Chapter 17
Deep Learning with Convolutional Neural Networks for Histopathology Image Analysis

Dragan Bošnački, Natal van Riel and Mitko Veta

Abstract In the recent years, deep learning based methods and, in particular, convolutional neural networks, have been dominating the arena of medical image analysis. This has been made possible both with the advent of new parallel hardware and the development of efficient algorithms. It is expected that future advances in both of these directions will increase this domination. The application of deep learning methods to medical image analysis has been shown to significantly improve the accuracy and efficiency of the diagnoses. In this chapter, we focus on applications of deep learning in microscopy image analysis and digital pathology, in particular. We provide an overview of the state-of-the-art methods in this area and exemplify some of the main techniques. Finally, we discuss some open challenges and avenues for future work.

17.1 Introduction

Imaging is an essential component of modern medicine. Routine clinical care results in large quantities of image data that is analyzed by medical experts for decisions regarding diagnosis and prognosis. However, these analyses are generally labor-intensive, subjective, and expensive tasks. The goal of medical image analysis is to develop automated methods that will enable a faster, more reliable, and quantitative analysis of medical images. At its core, medical image analysis is an interdisciplinary field combining knowledge from several disciplines, including medicine, physics, mathematics, and computer science.

D. Bošnački (✉) · N. van Riel · M. Veta
Biomedical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: D.Bosnacki@tue.nl

N. van Riel
e-mail: N.A.W.v.Riel@tue.nl

M. Veta
e-mail: M.Veta@tue.nl

P. Liò and P. Zuliani (eds.), *Automated Reasoning for Systems Biology and Medicine*, Computational Biology 30,
https://doi.org/10.1007/978-3-030-17297-8_17

In the coming decade, artificial intelligence will affect many aspects of human life and society ranging from home automation and self-driving cars to algorithmic detection of "fake news" on social media. This technology will also inevitably influence health care and in turn medical imaging. One of the most active fields of research in artificial intelligence is deep machine learning. This concept works by learning abstract, hierarchical feature representations of input data that are predictive of a certain target (e.g., a class label in the case of classification). In contrast to classical machine learning methods, deep learning eliminates the step of "manual" feature engineering that is a tedious process and suboptimal compared with learned representations.

In the field of medical image analysis, the use of deep learning methods has already resulted in notable progress for a variety of tasks. Some of the first applications where automated methods were shown to outperform medical experts have been in histopathology image analysis. In a recent work [5], it was demonstrated that state-of-the-art deep convolutional neural networks can outperform pathologists in detecting metastases in sentinel lymph nodes of breast cancer patients.

In recent years, methods based on Convolutional Neural Networks (CNNs) have been showing great benefits and offering promising perspectives. This remarkable progress was made possible with the advent of highly parallel processors, like General Purpose Graphics Processing Units (GPUs). Thanks to this massive parallelism for the masses, it became feasible to revive some old concepts as well as develop new approaches in CNNs. Another factor that boosted the deep learning, in general, was the massive data gathered throughout the years.

Here, we focus mostly on methods based on convolutional neural networks as a dominant image analysis concept. The first ideas on convolutional networks can be traced to Fukushima [19, 20] in the beginning of the 80s and the first application in medical image analysis is in the 90s [36]. The first landmark in the success story of CNNs is the 1998 paper by LeCun et al. [33] on applying CNNs for recognition of handwritten digits. Still another decade needed to pass and the emergence of the new parallel hardware to happen in order AlexNet by Krizhevsky et al. [31] to appear. With AlexNet they won the ImageNet competition [31], which triggered applications of convolutional networks in many areas including their rediscovery in medical image analysis. Since then an enormous proliferation of publications has happened, with CNN-based methods for medical image analysis dominating conferences and dedicated workshops.

In this work, we give an overview of various methods and applications of convolutional neural networks in digital pathology and microscopy. Considering the fact that similar reviews have appeared before, e.g., [35, 41] including some more general reviews that appeared while this work was already accepted for publication, like [48], our paper focuses on some most recent work, mostly in the last 3 years. This overview does not attempt to achieve a comprehensive cover of all new papers that have appeared after these reviews. Rather we have made a selection of works that have been the most interesting for us and relevant for our own research, but which hopefully can still be of a broader importance. For the papers that have been included in the previous reviews, we tried to give a more elaborate treatment as well as provide a view from another angle on the treated subjects.

Automated image analysis with artificial-intelligence-based methods will undoubtedly lead to better health care because of the more objective evaluation, standardization in quality, and reduction in errors. However, even with the recent advances and success of this methodology applied to medical images, important research questions still remain and need to be addressed. Most of the past efforts have been focused on automating routine image analysis procedures, such as the evaluation of known imaging biomarkers. While this is an important and still relevant application area, the dominantly data-driven nature of deep learning methods offers the possibility for discovery of novel imaging biomarkers. This can be achieved by training deep learning models that predict patient outcomes such as response to therapy and survival directly from the image data, without relying of prior knowledge. Furthermore, in order to translate these methods to the clinic, further improvements in the robustness across data sets from different centers are needed. We discuss these and some other avenues for future work and identify the most important challenges.

17.2 Preliminaries

In this section, we review the basic concepts related to neural networks and, in particular, convolutional neural networks.

Artificial neural networks are parallel distributed computational models. They consist of artificial neurons arranged in layers and connected with multiple other neurons. In recent decade, mostly due to the emergence of powerful parallel processor architectures, in particular, the General Purpose Graphics Processing Units (GPUs), neural networks consisting of thousands of neurons became feasible. The new developments allowed also networks with multiple layers which lead to the name "deep learning" to distinguish them from the previous "shallow" neural networks which usually featured one or two hidden layers.

Artificial neurons are the basic building blocks of the neural networks. An artificial neuron is a computing element inspired by the natural neurons that have the ability to react to external input signals (stimuli) and adapt its behavior depending on the environment. The weighted cumulative effect of the input signals is processed, for instance, by comparing it to a threshold.

More formally, we assume an input vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. The elements x_i are usually referred to as features. A neuron is defined with its parameters set $\theta = (\mathbf{w}, b)$ and a transfer function z. The first parameter $\mathbf{w} = (w_1, w_2, \ldots, w_p)$ is a vector of weights associated to each of the inputs, whereas b is a so-called bias.[1] The transfer function is obtained by combining an activation function with a nonlinear output function. The activation function is a linear combination of the input and the parameters, whereas the output function σ is a nonlinearity, for instance, a sigmoidal

[1] Actually b can be considered as a special weight w_0 associated with a special input x_0 which has a constant value 1. In this way the transfer function becomes slightly simpler $\sigma(\mathbf{w}^T\mathbf{x})$. However, for the sake of clarity here we keep these two parameters separately.

function like $1/(1+e^{-x})$. So, the usual form of the transfer function is

$$z(\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + b)$$

Such a neuron can be used a classifier which can distinguish between two classes of object, e.g., by comparing the output value z with a threshold.

In general, however we might have more than two classes. Therefore, we need a classifier which implements a function $f : \mathbb{R}^p \to \mathbb{R}^k$ which takes the input vector $\mathbf{x} \in \mathbb{R}^p$ as argument and returns an output vector $\mathbf{y} = (y_1, y_2, \ldots, y_k) \in \mathbb{R}^k$, where each element y_i corresponds to one of the k classes. The value of the element y_i of $\mathbf{y}$ can be seen as a confidence score that the input $\mathbf{x}$ belongs to class i. The output score can be processed further, for instance, to obtain a probability distribution $P(\mathbf{y} \mid \mathbf{x})$ for $\mathbf{x}$ over the classes.

Such classifiers can be implemented in a straightforward way by assigning one neuron per class and feeding the same input $\mathbf{x}$ to each of the neurons. Each neuron has a separate parameter set $\theta_i = (\mathbf{w}_i, b_i)$. The weight vectors can be put together into a matrix $\mathbf{W}$, each $\mathbf{w}_i$ being a row of the matrix. Similarly, a bias vector $\mathbf{b}$ is formed.

The drawback of such a simple network consisting of only one layer of neurons is that it can implement only a limited class of functions f, more precisely functions of the form

$$f(\mathbf{x}, \theta) = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

This constraint can be removed by pipelining several layers such that the input of one layer is the output of another. It can be shown that in this way one can implement any "reasonably" behaving function. The above-described network architectures are known as Multilayered Perceptrons (MLP). Hence, MLP consists of several layers connected with one another and implementing the superposition function

$$f(\mathbf{x}, \Theta) = \sigma_l(\mathbf{W}_l\sigma_{l-1}(\mathbf{W}_{l-1}\ldots\sigma_1(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)\ldots) + \mathbf{b}_l)$$

where $\mathbf{W}_i$, $\mathbf{b}_i$, and σ_i, are the corresponding parameters and nonlinearities of each of the layers, respectively, and $\Theta = \{(\mathbf{W}_1, \mathbf{b}_1), \ldots, (\mathbf{W}_l, \mathbf{b}_l)\}$. Multilayered networks are also called *deep* neural networks. The layers which are between the input and output layer are called *hidden* layers. The number of the layers defines the depth of the network.

There are several possibilities to obtain the probability distribution over the classes for a given input. One of the most used functions is *softmax*

$$\frac{e^{\mathbf{w}_i^T + b_i}}{\Sigma_{j=1}^k e^{\mathbf{w}_j^T + b_j}}$$

Such a function is implemented within the very last layer by choosing for the nonlinearity $\sigma(z) = e_i^z/(\Sigma_j e_j^z)$.

Based on difference between the actual and the desired outputs, the neural networks can be trained by adjusting the parameter values in Θ. Depending on the scores (distributions) and the methods, various *loss functions* can be defined. The most popular algorithms to fit the parameters and minimize the loss functions are based on *stochastic gradient descent* with momentum and the Adam optimizer [17, 30].

Convolutional Neural Networks (CNNs/ConvNets) are a class of neural networks which nowadays is the most used in biomedical applications and in particular in image processing. In a nutshell, the main difference between the MLPs and CNNs is that the latter have sparse connectivity and replication of the weights among different neurons. In this way, CNNs require smaller number of parameters than MLPs, and therefore less computational resources while retaining the same expressiveness, i.e., ability to implement a transfer function relating the input and the output vectors.

In the MLP architecture, the neuron layers are dense (fully connected) in the sense that the output of each neuron of a layer serves as input to each neuron of the next layer. In the CNNs, this is not the case. Instead, the neurons of the next layer use only a subset of input elements (neurons) to perform a convolution on the images.

CNNs exploit the fact that the input data, especially in image analysis, is a multidimensional array. A filter (convolution kernel) consisting of a neighborhood of input elements, i.e., a rectangular patch, is slid over the input image. The outputs of the picture elements "covered" by the patch are wired to a neuron in the next layer. The advantage of such floating neighborhoods is that they are not fixed anymore to a certain position in the image. Hence, one particular filter specialized in detecting a type of objects is sufficient to cover the whole image, instead of having multiple copies for each position as this would be the case in the MLPs. In that regard, we say that CNNs are translation invariant and thus able to detect the"object" regardless where and how it is positioned in the image.

For such multidimensional inputs, also called *tensors* or simply *3D volumes*, one can define the operation of *convolution*, denoted as "*", between the input and an analogous smaller structure—the filter. For instance, assuming a 2D input vector $\mathbf{x}$ and a 2D filter $\mathbf{w}$, we can define[2]

$$\mathbf{h}(i, j) = (\mathbf{x} * \mathbf{w})(i, j) = \Sigma_m \Sigma_n \mathbf{x}(i + m, j + m)\mathbf{w}(m, n)$$

Hence, to compute the modified input $\mathbf{h}$, for each original input element at position (i, j), a weighted sum of its $m \times n$ neighborhood is computed, where the weights are given by the filter. In that way, the convolution operation is analogous to the dot product of the input signals with the weights in the traditional neural networks, only in the latter the neighborhood consists of all neurons of the previous layer.[3]

[2] Actually this is a definition of a cross-correlation which is slightly different than the usual mathematical notion of convolution, but in the machine learning practice this is how the convolution operation is implemented [23].

[3] In principle, one can unfold the $m \times n$ covered rectangular patch of the input and the filter into l-dimensional vectors, where $l = m \times n$. In this way, "*" becomes real a dot product between the vectors. Also, a bias element can be added, like in the traditional neural networks.

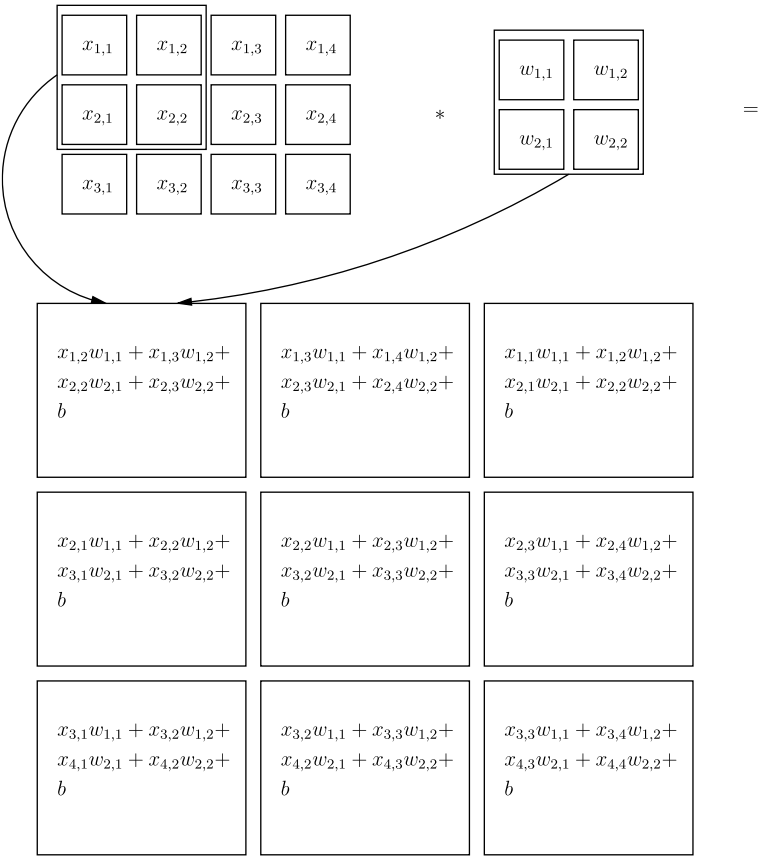

Fig. 17.1 An example of a 2D convolution (adapted from [23]). A 2×2 filter (kernel) is slid over a 3×4 image. To avoid out of range indexing the convolution is computed only when the filter is within the image. As a result, the obtained 2D array has smaller dimensions than the input. (The out of range indexing can be resolved by padding the input array with extra rows and columns, usually having some default value like 0.) Please note that in our example also a bias b is used in the convolution, similarly to the dot product in the traditional networks. Like in this example, in general, the output volumes have smaller width and height than the input arrays. The number of elements in the output array can be further reduced by using a stride parameter, i.e., skipping some of the values of the indices i and j in the sliding. In a similar way, also the max pooling op ration is done. The element-wise product between the input and the filter elements is replaced by the max function. For instance, for the top left entry of the resulting array is obtained as $max(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2})$. The output arrays obtained by convolving each of the filters associated with the convolution layer are stacked together in a 3D array. For instance, if we had five filters, the output array has dimensions $3 \times 3 \times 5$. In general, the dimensions of the output volume are given by the dimensions of the input, the filter, and possibly by the abovementioned stride parameter.

The "sliding" of the filter is achieved by ranging over i and j. In this way, a new transformed input array x' is obtained as illustrated in Fig. 17.1.

The intuition behind the convolution is that, by averaging the neighborhood of an input element, the filter acts as a noise reducing element. By sliding the filter over the input, a "smoothed" version of the input is obtained.

More formally, in a convolution layer, the input is convoluted with a set of filters (kernels) $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$.

With each $\mathbf{w}_i$, a corresponding 2D output is obtained $\mathbf{h_i} = \mathbf{x} * \mathbf{w}_i$, where the convolution operation $*$ is lifted to have the whole input array as argument. The outputs are put together to form a 3D array $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n)$.

The CNN architectures are obtained by using a combination of convolution layers and fully connected layers.[4] The convolutional layers are followed by a nonlinearity layer and optionally by *pooling layers*. The most preferred nonlinear function is the so-called *Rectified Linear Unit (ReLU)* function defined as $\max(0, x)$. The nonlinear function is applied to each element of the output $\mathbf{H}$ of the convolutional layer. The pooling layers are interspersed between series of convolutional (plus ReLU) layers. They also combine the outputs of several related neurons by sliding a small filter over the output of the previous layer (which is usually a ReLU); however, in a simpler way than the convolutional and the dense layers. Usually, a function like maximum is used to approximate the values of a set of input elements. The last layer, which computes the distribution, is a fully connected layer (see also Fig. 17.1). The net effect is that one obtains a transfer function analogous to the one of the traditional networks $\mathbf{h}_k^{l+1} = \sigma^l(\mathbf{H}^l * \mathbf{w}_k^l + \mathbf{b})$, where the subscripts $l, l+1$ denote an association with the l-the and $l+1$-th layer, respectively, and σ^l includes the effects of the nonlinear and pooling layers and $\mathbf{b}$ is the set of biases that are used in the convolution (see also the remarks in Fig. 17.1).

As mentioned above, the inputs of the CNNs are commonly seen as three-dimensional arrays, since for each 2D image element there is also a color (the depth dimension). The filters have the same depth as the original image—for instance, in the first layer this is usually three—but much smaller breadth and width. To obtain the modified input, i.e., the new feature vector, the processed input images are stacked on each other and the new depth of the input is equal to the number n of filters W_i associated with the layer. The breadth and width of the obtained output are usually smaller than in the input volume (see also the explanation in Fig. 17.1).

Recurrent Neural Networks (RNNs) are used for sequence analysis. Mathematically, they learn a distribution $P(y \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ over an input vector sequence instead of a single vector $\mathbf{x}$. In the activation function, besides the input $\mathbf{x}_i$ also the effect of its predecessors is taken into account. The effect of the previous state is recorded in the hidden states which are computed for each input of the sequence.

[4] In recent years, there is growing a trend to use fully convolutional networks in which the fully connected layers are implemented by means of convolutional layers.

17.3 Overview of Digital Pathology Techniques

In this section, we give an overview of different CNN-based models and algorithms for medical image analysis that have appeared mostly in the recent years. As already mentioned in the introduction, this is not an attempt to present a comprehensive overview of the new developments, but rather a selection which to a great extent was shaped by our own research interests. We begin the overview with the grand challenges which have pushed the field forward to a great extent.

17.3.1 Grand Challenges in Digital Pathology

One of the developments that brought about the success of deep learning methods in computer vision was the availability of large, public annotated image data sets. In medical image analysis, this was mirrored in the form of medical image analysis challenges—friendly competitions in which researchers worldwide evaluate their solutions on the same data with the same criteria, in a blinded manner. Organizing a challenge constitutes collecting and publicly distributing a data set that can be used to address a specific clinical task (e.g., assessment of breast tumor proliferation speed). Medical image analysis challenges have been a significant driver for improving the state of the art for a variety of different tasks. The continuation and, more importantly, the improvement in the quality of these efforts will continue to be an essential driver of the medical image analysis field in the future.

The challenges are often associated with one of the major conferences. Among others, such challenges include the 2012 EM segmentation challenge for the 2D segmentation of neuronal process, the ICPR 2012 and AMIDA 2013 mitosis detection challenges, GLAS for gland segmentation, and CAMELYON 16 and TUPAC for processing of breast cancer tissue.

17.3.2 Algorithms

The algorithm developed by IDSIA group won both the ICPR 2012 and AMIDA 13 challenges on mitosis detection with significant margin. In [13], they use deep max-pooling convolutional networks to detect mitosis in breast cancer. In a manner characteristic to the CNN approach, the network is trained to classify each pixel based on the neighboring pixels. The method was evaluated on the public MITOS data set that includes 50 images corresponding to 50 high-power fields in 5 different biopsy slides stained with Hematoxylin and Eosin.

Also, the same group had the best result in the ISBI EM 12 segmentation challenge for the 2D segmentation of neural process [12]. The automated segmentation of neuronal structures depicted in stacks of electron microscopy (EM) images is vital in efficient mapping of 3D brain structure connectivity. Again, each pixel is classified

based on the immediate neighborhood using CNNs. The approach was the only one at the competition which was able to outperform a second human observer.

Xu et al. [50] won the GLAS challenge on gland instance segmentation in colon histology images. The glands need not only be segmented from the background, but they are also identified individually. They proposed the method called Deep Multichannel Side Supervision (DMCS) involving three CNNs. The first one labels the pixels as gland and non-gland and the second one extracts edge information. The final segmentation is produced by the third CNN.

Wang et al. [49] won the CAMELYON16 challenge associated with the International Symposium on Biomedical Imaging (ISBI). This challenge was the first challenge involving whole-slide images (WSI). More precisely, the goal was an automated detection of metastatic breast cancer of WSIs sentinel lymph node biopsies. The challenge was characterized by the contestants using very deep models. Such was also the winning model of Wang et al. which was based on GoogleNet architectures [44]. Combined with human pathologists diagnosis their algorithm was able to increase the pathologist's AUC score and reduce the human error rate for 85%. A later standalone version of the algorithm [6] outperformed the pathologists score.

Paeng et al. [39] was the best-performing method of another WSI challenge Tumor Proliferation Assessment Challenge (TUPAC) associated with MICAI 2016. To predict proliferation in breast cancer, their framework combines three models: a WSI processing component, a mitosis detection based on CNNs, and a Support Vector Machine (SVM)-based proliferation module.

Mobadersany et al. [38] introduces Survival Convolutional Neural Networks (SCCNs) which combine CNNs with traditional survival models. The approach integrates information from histology images with genome biomarkers. The new concept is applied to brain tumor classification and prediction of patient outcomes. The obtained accuracy surpasses that of human experts. The information extracted from the SCNNs can be used to visualize and identify important structures, like microvascular proliferation that is crucial for the prognosis.

In [15], a CNN classifier is used for detecting the extent of invasive breast cancer regions on whole-slide images. Before that, a phase of tissue sampling and tile preprocessing are done. One of the main goals of the study was to evaluate the robustness of the classifiers with regard to various conditions, like staining, preparation, scanning platform, etc. To achieve robustness, the classifiers were trained on samples from different sources. Classifiers trained with different data cohorts showed highly correlated performance measures ($r \geq 0.8$) when tested with an independent data cohort. The approach shows limitations though when applied to in situ breast cancer which is different form the invasive cancer. In general, the presented classifiers are quite successful with more than 70% of positive predictive value and 90% of negative predictive value of pixel-by-pixel evaluation compared to the manually annotated samples.

Bejnordi et al. [7] presents context-aware CNNs for classifications of breast carcinomas in three classes: normal/benign, DCIS, and invasive ductal carcinoma. The approach consists of two stages. In the first one, cellular-level features are extracted using small high-resolution patches. The output of this stage is pipelined with a

fully convolutional network to facilitate predictions in local regions. The incorporation of more context results in improved performance. The obtained accuracy for three-classes classification is above 80%.

It is generally recognized that a successful training of deep learning networks requires many thousands of training samples, e.g., [45]. In image analysis, this is not always possible since the number of images is often in the order of magnitude of tens of hundreds. To alleviate this, problem augmentation techniques are proposed which are able to produce variations of the available training samples. Ronneberger et al. [40] propose a method that strongly relies on data augmentation. Their architecture, called U-Net, builds upon a fully convolutional architecture and features contracting paths which are followed by symmetric expanding paths. The contracting paths use max pooling and are used to capture context. The role of the expanding paths is to enable precise localization. The expanding paths use up-convolution which to certain extent are inverse of the max pooling since they increase the number of pixels. The corresponding opposing convolution and deconvolution layers are connected with skip connections. The networks are applied for various types of segmentation of light microscopy images with good results.

Several derivatives of the U-Net architecture have appeared mostly in the context of 3D image segmentation, like [11] and [37]. The latter introduce V-Net, a 3D variant of the U-Net architecture, which features 3D convolutional layers. Drodzal et al. [18] propose another variation that—inspired by ResNet [25]—adds to the exisiting long skip connections also short skip connections between the up- and downsampling layers.

Two variants of U-Net are evaluated in [14] within a segmentation study of diseased and healthy skin. It was found that Dense Residual U-Net [27] performs better when direct transfer learning is used, without fine-tuning. In densly reisdual U-Nets, the convolution layers are replaced with dense convolutions. In the dense convolutional networks, all convolutional layers are connected with each other. Also, additional residual layers are added [25]. It is observed that when direct training and fine-tuning are used U-Net ouperforms the dense residual networks.

Arbell and Riklin Raviv [4] propose to integrate the U-Net architecture with Convolutional Long Short Term Memory (C-LSTM) [42] with the U-Net. C-LSTM is a convolutional version of the fully connected LSTM [1, 16, 26] which have been developed originally for handling of temporal correlation. To accomodate locally spatial information in image sequences, C-LSTM replaces matrix multiplication with convolutions. This novel approach is applied to individual cells' segmentation from microscopy sequences.

Johnson [28] uses the Mask-RCNN [24] algorithm for segmentation of microscopy images of cell nuclei. Mask-RCNNs predicts a segmentation mask for each region of interest. Mask-RCNN was a convincing winner in the 2016 COCO Challenge [34], a large-scale object detection, segmentation, and captioning challenge. Unlike U-Net, Mask-RCNNs does not need a region proposal (mask). Instead, U-Net uses its encoder–decoder framework in which the convolutional network encodes the content of the network in the first phase and the deconvolutional part to construct the desired segmentation mask. In [28], it was shown that although Mask-RCNNs were

originally developed for applciations in object detection, object localization, and instance segmentation of natural images, with very little modification, can perform very well also segmentation of nuclei in a broad spectrum of microscopy images. Also, possible applications are discussed in segmentation of heart ventricles or liver and tumor segmentation [9].

Bekkers et al. [8] propose a framework that avoids the need for data augmentation by rotation. To this end, they use a generalization of CNNs called *SE(2) group CNNs (G-CNNs)*. SE(2) group is a formal descriptions of combinations of rotations and translations (roto-tranaslations). More precisely, the group describes how two roto-translations are cobmined in one resulting roto-translation. G-CNNs feature special SE(2) group convolutional layers which capture the symmetries of the SE(2) group. The usual translational convolutions with a filter (kernel) are replaced by SE(2) convolutions. The rotational covariance does not need to be learned since it is inherent to the model. Three layers are introduced: lifting layer transforms a 2D image into an SE(2) image, group convolution layer from and to an SE(2)-image, and a projection layer which returns the SE(2) image to a 2D image. The G-CNNs are evaluated experimentally on three different image analysis tasks: histopatology, retinal imaging, and elecron microscopy . The results show improvement of the G-CNNs with regard to the CNNs.

In [3], an approach is presented inspired by the Generative Adversarial Neural Networks (GANs) for cell segmentation. A pair of neural networks play a minimax game which results in segmentation. With such a game approach, a loss function is not needed any more. Another advantage of the approach is that it can work with a limited number of training smaples. The method is evaluated on fluorescent microscopy data, more precisely, on cell segmentation. GANs have recently been applied in microscopy imaging. In [10, 22], they applied in the context of fluorescence microscopy for morphological profiling human cultured cells. The reported results show that GANs are superior to autoencoder-based approaches. In [10], GANs are used for predicting fluorescent labels in unlabeled images over a broad range of labels (nuclei, cell type, cell state).

Lafarge et al. [32] consider an alternative to standard augmentation and normalization approaches—an adversarial approach for mitosis detection in breast cancer histopathology images. Domain-adversarial neural networks (DANN) [21] allow to learn a classification task, while ensuring that the domain of origin of any sample of the training data cannot be recovered from the learned feature representation. The color augmentation combined with DANN showed the best performance.

17.4 Discussion, Conclusions, and Outlook

We gave a selected overview of the recent developments in CNN-based methods for histopatology image analysis. Research on CNNs has seen enormous expansion in the recent years and novel network architectures are produced with a tremendous speed. From the overview, it can be seen that the medical image analysis community

quickly adapts these novel ideas. Although quite often the new architectures can be applied without much modification in the end this still requires fine-tuning of the parameters and often expert insight into the medical application area.

One of the main challenges in medical imaging is the limited amount of data. Unlike standard image analysis for which millions of images can be found on the Internet good training data is usually difficult to find for the medical applications. In particular, well-annotated data is scarce. Another problem is the class imbalance stemming from the fact that usually examples of healthy tissue are much more common. Because of these issues, the importance of data augmentation is increased.

In the future, it would be interesting to combine data obtained from the image analysis with other data characterizing tissues, like genome or metabolome data. Integration of image and molecular data can be beneficial in the elucidation of the disease mechanisms and medicine development. In that context, biomarker detection and discovery will play an important role.

Biomarker detection is defined as automated evaluation of known imaging biomarkers. The goal of applying image analysis is to make the assessment more reliable, quantitative, fast, and accurate. While the application of deep learning methods for biomarker detection has resulted in substantial improvements in accuracy, several important research questions still need to be addressed. This includes development of fast methods that can enable efficient analysis of larger images (up to several terapixels in size in histopathology) and training of models using smaller annotated data sets.

Biomarker discovery refers to the use of deep learning methods for identification of novel, previously unknown image features that are predictive of a certain clinical outcome. The clinical end points that can be used for biomarker discovery such as survival or recurrence, are global, patient-level annotations, i.e., they are not localized to a particular anatomical region or region of interest in the image. This constitutes a so-called weakly labeled learning scenario, and, while techniques that specifically address weakly labeled data have been proposed, it remains an open problem.

The generalization of data from medical centra that were not included in the training data set is crucial for the practical application of the CNN techniques. One of the major obstacles for wide implementation of image analysis methods in the clinic is the poor generalization on data sets from "external" medical centers that were not included in the training set. This drop in the performance can happen due to a variety of reasons, such as use of different imaging equipment and parameters or, in the case of histopathology, differences in the tissue preparation and staining processes. A trivial solution to this problem is to create center-specific data sets and retrain the deep learning models; however, this is not always feasible and can be very expensive. Thus, an important research question that needs to be addressed is the training of deep learning models that are robust w.r.t. such differences.

While ad hoc approaches such as appearance standardization algorithms can lead to some success, it would be preferable to address this issue in a more systematic way. One approach is through domain-adversarial training of deep neural networks. With domain-adversarial training, the optimization procedure is modified in such a way that irrelevant information such as the domain of origin of the training samples is

explicitly removed from the learned representation, which leads to better generalization on new image data sets. In [32], it has been already demonstrated that domain-adversarial approaches can improve the generalization ability for histopathology image analysis tasks.

Another important aspect that needs to be addressed is development mechanisms for continuous training of the models as they are implemented in the clinic and more training data becomes available.

It was already mentioned that the organization of medical image analysis challenges has had a tremendous impact on the advancement of the state of the art in medical image analysis. In the coming decade, such methods will continue to push the field forward. However, it is important that the quality of the challenge data sets keeps improving. Most of all, there needs to be a shift from small and single-center to large and multi-center data sets.

A successful transfer learning, i.e., taking over new CNN architectures that have been developed in other application areas, has been one of the reasons for the success of the CNN-based methods. However, as already mentioned, the basic architecture usually requires some fine-tuning. In this situation, human expert knowledge can be quite valuable. Therefore, it would be of enormous benefit if machine learning and CNN, in particular, can be made easier to use for nonexperts in image processing. There has been an ongoing research on developing online frameworks (c.f. [46]) for "democratization" of machine learning, i.e., to make machine learning methods more accessible. To this end, various techniques like Bayesian learning and genetic algorithms are used on different levels. For instance, to compose workflows of various methods for fine-tuning of the parameters of the methods within the workflow. It is interesting that to this end machine learning is used to improve machine learning and to replace the parameter optimization which is at the moment still a highly empirical exercise based on trial and error and adhocery. Such a prospective automated framework for machine learning can be refined by taking into account the specifics of a particular class of applications, in our case, image analysis.

A tremendous challenge remains to understand how and why neural networks and, in particular, CNNs actually work. Since humans are inherently good in sequential reasoning and algorithms, understanding the highly parallel and distributed models will always be coupled with difficulties. Also, in this direction, the automated trial by the open machine learning frameworks can help us to make some progress.

A related issue which could become quite relevant in the future, especially in the medical applications, is the question of reliability of the CNN-based methods. Many—if not most—of the deep learning methods are based on trial and error and intuition. Defining a rigorous basis for deep learning will help resolve this issue as well as other research questions like the limitations of the neural network models and the interpretability of the results. Interpretability is, in particular, relevant in diagnostics when a "black box" method could be often unsatisfying. Moreover, a deeper understanding of the workings of the neural networks can be helpful in the developing of new architectures. A recent effort toward formal verification can be found [29].

The outcomes of the empirical evaluation in CNNs and machine learning, in general, could be quite dependent on the data sets used for training and benchmarking. Therefore, the generalizability of the results will be a question which will need to be addressed in a more systematic manner in the future, e.g., [43].

In the future, there will be a strong demand for the integration of imaging and other types of information, like clinical data or gene expression studies. Cross-modal CNNs [47] could be a possible answer to this challenge. Cross-modal CNNs generalize the idea of neural network ensembles. The information is partitioned over several smaller CNNs. In [47], it is shown that CNNs can be used to cope with situations when "big data" is not available. Instead, the"breadth" of the data is used. An example of such situation is clinical studies when not many patients might be available, but the abundant heterogeneous information is distributed over the component CNNs.

Acknowledgements The authors would like to thank the anonymous reviewers as well as Stojan Trajanovski for their comments and suggestions that contributed to the final version of this paper.

References

1. 15th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2015, Seoul, South Korea, 3–5 November 2015. IEEE (2015). http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7349033
2. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015. IEEE Computer Society (2015). http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7293313
3. Arbelle A, Raviv TR (2018) Microscopy cell segmentation via adversarial neural networks. In: 15th IEEE International symposium on biomedical imaging, ISBI 2018, Washington, DC, USA, 4–7 April 2018. IEEE, pp 645–648. https://doi.org/10.1109/ISBI.2018.8363657
4. Arbelle A, Raviv TR (2018) Microscopy cell segmentation via convolutional LSTM networks. CoRR. arXiv:1895.11247
5. Ehteshami Bejnordi B, Veta M, Johannes van Diest P et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318(22):2199–2210. https://doi.org/10.1001/jama.2017.14585
6. Bejnordi BE, Litjens GJS, Timofeeva N, Otte-Holler I, Homeyer A, Karssemeijer N, van der Laak JAWM (2016) Stain specific standardization of whole-slide histopathological images. IEEE Trans Med Imag 35(2):404–415. https://doi.org/10.1109/TMI.2015.2476509
7. Bejnordi BE, Zuidhof G, Maschenka Balkenhol MH, Bult P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak J (2017) Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J Med Imag 4:4–8. https://doi.org/10.1117/1.JMI.4.4.044504
8. Bekkers EJ, Lafarge MW, Veta M, Eppenhof KAJ, Pluim JPW, Duits R (2018) Roto-translation covariant convolutional networks for medical image analysis. CoRR. arXiv:1804.03393
9. Christ PF, Ettlinger F, Grün F, Elshaer MEA, Lipková J, Schlecht S, Ahmaddy F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Hofmann F, D'Anastasi M, Ahmadi S, Kaissis G, Holch J, Sommer WH, Braren R, Heinemann V, Menze BH (2017) Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. CoRR arXiv:1702.05970
10. Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G, Lipnick S, Mount E, O'Neil A, Shah K, Lee AK, Goyal P, Fedus W, Poplin R, Esteva A, Berndl M, Rubin

LL, Nelson P, Finkbeiner S (2018) In silico labeling: predicting fluorescent labels in unlabeled images. Cell 173(3):792–803.e19. https://doi.org/10.1016/j.cell.2018.03.040, http://www.sciencedirect.com/science/article/pii/S0092867418303647
11. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3d u-net: learning dense volumetric segmentation from sparse annotation. CoRR. arXiv:1606.06650
12. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J (2012) Deep neural networks segment neuronal membranes in electron microscopy images. In: Bartlett PL, Pereira FCN, Burges CJC, Bottou L, Weinberger KO (eds.) Advances in Neural information processing systems 25: 26th annual conference on neural information processing systems 2012. Proceedings of a meeting held 3–6 December 2012, Lake Tahoe, Nevada, United States, pp 2852–2860. http://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images
13. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J (2013) Mitosis detection in breast cancer histology images with deep neural networks. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N (eds.) Medical image computing and computer-assisted intervention - MICCAI 2013 - 16th international conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II, Lecture Notes in Computer Science, vol 8150. Springer, pp 411–418. https://doi.org/10.1007/978-3-642-40763-5_51
14. Codella NCF, Anderson D, Philips T, Porto A, Massey K, Snowdon J, Feris RS, Smith JR (2018) Segmentation of both diseased and healthy skin from clinical photographs in a primary care setting. CoRR. arXiv:1804.05944
15. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, Tomaszewski J, González FA, Madabhushi A (2017) Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Scientif Rep 7:46450 EP. https://doi.org/10.1038/srep46450
16. Donahue J, Hendricks LA, Guadarrama S, Rohrbach M, Venugopalan S, Darrell T, Saenko K (2015) Long-term recurrent convolutional networks for visual recognition and description. In: IEEE conference on computer vision and pattern recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015, vol 2, pp 2625–2634. https://doi.org/10.1109/CVPR.2015.7298878
17. Dozat T (2015) Incorporating nesterov momentum into adam
18. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C (2016) The importance of skip connections in biomedical image segmentation. CoRR. arXiv:1608.04117
19. Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybernet 36:193–202
20. Fukushima K, Miyake S (1982) Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recogn 15(6):455–469. http://www.sciencedirect.com/science/article/B6V14-48MPJ6Y-F7/2/2588c38bc16488ae94fe2334068ed166
21. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky VS (2016) Domain-adversarial training of neural networks. J Mach Learn Res 17, 59:1–59:35. http://jmlr.org/papers/v17/15-239.html
22. Goldsborough P, Pawlowski N, Caicedo JC, Singh S, Carpenter A (2017) Cytogan: generative modeling of cell images. bioRxiv. https://doi.org/10.1101/227645, https://www.biorxiv.org/content/early/2017/12/02/227645
23. Goodfellow IJ, Bengio Y, Courville AC (2016) Deep learning: adaptive computation and machine learning. MIT Press. http://www.deeplearningbook.org/
24. He K, Gkioxari G, Dollár P, Girshick RB (2017) Mask R-CNN. In: IEEE international conference on computer vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp 2980–2988. IEEE Computer Society. https://doi.org/10.1109/ICCV.2017.322
25. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. CoRR. arXiv:1512.03385
26. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

27. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017. IEEE Computer Society, pp 2261–2269. https://doi.org/10.1109/CVPR.2017.243
28. Johnson JW (2018) Adapting mask-rcnn for automatic nucleus segmentation. CoRR. arXiv:1805.00500
29. Katz G, Barrett C, Dill DL, Julian K, Kochenderfer MJ (2017) Towards proving the adversarial robustness of deep neural networks. In: Bulwahn L, Kamali M, Linker S (eds.) Proceedings first workshop on formal verification of autonomous vehicles, FVAV@iFM 2017, Turin, Italy, 19th September 2017. EPTCS, vol 257, pp 19–26. https://doi.org/10.4204/EPTCS.257.3
30. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. CoRR arXiv:1412.6980
31. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th international conference on neural information processing systems, vol 1, NIPS'12, pp 1097–1105. Curran Associates Inc., USA. http://dl.acm.org/citation.cfm?id=2999134.2999257
32. Lafarge MW, Pluim JPW, Eppenhof KAJ, Moeskops P, Veta M (2017) Domain-adversarial neural networks to address the appearance variability of histopathology images. CoRR. arXiv:1707.06183
33. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.138.1115, http://www.cs.berkeley.edu/daf/appsem/Handwriting/papers/00726791.pdf
34. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer Vision - ECCV 2014. Springer International Publishing, Cham, pp 740–755
35. Litjens GJS, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Imag Anal 42:60–88
36. Lo SCB, Lou SLA, Lin JS, Freedman MT, Chien MV, Mun SK (1995) Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imag 14(4):711–718. https://doi.org/10.1109/42.476112
37. Milletari F, Navab N, Ahmadi S (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. CoRR arXiv:1606.04797
38. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, Brat DJ, Cooper LAD (2018) Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Nat Acad Sci 115(13):E2970–E2979. https://doi.org/10.1073/pnas.1717139115, http://www.pnas.org/content/115/13/E2970
39. Paeng K, Hwang S, Park S, Kim M (2017) A unified framework for tumor proliferation score prediction in breast histopathology. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood TF, Tavares JMRS, Moradi M, Bradley AP, Greenspan H, Papa JP, Madabhushi A, Nascimento JC, Cardoso JS, Belagiannis V, Lu Z (eds.) Deep learning in medical image analysis and multimodal learning for clinical decision support - Third international workshop, DLMIA 2017, and 7th international workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings, Lecture Notes in Computer Science, vol 10553. Springer, pp 231–239. https://doi.org/10.1007/978-3-319-67558-9_27
40. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. CoRR arXiv:1505.04597
41. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/28301734 [pmid]
42. Shi X, Chen Z, Wang H, Yeung D, Wong W, Woo W (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. CoRR. arXiv:1506.04214
43. Sundermann B, Feder S, Wersching H, Teuber A, Schwindt W, Kugel H, Heindel W, Arolt V, Berger K, Pfleiderer B (2017) Diagnostic classification of unipolar depression based on resting-state functional connectivity MRI: effects of generalization to a diverse sample. J Neural Trans 124(5):589–605. https://doi.org/10.1007/s00702-016-1673-8

44. Szegedy C, Liu W, Jia Y, Sermanet P, Reed SE, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: IEEE conference on computer vision and pattern recognition, CVPR 2015, Boston, MA, USA, June 7–12 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594
45. Trajanovski S, Mavroeidis D, Swisher CL, Gebre BG, Veeling B, Wiemker R, Klinder T, Tahmasebi A, Regis SM, Wald C, McKee BJ, MacMahon H, Pien H (2018) Towards radiologist-level cancer risk assessment in CT lung screening using deep learning. CoRR. arXiv:1804.01901
46. Vanschoren J, van Rijn JN, Bischl B (2015) Taking machine learning research online with openml. In: Proceedings of the 4th international workshop on big data, streams and heterogeneous source mining: algorithms, systems, programming models and applications, BigMine 2015, Sydney, Australia, August 10 2015. JMLR Workshop and Conference Proceedings, vol 41, pp 1–4. JMLR.org. http://jmlr.org/proceedings/papers/v41/vanschoren15.html
47. Veličković P, Wang D, Lane ND, Liò P (2016) X-cnn: cross-modal convolutional neural networks for sparse datasets. In: 2016 IEEE symposium series on computational intelligence (SSCI), pp 1–8. https://doi.org/10.1109/SSCI.2016.7849978
48. Wainberg M, Merico D, Delong A, Frey BJ (2018) Deep learning in biomedicine. Nat Biotechnol 36:829 EP. https://doi.org/10.1038/nbt.4233
49. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH (2016) Deep learning for identifying metastatic breast cancer. CoRR. arXiv:1606.05718
50. Xu Y, Li Y, Liu M, Wang Y, Lai M, Chang EI (2016) Gland instance segmentation by deep multichannel side supervision. CoRR. arXiv:1607.03222