microscopy images, several novel or improved CNN architectures have been proposed for object detection and segmentation [41], [62], [75]. We expect that this kind of architecture design will gain increased attention from the microscopy image analysis community in the future.

Since it is usually expensive to obtain sufficient annotated data in microscopy image computing, unsupervised or semi-supervised learning has been considered promising research directions. The recent breakthroughs in unsupervised deep learning like variational autoencoders (VAEs) [204] and generative adversarial networks (GANs) [205] provide a gateway to harness the huge amount of unlabeled medical data including microscopy images. The VAE arranges graphical models in an encoder–decoder framework and can be trained in an end-to-end manner. The GAN integrates two neural networks in a minimax two-player game. One is a generative network that attempts to generate fake samples to fool the other network called a discriminator, which aims to learn to differentiate the fake from true samples. The GAN can be trained end-to-end via minimizing the Jensen–Shannon divergence through backpropagation. These two unsupervised networks have proved to be more powerful than previous unsupervised methods in medical image computing [206], [207]. Although their advantages in microscopy image analysis have not been explored yet, they would play an important role in the field due to their ability of training with abundant unlabeled image data, which are relatively more easier to obtain than annotated data.

ACKNOWLEDGMENT

The authors would like to thank M. M. McGough for his suggestions for the manuscript writing.

REFERENCES

[1] C. Sommer and D. W. Gerlich, "Machine learning in cell biology—Teaching computers to recognize phenotypes," *J. Cell Sci.*, vol. 126, no. 24, pp. 5529–5539, 2013.
[2] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 25–38, Jul. 2010.
[3] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, and X. Li, "Modeling disease progression via multisource multitask learners: A case study with alzheimer's disease," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1508–1519, Jul. 2017.
[4] M. Graña and D. Chyzhyk, "Image understanding applications of lattice autoassociative memories," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 9, pp. 1920–1932, Sep. 2016.
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[7] Y. Yuan, L. Mou, and X. Lu, "Scene recognition by manifold regularized deep learning architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, Oct. 2015.
[8] R. Chalasani and J. C. Principe, "Context dependent encoding using convolutional dynamic networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1992–2004, Sep. 2015.
[9] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 3, nos. 3–4, pp. 197–387, 2014.
[10] S. M. Siniscalchi and V. M. Salerno, "Adaptation to new microphones using artificial neural networks with trainable activation functions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1959–1965, Aug. 2017.
[11] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, 2016.
[12] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings Bioinform.*, vol. 18, no. 5, pp. 851–869, 2017.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[14] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[15] I. Arganda-Carreras *et al.*, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers Neuroanatomy*, vol. 9, p. 142, Nov. 2015.
[16] M. Veta *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, 2015.
[17] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, "Deep neural nets as a method for quantitative structure-activity relationships," *J. Chem. Inf. Model.*, vol. 55, no. 2, pp. 263–274, 2015.

[18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[20] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

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

[22] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[23] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2010, pp. 253–256.

[24] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Int. Conf. Mach. Learn.*, 2009, pp. 609–616.

[25] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Commun. ACM*, vol. 54, no. 10, pp. 95–103, Oct. 2011.

[26] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[27] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015. [Online]. Available: http://neuralnetworksanddeeplearning.com/

[28] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning—A new frontier in artificial intelligence research [research frontier]," *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, Nov. 2010.

[29] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[30] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[31] G. Litjens *et al.*, "anchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[32] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[33] M. T. McCann, J. A. Ozolek, C. A. Castro, B. Parvin, and J. Kovacevic, "Automated histology analysis: Opportunities for signal processing," *IEEE Signal Process. Mag.*, vol. 32, no. 1, pp. 78–87, Jan. 2015.

[34] M. Veta, J. Pluim, P. van Diest, and M. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[35] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.

[36] M. Veta, P. J. van Diest, and J. P. W. Pluim, "Cutting out the middleman: Measuring nuclear area in histopathology slides without segmentation," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 632–639.

[37] Z. Xu and J. Huang, "Detecting 10,000 cells in one second," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 676–684.

[38] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 8150. 2013, pp. 411–418.

[39] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.

[40] F. Liu and L. Yang, "A novel cell detection method using deep convolutional neural network and maximum-weight independent set," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 349–357.

[41] Y. Xie, F. Xing, X. Kong, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 358–365.

[42] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting with fully convolutional regression networks," in *Proc. 1st Workshop Deep Learn. Med. Image Anal. (MICCAI)*, 2015, pp. 1–8.

[43] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 374–382.

[44] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1313–1321, May 2016.

[45] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[46] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.

[47] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[48] J. Wu, C. Leng, Y. Wang, Q. Hu, and J. Cheng, "Quantized convolutional neural networks for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4820–4828.

[49] J. Cong and B. Xiao, "Minimizing computation in convolutional neural networks," in *Proc. 24th Int. Conf. Artif. Neural Netw. Artif. Neural Netw. Mach. Learn. (ICANN)*, 2014, pp. 281–290.

[50] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," in *Proc. 30th Int. Conf. Int. Conf. Mach. Learn.*, vol. 28. 2013, pp. III-1319–III-1327.

[51] A. Graves, S. Fernández, and J. Schmidhuber, "Multi-dimensional recurrent neural networks," in *Proc. 17th Int. Conf. Artif. Neural Netw.*, 2007, pp. 549–558.

[52] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 545–552.

[53] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with LSTM recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3547–3555.

[54] Y. Xie, Z. Zhang, M. Sapkota, and L. Yang, "Spatial clockwork recurrent neural network for muscle perimysium segmentation," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9901. 2016, pp. 185–193.

[55] L. Roux *et al.*, "Mitosis detection in breast cancer histological images an ICPR 2012 contest," *J. Pathol. Inform.*, vol. 4, no. 1, pp. 1–8, May 2013.

[56] F. Xing and L. Yang, "Fast cell segmentation using scalable sparse manifold learning and affine transform-approximated active contour," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 332–339.

[57] F. Xing, X. Shi, Z. Zhang, J. Cai, Y. Xie, and L. Yang, "Transfer shape modeling towards high-throughput microscopy image segmentation," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 9902. 2016, pp. 183–190.

[58] Y. Mao, Z. Yin, and J. M. Schober, "Iteratively training classifiers for circulating tumor cell detection," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 190–194.

[59] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "Neutrophils identification by deep learning and Voronoi diagram of clusters," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 226–233.

[60] B. Dong, L. Shao, M. Da Costa, O. Bandmann, and A. F. Frangi, "Deep learning for automatic cell detection in wide-field microscopy zebrafish images," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 772–776.

[61] A. Shkolyar, A. Gefen, D. Benayahu, and H. Greenspan, "Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2015, pp. 743–746.

[62] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[63] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1324–1332.

[64] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1273–1284, May 2016.

[65] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 140–145, Sep. 2012.

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

[67] A. Fakhry, H. Peng, and S. Ji, "Deep models for brain EM image segmentation: Novel insights and improved performance," *Bioinformtics*, vol. 32, no. 15, pp. 2352–2358, 2016.

[68] T. Liu, M. Seyedhosseini, M. Ellisman, and T. Tasdizen, "Watershed merge forest classification for electron microscopy image stack segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4069–4073.

[69] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, "Toward automatic phenotyping of developing embryos from videos," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1360–1371, Sep. 2005.

[70] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.

[71] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[72] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 55–58.

[73] M. Sapkota, F. Xing, H. Su, and L. Yang, "Automatic muscle perimysium annotation using deep convolutional neural network," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 205–208.

[74] H. Su, F. Xing, X. Kong, Y. Xie, S. Zhang, and L. Yang, "Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 383–390.

[75] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 234–241.

[76] A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 766–774.

[77] M. Maška *et al.*, "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, Feb. 2014.

[78] H. Chen, X. Qi, J. Cheng, and P. A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1167–1173.

[79] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. 2007, pp. 153–160.

[80] C. Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. 18th Int. Conf. Artif. Intell. Statist.*, 2015, pp. 562–570.

[81] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 690–703, Mar. 2017.

[82] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai, "Deep direct reinforcement learning for financial signal representation and trading," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 653–664, Mar. 2017.

[83] Y. Zhang, R. Yamaguchi, S. Imoto and S. Miyano, "Sequence-specific bias correction for RNA-seq data using recurrent neural networks," *BMC Genomics*, vol. 18, no. 1, pp. 1–6, 2017.

[84] S. Hochreiter, M. Heusel, and K. Obermayer, "Fast model-based protein homology detection without alignment," *Bioinformatics*, vol. 23, no. 14, pp. 1728–1736, 2007.

[85] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.

[86] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28. 2015, pp. 2980–2988.

[87] J. Koutník, K. Greff, F. Gomez, and J. Schmidhuber, "A clockwork RNN," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 32. 2014, pp. 1863–1871.

[88] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[89] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[90] M. Ranzato, Y. Boureau, and Y. LeCun, "Sparse feature learning for deep belief networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 20. 2007, pp. 1185–1192.

[91] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 cell image classification with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 416–428, Mar. 2017.

[92] C. L. Chen *et al.*, "Deep learning in label-free cell classification," *Sci. Rep.*, vol. 6, Mar. 2016, Art. no. 21471.

[93] G. Carneiro, T. Peng, C. Bayer, and N. Navab, "Weakly-supervised structured output learning with flexible and latent graphs using high-order loss functions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 648–656.

[94] Y. Xu, Z. Jia, Y. Ai, F. Zhang, M. Lai, and E. I.-C. Chang, "Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 947–951.

[95] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[96] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 1–8.

[97] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I.-C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 1626–1630.

[98] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 10. 1998, pp. 570–576.

[99] P. A. Viola, J. C. Platt, and C. Zhang, "Multiple instance boosting for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18. 2007, pp. 1417–1426.

[100] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 403–410.

[101] K. Kavukcuoglu, M. Ranzato, and Y. LeCun. (2010). "Fast inference in sparse coding algorithms with applications to object recognition." [Online]. Available: https://arxiv.org/abs/1010.3467

[102] H. Chang, N. Nayak, P. T. Spellman, and B. Parvin, "Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching," in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 91–98.

[103] H. Chang, Y. Zhou, P. Spellman, and B. Parvin, "Stacked predictive sparse coding for classification of distinct regions of tumor histopathology," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 169–176.

[104] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. 1993, pp. 40–44.

[105] T. Nitta, "Resolution of singularities introduced by hierarchical structure in deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2282–2293, Oct. 2017.

[106] M. Bianchini and F. Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 8, pp. 1553–1565, Aug. 2014.

[107] L. Szymanski and B. McCane, "Deep networks are effective encoders of periodicity," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1816–1827, Oct. 2014.

[108] N. Le Roux and Y. Bengio, "Representational power of restricted Boltzmann machines and deep belief networks," *Neural Comput.*, vol. 20, no. 6, pp. 1631–1649, Jun. 2008.

[109] I. Sutskever and G. E. Hinton, "Deep, narrow sigmoid belief networks are universal approximators," *Neural Comput.*, vol. 20, no. 11, pp. 2629–2636, Nov. 2008.

[110] C.-H. Chang, "Deep and shallow architecture of multilayer neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2477–2486, Oct. 2015.

[111] P. P. Brahma, D. Wu, and Y. She, "Why deep learning works: A manifold disentanglement perspective," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 1997–2008, Oct. 2016.

[112] L. Shao, D. Wu, and X. Li, "Learning deep and wide: A spectral method for learning deep networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2303–2308, Dec. 2014.

[113] J. Li, T. Zhang, W. Luo, J. Yang, X.-T. Yuan, and J. Zhang, "Sparseness analysis in the pretraining of deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 6, pp. 1425–1438, Jun. 2017.

[114] M. Gong, J. Liu, H. Li, Q. Cai, and L. Su, "A multiobjective sparse feature learning model for deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3263–3277, Dec. 2015.

[115] D. Erhan, A. Courville, and Y. Bengio, "Understanding representations learned in deep architectures," Univ. Montréal/DIRO, Montreal, QC, Canada, Tech. Rep. 1355, 2010.

[116] C. Szegedy *et al.* (2014). "Intriguing properties of neural networks." [Online]. Available: https://arxiv.org/abs/1312.6199

[117] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5188–5196.

[118] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 818–833.

[119] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. Int. Conf. Learn. Represent. Workshop*, 2014, pp. 1–8.

[120] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, p. e0130140, 2015.

[121] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017.

[122] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 265–272.

[123] Y. Jia *et al.* (2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/1408.5093

[124] Theano Development Team. (2016). "Theano: A Python framework for fast computation of mathematical expressions." [Online]. Available: https://arxiv.org/abs/1605.02688

[125] M. Abadi *et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Accessed: 2015. [Online]. Available: http://tensorflow.org/

[126] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A MATLAB-like environment for machine learning," in *Proc. BigLearn, NIPS Workshop*, 2011, pp. 1–6.

[127] J. Dean *et al.*, "Large scale distributed deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25. 2012, pp. 1223–1231.

[128] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 873–880.

[129] M. Li *et al.*, "Scaling distributed machine learning with the parameter server," in *Proc. 11th USENIX Symp. Oper. Syst. Design Implement.*, Oct. 2014, pp. 583–598.

[130] B. Recht, C. Re, S. Wright, and F. Niu, "Hogwild: A lock-free approach to parallelizing stochastic gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24. 2011, pp. 693–701.

[131] A. Coates, B. Huval, T. Wang, D. J. Wu, A. Y. Ng, and B. Catanzaro, "Deep learning with COTS HPC systems," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 1337–1345.

[132] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA," *Queue*, vol. 6, no. 2, pp. 40–53, Mar. 2008.

[133] S. Chetlur *et al.* (2014). "cuDNN: Efficient primitives for deep learning." [Online]. Available: https://arxiv.org/abs/1410.0759

[134] S. Young, J. Lu, J. Holleman, and I. Arel, "On the impact of approximate computation in an analog DeSTIN architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 934–946, May 2014.

[135] C. Farabet, C. Poulet, and Y. LeCun, "An FPGA-based stream processor for embedded real-time vision with convolutional networks," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops*, Sep. 2009, pp. 878–885.

[136] C. Yan *et al.*, "Efficient parallel framework for HEVC motion estimation on many-core processors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2077–2089, Dec. 2014.

[137] C. Yan *et al.*, "A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 573–576, May 2014.

[138] C. Yan, Y. Zhang, F. Dai, X. Wang, L. Li, and Q. Dai, "Parallel deblocking filter for HEVC on many-core processor," *Electron. Lett.*, vol. 50, no. 5, pp. 367–368, Feb. 2014.

[139] C. Yan, Y. Zhang, F. Dai, J. Zhang, L. Li, and Q. Dai, "Efficient parallel HEVC intra-prediction on many-core processor," *Electron. Lett.*, vol. 50, no. 11, pp. 805–806, May 2014.

[140] C. Yan, Y. Zhang, F. Dai, and L. Li, "Highly parallel framework for HEVC motion estimation on many-core platform," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2013, pp. 63–72.

[141] G. Lacey, G. W. Taylor, and S. Areibi. (2016). "Deep learning on FPGAs: Past, present, and future." [Online]. Available: https://arxiv.org/abs/1602.04283

[142] A. Dundar, J. Jin, B. Martini, and E. Culurciello, "Embedded streaming deep neural networks accelerator with applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1572–1583, Jul. 2017.

[143] C. E. Cox and W. E. Blanz, "GANGLION—A fast field-programmable gate array implementation of a connectionist classifier," *IEEE J. Solid-State Circuits*, vol. 27, no. 3, pp. 288–299, Mar. 1992.

[144] K. Paul and S. Rajopadhye, "Back-propagation algorithm achieving 5 GOPS on the Virtex-E," in *FPGA Implementations of Neural Networks*. Boston, MA, USA: Springer, 2006, pp. 137–165.

[145] K. Ovtcharov, O. Ruwase, J.-Y. Kim, J. Fowers, K. Strauss, and E. S. Chung, "Accelerating deep convolutional neural networks using specialized hardware," Microsoft Res., Redmond, WA, USA, White Paper 2, 2015.

[146] C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, "Optimizing FPGA-based accelerator design for deep convolutional neural networks," in *Proc. ACM/SIGDA Int. Symp. Field-Programmable Gate Arrays (FPGA)*, 2015, pp. 161–170.

[147] J. Qiu *et al.*, "Going deeper with embedded FPGA platform for convolutional neural network," in *Proc. ACM/SIGDA Int. Symp. Field-Programmable Gate Arrays (FPGA)*, 2016, pp. 26–35.

[148] C. Farabet, B. Martini, P. Akselrod, S. Talay, Y. LeCun, and E. Culurciello, "Hardware accelerated convolutional neural networks for synthetic vision systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, May/Jun. 2010, pp. 257–260.

[149] B. Reagen *et al.*, "Minerva: Enabling low-power, highly-accurate deep neural network accelerators," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit.*, Jun. 2016, pp. 267–278.

[150] Y.-H. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan./Feb. 2016, pp. 262–263.

[151] E. Stromatias, D. Neil, F. Galluppi, M. Pfeiffer, S.-C. Liu, and S. Furber, "Scalable energy-efficient, low-latency implementations of trained spiking deep belief networks on SpiNNaker," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2015, pp. 1–8.

[152] S. Han *et al.*, "EIE: Efficient inference engine on compressed deep neural network," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit.*, Jun. 2016, pp. 243–254.

[153] Google. (2016). *Tensor Processing Unit*. https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html

[154] R. Krishna *et al.* (2016). "Visual genome: Connecting language and vision using crowdsourced dense image annotations." [Online]. Available: https://arxiv.org/abs/1602.07332

[155] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 3729–3759, Jan. 2016.

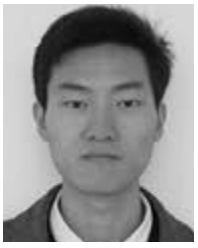
[156] D. Geman, S. Geman, N. Hallonquist, and L. Younes, "Visual Turing test for computer vision systems," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 12, pp. 3618–3623, 2015.

[157] M. Malinowski and M. Fritz, "A multi-world approach to question answering about real-world scenes based on uncertain input," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27. 2014, pp. 1682–1690.
[158] K. Tu, M. Meng, M. W. Lee, T. E. Choe, and S.-C. Zhu, "Joint video and text parsing for understanding events and answering queries," *IEEE Multimedia*, vol. 21, no. 2, pp. 42–70, Apr./Jun. 2014.
[159] S. Antol *et al.*, "VQA: Visual question answering," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2425–2433.
[160] M. Malinowski, M. Rohrbach, and M. Fritz, "Ask your neurons: A neural-based approach to answering questions about images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1–9.
[161] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Med. Image Anal.*, vol. 33, pp. 170–175, Oct. 2016.
[162] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang, "Pathology imaging informatics for quantitative analysis of whole-slide images," *J. Amer. Med. Inform. Assoc.*, vol. 20, no. 6, pp. 1099–1108, 2013.
[163] H.-S. Wu *et al.*, "Restoration of distorted colour microscopic images from transverse chromatic aberration of imperfect lenses," *J. Microsc.*, vol. 241, no. 2, pp. 125–131, Feb. 2011.
[164] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014.
[165] S. Kothari, J. H. Phan, and M. D. Wang, "Scale normalization of histopathological images for batch invariant cancer diagnostic models," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2012, pp. 4406–4409.
[166] B. E. Bejnordi *et al.*, "Stain specific standardization of whole-slide histopathological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 404–415, Feb. 2016.
[167] A. Janowczyk, A. Basavanhally, and A. Madabhushi, "Stain normalization using sparse AutoEncoders (StaNoSA): Application to digital pathology," *Comput. Med. Imag. Graph.*, vol. 57, pp. 50–61, Apr. 2017.
[168] F. Ciompi *et al.*, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 160–163.
[169] F. Xing and L. Yang, "Robust selection-based sparse shape model for lung cancer image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, vol. 8151. 2013, pp. 404–412.
[170] TCGA Research Network. (2017). *The Cancer Genome Altas*. [Online]. Available: http://cancergenome.nih.gov/
[171] J. Lonsdale *et al.*, "The genotype-tissue expression (GTEx) project," *Nature Genet.*, vol. 45, no. 6, pp. 580–585, 2013.
[172] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3642–3649.
[173] J. Weese and C. Lorenz, "Four challenges in medical image analysis from an industrial perspective," *Med. Image Anal.*, vol. 33, pp. 44–49, Oct. 2016.
[174] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, Jun. 2006.
[175] E. Estellés-Arolas, F. González-Ladrón-De-Guevara, "Towards an integrated crowdsourcing definition," *J. Inf. Sci.*, vol. 38, no. 2, pp. 189–200, Apr. 2012.
[176] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[177] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, Feb. 2010.
[178] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[179] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.
[180] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 596–608, Apr. 2012.
[181] H. Goh, N. Thome, M. Cord, and J.-H. Lim, "Learning deep hierarchical visual feature coding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2212–2225, Dec. 2014.
[182] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27. 2014, pp. 3320–3328.
[183] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[184] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, "Selective sampling using the query by committee algorithm," *Mach. Learn.*, vol. 28, no. 2, pp. 133–168, 1997.
[185] M. Lin, K. Tang, and X. Yao, "Dynamic sampling approach to training neural networks for multiclass imbalance classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 4, pp. 647–660, Apr. 2013.
[186] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. 17th Int. Joint Conf. Artif. Intell.*, vol. 2. 2001, pp. 973–978.
[187] M. Kukar and I. Kononenko, "Cost-sensitive learning with neural networks," in *Proc. 13th Eur. Conf. Artif. Intell.*, Aug. 1998, pp. 445–449.
[188] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.
[189] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 888–899, Jun. 2013.
[190] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[191] H. He and Y. Ma, Eds., *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed. New York, NY, USA: Wiley, 2013.
[192] D. C. Paech *et al.*, "A systematic review of the interobserver variability for histology in the differentiation between squamous and nonsquamous non-small cell lung cancer," *J. Thoracic Oncol.*, vol. 6, no. 1, pp. 55–63, Jan. 2011.
[193] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2424–2433.
[194] B. Korbar *et al.*, "Deep learning for classification of colorectal polyps on whole-slide images," *J. Pathol. Inform.*, vol. 8, no. 1, p. 30, 2017.
[195] X. Zhu, J. Yao, F. Zhu, and J. Huang, "WSISA: Making survival prediction from whole slide histopathological images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 7234–7242.
[196] A. D. Ward *et al.*, "Prostate: Registration of digital histopathologic images to *in vivo* MR images acquired by using endorectal receive coil," *Radiol.*, vol. 263, no. 3, pp. 856–864, 2012.
[197] R. S. Savage and Y. Yuan, "Predicting chemoinsensitivity in breast cancer with 'omics/digital pathology data fusion," *Roy. Soc. Open Sci.*, vol. 3, no. 2, p. 140501, 2016.
[198] T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth, and G. Langs, "Predicting semantic descriptions from medical images with convolutional neural networks," in *Proc. 24th Int. Conf. Inf. Process. Med. Imag.*, 2015, pp. 437–448.
[199] Z. Zhang, P. Chen, M. Sapkota, and L. Yang, "TandemNet: Distilling knowledge from medical images using diagnostic reports as optional semantic references," in *Proc. 20th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 320–328.
[200] Z. Zhang, Y. Xie, F. Xing, M. McGough, and L. Yang, "MDNet: A semantically and visually interpretable medical image diagnosis network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6428–6436.
[201] H. Wang *et al.*, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. Med. Imag.*, vol. 1, no. 3, p. 034003, 2014.
[202] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[203] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[204] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. Int. Conf. Learn. Represent.*, 2013.
[205] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[206] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *Proc. 20th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 417–425.

[207] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. 25th Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 146–157.

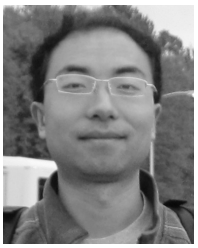
Fuyong Xing received the bachelor's degree from Xi'an Jiaotong University, Xi'an, China, the M.S. degree from Rutgers University, New Brunswick, NJ, USA, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2017.

He is currently an Assistant Professor with the Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado at Denver, Denver, CO, USA. His current research interests include biomedical image computing, imaging informatics, computer vision, machine learning, and deep learning. He is also involved in high performance computing for biomedical imaging informatics.



Yuanpu Xie received the bachelor's degree from the Dalian University of Technology, Dalian, China, in 2013. He is currently pursuing the Ph.D. degree in biomedical engineering from the University of Florida, Gainesville, FL, USA.

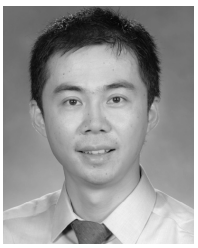
He has authored over ten peer-reviewed journal and conference proceeding articles. His current research interests include medical image analysis, biomedical imaging informatics, machine learning, and deep learning.



Hai Su received the B.E. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 2003 and 2006, respectively, and the M.S. degree in computer science from the University of Kentucky, Lexington, KY, USA. He is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA.

His current research interests include deep learning, imaging informatics, and biomedical image analysis.

Fujun Liu received the B.S. degree in communication engineering from Shandong University, Jinan, China, in 2007, the M.S. degree in information and communication engineering from the Chinese Academy of Sciences, Beijing, China, in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2017, where he is involved in medical image processing, computer vision, machine learning, and deep learning.



Lin Yang (M'09) is currently an Associate Professor with the J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, USA, where he is also an official affiliated Associate Professor with the Department of Electrical and Computer Engineering and the Department of Computer and Information Science and Engineering. He leads the Biomedical Image Computing and Imaging Informatics Laboratory. His current research interests include biomedical image analysis and imaging informatics, computer vision, biomedical informatics, and machine learning. He is also involved in high performance computing and computed aided health care and information technology using big data.

Dr. Yang is the co-author of the award winning papers for the 2008 ISBI NIH Young Investigator Best Paper and Travel Award, the 2014 NANETS Young Investigator Paper and Travel Award, the 2015 Young Scientist Best Paper Award Runner-up, and the 2015 MICCAI Young Scientist Best Paper Award.