Guest Editorial
Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique

I. Introduction

DEEP learning is a growing trend in general data analysis and has been termed one of the 10 breakthrough technologies of 2013 [1]. Deep learning is an improvement of artificial neural networks, consisting of more layers that permit higher levels of abstraction and improved predictions from data [2]. To date, it is emerging as the leading machine-learning tool in the general imaging and computer vision domains.

In particular, convolutional neural networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks. Deep CNNs automatically learn mid-level and high-level abstractions obtained from raw data (e.g., images). Recent results indicate that the generic descriptors extracted from CNNs are extremely effective in object recognition and localization in natural images. Medical image analysis groups across the world are quickly entering the field and applying CNNs and other deep learning methodologies to a wide variety of applications. Promising results are emerging.

In medical imaging, the accurate diagnosis and/or assessment of a disease depends on both image acquisition and image interpretation. Image acquisition has improved substantially over recent years, with devices acquiring data at faster rates and increased resolution. The image interpretation process, however, has only recently begun to benefit from computer technology. Most interpretations of medical images are performed by physicians; however, image interpretation by humans is limited due to its subjectivity, large variations across interpreters, and fatigue. Many diagnostic tasks require an initial search process to detect abnormalities, and to quantify measurements and changes over time. Computerized tools, specifically image analysis and machine learning, are the key enablers to improve diagnosis, by facilitating identification of the findings that require treatment and to support the expert's workflow. Among these tools, deep learning is rapidly proving to be the state-of-the-art foundation, leading to improved accuracy. It has also opened up new frontiers in data analysis with rates of progress not before experienced.

A. Historical Perspective on Networks

Neural networks and the basic ideas behind deep learning have been around for decades [3]. They have typically had just a few layers. The performance of neural networks improved markedly after the development of the backpropagation algorithm. However, performance was still insufficient. Other classifiers were subsequently developed, including decision trees, boosting and support vector machines. Each of these has been applied to medical image analysis, especially for detection of abnormalities, but also in related fields such as segmentation. Despite such developments, relatively high false positive rates for detection were the norm.

CNNs were applied to medical image processing as long ago as 1996 in work of Sahiner *et al.* [4]. In this work, ROIs containing either biopsy-proven masses or normal tissues were extracted from mammograms. The CNN consisted of an input layer, two hidden layers and an output layer and used backpropagation. Training times were described as "computationally intensive" in this pre-GPU era, but no times were given. In 1993, CNNs were applied to lung nodule detection [5]. A CNN was used to detect microcalcifications on mammography in 1995 [6].

The typical CNN architecture for image processing consists of a series of layers of convolution filters, interspersed with a series of data reduction or pooling layers. The convolution filters are applied to small patches of the input image. Like the low-level vision processing in the human brain, the convolution filters detect increasingly more relevant image features, for example lines or circles that may represent straight edges (such as for organ detection) or circles (such as for round objects like colonic polyps), and then higher order features like local and global shape and texture. The output of the CNN is typically one or more probabilities or class labels. The convolution filters are learned from training data. This is desirable because it reduces the necessity of the time-consuming hand-crafting of features that would otherwise be required to pre-process the images with application-specific filters or by calculating computable features. There are other network architecture variants, such as a deep recurrent neural network known as long short-term memory [7].

CNNs are highly parallelizable algorithms. Much of the practicality of using CNNs today lies in the vast acceleration (approximately 40 times) enabled by graphics processing unit (GPU) computer chips compared to CPU processing alone. An early paper describing the value of the GPU for training CNNs and other machine learning techniques was published in 2006 [8]. In medical image processing, GPUs were introduced first for segmentation, reconstruction and registration, and then much later for machine learning [9], [10]. Interestingly, while Eklund *et al.* [10] discuss convolutions extensively in their 2013 paper, convolutional neural networks and deep learning

Digital Object Identifier 10.1109/TMI.2016.2553401

are not mentioned at all. This highlights how rapidly the revolution of deep learning has refocused medical image processing research.

B. Networks Today

Deep neural networks have recently gained considerable commercial interest due to the development of new variants of CNNs and the advent of efficient parallel solvers optimized for modern GPUs. The main power of a CNN lies in its deep architecture, which allows for extracting a set of discriminating features at multiple levels of abstraction. Training a deep CNN from scratch (or full training) is a challenge. First, CNNs require a large amount of labeled training data, a requirement that may be difficult to meet in the medical domain where expert annotation is expensive and the diseases (e.g., lesions) are scarce. Second, training a deep CNN requires large computational and memory resources, without which the training process would be extremely time-consuming. Third, training a deep CNN is often complicated by overfitting and convergence issues, which often require repetitive adjustments in the architecture or learning parameters of the network to ensure that all layers are learning with comparable speed. Given these difficulties, several new learning schemes, termed "transfer learning" and "fine-tuning", are shown to provide solutions and are increasingly gaining popularity. These will be discussed further in Section II-C.

C. Networks in the Medical Domain

Deep learning methods are most effective when applied to large training sets, but in the medical domain large data sets are not always available. We are therefore faced with major challenges including (a) Can deep networks be used effectively for medical tasks? (b) Is transfer learning from general imagery to the medical domain relevant? (c) Can we rely on learned features alone or may we combine them with handcrafted features for the task? This special issue of IEEE-Transactions on Medical Imaging (IEEE-TMI) on deep learning in medical imaging focuses on progress in this new era of machine learning and its role in the medical imaging domain. This issue presents recent achievements of CNNs and other deep learning applications to medical tasks. It contains 18 papers by various investigators from around the world, selected from 50 submissions, an unusually high number for IEEE special issues, and this was achieved in a period between publication of the call for papers and the submission deadline that was shorter than usual. The papers focus on a variety of classical tasks, from detection to categorization (e.g., lesion detection, image segmentation, shape modeling, image registration), as well as opening up new and novel application domains. Also included are several works that focus on the exploration of the networks and provide insight on the architectures to be chosen for various tasks, parameters, training sets and more.

II. Overview of Topics and Papers in the Journal

A. Lesion Detection

Computer-aided detection (CAD) is a well established area of medical image analysis that is highly amenable to deep learning. In the standard approach to CAD [11] candidate lesions are detected, either by supervised methods or by classical image processing techniques such as filtering and mathematical morphology. Candidate lesions are often segmented, and described by an often large set of hand-crafted features. A classifier is used to map the feature vectors to the probability that the candidate is an actual lesion. The straightforward way to employ deep learning instead of hand-crafted features is to train a CNN operating on a patch of image data centered on the candidate lesion. Several articles in this issue use this approach. Setio *et al.* [12] combine three previously developed candidate detectors for pulmonary nodules in 3D chest CT scans and extract 2D patches centered on these candidates under nine different orientations. A combination of different CNNs is used to classify each candidate. A slight improvement compared to a previously published classical CAD system for the same task is reported.

Roth *et al.* [13] applied CNNs to improve three existing CAD systems for the detection of colonic polyps on CT colonography, sclerotic spine metastases on body CT and enlarged lymph nodes on body CT. They also used previously developed candidate detectors and 2D patches in three orthogonal directions, and up to 100 randomly rotated views. The randomly rotated "2.5D" views are a method of decompositional image representation from the original 3D data. The CNN predictions on these 2.5D views are later aggregated to attain additional gains in accuracy. The sensitivity for lesion detection improved 13 – 34% for all three CAD systems with the use of CNNs, indicating that the approach was both generalizable and scaleable. Improvements of this magnitude have been nearly impossible to achieve using non-deep learning classifiers such as committees of support vector machines.

Dou *et al.* [14] detected cerebral microbleeds from susceptibility weighted MRI scans. They use 3D CNNs and also replaced the candidate detection stage with a CNN, proposing a two stage approach. They report improved results with their 3D CNN compared to various classical and 2D CNN approaches from the literature that the authors re-implemented and trained and tested on the same data set.

Sirinukunwattana *et al.* [15] detected and classified nuclei in histopathological images. They use a CNN that takes small patches as input and instead of just predicting if the central pixel of the patch is a cell nucleus, they model the output as a high peak in the vicinity of the center of each nucleus and flat elsewhere. This spatially constrained CNN, in combination with a fusion of overlapping patches in the test stage, produces better results than previously proposed techniques for this task also based on CNNs and on classical feature-based approaches.

Anthimopoulos *et al.* [16] focus on detecting patterns of interstitial lung diseases from 2D patches of chest CT scans. They are one of three groups in this issue (Shin *et al.* [17] and van Tulder *et al.* [18]) using a public data set from [19]. They train a CNN to classify patches of 32×32 pixels into one of 7 classes and report higher accuracy than three previously published methods that employ hand-crafted features.

In several other articles in this issue, lesion detection is also the topic of interest, but the focus of these papers are broader or

zoom in on particular methodological issues. These papers will be briefly discussed below.

B. Segmentation and Shape Modeling

For a large dataset of 2891 cardiac ultrasound examinations, Ghesu *et al.* [20] combine deep learning and marginal space learning for object detection and segmentation. The combination of "efficient exploration of large parameter spaces" and a method to enforce sparsity in the deep networks, increased computational efficiency, and led to a 13.5% reduction in mean segmentation error compared to a reference method published by the same group.

Three studies focused on segmentation of brain structures or brain lesions. The problem of multiple sclerosis brain lesion segmentation on MRI was addressed by Brosch *et al.* [21]. The authors developed a 3D deep convolutional encoder network that combined interconnected convolutional and deconvolutional pathways. The convolutional pathway learned higher level features, and the deconvolutional pathway predicted the voxel level segmentation. They applied their network to two publicly available datasets and one clinical trial data set. They compared their method with 5 publicly available methods. The method was reported to perform "comparably to the best state-of-the-art methods".

Brain tumor segmentation on MRI was investigated in Pereira *et al.* [22]. The authors used small kernels, a deeper architecture, intensity normalization and data augmentation. Different CNN architectures were used for low and high grade tumors. The method separately segmented the enhancing part and the core of the tumor. They attained top results on a 2013 public challenge dataset, and second place in an on-site 2015 challenge.

For brain structure segmentation, CNNs performed well on five different datasets for patients in different age groups, spanning from pre-term infants to older adults in a study by Moeskops *et al.* [23]. A multi-scale approach was used to achieve robustness. The method attained good results on eight tissue classes, with Dice similarity coefficients averaging 0.82 to 0.91 for the five datasets.

C. Network Exploration

1) Data Dimensionality Issues- 2D vs 3D: In most works to date we see analysis performed in 2D. It is often questioned if the transition to 3D will be a key to major step forward in performance. Several variations exist in the data augmentation process, including 2.5D. For example, in Roth *et al.*, [13] axial, coronal and sagittal images were taken centered on a voxel in a colonic polyp or lymph node candidate and fed into the cuda-convnet CNN, which incorporates three channels normally used to represent the red, green and blue color channels of a natural light image. Explicitly 3D CNNs were used in this issue by Brosch *et al.*[21] and Dou *et al.* [14].

2) Learning Methodology- Unsupervised vs Supervised: When we look at the literature of networks, it is evident that the majority of works focus on the supervised CNNs in order to achieve categorization. Such networks are important for many applications, including detection, segmentation and labeling. Still, several works focus on unsupervised schemes which are mostly shown to be useful in image encoding, efficient image representation schemes and as a preprocessing step for further supervised schemes. Unsupervised representation learning methods such as Restricted Boltzmann Machines (RBM) may outperform standard filter banks because they learn a feature description directly from the training data. The RBM is trained with a generative learning objective; this enables the network to learn representations from unlabeled data, but does not necessarily produce features that are optimal for classification. Van Tulder *et al.*, [18] conducted an investigation to combine the advantages of both generative and a discriminative learning objectives in a convolutional classification restricted Boltzmann machine, which learns filters that are good both for describing the training data and for classification. It is shown that a combination of learning objectives outperforms purely discriminative or generative learning.

3) Training Data Considerations: CNNs enable learning data-driven, highly representative, layered hierarchical image features. These features have been demonstrated to be a very strong and robust representation in many application domains, as presented in this issue. In order to provide such a rich representation and successful classification, sufficient training data are needed. The amount of data required is a key question to be explored. Related questions include the following: How can we use the training data we have most efficiently? What can we do in cases in which data are not available? And finally, are there alternative methods for acquiring and medically annotating data?

Several of these questions are addressed by papers in this issue. Van Grinsven *et al.* [24] attempt to improve and speed-up the CNN training for medical image analysis tasks by dynamically selecting misclassified negative samples during training. The CNN training process is a sequential process requiring many iterations (or epochs) to optimize the network parameters. In every epoch, a subset of samples is randomly selected from the training data and is presented to the network to update its parameters through back-propagation, minimizing a cost function. Classification tasks in the medical domain are often a normal vs pathology discrimination task. In such a scenario, the normal class is extremely over-represented; moreover, the majority of normal training samples are highly correlated due to the repetitive pattern of normal tissues in each image. Only a small fraction of these samples are informative. Treating uniformly this data during the learning process leads to many training iterations wasted on non-informative samples, making the CNN training process unnecessarily time-consuming. An approach to identify informative normal samples, as shown in the work, increased the efficiency of the CNN learning process and reduced the training time.

4) Transfer Learning and Fine Tuning: Obtaining datasets in the medical imaging domain that are as comprehensively annotated as ImageNet remains a challenge. When sufficient data are not available, there are several ways to proceed: 1) Transfer learning: CNN models (supervised) pre-trained from natural image dataset or from a different medical domain are used for a new medical task at hand. In one scheme, a pre-trained

CNN is applied to an input image and then the outputs are extracted from layers of the network. The extracted outputs are considered features and are used to train a separate pattern classifier. For instance, in Bar *et al.* [25], [26] pre-trained CNNs were used as a feature generator for chest pathology identification. In Ginneken *et al.* [27] integration of CNN-based features with handcrafted features enabled improved performance in a nodule detection system. 2) Fine Tuning: When a medium sized dataset does exist for the task at hand, one suggested scheme is to use a pre-trained CNN as initialization of the network, following which further supervised training is conducted, of several (or all) the network layers, using the new data for the task at hand.

Transfer learning and fine tuning are key components in the use of deep CNNs in medical imaging applications. Two works that explore these issues are Shin *et al.* [17] and Tajbakhsh *et al.* [28]. The experiments conducted in these works consistently show that using pre-trained CNNs with fine-tuning achieved the strongest results, regardless of specific applications domain (Tajbakhsh *et al.*), and for all network architectures (Shin *et al.*). In Tajbakhsh *et al.*, [28] further analysis shows that deep fine-tuning led to improved performance over shallow fine-tuning, and the importance of using fine-tuning increases with reduced size training sets. In Shin *et al.*, [17] the GoogLeNet architecture led to state-of-the-art detection of mediastinal lymph nodes compared to other less deep architectures.

5) Ground Truth – From the Experts and the Non-Experts: The lack of publicly available ground-truth data, and the difficulty in collecting such data per medical task, both cost-wise as well as time-wise, is a prohibitively limiting factor in the medical domain. Though crowdsourcing has enabled annotation of large scale databases for real world images, its application for biomedical purposes requires a deeper understanding and hence, more precise definition of the actual annotation task Nguyen *et al.* [29], McKenna *et al.* [30]. The fact that expert tasks are being outsourced to non-expert users may lead to noisy annotations introducing disagreement between users. Many issues arise in combining the knowledge of medical experts with non-professionals, such as how to combine the information sources, how to assess and incorporate the inputs weighted by their prior-proved accuracy in performance and more. These issues are addressed in Albarqouni *et al.* [31] who present a network that combines an aggregation layer that is integrated into the CNN to enable learning inputs from the crowds as part of the network learning process. Results shown give valuable insights into the functionality of deep CNN learning from crowd annotations. The most astonishing fact about crowdsourcing studies in the medical domain, however, is the conclusion that a crowd of nonprofessional, inexperienced users can in fact perform as well as the medical experts. This has also been observed by Nguyen *et al.* [29] and McKenna *et al.* [30] for radiology images.

D. Novel Applications and Unique Use Cases

Unsupervised feature learning for mammography risk scoring is presented in Kallenberg *et al.* [32]. In this work, a method is shown that learns a feature hierarchy from unlabeled data. The learned features are then input to a simple classifier, and two different tasks are addressed: i) breast density segmentation, and ii) scoring of mammographic texture, with state-of-the-art results achieved. To control the model capacity a sparsity regularizer is introduced that incorporates both lifetime and population sparsity. The convolutional layers in the unsupervised parts are trained as autoencoders; In the supervised part the (pre-trained) weights and bias terms are fine-tuned using softmax regression.

Yan *et al.* [33] design a multi-stage deep learning framework for image classification and apply it on body part recognition. In the pre-train stage, a convolutional neural network (CNN) is learned using multi-instance learning to extract the most discriminative and non-informative local patches from the training slices. In the boosting stage, the pre-trained CNN is further boosted by these local patches for image classification. A hallmark of the method was that it automatically discovered the discriminative and non-informative local patches through multi-instance deep learning. Thus, no manual annotation was required.

Regression networks are not very common in the medical imaging domain. In Miao *et al.* [34], a CNN regression approach is presented, for real-time 2-D/3-D registration. Three algorithmic strategies are proposed to simplify the underlying mapping to be regressed, and to design a CNN regression model with strong non-linear modeling. Results show that the DL method is more accurate and robust than two state-of-the-art accelerated intensity-based 2-D/3-D registration methods.

We have yet to explore the domains to which NNs can be applied, and the applications and dimensionality of tasks to which they will provide a substantial contribution. In an exploratory work, Golkov *et al.* [35] provide an initial proof-of-concept, applying DL to reduce diffusion MRI data processing to a single optimized step. They show that this modification enables one to obtain scalar measures from advanced models at twelve-fold reduced scan time and to detect abnormalities without using diffusion models. The relationship between the diffusion-weighted signal and microstructural tissue properties is non-trivial. Golkov *et al.* [35] demonstrate that with the use of a DNN such relationships may in fact be revealed: DWIs are directly used as inputs rather than using scalar measures obtained from model fitting. The work shows microstructure prediction on a voxel-by-voxel basis as well as automated model-free segmentation from DWI values, into healthy tissues and MS lesions. Diffusion kurtosis is shown to be measured from only 12 data points and neurite orientation dispersion and density measures from only 8 data points. This may allow for fast and robust protocols facilitating clinical routine and demonstrates how classical data processing can be streamlined by means of deep learning.

III. Discussion: Key Issues and Main Promises

In the majority of works presented, use of a deep network is shown to improve over the state-of-the-art. As these improvements seem to be consistent across a large variety of domains, and as is usually the case, development of a deep learning solution is found to be relatively straight-forward, we can view this as a major step forward in the medical computing field. Still, a

major question remains as to how and when we can reach a substantial leap in performance – equivalent to the 10% increase in large scale category recognition of 2012. Are we asking the right questions and investigating the right tasks? Are we using strong enough input representations (e.g., 2D vs 3D)? Do we need to work on getting real Big Data for each medical task, or will transfer learning be sufficient? These issues and more are partly addressed in the papers of this issue (Section II), and mostly remain as the challenges to be answered in the future.

In the literature we can find both unsupervised as well as supervised learning via deep networks. It seems that the majority of works are in fact using supervised learning. The question arises if in the medical domain, where data numbers are a critical factor, the formalism should combine the benefits of both the unsupervised and the supervised. It is likely that to leverage really big data for which hand annotations are unavailable or intractable to obtain, the field will need to move more towards semi-supervised and unsupervised learning.

The literature contains many network architectures. The variability is large, as can be seen across the works in the current journal. Possibilities include selecting a known architecture, devising a task-specific architecture, fusing across architectures and more. One interesting question going forward is if the very deep residual networks that used 152 layers and performed best in the ILSVRC 2015 classification task [36] also obtain good results on medical tasks.

One key aspect of deep learning is that it can benefit from large amounts of training data. Breakthrough results in computer vision were obtained on the ILSVRC challanges based on the ImageNET data set (http://www.image-net.org/). This data set is very large compared to most of the training and test data sets used in the papers in this special issue (millions versus 100's or 1000's). Our community could likely benefit strongly if similarly large publicly available medical image data sets were constructed.

There are several reasons why this is challenging. First, it is difficult to obtain funding for the construction of data sets. Second, scarce and expensive medical expertise is needed for high quality annotation of medical imaging data. Third, privacy issues make it more difficult to share medical data than natural images. Fourth, the breadth of applications in medical imaging requires that many different data sets need to be collected. Despite these potential hurdles, we see rapid progress in the area of data collection and data sharing. Many public data sets have been released and studies today routinely use them in experimental validation. Examples include VISCERAL and The Cancer Imaging Archive (http://www.visceral.eu/ and http://www.cancerimagingarchive.net/). Roth *et al.* [13] and Shin *et al.* [17] analyze a dataset of enlarged lymph nodes on CT scans that they have made public on The Cancer Imaging Archive [37]. The same group has made a pancreas dataset available online [38].

Since 2007, it has become customary to organize challenge workshops at medical imaging conferences such as MICCAI, ISBI, and SPIE Medical Imaging. This has led to a large amount of data sets and ongoing benchmark studies, documented at the website http://www.grand-challenge.org/. Using these public benchmark data sets has a distinct advantage over using public data sets only: challenges provide a precise definition of a task to be solved and define one or more evaluation metrics that provide a fair and standardized comparison between proposed algorithmic. Without such a standardization, it is often difficult to compare different approaches to the same problem even if they use the same data set. This is illustrated in this issue where three studies (Anthimopoulos *et al.* [16], Shin *et al.* [17], and van Tulder *et al.* [18]) use the same data set of chest CT scans with annotations of interstitial lung disease patterns [19], but they all report results in a different form.

One study in this issue (Setio *et al.* [12]) has resulted in a challenge for pulmonary nodule detection (http://luna16.grand-challenge.org/), organized in conjunction with the IEEE ISBI conference, using the publicly available LIDC/IDRI data set, and thus the system described in this issue can be compared directly to alternative approaches.

Last year we have seen the first large challenges devoted to medical image analysis organized on major platforms traditionally focusing on other machine learning applications. Kaggle organized a competition on detection and staging of diabetic retinopathy from color fundus images for which 661 teams submitted results, $100,000 prize money was awarded, and around 80,000 images were made available (https://www.kaggle.com/c/diabetic-retinopathy-detection). This data was used in one study in this special issue (van Grinsven *et al.* [24]). Recently, a second medical image analysis competition was completed using MRI to measure cardiac volumes and derive ejection fractions in which 192 teams participated and $200,000 prize money was available (https://www.kaggle.com/c/second-annual-data-science-bowl). In both these competitions, the top contenders all used convolutional neural networks. We expect both trends to continue: challenges will use larger data sets and deep learning will be a dominant technology among the best performing solutions. In this context, the upcoming series of worldwide competitions to increase the accuracy of various imaging based cancer screenings (coding4cancer.org) could be of interest for readers of this special issue.

On-line platforms, such as the ones used for competitions, serve multiple purposes. They lead to new collaborations and joint solutions, could also be the most effective way to obtain large amounts of data annotations via crowd-sourcing, which was shown by the study from Albarqouni *et al.* [31] in this issue.

Finally, we are grateful for the guidance from the current Editor-in-Chief, the help from the TMI office, and most importantly the tremendous efforts by the authors and the reviewers. This Special issue provides a snapshot in time of a very fast moving field of medical image processing. We hope you will enjoy it and look forward to your future contributions to this dynamic field.

Acknowledgment

We would like to thank all the submissions to the journal. We value greatly the support from our large set of expert reviewers, who helped select the papers and shape this issue.

H. Greenspan is funded in part by INTEL Collaborative Research Institute for Computational Intelligence (ICRI-CI). This

work was funded in part by the Intramural Research Program of the National Institutes of Health, Clinical Center.

Disclaimer: The opinions expressed herein are those of the authors and do not necessarily represent those of the National Institutes of Health. Financial Disclosures: R. M. Summers receives patent royalties from iCAD Medical.

HAYIT GREENSPAN, *Guest Editor*
Biomedical Image Computing Lab
Department of Biomedical Engineering
Faculty of Engineering
Tel-Aviv University
Tel-Aviv, 69978 Israel

BRAM VAN GINNEKEN, *Guest Editor*
Diagnostic Image Analysis Group
Radboud University Medical Center
Nijmegen, 6525 GA The Netherlands

RONALD M. SUMMERS, *Guest Editor*
Imaging Biomarkers and Computer-Aided Diagnosis Lab
Radiology and Imaging Sciences
National Institutes of Health Clinical Center
Bethesda, MD 20892 USA

REFERENCES

[1] *MIT Technol. Rev.*, , 2013 [Online]. Available: https://www.technologyreview.com/s/513696/deep-learning
[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[3] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.
[4] B. Sahiner *et al.*, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 598–610, Oct. 1996.
[5] S. C. B. Lo, J. S. J. Lin, M. T. Freedman, and S. K. Mun, "Computer-assisted diagnosis of lung nodule detection using artificial convolution neural-network," *Proc. SPIE Med. Imag., Image Process.*, vol. 1898, pp. 859–869, 1993.
[6] H.-P. Chan, S.-C. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phys.*, vol. 22, no. 10, pp. 1555–67, 1995.
[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[8] 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.
[9] D. Castano-Diez, D. Moser, A. Schoenegger, S. Pruggnaller, and A. S. Frangakis, "Performance evaluation of image processing algorithms on the GPU," *J. Struct. Biol.*, vol. 164, no. 1, pp. 153–160, 2008.
[10] A. Eklund, P. Dufort, D. Forsberg, and S. M. LaConte, "Medical image processing on the GPU–Past, present and future," *Med. Image Anal.*, vol. 17, no. 8, pp. 1073–94, 2013.
[11] B. van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: How to move from the laboratory to the clinic," *Radiol.*, vol. 261, no. 3, pp. 719–732, 2011.
[12] A. Setio *et al.*, "Pulmonary nodule detection in CT images using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.
[13] H. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.
[14] Q. Dou *et al.*, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.
[15] K. Sirinukunwattana *et al.*, "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.
[16] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.
[17] 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.
[18] G. van Tulder and M. de Bruijne, "Combining generative and discriminative representation learning in convolutional restricted Boltzmann machines," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1262–1272, May 2016.
[19] A. Depeursinge *et al., Comput. Med. Imag. Graph.*, vol. 36, no. 3, pp. 227–238, 2012.
[20] F. Ghesu *et al.*, "Marginal space deep learning: Efficient architecture for volumetric image parsing," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1217–1228, May 2016.
[21] T. Brosch *et al.*, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1229–1239, May 2016.
[22] S. Pereira, A. Pinto, V. Alves, and C. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.
[23] P. Moeskops *et al.*, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.
[24] M. van Grinsven, B. van Ginneken, C. Hoyng, T. Theelen, and C. 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.
[25] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," *Proc. SPIE Med. Imag. Computer-Aided Diagnosis*, vol. 9414, 2015.
[26] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 294–297.
[27] B. van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 286–289.
[28] 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.
[29] T. B. Nguyen *et al.*, "Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography," *Radiology*, vol. 262, no. 3, pp. 824–833, 2012.
[30] M. T. McKenna *et al.*, "Strategies for improved interpretation of computer-aided detections for CT colonography utilizing distributed human intelligence," *Med. Image Anal.*, no. 6, pp. 1280–1292, 2012.
[31] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "Agg-Net: 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.
[32] M. Kallenberg *et al.*, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1322–1331, May 2016.
[33] Z. Yan *et al.*, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, May 2016.
[34] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.
[35] V. Golkov *et al.*, "q-Space deep learning: Twelve-fold shorter and model free diffusion MRI scans," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1344–1351, May 2016.
[36] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition ArXiv, 2015 [Online]. Available: arXiv:1512.03385, to be published
[37] H. R. Roth *et al.*, "A new 2.5 d representation for lymph node detection in CT," *Cancer Imag. Arch.*, 2015 [Online]. Available: http://dx.doi.org/10.7937/K9/TCIA.2015.AQIIDCNM
[38] H. R. Roth *et al.*, "Data from pancreas-CT," *Cancer Imag. Arch.*, 2016 [Online]. Available: http://dx.doi.org/10.7937/K9/TCIA.2016.tNB1kqBU

Hayit Greenspan received the B.S. and M.S. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, and the Ph.D. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA.

She was a Postdoc with the Computer Science Division at University of California Berkeley following which she joined Tel-Aviv University and founded the Biomedical Image Computing lab, which she heads. She is a Tenured Professor at the Biomedical Engineering Department in the Faculty of Engineering, Tel-Aviv University. She is affiliated with the International Computer Science Institute (ICSI), Berkeley, CA, USA. From 2008 until 2010, she was a visiting Professor at Stanford University, Department of Radiology, Faculty of Medicine. She has been conducting research in medical image analysis for the past 20 years, with a special focus on image modeling and analysis, deep learning, and content-based image retrieval. Among her current research projects: MRI resolution augmentation, brain MRI research (structural and DTI), CT and X-ray image analysis and medical imagery indexing and retrieval. She has over 100 publications in leading international journals and conference proceedings. She is a coauthor on several patents. She has received grants from several Israeli and U.S. agencies. Currently her Lab is funded for Deep Learning in Medical Imaging by the INTEL Collaborative Research Institute for Computational Intelligence (ICRI-CI).

Dr. Greenspan has received several awards. She is a member of several journal and conference program committees, including SPIE medical imaging, IEEE ISBI and MICCAI. She is an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING.

Bram van Ginneken studied physics at the Eindhoven University of Technology, Eindhoven, The Netherlands, and at Utrecht University, Utrecht, The Netherlands. He received the Ph.D. degree in computer-aided diagnosis in chest radiography from the Image Sciences Institute, Utrecht, The Netherlands, in March 2001.

He is Professor of Functional Image Analysis at Radboud University Medical Center Nijmegen. Since 2010, he is co-Chair of the Diagnostic Image Analysis Group within the Department of Radiology and Nuclear Medicine, together with N. Karssemeijer. He also works for Fraunhofer MEVIS, Bremen, Germany, and is one of the founders of Thirona, a company that provides quantitative analysis of chest CT scans. From 2001 through 2009, he led the Computer-Aided Diagnosis group at the Image Sciences Institute, where he still has an Associated Faculty position. He has (co-)authored over 150 publications in international journals. He is a member of the Editorial Board of *Medical Image Analysis*. He has pioneered the concept of challenges in medical image analysis, where multiple teams apply their algorithm to solve a particular medical image analysis task to the same test data set, so that a fair comparison between different approaches can be made.

Dr. van Ginneken is an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING.

Ronald M. Summers received the B.A. degree in physics and the M.D. and Ph.D. degrees in medicine/anatomy and cell biology from the University of Pennsylvania, Philadelphia, PA, USA. He completed a medical internship at the Presbyterian-University of Pennsylvania Hospital, Philadelphia, PA, USA, a radiology residency at the University of Michigan, Ann Arbor, MI, USA, and an MRI fellowship at Duke University, Durham, NC, USA.

In 1994, he joined the Diagnostic Radiology Department at the NIH Clinical Center, Bethesda, MD, USA, where he is now a tenured Senior Investigator and Staff Radiologist. He is currently Chief of the Clinical Image Processing Service and directs the Imaging Biomarkers and Computer-Aided Diagnosis (CAD) Laboratory. His research interests include deep learning, virtual colonoscopy, CAD and development of large radiologic image databases. His clinical areas of specialty are thoracic and abdominal radiology and body cross-sectional imaging. He is a member of the editorial boards of the journals *Radiology*, *Journal of Medical Imaging*, and *Academic Radiology*. He has co-authored over 400 journal, review and conference proceedings articles and is a coinventor on 14 patents.

Dr. Summers received the Presidential Early Career Award for Scientists and Engineers, presented by Dr. Neal Lane, President Clinton's science advisor, in 2000. In 2012, he received the NIH Director's Award, presented by NIH Director Dr. Francis Collins. He is a program committee member of the Computer-aided Diagnosis section of the annual SPIE Medical Imaging conference.