

Received June 10, 2019, accepted July 10, 2019, date of publication July 16, 2019, date of current version August 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929365

# Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions

**FOUZIA ALTAF<sup>ID</sup><sup>1</sup>, SYED M. S. ISLAM<sup>ID</sup><sup>1,2</sup>, NAVEED AKHTAR<sup>2</sup>, AND NAEEM KHALID JANJUA<sup>1</sup>**

<sup>1</sup>School of Science, Edith Cowan University, Joondalup, WA 6027, Australia

<sup>2</sup>Department of Computer Science, The University of Western Australia, Perth, WA 6009, Australia

Corresponding author: Fouzia Altaf (faltaf@our.ecu.edu.au)

This work was supported by the Australian Government Research Training Program Scholarship.

**ABSTRACT** Medical image analysis is currently experiencing a paradigm shift due to deep learning. This technology has recently attracted so much interest of the Medical Imaging Community that it led to a specialized conference in “Medical Imaging with Deep Learning” in the year 2018. This paper surveys the recent developments in this direction and provides a critical review of the related major aspects. We organize the reviewed literature according to the underlying pattern recognition tasks and further sub-categorize it following a taxonomy based on human anatomy. This paper does not assume prior knowledge of deep learning and makes a significant contribution in explaining the core deep learning concepts to the non-experts in the Medical Community. This paper provides a unique computer vision/machine learning perspective taken on the advances of deep learning in medical imaging. This enables us to single out “lack of appropriately annotated large-scale data sets” as the core challenge (among other challenges) in this research direction. We draw on the insights from the sister research fields of computer vision, pattern recognition, and machine learning, where the techniques of dealing with such challenges have already matured, to provide promising directions for the Medical Imaging Community to fully harness deep learning in the future.

**INDEX TERMS** Deep learning, medical imaging, artificial neural networks, survey, tutorial, data sets.

## I. INTRODUCTION

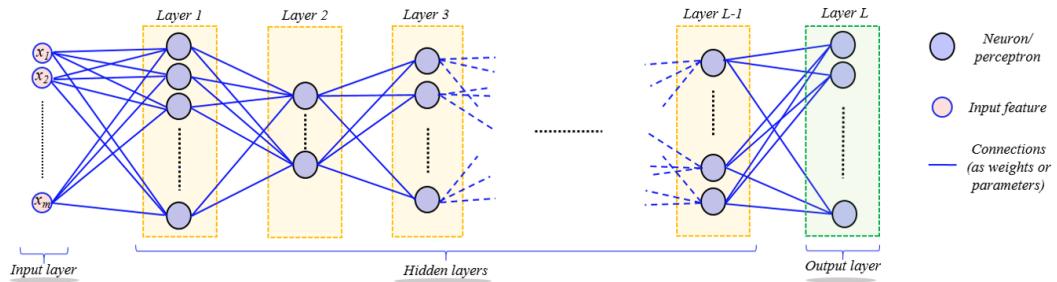
Deep Learning (DL) [1] is a major contributor of the contemporary rise of Artificial Intelligence in nearly all walks of life. This is a direct consequence of the recent breakthroughs resulting from its application across a wide variety of scientific fields; including Computer Vision [2], Natural Language Processing [3], Particle Physics [4], DNA analysis [5], brain circuits studies [6], and chemical structure analysis [7] etc. Very recently, it has also attracted a notable interest of researchers in Medical Imaging, holding great promises for the future of this field.

The DL framework allows machines to learn very complex mathematical models for data representation, that can subsequently be used to perform accurate data analysis. These models hierarchically compute non-linear and/or linear functions of the input data that is weighted by the model parameters. Treating these functions as data processing ‘layers’, the hierarchical use of a large number of such layers also

inspires the name ‘Deep’ Learning. The common goal of DL methods is to iteratively learn the parameters of the computational model using a training data set such that the model gradually gets better in performing a desired task, e.g. classification; over that data under a specified metric. The computational model itself generally takes the form of an Artificial Neural Network (ANN) [8] that consists of multiple layers of neurons/perceptrons [9] - basic computational blocks, whereas its parameters (a.k.a. network weights) specify the strength of the connections between the neurons of different layers. We depict a deep neural network in Fig. 1 for illustration.

Once trained for a particular task, the DL models are also able to perform the same task accurately using a variety of previously unseen data (i.e. testing data). This strong generalization ability of DL currently makes it stand out of the other Machine Learning techniques. Learning of the parameters of a deep model is carried out with the help of back-propagation strategy [10] that enables some form of the popular Gradient Descent technique [11], [12] to iteratively arrive at the desired parameter values. Updating the model

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Hugo Albuquerque.



**FIGURE 1.** Illustration of a deep neural network: The network consists of multiple layers of neurons/perceptrons that are connected in an inter-layer fashion. The neurons compute non-linear/linear transformations of their inputs. Each feature of the input signal is weighted by the network parameters and processed by the neuron layers hierarchically. A network connection performs the weighting operation. The strengths of the connections (i.e. network parameter values) are learned using training data. The network layers not seen by the input and output signals are often termed ‘hidden’ layers.

parameters using the complete training data *once* is known as a single epoch of network/model training. Contemporary DL models are normally trained for hundreds of epochs before they can be deployed.

Although the origins of Deep Learning can be traced back to 1940s [13], the sudden recent rise in its utilization for solving complex problems of the modern era results from three major phenomena. (1) Availability of large amount of training data: With ubiquitous digitization of information in recent times, very large amount of data is available to train complex computational models. Deep Learning has an intrinsic ability to model complex functions by simply stacking multiple layers of its basic computational blocks. Hence, it is a convenient choice for dealing with hard problems. Interestingly, this ability of deep models is known for over few decades now [14]. However, the bottleneck of relatively smaller training data sets had restricted the utility of Deep Learning until recently. (2) Availability of powerful computational resources: Learning complex functions over large amount of data results in immense computational requirements. Related research communities are able to fulfill such requirements only recently. (3) Availability of public libraries implementing DL algorithms: There is a growing recent trend in different research communities to publish the source codes on public platforms. Easy public access to DL algorithm implementations has exploded the use of this technique in many application domains.

The field of Medical Imaging has been exploiting Machine Learning since 1960s [15], [16]. However, the first notable contributions that relate to modern Deep Learning techniques appeared in the Medical Imaging literature in 1990s [17]–[21]. The relatedness of these methods to contemporary DL comes in the form of using ANNs to accomplish Medical Imaging tasks. Nevertheless, restricted by the amount of training data and computational resources, these works trained networks that were only two to three layers deep. This is no longer considered ‘deep’ in the modern era. The number of layers in the contemporary DL models generally ranges from a dozen to over one hundred [22]. In the context of image analysis, such models have mostly originated in Computer Vision literature [2].

The field of Computer Vision closely relates to Medical Imaging in analyzing digital images. Medical Imaging has a long tradition of profiting from the findings in Computer Vision. In 2012 [23], DL provided a major breakthrough in Computer Vision by performing a very hard image classification task with remarkable accuracy. Since then, the Computer Vision community has gradually shifted its main focus to DL. Consequently, Medical Imaging literature also started witnessing methods exploiting deep neural networks in around 2013, and now such methods are appearing at an ever increasing rate. Sahiner *et al.* [24] noted that the peer-reviewed publications that employed DL for radiological images tripled from 2016 ( $\sim 100$ ) to 2017 ( $\sim 300$ ), whereas the first quarter of 2018 alone saw over 100 such publications. Similarly, the main stream Medical Imaging conference, i.e. International Conference on ‘Medical Image Computing and Computer Assisted Intervention’ (MICCAI) published over 120 papers in its main proceedings in 2018 that employed Deep Learning for Medical Image Analysis tasks. The large inflow of contributions exploiting DL in Medical Imaging has also given rise to a specialized venue in the form of International Conference on ‘Medical Imaging with Deep Learning’ (MIDL) in 2018.<sup>1</sup>

We note that a few review articles for closely related research directions also exist at the time of this publication [24]–[29]. Among them, [25] only focuses on MRI, whereas [24], [26]–[29] mainly discuss the literature until the year 2017. However, the rapid advancements and breakthroughs in Deep Learning have already invalidated a few observations made in those reviews. Moreover, there is a large number of more recent Medical Imaging techniques employing DL technologies that were not invented (or at least not popular) at the time of those surveys. Besides reviewing the more recent literature as compared to the existing reviews, our survey also provides a unique Computer Vision and Machine Learning perspective on the Medical Imaging problems. Confined by their narrow perspective, the existing reviews also fall short on elaborating the root causes of the challenges faced by Deep Learning in Medical Imaging. Moreover, they are unable to provide insights for leveraging

<sup>1</sup><https://midl.amsterdam/>

the findings in other fields for addressing the challenges. These issues are specifically addressed in our review.

In this paper, we provide a comprehensive review of the recent DL techniques in Medical Imaging, focusing mainly on the very recent methods published in 2018 or after. We categorize these techniques under different pattern recognition tasks and further sub-categorize them following a taxonomy based on human anatomy. Analyzing the reviewed literature, we establish ‘lack of appropriately annotated large-scale datasets’ for Medical Imaging tasks as the fundamental challenge (among other challenges) in fully exploiting Deep Learning for those tasks. We then draw on the literature of Computer Vision, Pattern Recognition and Machine Learning in general; to provide guidelines to deal with this and other challenges in Medical Image Analysis using Deep Learning. This review also touches upon the available public datasets to train DL models for the medical imaging tasks. Considering the lack of in-depth comprehension of Deep Learning by the broader Medical community, this article also provides the understanding of the core technical concepts related to DL at an appropriate level. This is an intentional contribution of this work.

The remaining article is organized as follows. In Section II, we present the core Deep Learning concepts for the Medical community in an intuitive manner. The main body of the reviewed literature is presented in Section III. We touch upon the public data repositories for Medical Imaging in Section IV. In Section V, we highlight the major challenges faced by Deep Learning in Medical Image Analysis. Recommendations for dealing with these challenges are discussed in Section VI as future directions. The article concludes in Section VII. We provide the list of acronyms used in the remaining article in Table 1.

## II. BACKGROUND CONCEPTS

In this Section, we first briefly introduce the broader types of Machine Learning techniques and then focus on the Deep Learning framework. Machine Learning methods are broadly categorized as *supervised* or *unsupervised* based on the training data used to learn the computational models. Deep Learning based methods can fall in any of these categories.

### 1) SUPERVISED LEARNING

In supervised learning, it is assumed that the training data is available in the form of pairs  $(\mathbf{x}, \mathbf{y})$ , where  $\mathbf{x} \in \mathbb{R}^m$  is a *training example*, and  $\mathbf{y}$  is its *label*. The training examples generally belong to different, say  $C$  *classes* of data. In that case,  $\mathbf{y}$  is often represented as a binary vector living in  $\mathbb{R}^C$ , such that its  $c^{\text{th}}$  coefficient is ‘1’ if  $\mathbf{x}$  belongs to the  $c^{\text{th}}$  class, whereas all other coefficients are zero. A typical task for supervised learning is to find a computational model  $\mathcal{M}$  with the help of training data such that it is also able to correctly predict the labels of data samples that it had not seen previously during training. The unseen data samples are termed *test/testing samples* in the Machine Learning parlance. To learn a model that can perform successful classification of the test samples,

**TABLE 1.** List of acronyms used in the survey.

Acronym	Explanation
AAE	Adversarial Auto Encoder
AE	Auto Encoders
AD	Alzheimer Disease
AIR	Adversarial Image Registration
AMD	Age-related Macular Degeneration
ANN	Artificial Neural Network
ARD	Acute Respiratory Disease
ASD	Autism Spectrum Disorder
AUC	Area Under Curve
BIC	Breast Invasive Carcinoma
BTT	Back-propagation Through Time
BN	Batch Normalization
CAD	Computer Aided Detection
CIS	Carcinoma In situ
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
CRC	Colorectal Cancer
CRF	Conditional Random Field
CSF	Cerebro Spinal Fluid
CMR	Cardiac Magnetic Resonance
CNNs	Convolutional Neural Networks
CVPR	Computer Vision and Pattern Recognition
DBT	Digital Breast Tomosynthesis
DDH	Developmental Dysplasia of Hip
DL	Deep Learning
DR	Diabetic Retinopathy
DIR	Deformable Image Registration
DME	Diabetic Macular Edema
ER	Estrogen Receptor
FCNs	Fully convolutional Networks
GANs	Generative Adversarial Networks
GM	Grey Matter
GPU	Graphical Processing Unit
GON	Glaucomatous Optic Neuropathy
ICCV	International Conference on Computer Vision
ICML	International Conference on Machine Learning
ILD	Interstitial Lung Disease
IRC	Intrarenal Cystoid Fluid
LSTM	Long Short Term Memory
MILD	Medical Imaging with Deep Learning
MIL	Multiple Instance Learning
MLP	Multi Layer Perceptron
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
NLP	Natural Language Processing
OCT	Optical Coherence Tomography
PD	Percent Density
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SLN	Sentinel Lymph Nodes
TS	Tissue Segmentation
US	Ultrasound
VAE	Variational Auto Encoder
VGG	Visual Geometry Group
VOI	Volume Of Interest
WM	White Matter

we can formulate our learning problem as estimation of the parameters  $\Theta$  of our model that minimizes a specific loss  $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})$ , where  $\hat{\mathbf{y}}$  is the label vector *predicted* by the model for a given test sample. As can be guessed, the loss is defined such that it has a small value only if the learned parametric

model is able to predict the correct label of the data sample. Whereas the model loss has its scope limited to only a single data sample, we define a *cost* for the complete training data. The cost of a model is simply the Expected value of the losses computed for the individual data samples. Deep Learning allows us to learn model parameters  $\Theta$  that are able to achieve very low cost over very large data sets.

Whereas classification generally aims at learning computational models that map input signals to discrete output values, i.e. class labels. It is also possible to learn models that can map training examples to continuous output values. In that case,  $y$  is typically a real number scalar or vector. An example of such task is to learn a model that can predict the probability of a tumor being benign or malignant. In Machine Learning, such a task is seen as a *regression problem*. Similar to the classification problems, Deep Learning has been able to demonstrate excellent performance in learning computational models for the regression problem.

## 2) UNSUPERVISED LEARNING

Whereas supervised learning assumes that the training data also provides labels of the samples; *unsupervised* learning assume that sample labels are not available. In that case, the typical task of a computational model is to *cluster* the data samples into different groups based on the similarities of their intrinsic characteristics - e.g. clustering pixels of color images based on their RGB values. Similar to the supervised learning tasks, models for unsupervised learning tasks can also take advantage of minimizing a loss function. In the context of Deep Learning, this loss function is normally designed such that the model learns an accurate mapping of an input signal to itself. Once the mapping is learned, the model is used to compute compact representations of data samples that cluster well. Deep Learning framework has also been found very effective for unsupervised learning.

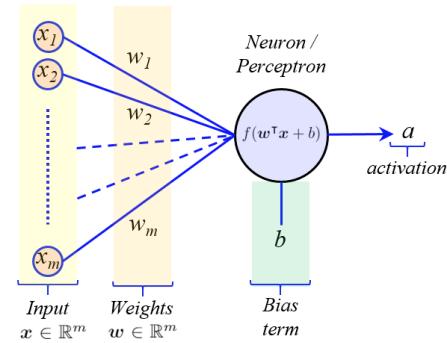
Along with supervised and unsupervised learning, other machine learning types include *semi-supervised learning* and *reinforcement learning*. Informally, semi-supervised learning computes models using the training data that provides labels only for its smaller subsets. On the other hand, reinforcement learning provides ‘a kind of’ supervision for the learning problem in terms of rewards or punishments to the algorithm. Due to their remote relevance to the tasks in Medical Imaging, we do not provide further discussion on these categories. Interested readers are directed to [30] for semi-supervised learning, and to [31] for reinforcement learning.

## A. STANDARD ARTIFICIAL NEURAL NETWORKS

An Artificial Neural Network (ANN) is a hierarchical composition of basic computational elements known as *neurons* (or *perceptrons* [9]). Multiple neurons exist at a single level of the hierarchy, forming a single *layer* of the network. Using many layers in an ANN makes it *deep*, see Fig. 1. A neuron performs the following simple computation:

$$a = f(\mathbf{w}^T \mathbf{x} + b), \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^m$  is the input signal,  $\mathbf{w} \in \mathbb{R}^m$  contains the neuron’s weights, and  $b \in \mathbb{R}$  is a *bias term*. The symbol  $f(\cdot)$  denotes an *activation function*, and the computed  $a \in \mathbb{R}$  is the neuron’s *activation signal* or simply its *activation*. Generally,  $f(\cdot)$  is kept non-linear to allow an ANN to induce complex non-linear computational models. The classic choices for  $f(\cdot)$  are the well-known ‘sigmoid’ and ‘hyperbolic tangent’ functions. We depict a single neuron/perceptron in Fig. 2.



**FIGURE 2.** Illustration of a single neuron/perceptron in a standard ANN. Each feature/coefficient ‘ $x_i \forall i \in \{1, \dots, m\}$ ’ of the input signal ‘ $x$ ’ gets weighted by a corresponding weight ‘ $w_i$ ’. A bias term ‘ $b$ ’ is added to the weighted sum ‘ $w^T x$ ’ and a non-linear/linear function  $f(\cdot)$  is applied to compute the activation ‘ $a$ ’.

A neural network must learn the weight and bias terms in Eq. (1). The strategy used to learn these parameters (i.e. back-propagation [10]) requires  $f(\cdot)$  to be a differentiable function of its inputs. In the modern Deep Learning era, Rectified Linear Unit (ReLU) [23] is widely used for this function, especially for non-standard ANNs such as CNNs (see Section II-B). ReLU is defined as  $a = \max(0, \mathbf{w}^T \mathbf{x} + b)$ . For the details beyond the scope of this discussion, ReLU allows more efficient and generally more effective learning of complex models as compared to the classic sigmoid and hyperbolic tangent activation functions.

It is possible to compactly represent the weights associated with all the neurons in a single layer of an ANN as a matrix  $\mathbf{W} \in \mathbb{R}^{p \times m}$ , where ‘ $p$ ’ is the total number of neurons in that layer. This allows us to compute the activations of all the neurons in the layer at once as follows:

$$\mathbf{a} = f(\mathbf{Wx} + \mathbf{b}), \quad (2)$$

where  $\mathbf{a} \in \mathbb{R}^p$  now stores the activation values of all the neurons in the layer under consideration. Noting the hierarchical nature of ANNs, it is easy to see that the functional form of a model induced by an  $L$ -layer network can be given as:

$$\mathcal{M}(\mathbf{x}, \Theta) = f_L(\mathbf{W}_L f_{L-1}(\mathbf{W}_{L-1} f_{L-2} \dots (\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \dots + \mathbf{b}_{L-1}) + \mathbf{b}_L), \quad (3)$$

where the subscripts denote the layer numbers. We collectively denote the parameters  $\mathbf{W}_i, \mathbf{b}_i, \forall i \in \{1, \dots, L\}$  as  $\Theta$ .

A neural network model can be composed by using different number of layers, having different number of neurons in each layer, and even having different activation functions

for different layers. Combined, these choices determine the *architecture* of a neural network. The design variables of the architecture and those of the learning algorithm are termed as *hyper-parameters* of the network. Whereas the model parameters (i.e.  $\Theta$ ) are learned automatically, finding the most suitable values of the hyper-parameter is usually a manual iterative process. Standard ANNs are also commonly known as Multi-Layer Perceptrons (MLPs), as their layers are generally composed of standard neurons/perceptrons. One notable exception to this layer composition is encountered at the very last layer, i.e. *softmax* layer used in classification. In contrast to ‘independent’ activation computation by each neuron in a standard perceptron layer, softmax neurons compute activations that are normalized across all the activation values of that layer. Mathematically, the  $i^{\text{th}}$  neuron of a softmax layer computes the activation value as:

$$a_i = \frac{e^{\mathbf{w}_i^\top \mathbf{a}_{L-1} + b_i}}{\sum_{j=1}^p e^{\mathbf{w}_j^\top \mathbf{a}_{L-1} + b_j}}. \quad (4)$$

The benefit of normalizing the activation signal is that the output of the softmax layer can be interpreted as a *probability* vector that encodes the *confidence* of the network that a given sample belongs to a particular class. This interpretation of softmax layer outputs is a widely used concept in the related literature.

## B. CONVOLUTIONAL NEURAL NETWORKS

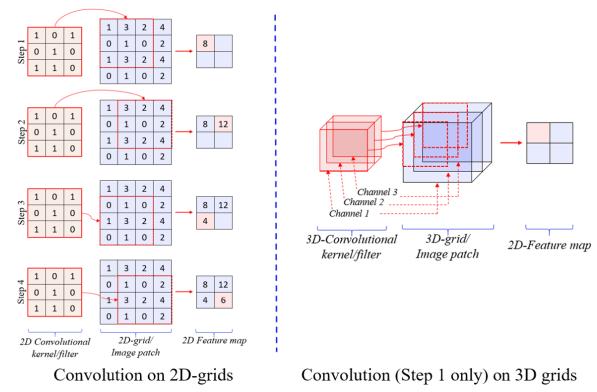
In the context of DL techniques for image analysis, Convolution Neural Networks (CNNs) [23], [32] are of the primary importance. Similar to the standard ANNs, CNNs consist of multiple layers. However, instead of simple perceptron layers, we encounter three different kinds of layers in these networks (a) Convolutional layers, (b) Pooling layers, and (c) Fully connected layers (often termed fc-layers). We describe these layers below, focusing mainly on the Convolutional layers that are the main source of strength for CNNs.

### 1) CONVOLUTIONAL LAYERS

The aim of convolution layers is to learn weights of the so-called<sup>2</sup> convolutional *kernels/filters* that can perform convolution operations on images. Traditional image analysis has a long history of using such filters to highlight/extract different image features, e.g. Sobel filter for detecting edges in images [33]. However, before CNNs, these filters needed to be designed by manually setting the weights of the kernel in a careful manner. The breakthrough that CNNs provided is in the automatic learning of these weights under the neural network settings.

We illustrate the convolution operation in Fig. 3. In 2D settings (e.g. grey-scale images), this operation involves moving a small window (i.e. kernel) over a 2D grid (i.e. image).

<sup>2</sup>Strictly speaking, the kernels in CNNs compute cross-correlations. However, they are always referred to as ‘convolutional’ by convention. Our subsequent explanation of the ‘convolution’ operation is in-line with the definition of this operation used in the context of CNNs.

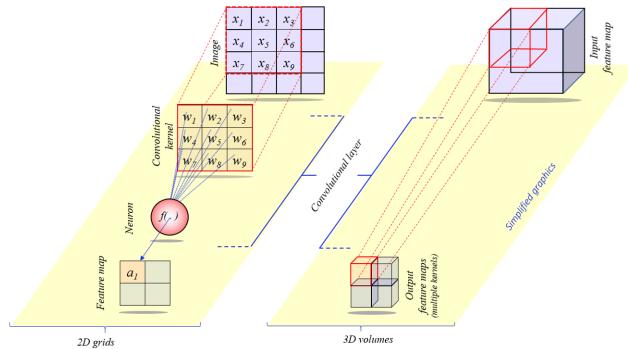


**FIGURE 3. Illustration of convolution operation for 2D-grids (left) and 3D-volumes (right). Complete steps of moving window are shown for the 2D case whereas only step-1 is shown for 3D. In 3D, the convolved channels are combined by simple addition, still resulting in 2D feature maps.**

In each moving step, the corresponding elements of the two grids get multiplied and summed up to compute a scalar value. Concluding the operation results in another 2D-grid, referred to as the *feature/activation* map in the CNN literature. In 3D settings, the same steps are performed for the individual pairs of the corresponding channels of the 3D volumes, and the resulting feature maps are simply added to compute a 2D map as the final output. Since color images have multiple channels, convolutions in 3D settings are more relevant for the modern CNNs. However, for the sake of better understanding, we often discuss the relevant concepts using the 2D grids. These concepts are readily transferable to the 3D volumes.

A convolutional layer of CNN forces the elements of kernels to become the network weights, see Fig. 4 for illustration. In the figure, we can directly compute the activation  $a_1$  (2D grids) using Eq. (1), where  $\mathbf{x}, \mathbf{w} \in \mathbb{R}^9$  are vectors formed by arranging  $w_1, \dots, w_9$  and  $x_1, \dots, x_9$  in the figure. The figure does not show the bias term, which is generally ignored in the convolutional layers. It is easy to see that under this setting, we can make use of the same tools to learn the convolutional kernel weights that we use to learn the weights of a standard ANN. The same concept applies to the 3D volumes, with a difference that we must use ‘multiple’ kernels to get a volume (instead of a 2D grid) at the output. Each feature map resulting from a kernel then acts as a separate channel of the output volume of a convolutional layer. It is a common practice in CNN literature to simplify the illustrations in 3D settings by only showing the input and output volumes for different layers, as we do in Fig. 4.

From the perspective presented above, a convolutional layer may look very similar to a standard perceptron layer, discussed in Section II-A. However, there are two major differences between the two. (1) Every input feature gets connected to its activation signal through the same kernel (i.e. weights). This implies that all input features *share* the kernel weights - called *parameter sharing*. Consequently,



**FIGURE 4.** Working of a convolutional layer. CNNs force kernel weights to become network parameters. (Left) In 2D grids, a single kernel moves over the image/input signal. (Right) A volume of multiple kernel moves over the input volume to result in an output volumes.

the kernels try to adjust their weights such that they resonate well to the basic building patterns of the whole input signal, e.g. edges for an image. (2) Since the same kernel connects all input features to output features/activation, convolutional layers have very few parameters to learn in the form of kernel weights. This *sparsity of connections* allows very efficient learning despite the high dimensionality of data - a feat not possible with standard *densely* connected perceptron layers.

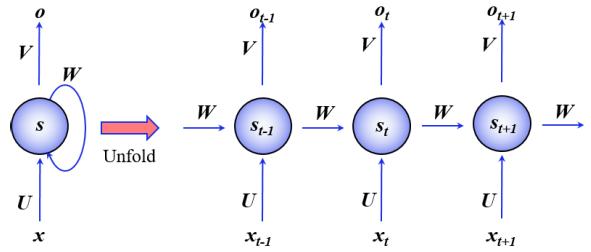
## 2) POOLING LAYERS

The main objective of a pooling layer is to reduce the width and height of the activation maps in CNNs. The basic concept is to compute a single output value ‘v’ for a small \$n\_p \times n\_p\$ grid in the activation map, where ‘v’ is simply the maximum or average value of that grid in the activation map. Based on the used operation, this layer is often referred as *max-pooling* or *average-pooling* layer. Interestingly, there are no learnable parameters associated with a pooling layer. Hence, this layer is sometimes seen as a part of Convolutional layer. For instance, the popular VGG-16 network [34] does not see pooling layer as a separate layer, hence the name VGG-16. On the other hand, other works, e.g. [35] that use VGG-16 often count more than 16 layers in this network by treating the pooling layer as a regular network layer.

## 3) FULLY CONNECTED LAYERS

These layers are the same as the perceptron layers encountered in the standard ANNs. The use of multiple Convolutional and Pooling layers in CNNs gradually reduces the size of resulting activation maps. Finally, the activation maps from a deeper layer are re-arranged into a vector which is then fed to the fully connected (fc) layers. It is a common knowledge now that the activation vectors of fc-layers often serve as very good compact representations of the input signals (e.g. images).

Other than the above mentioned three layers, *Batch Normalization* (BN) [36] is another layer that is now a days encountered more often in CNNs than in the standard ANNs. The main objective of this layer (with learnable parameters)



**FIGURE 5.** Illustration of unfolded RNN [1]. A state \$s\$ is shown with a single neuron for simplicity. At the \$t^{\text{th}}\$ time stamp, the network updates its state \$s\_t\$ based on the input \$x\_t\$ and the previous state. It optionally outputs a signal \$o\_t\$. \$U\$, \$V\$ and \$W\$ are network parameters.

is to control the mean and variance of the activation values of different network layers such that the induction of the overall model becomes more efficient. This idea is inspired by a long known fact that induction of ANN models generally becomes easier if the inputs are normalized to have zero mean and unit variance. The BN layer essentially applies a similar principle to the activations of deep neural networks.

## C. RECURRENT NEURAL NETWORKS

Standard neural networks assume that input signals are independent of each other. However, often this is not the case. For instance, a word appearing in a sentence generally depends on sequence of the words preceding it. Recurrent Neural Networks (RNNs) are designed to model such sequences. An RNN can be thought to maintain a ‘memory’ of the sequence with the help of its internal states. In Fig. 5, we show a typical RNN that is *unfolded* - complete network is shown for the sequence. If the RNN has three layers, it can model e.g. sentences that are three words long. In the figure, \$x\_t\$ is the input at the \$t^{\text{th}}\$ time stamp. For instance, \$x\_t\$ can be some quantitative representation of the \$t^{\text{th}}\$ word in a sentence. The memory of the network is maintained by the state \$s\_t\$ that is computed as:

$$s_t = f(Ux_t + Vs_{t-1}), \quad (5)$$

where \$f(\cdot)\$ is typically a non-linear activation function, e.g. ReLU. The output at a given time stamp \$o\_t\$ is a function of a weighted version the network state at that time. For instance, predicting probability of the next word in a sentence can assume the output form \$o\_t = \text{softmax}(Vs\_t)\$, where ‘softmax’ is the same operation discussed in Section II-A.

One aspect to notice in the above equations is that we use the same weight matrices \$U\$, \$V\$, \$W\$ at all time stamps. Thus, we are recursively performing the same operations over an input sequence at multiple time stamps. This fact also inspires the name ‘Recursive’ NN. It also has a significant implication that for an RNN we need a special kind of back-propagation algorithm, known as *back-propagation through time* (BTT). As compared to the regular back-propagation, BTT must propagate error recursively back to the previous time stamps. This becomes problematic for long sequences that involve too many time stamps. A phenomenon known as

*vanishing/exploding gradient* is the root cause of this problem. This has lead RNN researcher to focus on designing networks that can handle longer sequences. Long Short-Term Memory (LSTM) network [37] is currently a popular type of RNN that is found be reasonably effective for dealing with long sequences.

LSTM networks have the same fundamental architecture of an RNN, however their hidden states are computed differently. The hidden units are commonly known as *cells* in the context of LSTMs. Informally, a cell takes in the previous state and the input at a given time stamp and decides on what to remember and what to erase from its memory. The previous state, current input and the memory is then combined for the next time stamp.

#### D. USING NEURAL NETWORKS FOR UNSUPERVISED LEARNING

Whereas we assumed availability of the label for each data sample while discussing the basic concepts of neural networks in the preceding subsections, those concepts can also be applied readily to construct neural networks to model data without labels. Here, we briefly discuss the mainstream frameworks that allow to do so. It is noteworthy that this article does not present neural networks as ‘supervised vs unsupervised’ intentionally. This is because the core concepts of neural networks are generally better understood in supervised settings. Unsupervised use of neural networks simply requires to employ the same ideas under different overall frameworks.

##### 1) AUTOENCODERS

The main idea behind autoencoders is to map an input signal (e.g. image, feature vector) to itself using a neural network. In this process, we aim to learn a latent representation of the data that is more powerful for a particular task than the raw data itself. For instance, the learned representation could cluster better than the original data. Theoretically, it is possible to use any kind of network layers in autoencoders that are used for supervised neural networks. The uniqueness of autoencoders comes in the output layer where the signal is the same as the input signal to the network instead of e.g. a label vector in classification task.

Mapping a signal to itself can result in trivial models (learning identity mapping). Several techniques have been adopted in the literature to preclude this possibility, leading to different kinds of autoencoders. For instance, *under-complete* autoencoders ensure that the dimensionality of the latent representation is much smaller than the data dimension. In MLP settings, this can be done by using a small number (as compared to the input signal’s dimension) of neurons in a hidden layer of the network, and use the activations of that layer as the latent representation. *Regularized* autoencoders also impose sparsity on neuron connections [38] and reconstruction of the original signal from its noisy version [39] to ensure learning of useful latent representation instead of identity mapping. Variational autoencoders [40] and contractive

autoencoders [41] are also among the other popular types of autoencoders.

##### 2) GENERATIVE ADVERSARIAL NETWORKS

Recent years have seen an extensive use of Generative Adversarial Networks (GANs) [42] in natural image analysis. GANs can be considered a variation of autoencoders that aim at mimicking the distribution generating the data. GANs are composed of two parts that are neural networks. The first part, termed *generator*, has the ability to generate a sample whereas the other, called *discriminator* can classify the sample as a real or fake. Here, a ‘real’ sample means that it is actually coming from the training data. The two networks essentially play a game where the generator tries to fool the discriminator by generating more and more realistic samples. In the process the generator keeps updating its parameters to produce better samples. The adversarial objective of the generator to fool the discriminator also inspires the name of GANs. In natural image analysis, GANs have been successfully applied for many tasks, e.g. inducing realism in synthetic images [43], domain adaption [44] and data completion [45]. Such successful applications of GANs to image processing tasks also open new directions for medical image analysis tasks. In their recent surveys of the application of GANs in Medical Imaging domain, Yi et al. [46] and Kazemini et al. [47] provide reviews of the successful GAN based medical image analysis techniques. For the interested readers, Yi et al. also provide a live Github page that includes online links to the literature reviewed by the authors.

##### E. BEST PRACTICES IN USING CNNS FOR IMAGE ANALYSIS

Convolutional Neural Networks (CNNs) form the backbone of the recent breakthroughs in image analysis. To solve different problems in this area, CNN based models are normally used in three different ways. (1) A network architecture is chosen and trained *from scratch* using the available training dataset in an *end-to-end* manner. (2) A CNN model *pre-trained* on some large-scale dataset is *fine-tuned* by further training the model for a few epochs using the data available for the problem at hands. This approach is more suitable when limited training data is available for the problem under consideration. It is often termed *transfer learning* in the literature. (3) Use a model as a *feature extractor* for the available images. In this case, training/testing images are passed through the network and the activations of a specific layer (or a combination of layers) are considered as image features. Further analysis is performed using those features.

Computer Vision literature provides extensive studies to reflect on the best practices of exploiting CNNs in any of the aforementioned three manners. We can summarize the crux of these practices as follows. One should only consider training a model from scratch if the available training data size is very large, e.g. 50K image or more. If this is not the case, use transfer learning. If the training data is even smaller, e.g. few hundred images, it may be better to use

CNN only as a feature extractor. No matter which approach is adopted, it is better that the underlying CNN is inspired by a model that has already proven its effectiveness for a similar task. This is especially true for the ‘training from scratch’ approach. We refer to the most successful recent CNN models in the Computer Vision literature in the paragraphs to follow. For transfer learning, it is better to use a model that is pre-trained on data/problem that is as similar as possible to the data/problem at hands. In the case of using CNN as a feature extractor, one should prefer a network with more representation power. Normally, deeper networks that are trained on very large datasets have this property. Due to their discriminative abilities, features extracted from such models are especially useful for classification tasks.

Starting from AlexNet in 2012 [23], many complex CNN models have been developed in the last seven years. Whereas still useful, AlexNet is no longer considered a state-of-the-art network. A network still applied frequently is VGG-16 [34] that was proposed in 2014 by the Visual Geometry Group (VGG) of Oxford university. A later version of VGG-16 is VGG-19 that uses 19 instead of 16 layers of the learnable parameters. Normally, the representation power of both versions are considered similar. Another popular network is GoogLeNet [48] that is also commonly known as ‘Inception’ network. This network uses a unique type of layer called inception layer/block from which it drives its main strength. To date, four different versions of Inception [49], [50] have been introduced by the original authors, with each subsequent version having slightly better representation power (under a certain perspective) than its predecessor. ResNet [22] is another popular network that enables deep learning with models having more than hundred layers. It is based on a concept known as ‘residual learning’, which is currently highly favored by Pattern Recognition community because it enables very deep networks. DenseNet [51] also exploits the insights of residual learning to achieve the representation power similar to ResNet, but with a more compact network.

Whereas the above-mentioned CNNs are mainly trained for image classification tasks, Fully Convolutional Networks (FCNs) [52] and U-Net [53] are among the most popular networks for the task of image segmentation. Analyzing the architectures and hyperparameter settings of these networks can often reveal useful insights for developing new networks. In fact, some of these networks (e.g. Inception-v4/ResNet [50]) already rely on the insights from others (e.g. ResNet [22]). The same practice can yield popular networks in the future as well. We draw further on the best practices of using CNNs for image analysis in Section VI.

#### F. DEEP LEARNING PROGRAMMING FRAMEWORKS

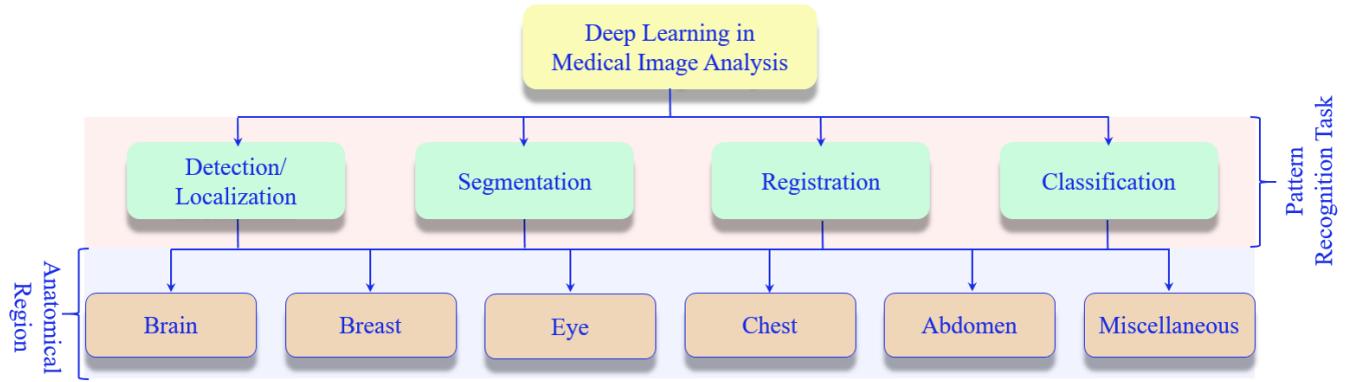
The rise of Deep Learning has been partially enabled by the public access to programming frameworks that implement the core techniques in this area in high-level programming languages. Currently, many of these frameworks are being continuously maintained by software developers and most of the new findings are incorporated in them rapidly. Whereas

availability of appropriate Graphical Processing Unit (GPU) is desired to fully exploit these modern frameworks, CPU support is also provided with most of them to train and test small models. The frameworks allow their users to directly test different network architectures and their hyperparameter settings etc. without the need of actually implementing the operations performed by the layers and the algorithms that train them. The layers and related algorithms come pre-implemented in the libraries of the frameworks.

Below we list the popular Deep Learning frameworks in use now a days. We order them based on their current popularity in Computer Vision and Pattern Recognition community for the problems of image analysis, starting from the most popular. Note that, we provide embedded links to the URLs of homepages for each of these frameworks. To reach those URLs, click on the red highlighted name of the framework. Also note that all red highlighted text in this document provides links to the relevant URLs for the convenience of its readers.

- *Tensorflow* [54] is originally developed by Google Brain, it is fast becoming the most popular deep learning framework due to its continuous development. It provides Python and C++ interface.
- *PyTorch* [55] is a Python based library supported by Facebook’s AI Research. It is currently receiving significant attention due to its ability to implement dynamic graphs.
- *Caffe2* [56] builds on Caffe (see below) and provides C++ and Python interface.
- *Keras* [57] can be seen as a high level programming interface that can run on top of Tensorflow and Theano [58]. Although not as flexible as other frameworks, Keras is particularly popular for quickly developing and testing networks using common network layers and algorithms. It is often seen as a gateway to deep learning for new users.
- *MatConvNet* [59] is the most commonly used public deep learning library for Matlab.
- *Caffe* [56] was originally developed by UC Berkeley, providing C++ and Python interface. Whereas Caffe2 is now fast replacing Caffe, this framework is still in use because public implementations of many popular networks are available in Caffe.
- *Theano* [58] is a library of Python to implement deep learning techniques that is developed and supported by MILA, University of Montreal.
- *Torch* [60] is a library and scripting language based on Lua. Due to its first release in 2011, many first generation deep learning models were implemented using Torch.

The above is not an exhaustive list of the frameworks for Deep Learning. However, it covers those frameworks that are currently widely used in image analysis. It should be noted, whereas we order the above list in terms of the ‘current’ trend in popularity of the frameworks, public implementations of many networks proposed in 2012 - 2015 are originally found



**FIGURE 6.** Taxonomy of literature review: The contributions exploiting deep learning technology are first categorized according to the underlying pattern recognition tasks. Each category is then sub-divided based on the human anatomical region studied in the papers.

in e.g. Torch, Theano or Caffe. This also makes those frameworks equally important. However, it is often possible to find public implementations of legacy networks in e.g. Tensorflow, albeit not by the original authors.

#### G. PRE-PROCESSING FOR DEEP LEARNING

Pre-processing of images is often considered an important component of image analysis pipeline. Pre-processing steps like image filtering, de-noising and dimensionality reduction etc. have been known to play an important role in the performance of conventional image analysis methods. However, one of the major advantages of deep learning methods is that appropriately learned deep models are able to perform many of these steps implicitly. For instance, in the case of CNNs, the learned feature maps are actually filtered forms of the input, where the filters are also learned by the learning algorithm. Since the filters used by a CNN are tailored to the task at hand (due to learning), they are able to easily outperform the generic hand-crafted filters used by the conventional pre-processing methods. Nevertheless, appropriately filtered data still remains preferable for deep learning. Unfiltered data can cause the induced models to learn idiosyncrasies of the unwanted signals in the input. This can lead to sub-optimal model performance if the same peculiarities are not present in the testing data.

Since learning deep models is generally computationally expensive, one may be tempted to employ input dimensionality reduction during the model training. Reducing the data dimensionality with lossy methods can result in the induction of inferior deep models. A more preferable choice for deep learning is to simply reduce the input size, e.g. by using overlapping small patches of images instead of larger images. Due to their ability to learn complex functions, deep networks are able to learn even minute data details. Lossy dimensionality reduction can deprive deep learning of this property.

As a general rule of thumb for the data pre-processing in deep learning, it should be ensured that training and testing data goes through the exact same pre-processing steps. This is important because a deep model essentially learns a mapping from an input data distribution to its outputs. Any

dissimilarity in the data-preprocessing can change the data distribution, leading to incorrect output predictions. Another rule for deep models is that avoiding data pre-processing generally comes at the cost of training more complex networks, as compared to the models required for clean data.

### III. DEEP LEARNING METHODS IN MEDICAL IMAGE ANALYSIS

In this Section, we review the recent contributions in Medical Image Analysis that exploit the Deep Learning technology. We mainly focus on the research papers published after December 2017, while briefly mentioning the more influential contributions from the earlier years. For a comprehensive review of the literature earlier than the year 2018, we collectively recommend the following articles [24], [28], [29]. Taking more of a Computer Vision/Machine Learning perspective, we first categorize the existing literature under ‘Pattern Recognition’ tasks. The literature pertaining to each task is then further sub-categorized based on the human anatomical regions. The taxonomy of our literature review is depicted in Fig. 6.

#### A. DETECTION/LOCALIZATION

The main aim of *detection* is to identify a particular region of interest in an image and draw a bounding box around it, e.g. brain tumor in MRI scans. Hence, *localization* is also another term used for the detection task. In Medical Image Analysis, detection is more commonly referred to as Computer Aided Detection (CAD). CAD systems are aimed at detecting the earliest signs of abnormality in patients. Lung and breast cancer detection can be considered as the common applications of CAD.

##### 1) BRAIN

For the anatomic region of brain, Islam and Zhang *et al.* [61] employed a CNN for the detection of Alzheimer’s Disease (AD) using the MRI images of OASIS data set [62]. The authors built on two baseline CNN networks, namely Inception-v4 [50] and ResNet [22], to categorize four classes of AD. These classes include moderate, mild, very mild

and non-demented patients. The accuracies reported by the authors for these classes are 33%, 62%, 75%, and 99%, respectively. It is claimed that the proposed method does not only perform well on the used dataset, but it also has the potential to generalize to ADNI dataset [63]. Chen and Konukoglu [64] proposed an unsupervised learning approach using an Auto-Encoders (AE). The authors investigated lesion detection using Variational Auto Encoder (VAE) [40] and Adversarial Auto Encoder (AAE) [65]. The analysis is carried out on BRATS 2015 datasets, demonstrating good results for the Area Under Curve (AUC) metric.

Alaverdyan *et al.* [66] used a deep neural network for epilepsy lesion detection in multiparametric MRI images. They also stacked convolution layers in an auto-encoders fashion and trained their network using the patches of the original images. Their model was trained using the data from 75 healthy subjects in an unsupervised manner. For the automated brain tumor detection in MR images Panda *et al.* [67] used discriminative clustering method to segregate the vital regions of brain such as Cerebro Spinal Fluid (CSF), White Matter (WM) and Gray Matter (GM). In another study of automatic detection in MR images [68], Laukampet *et al.* used multi-parametric deep learning model for the detection of meningiomas in brain.

## 2) BREAST

In assessing cancer spread, histopathological analysis of Sentinel Lymph Nodes (SLNs) becomes important for the task of cancer staging. Bejnordi *et al.* [69] analyzed deep learning techniques for metastases detection in eosin-stained tissues and hematoxylin tissue sections of lymph nodes of the subjects with cancer. The computational results are compared with human pathologist diagnoses. Interestingly, out of the 32 techniques analyzed, the top 5 deep learning algorithms arguably out-performed eleven pathologists.

Chiang *et al.* [70] developed a CAD technique based on a 3D CNN for breast cancer detection using Automated whole Breast Ultrasound (ABUS) imaging modality. In their approach, they first extracted Volumes of Interest (VOIs) through a sliding window technique, then the 3D CNN was applied and tumor candidates were selected based on the probability resulting from the application of 3D CNN to VOIs. In the experiments 171 tumors are used for testing, achieving sensitivities of up to 95%. Dalmış *et al.* [71] proposed a CNN based CAD system to detect breast cancer in MRI images. They used 365 MRI scans for training and testing, out of which 161 were malignant lesions. They claimed the achieved sensitivity obtained by their technique to be better than the existing CAD systems. For the detection of breast mass in mammography images, Zhang *et al.* [72] developed a Fully Convolutional Network (FCN) based end-to-end heatmap regression technique. They demonstrated that mammography data could be used for digital breast tomosynthesis (DBT) to improve the detection model. They used transfer learning by fine-tuning an FCN model on mammography images. The approach is tested on tomosynthesis

data with 40 subjects, demonstrating better performance as compared to the model trained from scratch on the same data.

## 3) EYE

For the anatomical region of eye, Li *et al.* [73] recently employed a deep transfer learning approach which fine tunes the VGG-16 model [34] that is pre-trained on ImageNet [74] dataset. To detect and classify Age-related Macular Degeneration (AMD) and Diabetic Macular Edema (DME) diseases in eye, they used 207,130 retinal Optical Coherence Tomography (OCT) images. The proposed method achieved 98.6% prediction detection accuracy in retinal images with 100%. Abràmoff *et al.* [75] used a CNN based technique to detect Diabetic Retinopathy (DR) in fundus images. They assessed the device IDx-DR X 2.1 in their study using a public dataset [76] and achieve an AUC score of 0.98. Schlegl *et al.* [77] employed deep learning for the detection and quantification of Intraretinal Cystoid Fluid (IRC) and Subretinal Fluid (SRF) in retinal images. They employed an auto encoder-decoder formation of CNNs, and used 1,200 OCT retinal images for the experiments, achieving AUC of 0.92 for SRF and AUC of 0.94 for IRC.

Deep learning is also being increasingly used for diagnosing retinal diseases [78], [79]. Li *et al.* [80] trained a deep learning model based on the Inception architecture [48] for the identification of Glaucomatous Optic Neuropathy (GON) in retinal images. Their model achieved AUC of 0.986 for distinguishing healthy from GON eyes. Recently, Christopher *et al.* [81] also used transfer learning with VGG16, Inception v3, and ResNet50 models for the identification of GON. They used pre-trained models of ImageNet. For their experiments, they used 14,822 Optic Nerve Head (ONH) fundus images of GON or healthy eyes. The achieved best performance for identifying moderate-to-severe GON in the eyes was reported to be AUC value 0.97 with 90% sensitivity and 93% specificity. Khojasteh *et al.* [82] used pre-trained ResNet-50 on DIARETDB1 [83] and e-Ophtha [84] datasets for the detection of exudates in the retinal images. They reported an accuracy of 98% with 99% sensitivity of detection on the used data.

## 4) CHEST

For the pulmonary nodule detection in lungs in Computed Tomography (CT) images, Zhu *et al.* [85] proposed a deep network called DeepEM. This network uses a 3D CNN architecture that is augmented with an Expectation-Maximization (EM) technique for the noisily labeled data of Electronic Medical Records (EMRs). They used the EM technique to train their model in an end-to-end manner. Three datasets were used in their study, including; the LUNA16 dataset [86] - the largest publicly available dataset for supervised nodule detection, NCI NLST dataset<sup>3</sup> for weakly supervised detection and Tianchi Lung Nodule Detection dataset.

<sup>3</sup><https://biometry.nci.nih.gov/cdas/datasets/nlst/>

For the detection of artifacts in Cardiac Magnetic Resonance (CMR) imaging, Oksuz *et al.* [87] also proposed a CNN based technique. Before training the model, they performed image pre-processing by normalization and region of interest (ROI) extraction. The authors used a CNN architecture with 6-convolutional layers (ReLU activations) followed by 4-pooling layers, 2 fc layers and a softmax layer to estimate the motion artifact labels. They showed good performance for the classification of motion artifacts in videos. The authors essentially built on the insights of [88] in which video classification is done using a spatio-temporal 3D CNN.

Li *et al.* [89] proposed a technique for the localization and identification of thoracic diseases in public database NIH X-ray<sup>4</sup> that comprises 120 frontal view X-ray images with 14 labels. Their model performs the tasks of localization and identification simultaneously. They used the popular ResNet [22] architecture to build the computational model. In their model, an input image is passed through the CNN for feature map extraction, then a max pooling or bi-linear interpolation layer is used for resizing the input image by a patch slicing layer. Afterwards, fully convolutional layers are used to eventually perform the recognition. For training, the authors exploit the framework of Multi-Instance Learning (MIL), and in the testing phase, the model predicts both labels and class specific localization details. Yi *et al.* [90] presented a scale recurrent network for the detection of catheter in X-ray images. Their network architecture is organized in an auto encoder-decoder manner. In another study [91], Masood *et al.* proposed a deep network, termed DFCNet, for the automatic computer aided lung pulmonary detection.

González *et al.* [92] proposed a deep network for the detection of Chronic Obstructive Pulmonary Disease (COPD) and Acute Respiratory Disease (ARD) prediction in CT images of smokers. They trained a CNN using 7,983 COPDGene cases and used logistic regression for COPD detection and ARD prediction. In another study [93], the same group of researchers used deep learning for weakly supervised lesion localization. Recently, Winkels and Cohen [94] used NLST and LDIC/IDRI [95] datasets for lung nodule detection in CT images. They proposed a 3D Group-equivariant Convolutional Neural Network (G-CNN) technique for that purpose. The proposed method was exploited for fast positive reduction in pulmonary lung nodule detection. The authors claim their method performs on-par with standard CNNs while trained using ten times less data.

## 5) ABDOMEN

Alensary *et al.* [96] proposed a deep reinforcement learning technique for the detection of multiple landmarks with ROIs in 3D fetal head scans. Ferlaino *et al.* [97] worked on placental histology using deep learning. They classified five different classes with an accuracy of 89%. Their model also learns deep embedding encoding phenotypic knowledge that classifies five different cell populations and learns inter-class

variances of phenotype. Ghesu *et al.* [98] used a large data of 1,487 3D CT scans for the detection of anatomic sites, exploiting multi-scale deep reinforcement learning.

Katzmann *et al.* [110] proposed a deep learning based technique for the estimation of Colorectal Cancer (CRC) in CT tumor images for early treatment. Their model achieved high accuracies in growth and survival prediction. Meng *et al.* [111] formulated an automatic shadow detection technique in 2D ultrasound images using weakly supervised annotations. Their method highlights the shadow regions which is particularly useful for the segmentation task. Horie *et al.* [112] recently applied a CNN technique for esophageal cancer detection. They used 8,428 WGD images and attained 98% results for the sensitivity. Yasaka *et al.* [113] used a deep CNN architecture for the diagnosis of three different phases (noncontrast-agent enhanced, arterial, and delayed) of masses of liver in dynamic CT images.

## 6) MISCELLANEOUS

Zhang *et al.* [114] achieved 98.51% accuracy and a localization error 2.45mm for the detection of inner ear in CT images. They used 3D U-Net [115] to map the whole 3D image which consists of multiple convolution-pooling layers that convert the raw input image into the low resolution and highly abstracted feature maps. They applied false positive suppression technique in the training process and used a shape based constraint during training. Rajpurkar *et al.* [116] recently released a data set MURA which consists of 40,561 images from 14,863 musculoskeletal studies labeled by radiologists as either normal or abnormal. The authors used CNN with 169-layers for the detection of normality and abnormality in each image study. Li and Ping [117] proposed a Neural Conditional Random Field (NCRF) technique for the metastasis cancer detection in whole slide images. Their model was trained end-to-end using back-propagation and it obtained successful FROC score of 0.8096 in testing using Camelyon16 dataset [118]. Codella *et al.* [119] recently organized a challenge at the International Symposium on Biomedical Imaging (ISBI), 2017 for skin lesion analysis for melanoma detection. The challenge task provided 2,000 training images, 150 validation images, and 600 images for testing. It eventually published the results of 46 submission. We refer to [119] for further details on the challenge itself and the submissions. We also mention few techniques in Table 2 related to the task of detection/localization. These methods appeared in the literature in the years 2016-17. Based on Google Scholar's citation index, these methods are among the highly influential techniques in the current related literature. This article also provides similar summaries of the highly influential papers from the years 2016-17 for each pattern recognition task considered in the Sections to follow.

## B. SEGMENTATION

In Medical Image Analysis, deep learning is being extensively used for image segmentation with different

<sup>4</sup><https://www.kaggle.com/nih-chest-xrays/data>

**TABLE 2.** Summary of highly influential papers appeared in 2016 and 2017 (based on Google Scholar citation index in January 2019) that exploit deep learning for detection/localization in medical image analysis.

Reference	Anatomic Site	Image Modality	Network type	Data	Citations
Shin <i>et al.</i> [99] (2016)	Lung	CT	CNN	Private	783
Sirinukunwattana <i>et al.</i> [100] (2016)	Abdomen	Histopathology	CNN	20000+ images	252
Setio <i>et al.</i> [101] (2016)	Lung	CT	CNN	888 images	247
Xu <i>et al.</i> [102] (2016)	Breast	Histopathology	AE	500 images	220
Wang <i>et al.</i> [103] (2016)	Breast	Histopathology	CNN	400 images	182
Kooi <i>et al.</i> [104] (2017)	Breast	Mammography	CNN	45000 images	162
Rajpurkar <i>et al.</i> [105] (2017)	Chest	X-ray	CNN	100000+ images	139
Liu <i>et al.</i> [106] (2017)	Breast	Histopathology	CNN	400 slides	98
Ghafoorian <i>et al.</i> [107] (2017)	Brain	MRI	CNN	LUNA16 ISBI 2016	90
Dou <i>et al.</i> [108] (2017)	Lung	CT	3D CNN	1075 images	40
Zhang <i>et al.</i> [109] (2017)	Brain	MRI	FCN	700 subjects	32

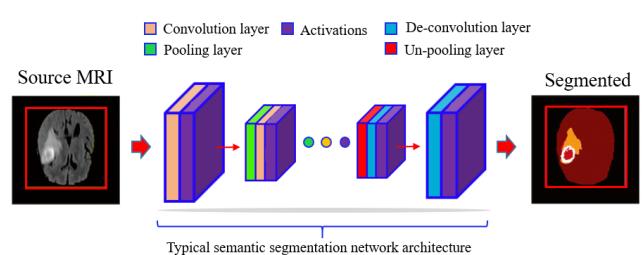
modalities, including Computed Tomography (CT), X-ray, Positron-Emission Tomography (PET), Ultrasound, Magnetic Resonance Imaging (MRI) and Optical Coherence Tomography (OCT) etc. Segmentation is the process of partitioning an image into different meaningful segments (that share similar characteristics) through automatic or semi-automatic outlining of the boundaries within the image. In medical imaging, these segments usually commensurate to different tissue classes, pathologies, organs or some other biological structure [120].

Before the popularity of deep learning, many conventional techniques were commonly used for the segmentation task in Medical Image Analysis, including region growing [121], [122], graph partitioning [123], [124] and genetic algorithms [125]. However, modern deep learning techniques are able to consistently outperform the conventional techniques. The main strength of deep learning based segmentation methods comes from the fact that deep networks are able to learn very complex functions required for the task. Moreover, application of the learned computational models needs minimal human intervention and heuristics that are commonly required by the other segmentation techniques. This makes deep learning particularly suitable for medical image segmentation.

In Fig. 7, we show a typical process of segmentation using a deep learning model. The models used for segmentation first compress the input image dimensions with the help of a stack of convolutional and pooling layers. This is done to learn a compact latent space where the regions to be segmented enter a more discriminative representation. The latter half of the networks expands the feature maps to the original image dimensions, and provides the segmentation labels for the region of interest.

## 1) BRAIN

Related to the anatomical region of brain, Dey and Hong [126] trained a complementary segmentation network, termed CompNet, for skull stripping in MRI scans for normal and pathological brain images. The OASIS dataset [127] was used for the training purpose. In their approach, the features used for segmentation are learned using an encoder-decoder



**FIGURE 7.** Typical process of segmentation with deep learning: A convolutional neural network (CNN) based model is learned that first compresses the source image with a stack of different convolution, activation and pooling layers. The compressed latent representation is expanded by the inverse operations. The network is kept end-to-end trainable. At the test time, a forward pass provides the segmentation labels.

network trained from the images of brain tissues and its complimentary part outside the brain. The approach is compared with a plain U-Net and a dense U-Net [128]. The accuracy achieved by the CompNet for the normal images is 98.27%, and for the pathological images is 97.62%. These results are better than those achieved by [128].

Zhao *et al.* [129] proposed a deep learning technique for brain tumor segmentation by integrating Fully Convolutional Networks (FCNs) and Conditional Random Fields (CRFs) in a combined framework to achieve segmentation with appearance and spatial consistency. They trained 3 segmentation models using 2D image patches and slices. First, the training is performed for the FCN using image patches, then CRF is trained with a Recurrent Neural Network (CRF-RNN) using image slices. During this phase, the parameters of the FCN are fixed. Afterwards, FCN and CRF-RNN parameters are jointly fine tuned using the image slice. The authors used the MRI image data provided by Multimodal Brain Tumor Image Segmentation Challenge (BRATS) 2013, BRATS 2015 and BRATS 2016. In their work, Nair *et al.* [130] used a 3D CNN approach for the segmentation and detection of Multiple Sclerosis (MS) lesions in MRI sequences. Roy *et al.* [131] used voxel-wise Bayesian FCN for the whole brain segmentation by using Monte Carlo sampling. They demonstrated high accuracies on four datasets, namely MALC, ADNI-29, CANDI-13, and IBSSR-18. Robinson *et al.* [132]

also proposed a real-time deep learning approach for the cadaver MR segmentation.

## 2) BREAST

In their study, Singh *et al.* [133] described a conditional Generative Adversarial Networks (cGAN) model for breast mass segmentation in mammography. Experiments were conducted on Digital Database for Screening Mammography (DDSM) public dataset and private dataset of mammograms from Hospital Universitari Sant Joan de Reus-Spai. They additionally used a simpler CNN to classify the segmented tumor area into different shapes (irregular, lobular, oval and round) and achieved an accuracy of 72% on DDSM dataset. Zhang *et al.* [134] exploited deep learning for image intensity normalization in breast segmentation task. They used 460 subjects from Dynamic Contrast-enhanced Magnetic Resonance Imaging (DCEMRI). Each subject contained one T1 weighted pre-contrast and three T1 weighted post-contrast images. Men *et al.* [135] trained a deep dilated ResNet for segmentation of Clinical Target Volume (CTV) in breasts. Lee and Nishikawa [136] based on fully convolution neural network proposed an automated segmentation technique for the breast density estimation in mammograms. For the evaluation of their approach, they used full-field digital screening mammograms of 604 subjects. They fine tuned the pre-trained network for breast density segmentation and estimation. The Percent Density (PD) estimation by their approach showed similarities with BI-RADS density assessment by radiologists and outperformed the then state-of-the-art computational approaches.

## 3) EYE

Retinal blood image segmentation is considered an important and a challenging task in retinopathy. Zhang and Chung [137] used a deep neural network for this purpose by exploiting the U-Net architecture [53] with residual connection. They shown their results on three public datasets STARE [138], CHASEDB1 [139] and DRIVE [140] and achieved an AUC value of 97.99% for the DRIVE dataset. De Fauw *et al.* [141] applied deep learning for the retinal tissue segmentation. They used 14,884 three dimensional OCT images for training their network. Their approach is claimed to be device independent - it maintains segmentation accuracy while using different device data. In another study of retinal blood vessels, Jebaseeli *et al.* [142] proposed a method to enhance the quality of retinal vessel segmentation. They analyzed the severity level of diabetic retinopathy. Their proposed method, Deep Learning Based SVM (DLBSVM) model uses DRIVE, STARE, REVIEW, HRF, and DRIONS databases for training. Liu *et al.* [143] proposed a semi-supervised learning for retinal layer and fluid region segmentation in retinal OCT B-scans. Adversarial technique was exploited for the unlabeled data. Their technique resembles the U-Net fully convolutional architecture.

## 4) CHEST

Duan *et al.* [144] proposed a Deep Nested Level Set (DNLS) technique for the multi-region segmentation of cardiac MR images in patients with Pulmonary Hypertension (PH). They compared their approach with a CNN method [145] and the conditional random field (CRF) CRF-CNN approach [146]. DNLS is shown to outperform those techniques for all anatomical structures, specifically for myocardium. However, it requires more computations than [145] which is the fastest method among the tested approaches. Bai *et al.* [147] used FCN and RNN for the pixel-wise segmentation of Aortic sequences in MR images. They trained their model in an end-to-end manner from sparse annotations by using a weighted loss function. The proposed method consists of two parts, the first extracts features from the FCN using a U-Net architecture [53]. The second feeds these features to an RNN for segmentation. Among the used 500 Aortic MR images (provided by the UK Biobank), the study used random 400 images for training, and the rest were used for testing the models. Another recent study on semi supervised myocardial segmentation has been conducted by Chartsias *et al.* [148], which was presented as an oral paper in MICCAI 2018. Their proposed network, called Spatial Decomposition Network (SDNet), model 2D input images in two representations, namely spatial representation of myocardial as a binary mask and a latent representation of the remaining features in a vector form. While not being a fully supervised techniques, their method still achieves remarkable results for the segmentation task.

Kervadec *et al.* [149] proposed a CNN based ENet [150] constrained loss function for segmentation of weakly supervised Cardiac images. They achieved 90% accuracy on the public datasets of 2017 ACDC challenge<sup>5</sup>. Their approach is closes the gap between weakly and fully supervised segmentation in semantic medical imaging. In another study of cariac CT and MRI images for multi-class image segmentation, Joyce *et al.* [151] proposed an adversarial approach consisting of a shallow UNet like method. They also demonstrated improved segmentation with an unsupervised cost.

LanLonde and Bageci [152] introduced a CNN based technique, termed SegCaps, by exploiting the capsule networks [153] for object segmentation. The authors exploited the LUNA16 subset of the LIDC-IDRI database and demonstrated the effectiveness of their method for analyzing CT lung scans. It is shown that the method achieves better segmentation performance as compared to the popular U-Net. The SegCaps are able to handle large images with size 512×512. In another study [154] of lung cancer segmentation using the LUNA16 dataset, Nam *et al.* proposed a CNN model using 24 convolution layers, 2 pooling, 2 deconvolutional layers and one fully connected layer. Similarly, Burlutskiy *et al.* [155] developed a deep learning framework for lung cancer segmentation. They trained their model using the scans of 712 patients and tested on the scans

<sup>5</sup><https://www.creatis.insa-lyon.fr/Challenge/acdc/>

of 178 patients of fully annotated Tissue Micro-Arrays (TMAs). Their model is aimed at finding high potential cancer areas in TMA cores.

## 5) ABDOMEN

Roth *et al.* [156] built a 3D FCN model for automatic semantic segmentation of 3D images. The model is trained on clinical Computed Tomography (CT) data, and it is shown to perform automated multi-organ segmentation of abdominal CT with 90% average Dice score across all targeted organs. A CNN method, termed Kid-Net, is proposed for kidney vessels; artery, vein and collecting system (ureter) segmentation by Taha *et al.* [157]. Their model is trained in an end-to-end fashion using 3D CT-volume patches. One promising claim made by the authors is that their method reduces kidney vessels segmentation time from hours to minutes. Their approach uses feature down-sampling and up-sampling to achieve higher classification and localization accuracies. Their network training methodology also handles unbalanced data, and focuses on reducing false positives. It is also claimed that the proposed method enables high-resolution segmentation with a limited memory budget. The authors exploit the findings in [158] for that purpose.

Oktay *et al.* [159] recently presented an ‘attention gate’ model to automatically find the target anatomy of different shapes and sizes. They essentially extended the U-Net model to an attention U-Net model for pancreas segmentation. Their model can be utilized for organ localization and detection tasks. They used 120 images of CT for training their model, and 30 images for testing. Overall, the algorithm achieves good performance with 2 to 3% increase in dice score as compared to the existing methods. A related research on pancreas segmentation had been conducted previously using dense connection by Gibson *et al.* [160] and sparse convolutions by Heinrich *et al.* [161]–[163]. For multi-organ segmentation (i.e lung, heart, liver, bone) in unlabeled X-ray images, Zhang *et al.* [164] proposed a Task Driven Generative Adversarial Network (TD-GAN) automated technique. This is an unsupervised end-to-end method for medical image segmentation. They fine tuned a dense image-to-image network (DI2I) [51], [165] on synthetic Digitally Reconstructed Radiographs (DRRs) and X-ray images. In another study of multi organ segmentation, Tong *et al.* [166] proposed an FCN with a shape representation model. Their experiments were carried out on H&N datasets of volumetric CT scans.

Yang *et al.* [167] used a conditional Generative Adversarial Network (cGAN) to segment the human liver in 3D CT images. Lessmann *et al.* [168] proposed an FCN based technique for the automatic vertebra segmentation in CT images. The underlying architecture of their network is inspired by U-Net. Their model is able to process a patch size of  $128 \times 128 \times 128$  voxels. It achieves 95.8% accuracy for classification and 92.1% for segmentation in the spinal images used by the authors. Jin *et al.* [169] proposed a 3D CGAN to learn lung nodules conditioned on a Volume Of Interest (VOI) with an erased central region in 3D CT images. They trained their

model on 1,000 nodules taken from LIDC dataset. The proposed CGAN was further used to generate a dataset for Progressive Holistically Nested Network (P-HNN) model [170] which demonstrates improved segmentation performance.

## 6) MISCELLANEOUS

For memory and computational efficiency, Xu *et al.* [171] applied a quantization mechanism to FCNs for the segmentation tasks in Medical Image Analysis. They also used quantization to mitigate the over fitting issue for better performance. The effectiveness of the developed method is demonstrated for 2015 MICCAI Gland Challenge dataset [172]. As compared to [173] their method improves the results by up to 1% with  $6.4 \times$  reduction in the memory usage. Recently, Zhao *et al.* [129] proposed a deep learning technique for 3D image instance segmentation. Their model is trainable with weak annotations that needs 3D bounding boxes for all instances and full voxel annotations for only a small fractions of instances. Liu and Jiang [174] employed a novel deep reinforcement learning approach for the segmentation and classification of surgical gesture. Their approach performs well on JIGSAW dataset in terms of edit score as compared to previous similar works. Al Arif *et al.* [175] presented a deep FCN model called SPNet, as shape predictor for object segmentation. The X-ray images used in their study are of cervical vertebra. The dataset used in their experiments included 124 training and 172 test images. Their SPNet was trained for 30 epochs with a batch size of 50 images.

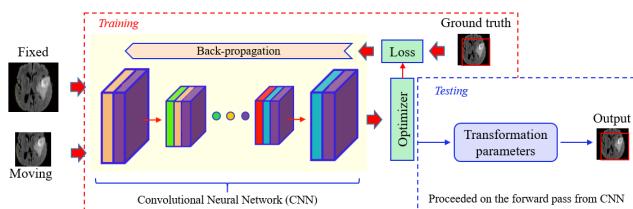
Sarker *et al.* [184] analyzed skin lesion segmentation with deep learning. They used autoencoder and decoder networks for feature extraction. The loss function is minimized in their work by combining negative Log Likelihood and end-point-error for the segmentation of melanoma regions with sharp edges. They evaluated their method SLSDeep on ISBI datasets [119], [185] for skin lesion detection, achieving encouraging segmentation results. In another related study of skin cancer, Mirikharaji and Hamarneh [186] also developed a deep FCN for skin lesion segmentation. They presented good results on ISBI 2017 dataset of dermoscopy images. They used two fully convolutional networks based on U-Net [53] and ResNet-DUC [22] in their technique. Yuan [187] proposed a deep fully convolutional-deconvolutional neural network (CDNN) for the automatic skin lesion segmentation, and acquired Jaccard index of 0.784 on the validation set of ISBI. Ambellan *et al.* [188] proposed a CNN based on 3D Statistical Shape Models (SSMs) for the segmentation of knee and cartilage in MRI images. In Table 3, we also summarize few popular methods in medical image segmentation that appeared prior to the year 2018.

## C. REGISTRATION

Image registration is a common task in medical image analysis that allows spatial alignment of images to a common anatomical space [189]. It aims at aligning a source image with a target image through transformations. Image

**TABLE 3.** Summary of influential papers appeared in 2016 and 2017 (based on Google Scholar citation index in January 2019) that exploit deep learning for the segmentation tasks in medical image analysis.

Reference	Anatomic Site	Image Modality	Network type	Data	Citations
Milletari et al. [158] (2016)	Brain	MRI	FCN	Private data	550
Cicek et al. [115] (2016)	Kidney	CT	3D U-Net	Private data	460
Pereira et al. [176] (2016)	Brain	MRI	CNN	BRATS2013, 2015	411
Moeskops et al. [177] (2016)	Brain	MRI	CNN	5 datasets	242
Liskowski et al. [178] (2016)	Eye	Ophthalmology	DNN	400 images, DRIVE, STARE, CHASE	177
Ibragimov et al. [179] (2016)	Breast	CT	AE	1400 images	174
Havaei et al. [180] (2017)	Brain	MRI	CNN	2013 BRATS	596
Kamnistas et al. [181] (2017)	Brain	MRI	11 layers 3D-CNN	BRATS 2015 and ISLES 2015	517
Fang et al. [182] (2017)	Eye	OCT	CNN	60 volumes	86
Ibragimov et al. [183] (2017)	Liver	CT	CNN	50 images	56



**FIGURE 8.** Common registration pipeline for using convolutional neural networks to quantify image similarity in intensity based image registration.

registration is one of the main stream tasks in medical image analysis that has received ample attention even before the deep learning era [190]–[196]. Advent of deep learning has also caused neural networks to penetrate in medical image registration [197]–[200].

In Fig. 8, we illustrate a generic framework that inserts a similarity metric into the intensity-based registration along with an interpolation strategy, transformation model, and an optimizer. Normally, a Convolutional Neural Network (CNN) is used for image representation that is fed with both the fixed and moving images in the context of image registration. During training, the network weights are updated with the ground truth annotations by back-projecting the loss computed by the optimizer. At the test time, the CNN output is processed further to compute the transformation parameters for the registration.

## 1) BRAIN

Deep learning based deformable image registration has also been recently performed by Sheikh et al. [201]. They used deep FCN to generate spatial transformations through under feed forward networks. In their experiments, they used cardiac MRI images from ACDC 2017 dataset and showed promising results in comparison to a moving mesh registration technique. Hou et al. [202] also proposed a learning based image registration technique using CNNs. They used 2D image slice transformation to construct 3D images using a canonical co-ordinate system. First, they simulated their approach on fetal MRI images and then used real fetal brain MRI images for the experiments. Their work is also claimed to be promising for computational efficiency. In another study

of image based registration [203], the same group of authors evaluated their technique on CT and MRI datasets for different loss functions using SE(3) as a benchmark. They trained CNN directly on SE(3) and proposed a Riemannian manifold based formulation for pose estimation problem. The registration accuracy with their approach increased from 2D to 3D image based registration as compared to previous methods. The authors further showed in [204] that CNN can reliably reconstruct 3D images using 2D image slices.

Recently, Balakrishnan et al. [205] worked on 3D pairwise MR brain image registration. They proposed an unsupervised learning technique named VoxelMorph CNN. They used a pair of two 3D images as input, with dimensions  $160 \times 192 \times 224$ ; and learned shared parameters in their network for convolution layers. They demonstrated their method on 8 publicly available datasets of brain MRI images. On ABIDE dataset their model achieved 1.5% improvement in the dice score. It is claimed that their method is also computationally more efficient than the exiting techniques for this problem.

## 2) EYE

Costa et al. [206] used adversarial autoencoders for the synthesis of retinal colored images. They trained a generative model to generate synthetic images and another model to classify its output into a real or synthetic. The model results in an end-to-end retinal image synthesis system and generates as many images as required by its users. It is demonstrated that the image space learned by the model has an arguably well defined semantic structure. The synthesized images were shown to be visually and quantitatively different from the images used for training their model. The shown images reflect good visual quality. Mahapatra et al. [207] proposed an end-to-end deep learning method using generative adversarial networks for multimodal image registration. They used retinal and cardiac images for registration. Tang et al. [208] demonstrated a robust image registration approach based on mixture feature and structure preservation (MFSP) non rigid point matching method. In their method they first extracted feature points by speed up robust feature (SURF) detector and partial intensity invariant feature descriptor (PIIFD) from model and target retinal image. Then they used MFSP for feature map detection.

Pan *et al.* [209] developed a deep learning technique to remove eye position difference for longitudinal 3D retinal OCT images. In their method, they first perform pre-processing for projection image then, to detect vessel shadows, they apply enhancement filters. The SURF algorithm [210] is then used to extract the feature points, whereas RANSAC [211] is applied for cleaning the outliers. Mahapatra [212] also proposed an end-to-end deep learning technique for image registration. They used GANs which registered images in a single pass with deformation field. They used ADAM optimizer [11] for minimizing the network cost and trained the model using the Mean Square Error (MSE) loss.

### 3) CHEST

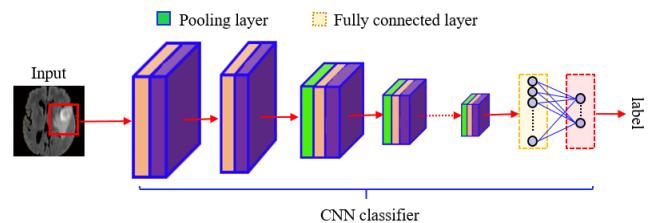
Relating to the anatomical region of chest, Eppenhof *et al.* [213] proposed a 3D FCN based technique for the registration of CT lung inspiration-expiration image pairs. They validated the performance of their method using two datasets, namely DIRLAB [214] and CREATIS [215]. In general, there is a growing perception in the Medical Imaging community that Deep learning is a promising tool for 2D and 3D image registration for the chest regions. de Vos *et al.* [216] also trained a CNN model for the affine and deformable image registration. Their technique allows registration of the pairs of unseen images in a single pass. They applied their technique to cardiac cine MRI and chest CT images for registration. Zheng *et al.* [217] trained a CNN model for 2D/3D image registration problem under a Pairwise Domain Adaptation (PDA) technique that uses synthetic data. It is claimed that their method can learn effective representations for image registration with only a limited number of training images. They demonstrated generalization and flexibility of their method for clinical applications. Their PDA method can be specially suitable where small training data is available.

### 4) ABDOMEN

Lv *et al.* [218] proposed a CNN based technique for the 3D MRI abdomen image registration. They trained their model for the spatial transformation analysis of different images. To demonstrate the effectiveness of their technique, they compared their method with three other approaches and claimed a reduction in the reconstruction time from 1h to 1 minute. In another related study, Lv *et al.* [219] proposed a deep learning framework based on the popular U-net architecture. To evaluate the performance of their technique they used 8 ROI's from cortex and medulla of segmented kidney. It is demonstrated by the authors that during free breathing measurements, their normalized root-mean-square error (NRMSE) values for cortex and medulla were significantly lower after registration.

### 5) MISCELLANEOUS

Yan *et al.* [220] presented an Adversarial Image Registration (AIR) method for multi-modal image MR-TRUS



**FIGURE 9.** Classification of image/patches is commonly performed by training a CNN classifier that stacks multiple convolution, activation and pooling layers, followed by fully-connected layers and a softmax layer.

registration [221]. They trained two deep networks concurrently, one for the generator component of the adversarial framework, and the other for the discriminator component. In their work, the authors learned not only an image registration network but also a so-called metric network which computes the quality of image registration. The data used in their experimentation consists of 763 sets of 3D TRUS volume and 2D MR volume with  $512 \times 512 \times 26$  voxels. The developed AIR network is also evaluated on clinical datasets acquired through image-fusion guided prostate biopsy procedures. For the visualization of 3D medical image data Zhao *et al.* [222] recently proposed a deep learning based technique, named Respond-weighted Class Activation Mapping (Respond-CAM). As compared to Grade-CAM [223] they claim better performance. Elss *et al.* [224] also employed Convolutional networks for single phase image motion in cardiac CT 2D/3D images. They trained regression network to successfully learn 2D motion estimation vectors. We also summarize few worth noting contributions from the years 2016 and 2017 in Table 4.

### D. CLASSIFICATION

Classification of images is a long standing problem in Medical Image Analysis and other related fields, e.g. Computer Vision. In the context of medical imaging, classification becomes a fundamental task for Computer Aided Diagnosis (CAD). Hence, it is no surprise that many researchers have recently tried to exploit the advances of deep learning for this task in medical imaging. In Fig. 9, we illustrate a generic image classifier based on deep learning. To learn the classifier, multiple convolutional, activation and pooling layers are stacked such that the feature map dimensions keep reducing as we go deeper into the architecture. After a considerable compression along the spatial dimensions (and expansion in terms of the number of learned feature maps), the features are mapped onto fully-connected layers. The activations of the last fully connected layer are provided to a softmax layer to predict the probability of the class labels. The classifier is learned by training on a large number of images. During the test time, the labels are predicted by simply forward passing the image through the learned classifier.

#### 1) BRAIN

Relating to the anatomical region of Brain, Li *et al.* [235] used deep learning to detect Autism Spectrum disorder (ASD) in

**TABLE 4.** Summary of influential papers appeared in 2016 and 2017 (based on Google Scholar citation index in January 2019) that exploit deep learning for the registration task in medical image analysis.

Reference	Anatomic Site	Imaging Modality	Network type	Data	Citations
Miao et al. [225] (2016)	Chest	X-ray	CNN regression	Synthetic	101
Wu et al. [226] (2016)	Brain	MRI	CNN	LONI, ADNI databases	58
Simonovsky et al. [227] (2016)	-	Multiple	CNN	IXI	57
Yang et al. [228] (2016)	Brain	MRI	Encoder-decoder	OASIS dataet	40
Barahmi et al. [229] (2016)	Brain	MRI	CNN	15 subjects	36
Zhang et al. [230] (2016)	Head, Abdomen, Chest	CT	CNN	Private	30
Nie et al. [231] (2017)	Multi task	MRI, CT	FCN	Private	110
Kang et al. [232] (2017)	Abdomen	CT	CNN	CT low-dose Grand Challenge	98
Yang et al. [200] (2017)	Brain	MRI	Encoder-decoder	373 OASIS and 375 IBJS images	64
De et al. [233] (2017)	Brain	MRI	CNN	Sunnybrook Cardiac Data [234]	50

functional Magnetic Resonance Imaging (fMRI). They developed a 2-stage neural network method. For the first stage, they trained a CNN (2CC3D) with 6 convolutional layers, 4 max-pooling layers and 2 fully connected layers. Their network uses a sigmoid output layer. For the second stage, in order to detect biomarkers for ASD, they took advantage of the anatomical structure of brain fMRI. They developed a frequency normalized sampling method for that purpose. Their method is evaluated using multiple databases, showing robust results for neurological function of biomarkers.

In their work, Hosseini-Asl et al. [236] employed an auto-encoder architecture for diagnosing Alzheimer's Disease (AD) patients. They reported up to 99% accuracy on ADNI dataset. They exploited Transfer Learning to handle the data scarcity issue, and used a model that is pre-trained with the help of CAD Dementia dataset. Their network architecture is based on 3D convolutional kernels that models generic brain features from sMRI data. The overall classification process in their technique first spatially normalizes the brain sMRI data, then it learns the 3D CNN model using the normalized data. The model is eventually fine-tuned on the target domain, where the fine-tuning is performed in a supervised manner.

Recently, Yan et al. [237] proposed a deep chronectome learning framework for the classification of MCI in brain using Full Bidirectional Long Short-Term Memory (Full-BiLSTM) networks. Their method can be divided into two parts, firstly a Full-LSTM is used to gather time varying information in brain for which MCI can be diagnosed. Secondly, to mine the contextual information hidden in dFC, they applied BiLSTM to access long range context in both directions. They reported the performance of their model on public dataset ADNI-2, achieving 73.6% accuracy. Heinsfeld et al. [238] also proposed a deep learning algorithm for the Autism Spectrum disorder (ASD) classification in rs-fMRI images on multi-site database ABIDE. They used denoising autoencoders for unsupervised pretraining. The classification accuracy achieved by their algorithm on the said dataset is 70%.

In the context of classification related to the anatomical region of brain, Soussia and Rekik [239] provided a review of 28 papers from 2010 to 2016 published in MICCAI. They

reviewed neuroimaging-based technical methods developed for the Alzheimer Disease (AD) and Mild-Cognitive Impairment (MCI) classification tasks. The majority of papers used MRI for dementia classification and few worked to predict MCI conversion to AD at later observations. We refer to [239] for the detailed discussions on the contributions reviewed by this article. Gutierrez-Becker and Wachinger et al. [240] proposed a deep neural network, termed Multi-structure point network (MSPNet), for the shape analysis on multiple structures. This network is inspired by PointNet [241] that can directly process point clouds. MSPNet achieves good classification accuracy for AD and MCI for the ADNI database.

## 2) BREAST

Awan et al. [242] proposed to use more context information for breast image classification. They used features of a CNN that is pre-trained on ICIAR 2018 dataset for histological images [243], and classified breast cancer as benign, carcinoma insitu (CIS) or breast invasive carcinoma (BIC). Their technique performs patch based and context aware image classification. They used ResNet50 architecture and overlapping patches of size  $512 \times 512$ . The extracted features are classified using a Support Vector Machine in their approach. Due to the unavailability of large-scale data, they used random rotation and flipping data augmentation techniques during the training process. It is claimed that their trained model can also be applied to other tasks where contextual information and high resolution are required for optimal prediction.

When only weak annotations are available for images, such as in heterogeneous images, it is often useful to turn to multiple instance learning (MIL). Courture et al. [244] described a CNN using quantile function for the classification of 5 types of breast tumor histology. They fine-tuned AlexNet [23]. The data used in their study consists of 1,713 images from the Carolina Breast Cancer Study, Phase 3 [245]. They improved the classification accuracy on this dataset from 68.6 to 85.6 for estrogen receptor (ER) task in breast images. Recently, MIL has also been used for breast cancer classification in [246] and [247] that perform patch based classification of histopathology images. Antropova et al. [248] used 690 cases with 5 fold cross-validation of MRI maximum intensity projection for breast lesion classification. They used

a pre-trained VGGNet [34] for feature extraction, followed by an SVM classifier. Ribli *et al.* [249] applied a CNN based on VGG16 for lesion classification in mammograms. They trained their model using DDSM dataset and tested it on INbreast [250]. They achieved the second best score for Mammography DREAM Challenge, with AUC of 0.95. Zheng *et al.* [251] proposed a CAD technique for breast cancer classification using CNN based on pre-trained VGG-19 model. They evaluated their technique's performance on digital mammograms of pairs of 69 cancerous and 27 healthy subjects. They achieved the values of 0.928 for sensitivity and 0.991 for specificity of classification.

### 3) EYE

Pertaining to the region of eye, Gergeya and Leng [76] took a data driven approach using deep learning to classify Diabetic retinopathy (DR) in color fundus images. The authors used public databases MESSIDOR 2 and E-ophtha to train and test their models and achieved 0.94 and 0.95 AUC score respectively on the test partitions of these datasets. A convolutional network is also employed by Pratt *et al.* [252] for diagnosing and classifying the severity of DR in color fundus images. Their model is trained using the Kaggle datasets, and it achieved 75% DR severity accuracy. Similarly, Ayhan and Berens [253] also exploited the deep CNN architecture of ResNet50 [254] for the fundus image classification. Mateen *et al.* [255] proposed a DR classification system based on VGG-19. They also performed evaluation using Kaggle dataset of 35,126 fundus images. It is claimed that their model outperforms the more conventional techniques, e.g. SIFT as well as earlier deep networks, e.g. AlexNet in terms of accuracy for the same task.

### 4) CHEST

Dey *et al.* [256] studied 3D CNNs for the diagnostic classification of lung cancer between benign and malignant in CT images. Four networks were analyzed for their classification task, namely a basic 3D CNN; a multi-output CNN; a 3D DenseNet, and an augmented 3D DenseNet with multi-outputs. They employed the public dataset LIDC-IDRI with 1,010 CT images and a private dataset of 47 CT images with both malignant and benign in this study. The best results are achieved by the 3D multi-output DenseNet (MoDenseNet) on both datasets, having accuracy 90.40% as compared to previously reported accuracy of 89.90% [257]. Gao *et al.* [258] proposed a deep CNN for the classification of Interstitial Lung Disease (IDL) patterns on CT images. Previously, batch based algorithms were being used for this purpose [259], [260]. In contrast, Gao *et al.* performed holistic classification using the complete image as network input. Their experiments used a public data [261] on which the classification accuracy improved to 87.9% from the previous results of 86.1% [260]. For the holistic image classification, the overall accuracy of 68.8% was achieved. In another work, Hoo-Chang *et al.* [99] et al. also analyzed three CNN

architectures, namely CifarNet, AlexNet, and GoogLeNet, for interstitial lung disease classification.

Biffi *et al.* [262] proposed a 3D convolutional generative model for the classification of cardiac diseases. They achieved impressive performance for the classification of healthy and hypertrophic cardiomyopathy MR images. For ACDC MICCAI 2017 dataset they were able to achieve 90% accuracy for classification of healthy subjects. Brestel *et al.* [263] proposed a CNN based technique RadBot-CXR to categorize focal lung opacities, diffuse lung opacity, cardiomegaly, and abnormal hilar prominence in chest X-ray images. They claim that their algorithm showed radiologists level performance for this task. Wang *et al.* [264] used deep learning in analyzing histopathology images for the whole slide lung cancer classification. Coudray *et al.* [265] used Inception3 CNN model to analyze whole slide images to classify lung cancer between adenocarcinoma (LUAD), squamous cell carcinoma (LUSC) or normal tissues. Moreover, they also trained their model for the prediction of ten most common mutated genes in LUAD and achieved good accuracies. Masood *et al.* [91] proposed a deep learning approach DFCNet based on FCN, which is used to classify the four stages of detected pulmonary lung cancer nodule.

### 5) ABDOMEN

Relating to abdomen, Tomczak *et al.* [266] employed deep Multiple Instance Learning (MIL) framework [267] for the classification of esophageal cancer in histopathology images. In another contribution, Frid-Adar *et al.* [268] used GANs for the synthetic medical image data generation. They made classification of CT liver images as their test bed and performed classification of 182 lesions. The authors demonstrated that by using augmented data with the GAN framework, up to 7% improvement is possible in classification accuracy. For automatic classification of ultrasound abdominal images, Xu *et al.* [269] proposed a multi-task learning framework based on CNN. For the experiments they used 187,219 ultrasound images and claimed better classification accuracy than human clinical experts.

### 6) MISCELLANEOUS

Esteva *et al.* [274] presented a CNN model to classify skin cancer lesions. Their model is trained in an end-to-end manner directly from the images, taking pixels and disease labels as inputs. They used datasets of 129,450 clinical images consisting of 2,032 different diseases to train CNN. They classified two most common cancer diseases; keratinocyte carcinomous versus benign seborrheic keratosis, and the deadliest skin cancer; malignant melanomas versus benign nevi. For skin cancer sun exposure classification, Combalia and Vilaplana [278] also applied a Monte Carlo method to only highlight the most elastic regions. Antony *et al.* [279] employed a deep CNN model for automatic quantification of severity of knee osteoarthritis (OA). They used Kellgren and Lawrence (KL) grades to assess the severity of knee.

**TABLE 5.** Summary of notable contributions appearing in 2016 and 2017 that exploit deep learning for the classification task in medical image analysis. The citation index is based on Google scholar (January 2019).

Reference	Anatomic Site	Image Modality	Network type	Data	Citations
Anthimopoulos et al. [270] (2016)	Lung	CT	CNN	Uni. Hospital of Geneva [261]& Inselspital	253
Kallenberg et al. [271] (2016)	Breast	Mammography	CNN	3 different databases	111
Huynh et al. [272] (2016)	Breast	Mammography	CNN	Private data (219 lesions)	76
Yan et al. [273] (2016)	12 regions	CT	CNN	Synthetic & Private	73
Zhang et al. [230] (2016)	Breast	Elastography	MLP	Private data (227 images)	50
Esteva et al. [274] (2017)	Skin	Histopathology	CNN	Private large-scale	1386
Sun et al. [275] (2017)	Breast	Mammography	CNN	Private data	40
Christodoulidis et al. [276] (2017)	Lung	CT	CNN	Texture data as Transfer learning source	36
Lekadir et al. [277] (2017)	Heart	US	CNN	Private data	26
Nibali et al. [257] (2017)	Lung	CT	CNN	LIDC/IDRI dataset	22

In their work, using deep CNN pre-trained on ImageNet and fine-tuned on knee OA images resulted in good classification performance. Paserin et al. [280] recently worked on diagnosis and classification of developmental dysplasia of hip (DDH). They proposed a CNN-RNN technique for 3D ultrasound volumes for DDH. Their model consists of convolutional layers for feature learning followed by recurrent layers for spatial relationship of their responses. Inspired by VGG network [34], they used CNN with 5 convolutional layers for feature extraction with ReLU activations, and  $2 \times 2$  max-pooling with a stride of 2. Finally, they used LSTM network that has 256 units. They achieved 82% accuracy with AUC 0.83 for 20 test volumes. Few notable contributions from the years 2016 and 2017 are also summarized in Table 5.

We also provide a comparative review of the techniques in Table 6. The table includes more influential papers for each image processing task considered above. It further divides the contributions based on the anatomical site of interest division used in our survey. The table summarizes the information on ‘imaging modality’, ‘dataset size’, ‘performance measure’ and the ‘network bias’ used in the referred papers.

#### IV. DATASETS

Good quality data has always remained the primary requirement for learning reliable computational models. This is also true for deep models that also have the additional requirement of consuming large amount of training data. Recently, many public datasets for medical imaging tasks have started to emerge. There is also a growing trend in the research community to compile lists of these datasets. For instance we can find few useful compilation of public dataset lists at Github repositories and other webpages. Few medical image analysis products are also helping in providing public datasets. Whereas detailed discussion on the currently available public datasets for medical imaging tasks is outside the scope of this article, we provide typical examples of the commonly used datasets in medical imaging by deep learning approaches in Table 7. The Table is not intended to provide an exhaustive list. We recommend the readers internet search for that purpose. A brief search can result in a long compilation of medical imaging datasets. However, we summarize few examples of contemporary datasets in Table 7 to make an important

point regarding deep learning research in the context of Medical Image Analysis. With the exception of few datasets, the public datasets currently available for medical imaging tasks are small in terms of the number of samples and patients. As compared to the datasets for general Computer Vision problems, where datasets typically range from few hundred thousand to millions of annotated images, the dataset sizes for Medical imaging tasks are too small. On the other hand, we can see the emerging trend in Medical Imaging community of adopting the practices of broader Pattern Recognition community, and aiming at learning deep models in end-to-end fashion. However, the broader community has generally adopted such practices based on the availability of large-scale annotated datasets, which is an important requirement for inducing reliable deep models. Hence, it remains to be seen that how effectively end-to-end trained models can really perform the medical image analysis tasks without over-fitting to the training datasets.

#### V. CHALLENGES IN GOING DEEP

In this Section, we discuss the major challenges faced in fully exploiting the powers of Deep Learning in Medical Image Analysis. Instead of describing the issues encountered in specific tasks, we focus more on the fundamental challenges and explain the root causes of these problems for the Medical Imaging community that can also help in understanding the task-specific challenges. Dealing with these challenges is the topic of discussion in Section VI.

##### A. LACK OF APPROPRIATELY ANNOTATED DATA

It can be argued that the single aspect of Deep Learning that sets it apart from the rest of Machine Learning techniques is its ability to model extremely complex mathematical functions. Generally, we introduce more layers to learn more complex models - i.e. go deep. However, a deeper network must also learn more model parameters. A model with a large number of parameters can only generalize well if we correspondingly use a very large amount of data to infer the parameter values. This phenomenon is fundamental to any Machine Learning technique. A complex model inferred using a limited amount of data normally over-fits to the used data and performs poorly on any other data. Such modeling

**TABLE 6.** A comparative summary of representative techniques, highlighting the used datasets, performance measures and the network bias in more influential works.

Task	Anatomic Site	Image Modality	Subjects/dataset size	Performance measure	Network bias
Detection	Brain	MRI [64] MRI [66]	Training 35, testing 45 subjects Training 75, validation 21	AUC True positive rate VOI, False positive	VAE,AEE sCAE 3D-CNN CNN
	Breast	ABUS [70] MRI [71]	171 tumors 385 scans	Sensitivity	CNN (AE,AD)
	Eye	OCT [77] OCT [73]	1200 images 207,103 images	AUC area under ROC	VGG-16
	Chest	X-ray [89] CMR [87]	120 images 3465 images	AUC score F1 score	CNN CNN
	Abdomen	WGD [112] US [111]	training 8428, testing 1118 images Private data	Sensitivity Shadow Confidence	CNN GAN
Segmentation	Brain	MRI [126] MRI [129]	406 subjects	Dice,Sensitivity,Specificity Dice,PPV,Sensitivity	AE,AD
	Breast	DCE-MRI [134] Mammograms [133]	BRATS 2013,2015,2016 training 300, testing 160 567 for training and testing	Runtime,TS,Intensity normalization Accuracy,Dice,Sensitivity,Specificity	FCNN+CRF FCN cGAN
	Eye	OCT [143] OCT [141]	110 images for training and testing 14,884 training images	SP,Runtime TS	GANs CNN
	Chest	MRI [147] CT,MRI [151]	400 training,100 testing images 20 CT,20 MRI scans	Accuracy,Dice Score Dice Score	CNN GANs
	Abdomen	CT [156] CT,MRI [168]	377 Scans CT Training 60,Testing 30,MRI 23	Dice Score ASSD,Dice score	FCN FCN
Registration	Brain	MRI [205] MRI [201]	7829 T1 weighted 100 exams	Dice Score,Computational time Dice score,Reliability	CNN
	Eye	OCT [207]	Private data	Runtime,deformation field	DFCN
	Chest	MR-CT [281] CT [213]	17 Unpaired and 3 paired subjects 17 pairs	DIR,Runtime	GANs
	Abdomen	CT-MRI [216] MRI [218] MRI [219]	45 MRI,2060 CT 10 Subjects 10 Subjects	Runtime,minimal parameter Dice, runtime,HD,ASD Runtime NRMSE	3D-CNN CNN CNN U-Net
Classification	Brain	fMRI [235] MRI [236]	82 children,42 age ADNI MRI data	ROIs,Classify ASD Accuracy,Robustness	DNN
	Breast	Histopathology [244] Mammography [251]	1713 images 96 Subjects	Classification accuracy,ER Sensitivity,Specificity	3D CNN CNN,AlexNET
	Eye	OCT [255] OCT [252]	35,126 images	Classify DR	VGG-19
	Chest	CT [256] MRI [262]	Kaggle dataset 1010+47 images	DR Severcity accuracy Classification accuracy	VGG-19 CNN
	Abdomen	CT [268] Histopatgolgy [266]	686 HCM,679 healthy 182 lessions 21 slides	Classification accuracy,VOIs Classification accuracy Classification ccuracy	AE,AD GANs MIL

**TABLE 7.** Examples of popular databases used by medical image analysis techniques that exploit deep learning.

Sr.	Database	Anatomic site	Image modality	Main task	Patients/Images
1	ILD [261]	Lung	CT	Classification	120 patients
2	LIDC-IDRI [95]	Lung	CT	Detection	1,018 patients
3	ADNI [63]	Brain	MRI	Classification	>800 patients
4	MURA [116]	Musculoskeletal	X-ray	Detection	40,561 images
5	TCIA	Multiple	Multiple	Multiple	>35,000 patients
6	BRATS [282]	Brain	MRI	Segmentation	-
7	DDSM [283]	Breast	Mammography	Segmentation	2,620 patients
8	MESSIDOR-2 [284], [285]	Eye	OCT	Classification	1,748 images
9	ChestX-ray14 [286]	Chest	X-ray	Multiple	>100,000 images
10	ACDC 2017	Brain	MRI	Classification	150 patients
11	2015 MICCAI Gland Challenge	Glands	Histopathology	Segmentation	165 images
12	OAI	Knee	X-ray, MRI	Multiple	4,796 patients
13	DRIVE [138], [287]	Eye	SLO	Segmentation	400 patients
14	STARE [140]	Eye	SLO	Segmentation	400 images
15	CHASEDB1 [139]	Eye	SLO	Segmentation	28 images
16	OASIS-3 [62], [288]–[292], [127]	Brain	MRI, PET	Segmentation	1,098 patients
17	MIAS [293]	Breast	Mammography	Classification	322 patients
18	ISLES 2018	Brain	MRI	Segmentation	103 patients
19	HVS MR 2018 [294]	Heart	CMR	Segmentation	4 patients
20	CAMELYON17 [118]	Breast	WSI	Segmentation	899 images
21	ISIC 2018	Skin	JPEG	Detection	2,600 images
22	OpenNeuro	Brain	Multiple	Classification	4,718 patients
23	ABIDE	Brain	MRI	Disease Diagnosis	1,114 patients
24	INbreast [250]	Breast	Mammography	Detection/Classification	410 images

is highly undesirable because it gives a false impression of learning the actual data distribution whereas the model is only learning the peculiarities of the used training data.

Learning deep models is inherently unsuitable for the domains where only limited amount of training data is available. Unfortunately, Medical Imaging is one such domain. For most of the problems in Medical Image Analysis, there is only a limited amount of data that is annotated in a manner that is suitable to learn powerful deep models. We encounter the problem of ‘lack of appropriately annotated data’ so frequently in the current Deep Learning related Medical Imaging literature that it is not difficult to single out this problem as ‘the fundamental challenge’ that Medical Imaging community is currently facing in fully exploiting the advances in Deep Learning.

The Computer Vision community has been able to take full advantage of Deep Learning because data annotation is relatively straightforward in that domain. Simple crowd sourcing can yield millions of accurately annotated images. This is not possible for Medical Images that require high level of specific expertise for annotation. Moreover, the stakes are also very high due to the nature of medical application, requiring extra care in annotation. Although we can also find large number of images in medical domain via systems like PACS and OIS, however using them to train deep models is still not easy because they lack appropriate level of annotations that is generally required for training useful deep models.

With only a few exceptions, e.g. [286] the public datasets available in the Medical Imaging domain are not large-scale - a requirement for training effective deep models. In addition to the issues of hiding patient’s privacy, one major problem in forming large-scale public datasets is that the concrete labels required for computational modeling can often not be easily inferred from medical reports. This is problematic if inter-observers are used to create large-scale datasets. Moreover, the required annotations for deep models often do not perfectly align with the general medical routines. This becomes an additional problem even for the experts to provide noise-free annotations.

Due to the primary importance of large-scale training datasets in Deep Learning there is an obvious need to develop such public datasets for Medical Imaging tasks. However, considering the practical challenges in accomplishing this goal it is also imperative to simultaneously develop techniques of exploiting Deep Learning with less amount of data. We provide discussion on future directions along both of these dimension in Section VI.

## B. IMBALANCED DATA

One problem that occurs much more commonly in Medical Imaging tasks as compared to general Computer Vision tasks is the imbalance of samples in datasets. For instance, a dataset to train a model for detecting breast cancer in mammograms may contain only a limited number of positive samples but a very large number of negative samples. Training deep networks with imbalanced data can induce models that are

biased. Considering the low frequency of occurrences of positive samples for many Medical Imaging tasks, balancing out the original data can become as hard as developing large-scale dataset. Hence, extra care must be taken in inducing deep models for Medical Imaging tasks.

## C. LACK OF CONFIDENCE INTERVAL

Whereas the Deep Learning literature often refers to the output of a model as ‘prediction confidence’; the output signal of a neuron can only be interpreted as a *single* probability value. The lack of provision of confidence interval around a predicted value is generally not desirable in Medical Imaging tasks. Litjens *et al.* [28] has noted that an increasing number of deep learning methods in Medical Imaging are striving to learn deep models in an end-to-end manner. Whereas end-to-end learning is the epitome of Deep Learning, it is not certain if this is the right way to exploit this technology in Medical Imaging. To an extent, this conundrum is also hindering the widespread use of Deep Learning in Medical Imaging.

## VI. FUTURE DIRECTIONS

With the recent increasing trend of exploiting Deep Learning in Medical Imaging tasks, we are likely to see a large influx of papers in this area in the near future. Here, we provide guidelines and directions to help those works in dealing with the inherent challenges faced by Deep Learning in Medical Image Analysis. We draw our insights from the reviewed literature and the literature in the sister fields of Computer Vision, Pattern Recognition and Machine Learning. Due to the earlier use of Deep Learning in those fields, the techniques of dealing with the related challenges have considerably matured in those areas. Hence, Medical Image Analysis can readily benefit from those findings in setting fruitful future directions.

Our discussion in this Section is primarily aimed at providing guiding principles for the Medical Imaging community. Therefore, we limit it to the fundamental issues in Deep Learning. Based on the challenges mentioned in the preceding Section, and the insights from the parallel scientific fields, we present our discussion along three directions, addressing the following major questions. (1) How can Medical Image Analysis still benefit from Deep Learning in the absence of large-scale annotated datasets? (2) What can be done for developing large-scale Medical Imaging datasets. (3) What should be the broader outlook of this research direction to catapult it in taking full advantage of the advances in Deep Learning?

## A. DEALING WITH SMALLER DATA SIZE

### 1) DISENTANGLING MEDICAL TASK TRANSFER LEARNING

Considering the obvious lack of large-scale annotated datasets, Medical Imaging community has already started exploiting ‘transfer learning’ [295]–[297]. In transfer learning, one can learn a complex model using data from a *source* domain where large-scale annotated images are available

(e.g. natural images). Then, the model is further fine-tuned with the data of *target* domain where only a small number of annotated images are available (e.g. medical images). It is clear from the literature that transfer learning is proving advantageous for Medical Image Analysis. Nevertheless, one promising recent development in transfer learning [298] in Computer Vision literature remains completely unexplored for Medical Image Analysis.

Zamir *et al.* [298] recently showed that performance of transfer learning can be improved by carefully selecting the source and target domains/tasks. By organizing different tasks that let the deep models transfer well between themselves, they developed a so-called ‘taskonomy’ to guide the use of transfer learning for natural images. This concept has received significant appreciation in the Computer Vision community, resulting in the ‘best paper’ award for the authors at the prestigious IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. A similar concept is worth exploring for the data deprived Medical Imaging tasks. Disentangling medical tasks for transfer learning may prove very beneficial. The core concept of disentanglement is to specify the source tasks for which relatively larger annotated data is available, and find the target task for which the source models can provide stronger results once fine-tuned. The source and target tasks may not necessarily be restricted to the same anatomical regions. Once a taxonomy of such tasks is established, it can consistently serve to achieve better performance with the fine-tuned models.

Another related direction that can help in dealing with smaller data size is to quantify the suitability of transfer learning between medical imaging and natural imaging tasks. In other words, it requires establishing the taskonomy between medical imaging tasks and the tasks performed for natural image processing. Understanding of the knowledge transfer abilities of the existing natural image models to the medical tasks can have a huge impact in Medical Image Analysis using Deep Learning.

## 2) WRAPPING DEEP FEATURES FOR MEDICAL IMAGING TASKS

The existing literature shows an increasing trend of training deep models for Medical tasks in an ‘end-to-end’ manner. For Deep Learning, end-to-end modeling is generally more promising for the domains where large-scale annotated data is available. Exploiting the existing deep models as feature extractors and then performing further learning on those features is a much more promising direction in the absence of large-scale training datasets. There is a considerable evidence in the Pattern Recognition literature that activation signals of deeper layers in neural networks often form highly expressive image features. For natural images, Akhtar *et al.* [299] demonstrated that features extracted from deep models can be used to learn further effective higher level features using the techniques that require less training samples. They used Dictionary Learning framework [300] to further wrap the deep features before using them with a classifier. In that

case, deep features become input samples to the wrapping technique. Another representation model for those features is then learned, and if required, a customized classifier can further be trained. A key advantage of following this strategy is that it can easily avoid over-fitting. For deep neural networks, a high percentage of the parameters belong to the fully-connected layers. Feature extraction performed before the fully-connected layers significantly reduces the effective model complexity. Other techniques that inherently learn less complex models can then take full advantage of the discriminative nature of deep learning features. For example, dictionary learning has been shown to exploit those feature very well under the sparse representation framework [299].

We note that Medical Image Analysis literature has already seen reasonably successful attempts of using the existing natural image deep models as feature extractors e.g. [301]–[303]. However, such attempts generally directly feed the features extracted from the pre-trained models to a classifier. The direction we point towards entail post-processing of deep features to better suit the requirements of the underlying Medical Image Analysis task.

## 3) TRAINING PARTIALLY FROZEN DEEP NETWORKS

As a general principle in Machine Learning, more amount of training data is required to train more complex computational models. In Deep Learning, the network depth normally governs the model complexity, whereas deeper networks also have more parameters that require large-scale datasets for training. It is known that the layers of CNNs - the most relevant neural networks for image analysis - systematically break down images into their features from lower level of abstraction to higher level of abstraction [304]. It is also known that the initial layers of CNNs learn very similar filters for a variety of natural images. These observations point towards the possibility of reducing the number of learnable parameters in a network by freezing few of its layers to the parameter values that are likely to be similar for a variety of images. Those parameter values can be directly borrowed from other networks trained on similar tasks. The remainder of the network - that now has less parameters but has the same complexity - can then be trained for the target task as normal. Training partially frozen networks for Medical Imaging task can mitigate the issues caused by the lack of large-scale annotated datasets.

## 4) ENSEMBLE DEEP MODELS AND MULTI-TASK LEARNING

It is generally the case that limited training data is provided for on-line competitions in Computer Vision. A common strategy is often adopted by the top performing techniques for such competitions. That is, instead of training a single deep model, multiple deep models are trained and combined to compute the final results. The used models normally differ in terms of network architectures, their complexity and loss functions to fully exploit different characteristics of the available data. Networks are combined under different methods, e.g. by using joint logit layer or combining the networks at the

output layer by computing losses while accounting for all the networks. It is a common observation that network ensembles provide significantly better results than their individual components alone. However, the larger network size for ensembles is not attractive for many applications. Nevertheless, for medical image analysis, this remains a viable option, where accuracy generally out-weights the network size in majority of the applications.

On similar lines, multi-task learning under the deep learning framework is also known to be effective for the cases where the training data is scarce. It allows to learn models for multiple tasks where the training data for each task can be limited, however the tasks are related to each other. Hence, different tasks are coupled to take advantage of each other's data annotations. However, multi-task learning can often be more complex than learning for a single task at a time.

## 5) USING GANS FOR SYNTHETIC DATA GENERATION

Generative Adversarial Networks (GANs) [42] are currently receiving tremendous attention of Computer Vision community for their ability to mimic the distributions from which images are sampled. Among other uses of GANs, one can use the GAN framework to generate realistic synthetic images for any domain. These images can then be used to train deeper models for that domain that generally outperform the models trained with only (limited) original data. This property of GANs is of particular interest for Medical Image Analysis. Therefore, we can expect to see a large number of future contributions in Medical Imaging that will exploit GANs. In fact, our literature review already found few recent applications of GANs in medical image analysis [206], [305]–[307]. Moreover, we also found recent surveys [46], [47] that specifically review the technique in Medical Image Analysis that exploit GANs. Besides healthy prospects of GANs as a data generation strategy for Medical Image Analysis tasks, these articles also indicate numerous future challenges in this direction. We also recommend that extra care is required while exploiting the GAN framework in medical imaging. It should be noted that GANs do not actually learn the original distribution of images, rather they only mimic it. Hence, the synthetic images generated from GANs can still be very different from the original images. Therefore, instead of training the final model with the data that includes GAN-generated data, it is often better to finally fine-tune such model with only the original images.

## 6) MISCELLANEOUS DATA AUGMENTATION TECHNIQUES

In general, Computer Vision and Pattern Recognition literature has also developed few elementary data augmentation techniques that have shown improvement in the performance of deep models. Whereas these techniques are generally not as effective as sophisticated methods, such as using GANs to increase data samples; they are still worth taking advantage of. We list the most successful techniques below. Again, we note that some of these methods have already proven their effectiveness in the context of Medical Image Analysis:

- *Image flipping:* A simple sideways flip of images doubles the number of training samples, that often results in a better model. For medical images, top-down flip is also a possibility due to the nature of images.
- *Image cropping:* Cropping different areas of a larger image into smaller images and treating each one of the cropped versions as an original image also benefits deep models. Five crops of equal dimensions from an image is a popular strategy in Computer Vision literature. The crops are made using the four corners and the central region of the image.
- *Adversarial training:* Very recently, it is discovered that we can ‘fool’ deep models using *adversarial images* [2]. These images are carefully computed such that they appear the same as the original images to humans, however, a deep model is not able to recognize them. Whereas developing such images is a different research direction, one finding from that direction is that including those images in training data can improve the performance of deep models [308]. Since adversarial examples are generated from the original images, they provide a useful data augmentation method that can be harnessed for Medical Imaging tasks.
- *Rotation and random noise addition:* In the context of 3D data, rotating the 3D scans and adding small amount of random noise (emulating jitters) are also considered useful data augmentation strategies [241].

## B. ENHANCING DATASET SIZES

Whereas the techniques discussed in Section VI-A can alleviate the issues caused by smaller training datasets, the root cause of those problems can only be eliminated by acquiring Deep Learning compatible large-scale annotated datasets for Medical Image Analysis tasks. Considering that Deep Learning has started outperforming human experts in Medical Image Analysis tasks [309], there is a strong need to implement protocols that make medical reports readily convertible to the formats useful for training computational models, especially Deep Learning models. In this context, techniques from the fields of Document Analysis [310] and Natural Language Processing (NLP) [311] can be used to alleviate the extra burden falling on the medical experts due to the implementations of such protocols.

Besides generating the new data at large-scales that is useful in learning computational models, it is also important to take advantage of the existing medical records for exploiting the current advances in Deep Learning. To handle the large volume of un-organized data (in terms of compatibility with Machine Learning), data mining with humans-in-the-loop [312] and active learning [30] can prove beneficial. Advances in Document Analysis and NLP can also be exploited for this task.

## C. BROADER OUTLOOK

We can make one important observation regarding Deep Learning research by exploring the literature of different

research fields. That is, the advancement in Deep Learning research has often experienced a quantum leap under the breakthroughs provided by different sister fields. For example, the ‘residual learning’ concept [22] that enabled very deep networks was first introduced in the literature of Computer Vision. This idea (along with other breakthroughs in core Machine Learning research) eventually enabled the tabula rasa algorithm of AlphaGo Zero [313]. Following up on this observation, we can argue that significant advances can be made in Deep Learning research in the context of Medical Image Analysis if researchers from the sister fields of Computer Vision and Machine Learning are able to better understand the Medical Image Analysis tasks.

Indeed, Medical Imaging community already involves experts from other related fields. However, this involvement is at a smaller scale. For the involvement of broader Machine Learning and Computer Vision communities, a major hindrance is the Medical Imaging literature jargon. Medical literature is not easily understood by the experts of other fields. One effective method to mitigate this problem can be regular organization of Medical Imaging Workshops and Tutorials in the reputed Computer Vision and Machine Learning Conferences, e.g. IEEE CVPR, ICCV, NeurIPS and ICML. These events should particularly focus on playing the role of translating the Medical Imaging problems to other communities in terms of their topics of interest.

Another effective strategy to take advantage of Deep Learning advances is to outsource the Medical Imaging problems by organizing online challenges, e.g. Kaggle competitions. The authors are already aware of few Kaggle competitions related to Medical Imaging, e.g. Histopathologic cancer detection. However, we can easily notice that Medical Imaging competitions are normally attracting fewer teams as compared to other competitions - currently 361 for Histopathologic cancer detection. Generally, the number of teams are orders of magnitude lower for the Medical Imaging competitions than those for the typical imaging competitions. In authors’ opinion, strict Medical parlance adopted in organizing such competitions is the source of this problem. Explanation of Medical Imaging tasks using the terms more common among Computer Vision and Machine Learning communities can greatly help in improving the popularity of Medical Imaging in those communities.

In short, one of the key strategies to fully exploit Deep Learning advances in Medical Imaging is to get the experts from other fields, especially Computer Vision and Machine Learning; to involve in solving Medical Imaging tasks. To that end, the Medical Imaging community must put an extra effort in making its literature, online competitions and the overall outlook of the field more understandable to the experts from the other fields. Deep Learning is being dubbed as ‘modern electricity’ by the experts. In the future, its ubiquitous nature will benefit those fields the most that are better understood by the wider communities.

## VII. CONCLUSION

This article presented a review of the recent literature in Deep Learning for Medical Imaging. It contributed along three major directions. First, we presented an instructive introduction to the core concepts of Deep Learning. Keeping in view the general lack of understanding of Deep Learning framework among Medical Imaging researchers, we kept our discussion intuitive. This part of the paper can be understood as a tutorial of Deep Learning concepts commonly used in Medical Imaging. The second part of the paper presented a comprehensive overview of the approaches in Medical Imaging that employ Deep Learning. Due the availability of other review articles until the year 2017, we mainly focused on the literature published in the year 2018. The third major part of the article discussed the major challenges faced by Deep Learning in Medical Image Analysis, and discussed the future directions to address those challenges.

Beside focusing on the very recent literature, this article is also different from the existing related literature surveys in that it provides a Computer Vision/Machine Learning perspective to the use of Deep Learning in Medical Image Analysis. Using that perspective, we are not only able to provide an intuitive understanding of the core concepts in Deep Learning for the Medical community, we also highlight the root cause of the challenges faced in this direction and recommend fruitful future directions by drawing on the insights from multiple scientific fields.

From the reviewed literature, we can single out the ‘lack of large-scale annotated datasets’ as the major problem for Deep Learning in Medical Image Analysis. We have discussed and recommended multiple strategies for the Medical Imaging community that are adopted to address similar problems in the sister scientific fields. We can conclude that Medical Imaging can benefit significantly more from Deep Learning by encouraging collaborative research with Computer Vision and Machine Learning research communities.

## REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [2] N. Akhtar and A. Mian, “Threat of adversarial attacks on deep learning in computer vision: A survey,” *IEEE Access*, vol. 6, pp. 14410–14430, 2018.
- [3] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.
- [4] T. Ciodaro, D. Deva, J. M. de Seixas, and D. Damazio, “Online particle detection with neural networks based on topological calorimetry information,” *J. Phys., Conf. Series*, vol. 368, no. 1, 2012, Art. no. 012030.
- [5] H. Y. Xiong, B. Alipanahi, L. J. Lee, H. Bretschneider, D. Merico, R. K. Yuen, Y. Hua, S. Gueroussov, H. S. Najafabadi, and T. R. Hughes, “The human splicing code reveals new insights into the genetic determinants of disease,” *Science*, vol. 347, no. 6218, 2015, Art. no. 1254806.
- [6] M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk, “Connectomic reconstruction of the inner plexiform layer in the mouse retina,” *Nature*, vol. 500, no. 7461, pp. 168–174, 2013.
- [7] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, “Deep neural nets as a method for quantitative structure–activity relationships,” *J. Chem. Inf. Model.*, vol. 55, no. 2, pp. 263–274, 2015.
- [8] R. J. Schalkoff, *Artificial Neural Networks*, vol. 1. New York, NY, USA: McGraw-Hill, 1997.

- [9] B. W. White and F. Rosenblatt, "Principles of neurodynamics: Perceptrons and the theory of brain mechanisms," *Amer. J. Psychol.*, vol. 76, no. 4, p. 705, Dec. 1963.
- [10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, p. 533, 1986.
- [11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [12] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.
- [13] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
- [14] R. Beale and T. Jackson, *Neural Computing—An Introduction*. Boca Raton, FL, USA: CRC Press, 1990.
- [15] H. C. Becker, W. J. Nettleton, P. H. Meyers, J. W. Sweeney, and C. M. Nice, "Digital computer determination of a medical diagnostic index directly from chest X-ray images," *IEEE Trans. Biomed. Eng.*, vol. BME-11, no. 3, pp. 67–72, Jul. 1964.
- [16] G. S. Lodwick, T. E. Keats, and J. P. Dorst, "The coding of roentgen images for computer analysis as applied to lung cancer," *Radiology*, vol. 81, no. 2, pp. 185–200, 1963.
- [17] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," *Radiology*, vol. 187, no. 1, pp. 81–87, 1993.
- [18] S. C. B. Lo, S. L. A. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 711–718, Dec. 1995.
- [19] B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "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.
- [20] H.-P. Chan, S.-C. B. 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. 1555–1567, 1995.
- [21] W. Zhang, K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt, "An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms," *Med. Phys.*, vol. 23, no. 4, pp. 595–601, 1996.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [24] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," *Med. Phys.*, vol. 46, no. 1, pp. e1–e36, 2018.
- [25] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
- [26] M. I. Razzaq, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification BioApps*. Cham, Switzerland: Springer, 2018, pp. 323–350.
- [27] M. Bakator and D. Radovac, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, 2018.
- [28] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [29] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
- [30] X. Zhu, "Semi-supervised learning literature survey," *Comput. Sci., Univ. Wisconsin-Madison*, vol. 2, no. 3, p. 4, 2006.
- [31] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.
- [32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [33] I. Sobel and G. Feldman, "A  $3 \times 3$  isotropic gradient operator for image processing," in *Proc. Talk Stanford Artif. Project*, 1968, pp. 271–272.
- [34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [35] S. Sun, N. Akhtar, H. Song, A. Mian, and M. Shah, "Deep affinity network for multiple object tracking," 2018, *arXiv:1810.11780*. [Online]. Available: <https://arxiv.org/abs/1810.11780>
- [36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [38] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1137–1144.
- [39] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
- [40] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: <https://arxiv.org/abs/1312.6114>
- [41] S. Rifai, P. Vincent, X. Müller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn.*, 2011, pp. 833–840.
- [42] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [43] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. CVPR*, Jul. 2017, vol. 2, no. 4, pp. 2242–2251.
- [44] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, vol. 1, no. 2, pp. 95–104.
- [45] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. CVPR*, Jul. 2017, vol. 2, no. 3, pp. 6882–6890.
- [46] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," 2018, *arXiv:1809.07294*. [Online]. Available: <https://arxiv.org/abs/1809.07294>
- [47] S. Kazemina, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," 2018, *arXiv:1809.06222*. [Online]. Available: <https://arxiv.org/abs/1809.06222>
- [48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [49] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
- [50] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, vol. 4, 2017, p. 12.
- [51] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Jul. 2017, vol. 1, no. 2, pp. 2261–2269.
- [52] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
- [53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [54] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.
- [55] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. NIPS Autodiff Workshop, Future Gradient-Based Mach. Learn. Softw. Techn.*, Long Beach, CA, USA, Dec. 2017.
- [56] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

- [57] F. Chollet. (2015). *Keras*. [Online]. Available: <https://keras.io>
- [58] R. Al-Rfou, "Theano: A Python framework for fast computation of mathematical expressions," May 2016, *arXiv:1605.02688*. [Online]. Available: <https://arxiv.org/abs/1605.02688>
- [59] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 689–692.
- [60] R. Collobert, S. Bengio, and J. Mariéthoz, "Torch: A modular machine learning software library," Idiap, Martigny, Switzerland, Tech. Rep. Idiap-RR-46-2002, 2002.
- [61] J. Islam and Y. Zhang, "Early diagnosis of alzheimer's disease: A neuroimaging study with deep learning architectures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1881–1883.
- [62] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2677–2684, 2010.
- [63] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, and A. W. Toga, "Alzheimer's disease neuroimaging initiative (ADNI): Clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.
- [64] X. Chen and E. Konukoglu, "Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders," 2018, *arXiv:1806.04972*. [Online]. Available: <https://arxiv.org/abs/1806.04972>
- [65] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, *arXiv:1511.05644*. [Online]. Available: <https://arxiv.org/abs/1511.05644>
- [66] Z. Alaverdyan, J. Jung, R. Bouet, and C. Lartizien, "Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: Application to epilepsy lesion screening," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [67] A. Panda, T. K. Mishra, and V. G. Phaniraham, "Automated brain tumor detection using discriminative clustering based MRI segmentation," in *Smart Innovations in Communication and Computational Sciences*. Singapore: Springer, 2019, pp. 117–126.
- [68] K. R. Laukamp, F. Thiele, G. Shakirin, D. Zopfs, A. Faymonville, M. Timmer, D. Maintz, M. Perkuhn, and J. Borggrefe, "Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI," *Eur. Radiol.*, vol. 29, no. 1, pp. 124–132, 2019.
- [69] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, and M. Balkenhol, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.
- [70] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang, "Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 240–249, Jan. 2019.
- [71] M. U. Dalmiş, S. Vreemann, T. Kooi, R. M. Mann, N. Karssemeijer, and A. Gubern-Mérida, "Fully automated detection of breast cancer in screening MRI using convolutional neural networks," *J. Med. Imag.*, vol. 5, no. 1, 2018, Art. no. 014502.
- [72] J. Zhang, E. H. Cain, A. Saha, Z. Zhu, and M. A. Mazurowski, "Breast mass detection in mammography and tomosynthesis via fully convolutional network-based heatmap regression," *Proc. SPIE*, vol. 10575, Feb. 2018, Art. no. 1057525.
- [73] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 257, no. 3, pp. 495–505, 2019.
- [74] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.
- [75] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Invest. ophthalmol. Vis. Sci.*, vol. 57, no. 13, pp. 5200–5206, 2016.
- [76] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.
- [77] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraßer, A. Sadeghipour, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Fully automated detection and quantification of macular fluid in OCT using deep learning," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, 2017.
- [78] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, and F. Yan, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [79] The U.S. Food and Drug Administration, "FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems," News Release, Apr. 2018.
- [80] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, 2018.
- [81] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 16685.
- [82] P. Khojasteh, L. A. P. Júnior, T. Carvalho, E. Rezende, B. Aliahmad, J. P. Papa, and D. K. Kumar, "Exudate detection in fundus images using deeply-learnable features," *Comput. Biol. Med.*, vol. 104, pp. 62–69, Jan. 2019.
- [83] R. Kälviäinen and H. Uusitalo, "DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proc. Med. Image Understand. Anal.*, 2007, p. 61.
- [84] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Lay, and A. Chabouis, "TeleOphta: Machine learning and image processing methods for teleophthalmology," *IRBM*, vol. 34, no. 2, pp. 196–203, 2013.
- [85] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, "DeepEM: Deep 3D convnets with EM for weakly supervised pulmonary nodule detection," 2018, *arXiv:1805.05373*. [Online]. Available: <https://arxiv.org/abs/1805.05373>
- [86] A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, and B. Geurts, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.
- [87] I. Oksuz, B. Ruijsink, E. Puyol-Antón, A. Bustin, G. Cruz, C. Prieto, D. Rueckert, J. A. Schnabel, and A. P. King, "Deep learning using K-space based data augmentation for automated cardiac MR motion artefact detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 250–258.
- [88] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4489–4497.
- [89] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," 2017, *arXiv:1711.06373*. [Online]. Available: <https://arxiv.org/abs/1711.06373>
- [90] X. Yi, S. Adams, P. Babyn, and A. Elnajmi, "Automatic catheter detection in pediatric X-ray images using a scale-recurrent network and synthetic data," 2018, *arXiv:1806.00921*. [Online]. Available: <https://arxiv.org/abs/1806.00921>
- [91] A. Masood, B. Sheng, P. Li, X. Hou, X. Wei, J. Qin, and D. Feng, "Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images," *J. Biomed. Inf.*, vol. 79, pp. 117–128, Mar. 2018.
- [92] G. González, S. Y. Ash, G. Vegas-Sánchez-Ferrero, J. O. Onieva, F. N. Rahaghi, J. C. Ross, A. Díaz, R. S. J. Estépar, and G. R. Washko, "Disease staging and prognosis in smokers using deep learning in chest computed tomography," *Amer. J. Respiratory Crit. Care Med.*, vol. 197, no. 2, pp. 193–203, 2018.
- [93] C. González-Gonzalo, B. Liefers, B. van Ginneken, and C. I. Sánchez, "Improving weakly-supervised lesion localization with iterative saliency map refinement," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [94] M. Winkels and T. S. Cohen, "3D G-CNNs for pulmonary nodule detection," 2018, *arXiv:1804.04656*. [Online]. Available: <https://arxiv.org/abs/1804.04656>
- [95] S. G. Armato, III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, and E. A. Hoffman, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011.

- [96] A. Alansary, O. Oktay, Y. Li, L. Le Folgoc, B. Hou, G. Vaillant, K. Kamnitsas, A. Vlontzos, B. Glocker, B. Kainz, and D. Rueckert, "Evaluating reinforcement learning agents for anatomical landmark detection," *Med. Image Anal.*, vol. 53, pp. 156–164, Apr. 2019.
- [97] M. Ferlaino, C. A. Glastonbury, C. Motta-Mejia, M. Vatish, I. Granne, S. Kennedy, C. M. Lindgren, and C. Nellåker, "Towards deep cellular phenotyping in placental histology," 2018, *arXiv:1804.03270*. [Online]. Available: <https://arxiv.org/abs/1804.03270>
- [98] F.-C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 176–189, Jan. 2019.
- [99] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "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.
- [100] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
- [101] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.
- [102] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.
- [103] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*. [Online]. Available: <https://arxiv.org/abs/1606.05718>
- [104] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer, "Large scale deep learning for computer aided detection of mammographic lesions," *Med. Image Anal.*, vol. 35, pp. 303–312, Jan. 2017.
- [105] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: <https://arxiv.org/abs/1711.05225>
- [106] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," 2017, *arXiv:1703.02442*. [Online]. Available: <https://arxiv.org/abs/1703.02442>
- [107] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F.-E. de Leeuw, B. van Ginneken, and E. Marchiori, "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage, Clin.*, vol. 14, pp. 391–399, Feb. 2017.
- [108] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1558–1567, Jul. 2017.
- [109] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4753–4764, Oct. 2017.
- [110] A. Katzmann, A. Muehlberg, M. Suehling, D. Noerenberg, J. W. Holch, V. Heinemann, and H.-M. Gross, "Predicting lesion growth and patient survival in colorectal cancer patients using deep neural networks," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [111] Q. Meng, C. Baumgartner, M. Sinclair, J. Housden, M. Rajchl, A. Gomez, B. Hou, N. Toussaint, V. Zimmer, J. Tan, J. Matthew, D. Rueckert, J. Schnabel, and B. Kainz, "Automatic shadow detection in 2D ultrasound images," in *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 66–75.
- [112] Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Hirasawa, T. Tsuchida, T. Ozawa, S. Ishihara, Y. Kumagai, M. Fujishiro, I. Maetani, J. Fujisaki, and T. Tada, "Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks," *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 25–32, 2019.
- [113] K. Yasaka, H. Akai, O. Abe, and S. Kiryu, "Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study," *Radiology*, vol. 286, no. 3, pp. 887–896, 2017.
- [114] D. Zhang, J. Wang, J. H. Noble, and B. M. Dawant, "Accurate detection of inner ears in head CTs using a deep volume-to-volume regression network with false positive suppression and a shape-based constraint," 2018, *arXiv:1806.04725*. [Online]. Available: <https://arxiv.org/abs/1806.04725>
- [115] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.
- [116] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," 2017, *arXiv:1712.06957*. [Online]. Available: <https://arxiv.org/abs/1712.06957>
- [117] Y. Li and W. Ping, "Cancer metastasis detection with neural conditional random field," 2018, *arXiv:1806.07064*. [Online]. Available: <https://arxiv.org/abs/1806.07064>
- [118] P. Bändi et al., "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 550–560, Feb. 2019.
- [119] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
- [120] M. Forouzanfar, N. Forghani, and M. Teshnehlab, "Parameter optimization of improved fuzzy C-means clustering algorithm for brain MR image segmentation," *Eng. Appl. Artif. Intell.*, vol. 23, no. 2, pp. 160–168, 2010.
- [121] R. K. Justice, E. M. Stokely, J. S. Strobel, R. E. Ideker, and W. M. Smith, "Medical image segmentation using 3D seeded region growing," *Proc. SPIE*, vol. 3034, Apr. 1997, pp. 900–911.
- [122] X. Zhang, X. Li, and Y. Feng, "A medical image segmentation algorithm based on bi-directional region growing," *Optik*, vol. 126, no. 20, pp. 2398–2404, 2015.
- [123] W. H. Elmasry, H. M. Moftah, N. El-Bendary, and A. E. Hassanien, "Graph partitioning based automatic segmentation approach for CT scan liver images," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2012, pp. 183–186.
- [124] L. Grady and E. Schwartz, "Isoperimetric graph partitioning for data clustering and image segmentation," Dept. Cogn. Neural Syst., Boston Univ. Center Adapt. Syst., Boston, MA, USA, Tech. Rep. CAS/CNS-TR-03-015, 2003.
- [125] U. Maulik, "Medical image segmentation using genetic algorithms," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 166–173, Mar. 2009.
- [126] R. Dey and Y. Hong, "CompNet: Complementary segmentation network for brain MRI extraction," 2018, *arXiv:1804.00521*. [Online]. Available: <https://arxiv.org/abs/1804.00521>
- [127] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.
- [128] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.
- [129] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2017.
- [130] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 655–663.
- [131] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Inherent brain segmentation quality control from fully convnet Monte Carlo sampling," 2018, *arXiv:1804.07046*. [Online]. Available: <https://arxiv.org/abs/1804.07046>

- [132] R. Robinson, O. Oktay, W. Bai, V. V. Valindria, M. M. Sanghvi, N. Aung, J. M. Paiva, F. Zemrak, K. Fung, E. Lukaschuk, A. M. Lee, V. Carapella, Y. J. Kimm, B. Kainz, S. K. Piechnik, S. Neubauer, S. E. Petersen, C. Page, D. Rueckert, and B. Glocker, "Subject-level prediction of segmentation failure using real-time convolutional neural nets," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [133] V. K. Singh, S. Romani, H. A. Rashwan, F. Akram, N. Pandey, M. M. K. Sarker, J. T. Barrena, S. Abdulwahab, A. Saleh, M. Arquez, M. Arenas, and D. Puig, "Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification," 2018, *arXiv:1805.10207*. [Online]. Available: <https://arxiv.org/abs/1805.10207>
- [134] J. Zhang, A. Saha, B. J. Soher, and M. A. Mazurowski, "Automatic deep learning-based normalization of breast dynamic contrast-enhanced magnetic resonance images," 2018, *arXiv:1807.02152*. [Online]. Available: <https://arxiv.org/abs/1807.02152>
- [135] K. Men, T. Zhang, X. Chen, B. Chen, Y. Tang, S. Wang, Y. Li, and J. Dai, "Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning," *Phys. Medica*, vol. 50, pp. 13–19, Jun. 2018.
- [136] J. Lee and R. M. Nishikawa, "Automated mammographic breast density estimation using a fully convolutional network," *Med. Phys.*, vol. 45, no. 3, pp. 1178–1190, 2018.
- [137] Y. Zhang and A. C. S. Chung, "Deep supervision with additional labels for retinal vessel segmentation task," 2018, *arXiv:1806.02132*. [Online]. Available: <https://arxiv.org/abs/1806.02132>
- [138] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
- [139] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.
- [140] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
- [141] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, pp. 1342–1350, 2018.
- [142] T. J. Jebaseeli, C. A. D. Durai, and J. D. Peter, "Segmentation of retinal blood vessels from ophthalmologic diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 73, pp. 245–258, Jan. 2019.
- [143] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2018.
- [144] J. Duan, J. Schlemper, W. Bai, T. J. W. Dawes, G. Bello, G. Doumou, A. De Marvao, D. P. O'Regan, and D. Rueckert, "Deep nested level sets: Fully automated segmentation of cardiac MR images in patients with pulmonary hypertension," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 595–603.
- [145] W. Bai et al., "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," 2017, *arXiv:1710.09289*. [Online]. Available: <https://arxiv.org/abs/1710.09289>
- [146] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
- [147] W. Bai, H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews, and D. Rueckert, "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 586–594.
- [148] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. Newby, R. Dharmakumar, and S. A. Tsafaris, "Factorised spatial representation learning: Application in semi-supervised myocardial segmentation," 2018, *arXiv:1803.07031*. [Online]. Available: <https://arxiv.org/abs/1803.07031>
- [149] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed, "Constrained-CNN losses for weakly supervised segmentation," 2018, *arXiv:1805.04628*. [Online]. Available: <https://arxiv.org/abs/1805.04628>
- [150] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*. [Online]. Available: <https://arxiv.org/abs/1606.02147>
- [151] T. Joyce, A. Chartsias, and S. A. Tsafaris, "Deep multi-class segmentation without ground-truth labels," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [152] R. LaLonde and U. Bagci, "Capsules for object segmentation," 2018, *arXiv:1804.04241*. [Online]. Available: <https://arxiv.org/abs/1804.04241>
- [153] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.
- [154] C.-M. Nam, J. Kim, and K. J. Lee, "Lung nodule segmentation with convolutional neural network trained by simple diameter information," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [155] N. Burlutskiy, F. Gu, L. K. Wilen, M. Backman, and P. Micke, "A deep learning framework for automatic diagnosis in lung cancer," 2018, *arXiv:1807.10466*. [Online]. Available: <https://arxiv.org/abs/1807.10466>
- [156] H. R. Roth, C. Shen, H. Oda, M. Oda, Y. Hayashi, K. Misawa, and K. Mori, "Deep learning and its application to medical image segmentation," *Med. Imag. Technol.*, vol. 36, no. 2, pp. 63–71, Mar. 2018.
- [157] A. Taha, P. Lo, J. Li, and T. Zhao, "Kid-net: Convolution networks for kidney vessels segmentation from CT-volumes," 2018, *arXiv:1806.06769*. [Online]. Available: <https://arxiv.org/abs/1806.06769>
- [158] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
- [159] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: <https://arxiv.org/abs/1804.03999>
- [160] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. R. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Towards image-guided pancreas and biliary endoscopy: Automatic multi-organ segmentation on abdominal CT with dense dilated networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 728–736.
- [161] M. P. Heinrich, O. Oktay, and N. Bouteldja, "OBELISK—One kernel to solve nearly everything: Unified 3D binary convolutions for image analysis," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [162] M. P. Heinrich and O. Oktay, "BRIEFnet: Deep pancreas segmentation using binary sparse convolutions," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 329–337.
- [163] M. P. Heinrich, M. Blendowski, and O. Oktay, "Ternarynet: Faster deep model inference without gpus for medical 3D segmentation using sparse and binary convolutions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 9, pp. 1311–1320, 2018.
- [164] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation," 2018, *arXiv:1806.07201*. [Online]. Available: <https://arxiv.org/abs/1806.07201>
- [165] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*. [Online]. Available: <https://arxiv.org/abs/1703.10593>
- [166] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, "Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks," *Med. Phys.*, vol. 45, no. 10, pp. 4558–4567, 2018.
- [167] D. Yang, D. Xu, S. K. Zhou, B. Georgescu, M. Chen, S. Grbic, D. Metaxas, and D. Comaniciu, "Automatic liver segmentation using an adversarial image-to-image network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 507–515.
- [168] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Işgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," 2018, *arXiv:1804.04383*. [Online]. Available: <https://arxiv.org/abs/1804.04383>
- [169] D. Jin, Z. Xu, Y. Tang, A. P. Harrison, and D. J. Mollura, "CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation," 2018, *arXiv:1806.04051*. [Online]. Available: <https://arxiv.org/abs/1806.04051>

- [170] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, and D. J. Mollura, "Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 621–629.
- [171] X. Xu, Q. Lu, Y. Hu, L. Yang, S. Hu, D. Chen, and Y. Shi, "Quantization of fully convolutional networks for accurate biomedical image segmentation," 2018, *arXiv:1803.04907*. [Online]. Available: <https://arxiv.org/abs/1803.04907>
- [172] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, Jan. 2017.
- [173] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D. Z. Chen, "Suggestive annotation: A deep active learning framework for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 399–407.
- [174] D. Liu and T. Jiang, "Deep reinforcement learning for surgical gesture segmentation and classification," 2018, *arXiv:1806.08089*. [Online]. Available: <https://arxiv.org/abs/1806.08089>
- [175] S. M. R. Al Arif, K. Knapp, and G. Slabaugh, "SPNet: Shape prediction using a fully convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 430–439.
- [176] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.
- [177] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.
- [178] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.
- [179] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans," *Sci. Rep.*, vol. 6, Apr. 2016, Art. no. 24454.
- [180] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [181] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
- [182] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, 2017.
- [183] B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks," *Med. Phys.*, vol. 44, no. 2, pp. 547–557, 2017.
- [184] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh, F. U. H. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, and D. Puig, "SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," 2018, *arXiv:1805.10241*. [Online]. Available: <https://arxiv.org/abs/1805.10241>
- [185] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016, *arXiv:1605.01397*. [Online]. Available: <https://arxiv.org/abs/1605.01397>
- [186] Z. Mirikharaji and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 737–745.
- [187] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," 2017, *arXiv:1703.05165*. [Online]. Available: <https://arxiv.org/abs/1703.05165>
- [188] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Med. Image Anal.*, vol. 52, pp. 109–118, Feb. 2019.
- [189] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. Woods, J. J. Mann, and R. V. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, 2009.
- [190] R. Bajcsy and S. Kovačić, "Multiresolution elastic matching," *Comput. Vis., Graph., Image Process.*, vol. 46, no. 1, pp. 1–21, 1989.
- [191] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.
- [192] M. F. Beg, M. I. Miller, and A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vis.*, vol. 61, no. 2, pp. 139–157, 2005.
- [193] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007.
- [194] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
- [195] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through MRFs and efficient linear programming," *Med. Image Anal.*, vol. 12, no. 6, pp. 731–741, 2008.
- [196] A. V. Dalca, A. Bobu, N. S. Rost, and P. Golland, "Patch-based discrete registration of clinical brain images," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.* Cham, Switzerland: Springer, 2016, pp. 60–67.
- [197] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 344–352.
- [198] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "SVF-Net: Learning deformable image registration using shape matching," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 266–274.
- [199] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 232–239.
- [200] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—A deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, Sep. 2017.
- [201] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "Unsupervised deformable image registration with fully connected generative neural network," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018.
- [202] B. Hou, N. Miolane, B. Khanal, M. Lee, A. Alansary, S. McDonagh, J. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, "Deep pose estimation for image-based registration," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–3.
- [203] B. Hou, B. Khanal, A. Alansary, S. McDonagh, A. Davidson, M. Rutherford, J. V. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, "Image-based registration in canonical atlas space," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–3.
- [204] B. Hou, B. Khanal, A. Alansary, S. McDonagh, A. Davidson, M. Rutherford, J. V. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, "3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2-D images," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1737–1750, Aug. 2018.
- [205] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9252–9260.
- [206] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 781–791, Mar. 2018.
- [207] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1449–1453.

- [208] H. Tang, A. Pan, Y. Yang, K. Yang, Y. Luo, S. Zhang, and S. H. Ong, “Retinal image registration based on robust non-rigid point matching method,” *J. Med. Imag. Health Inform.*, vol. 8, no. 2, pp. 240–249, 2018.
- [209] L. Pan, F. Shi, W. Zhu, B. Nie, L. Guan, and X. Chen, “Detection and registration of vessels for longitudinal 3D retinal OCT images using SURF,” *Proc. SPIE*, vol. 10578, Mar. 2018, Art. no. 105782P.
- [210] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded up robust features,” in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
- [211] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [212] D. Mahapatra, S. Sedai, and R. Garnavi, “Elastic registration of medical images with GANs,” 2018, *arXiv:1805.02369*. [Online]. Available: <https://arxiv.org/abs/1805.02369>
- [213] K. A. J. Eppenhof, M. W. Lafarge, P. Moeskops, M. Veta, and J. P. W. Pluim, “Deformable image registration using convolutional neural networks,” *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105740S.
- [214] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, “Four-dimensional deformable image registration using trajectory modeling,” *Phys. Med. Biol.*, vol. 55, no. 1, p. 305, 2010.
- [215] J. Vandemeulebroucke, O. Bernard, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut, “Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT,” *Med. Phys.*, vol. 39, no. 2, pp. 1006–1015, 2012.
- [216] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, “A deep learning framework for unsupervised affine and deformable image registration,” *Med. Image Anal.*, vol. 52, pp. 128–143, Feb. 2018.
- [217] J. Zheng, S. Miao, Z. J. Wang, and R. Liao, “Pairwise domain adaptation module for CNN-based 2-D/3-D registration,” *Proc. SPIE*, vol. 5, no. 2, 2018, Art. no. 021204.
- [218] J. Lv, M. Yang, J. Zhang, and X. Wang, “Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: A feasibility study,” *Brit. J. Radiol.*, vol. 91, no. 1083, 2018, Art. no. 20170788.
- [219] J. Lv, W. Huang, J. Zhang, and X. Wang, “Performance of U-net based pyramidal lucas-kanade registration on free-breathing multi-B-value diffusion MRI of the kidney,” *Brit. J. Radiol.*, vol. 91, no. 1086, 2018, Art. no. 20170813.
- [220] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, “Adversarial image registration with application for MR and TRUS image fusion,” 2018, *arXiv:1804.11024*. [Online]. Available: <https://arxiv.org/abs/1804.11024>
- [221] Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. A. Noble, D. C. Barratt, and T. Vercauteren, “Label-driven weakly-supervised learning for multimodal deformable image registration,” in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1070–1074.
- [222] G. Zhao, B. Zhou, K. Wang, R. Jiang, and M. Xu, “Respond-CAM: Analyzing deep models for 3D imaging data by visualizations,” 2018, *arXiv:1806.00102*. [Online]. Available: <https://arxiv.org/abs/1806.00102>
- [223] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. ICCV*, Oct. 2017, pp. 618–626.
- [224] T. Elss, H. Nickisch, T. Wissel, R. Bippus, M. Morlock, and M. Grass, “Motion estimation in coronary ct angiography images using convolutional neural networks,” in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–9.
- [225] 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.
- [226] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, “Scalable high-performance image registration framework by unsupervised deep feature representations learning,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.
- [227] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis, “A deep metric for multimodal registration,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 10–18.
- [228] X. Yang, R. Kwitt, and M. Niethammer, “Fast predictive image registration,” in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 48–57.
- [229] K. Bahrami, F. Shi, I. Rekik, and D. Shen, “Convolutional neural network for reconstruction of 7T-like images from 3T MRI using appearance and anatomical features,” in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 39–47.
- [230] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, and H. Zheng, “Deep learning based classification of breast tumors with shear-wave elastography,” *Ultrasonics*, vol. 72, pp. 150–157, Dec. 2016.
- [231] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 417–425.
- [232] E. Kang, J. Min, and J. C. Ye, “A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction,” *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.
- [233] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, “End-to-end unsupervised deformable image registration with a convolutional neural network,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 204–212.
- [234] P. Radau, Y. Lu, K. Connelly, G. Paul, A. J. Dick, and G. A. Wright, “Evaluation framework for algorithms segmenting short axis cardiac MRI,” *MIDAS J.-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, Jul. 2009, p. 1.
- [235] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, “Brain biomarker interpretation in ASD using deep learning and fMRI,” 2018, *arXiv:1808.08296*. [Online]. Available: <https://arxiv.org/abs/1808.08296>
- [236] E. H. Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. Shalaby, M. F. Casanova, G. Barnes, G. Gimel’farb, R. S. Keynton, and A. S. El-Baz, “Alzheimer’s disease diagnostics by a 3D deeply supervised adaptable convolutional network,” *Frontiers Biosci.*, vol. 23, no. 3, pp. 584–596, 2018.
- [237] W. Yan, H. Zhang, J. Sui, and D. Shen, “Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 249–257.
- [238] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, “Identification of autism spectrum disorder using deep learning and the ABIDE dataset,” *NeuroImage, Clin.*, vol. 17, pp. 16–23, Jan. 2018.
- [239] M. Soussia and I. Rekik, “A review on image- and network-based brain data analysis techniques for Alzheimer’s disease diagnosis reveals a gap in developing predictive methods for prognosis,” 2018, *arXiv:1808.01951*. [Online]. Available: <https://arxiv.org/abs/1808.01951>
- [240] B. Gutierrez-Becker and C. Wachinger, “Deep multi-structural shape analysis: Application to neuroanatomy,” 2018, *arXiv:1806.01069*. [Online]. Available: <https://arxiv.org/abs/1806.01069>
- [241] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, vol. 1, no. 2, p. 4.
- [242] R. Awan, N. A. Koobanani, M. Shaban, A. Lisowska, and N. Rajpoot, “Context-aware learning using transferable features for classification of breast cancer histology images,” in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 788–795.
- [243] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, “Classification of breast cancer histology images using convolutional neural networks,” *PLoS ONE*, vol. 12, no. 6, 2017, Art. no. e0177544.
- [244] H. D. Couture, J. S. Marron, C. M. Perou, M. A. Troester, and M. Niethammer, “Multiple instance learning for heterogeneous images: Training a CNN for histopathology,” 2018, *arXiv:1806.05083*. [Online]. Available: <https://arxiv.org/abs/1806.05083>
- [245] M. A. Troester, X. Sun, E. H. Allott, J. Geradts, S. M. Cohen, C.-K. Tse, E. L. Kirk, L. B. Thorne, M. Mathews, Y. Li, and Z. Hu, “Racial differences in PAM50 subtypes in the carolina breast cancer study,” *JNCI: J. Nat. Cancer Inst.*, vol. 110, no. 2, pp. 176–182, 2018.
- [246] P. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, “Multiple instance learning for histopathological breast cancer image classification,” *Expert Syst. Appl.*, vol. 117, pp. 103–111, Mar. 2019.

- [247] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for classification of breast histology images using deep learning," *Comput. Med. Imag. Graph.*, vol. 71, pp. 90–103, Jan. 2019.
- [248] N. Antropova, H. Abe, and M. Giger, "Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks," *Proc. SPIE*, vol. 5, no. 1, 2018, Art. no. 014503.
- [249] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 4165.
- [250] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012.
- [251] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," *Proc. SPIE*, vol. 10669, May 2018, Art. no. 1066905.
- [252] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Comput. Sci.*, vol. 90, pp. 200–205, Jan. 2016.
- [253] M. S. Ayhan and P. Berens, "Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks," in *Proc. 1st Conf. Med. Imag. Deep Learn.*, 2018, pp. 1–9.
- [254] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.
- [255] M. Mateen, J. Wen, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, 2019.
- [256] R. Dey, Z. Lu, and Y. Hong, "Diagnostic classification of lung nodules using 3D neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 774–778.
- [257] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 10, pp. 1799–1808, 2017.
- [258] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, and R. M. Summers, Z. Xu, and D. J. Mollura, "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Comput. Methods Biomed. Biomed. Eng., Imag. Vis.*, vol. 6, no. 1, pp. 1–6, 2018.
- [259] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, and M. Chen, "Large margin local estimate with applications to medical image classification," *IEEE Trans. Med. Imag.*, vol. 34, no. 6, pp. 1362–1377, Jun. 2015.
- [260] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 797–808, Apr. 2013.
- [261] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Comput. Med. Imag. Graph.*, vol. 36, no. 3, pp. 227–238, 2012.
- [262] C. Biffi, O. Oktay, G. Tarroni, W. Bai, A. De Marvao, G. Doumou, M. Rajchl, R. Bedair, S. Prasad, S. Cook, D. O'Regan, and D. Rueckert, "Learning interpretable anatomical features through deep generative models: Application to cardiac remodeling," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 464–471.
- [263] C. Brestel, R. Shadmi, I. Tamir, M. Cohen-Sfaty, J. Laserson, E. Goz, and E. Elnekave, "RadBot-CXR: Classification of four clinical finding categories in chest X-ray using deep learning," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–9.
- [264] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, Q. Huang, M. Cai, and P.-A. Heng, "Weakly supervised learning for whole slide lung cancer image classification," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–10.
- [265] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, p. 1559, 2018.
- [266] J. M. Tomczak, M. Ilse, M. Welling, M. Jansen, H. G. Coleman, M. Lucas, K. de Laat, M. de Bruin, H. Marquering, M. J. van der Wel, O. J. de Boer, C. D. S. Heijink, and S. L. Meijer, "Histopathological classification of precursor lesions of esophageal adenocarcinoma: A deep multiple instance learning approach," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–3.
- [267] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," 2018, *arXiv:1802.04712*. [Online]. Available: <https://arxiv.org/abs/1802.04712>
- [268] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 289–293.
- [269] Z. Xu, Y. Huo, J. Park, B. Landman, A. Milkowski, S. Grbic, and S. Zhou, "Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images," 2018, *arXiv:1805.10376*. [Online]. Available: <https://arxiv.org/abs/1805.10376>
- [270] M. Anthimopoulos, S. Christodoulidis, L. Ebner, 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, 2016.
- [271] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, and N. Karssemeijer, "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.
- [272] B. Q. Huynh, H. Li, and M. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *J. Med. Imag.*, vol. 3, no. 3, p. 034501, Aug. 2016.
- [273] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. N. Metaxas, and X. S. Zhou, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, May 2016.
- [274] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.
- [275] W. Sun, T. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Comput. Med. Imag. Graph.*, vol. 57, pp. 4–9, Apr. 2017.
- [276] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multi-source transfer learning with convolutional neural networks for lung pattern analysis," 2016, *arXiv:1612.02589*. [Online]. Available: <https://arxiv.org/abs/1612.02589>
- [277] K. Lekadir, A. Galimzianova, A. Betriu, M. del Mar Vila, L. Igual, D. L. Rubin, and E. Fernández, P. Radeva, and S. Napel, "A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 48–55, Jan. 2017.
- [278] M. Combalia and V. Vilaplana, "Monte-carlo sampling applied to multiple instance learning for whole slide image classification," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2018, pp. 1–3.
- [279] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1195–1200.
- [280] O. Pasarin, K. Mulpuri, A. Cooper, A. J. Hodgson, and R. Garbi, "Real time RNN based 3D ultrasound scan adequacy for developmental dysplasia of the hip," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 365–373.
- [281] C. Tanner, F. Ozdemir, R. Profanter, V. Vishnevsky, E. Konukoglu, and O. Goksel, "Generative adversarial networks for MR-CT deformable image registration," 2018, *arXiv:1807.07349*. [Online]. Available: <https://arxiv.org/abs/1807.07349>
- [282] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
- [283] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digit. Mammogr.*, 2000, pp. 212–218.
- [284] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008.
- [285] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordóñez, P. Massin, and A. Erginay, "Feedback on a publicly distributed image database: The messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

- [286] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.
- [287] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. Abramoff, “Comparative study of retinal vessel segmentation methods on a new publicly available database,” *Proc. SPIE*, vol. 5370, pp. 648–657, May 2004.
- [288] A. F. Fotenos, A. Z. Snyder, L. E. Girton, J. C. Morris, and R. L. Buckner, “Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD,” *Neurology*, vol. 64, no. 6, pp. 1032–1039, 2005.
- [289] J. C. Morris, “The clinical dementia rating (CDR): Current version and scoring rules,” *Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993.
- [290] R. L. Buckner, D. Head, J. Parker, A. F. Fotenos, D. Marcus, J. C. Morris, and A. Z. Snyder, “A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume,” *NeuroImage*, vol. 23, no. 2, pp. 724–738, 2004.
- [291] E. H. Rubin, M. Storandt, J. P. Miller, D. A. Kinscherf, E. A. Grant, J. C. Morris, and L. Berg, “A prospective study of cognitive function and onset of dementia in cognitively healthy elders,” *Arch. Neurol.*, vol. 55, no. 3, pp. 395–401, 1998.
- [292] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [293] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, and P. Taylor, “Mammographic image analysis society (MIAS) database v1.21,” in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2015. [Online]. Available: <https://www.repository.cam.ac.uk/handle/1810/250394>
- [294] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, “Interactive whole-heart segmentation in congenital heart disease,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 80–88.
- [295] S. Azizi, P. Mousavi, P. Yan, A. Tahmasebi, J. T. Kwak, S. Xu, B. Turkbey, P. Choyke, P. Pinto, and B. Wood, “Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 7, pp. 1111–1121, 2017.
- [296] T. Kooi, B. van Ginneken, N. Karssemeijer, and A. den Heeten, “Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network,” *Med. Phys.*, vol. 44, no. 3, pp. 1017–1027, 2017.
- [297] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, K. H. Cha, and C. D. Richter, “Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms,” *Phys. Med. Biol.*, vol. 62, no. 23, p. 8894, 2017.
- [298] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese, “Taskonomy: Disentangling task transfer learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3712–3722.
- [299] N. Akhtar, A. Mian, and F. Porikli, “Joint discriminative Bayesian dictionary and classifier learning,” in *Proc. CVPR*, vol. 4, Jul. 2017, pp. 1193–1202.
- [300] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 689–696.
- [301] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, “Deep learning with non-medical training used for chest pathology identification,” *Proc. SPIE*, vol. 9414, Mar. 2015, Art. no. 94140V.
- [302] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, “Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box,” *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, Dec. 2015.
- [303] U. K. Lopes and J. F. Valiati, “Pre-trained convolutional neural networks as feature extractors for tuberculosis detection,” *Comput. Biol. Med.*, vol. 89, pp. 135–143, Oct. 2017.
- [304] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, “Understanding neural networks through deep visualization,” 2015, *arXiv:1506.06579*. [Online]. Available: <https://arxiv.org/abs/1506.06579>
- [305] L. Zhang, A. Gooya, and A. F. Frangi, “Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets,” in *Proc. Int. Workshop Simulation Synth. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 61–68.
- [306] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, “Biomedical data augmentation using generative adversarial neural networks,” in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2017, pp. 626–634.
- [307] A. Lahiri, K. Ayush, P. K. Biswas, and P. Mitra, “Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy images: Automated vessel segmentation in retinal fundus image as test case,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jul. 2017, pp. 42–48.
- [308] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” 2014, *arXiv:1412.6572*. [Online]. Available: <https://arxiv.org/abs/1412.6572>
- [309] H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk, and L. Uhmann, “Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018.
- [310] G. A. Bowen, “Document analysis as a qualitative research method,” *Qualitative Res. J.*, vol. 9, no. 2, pp. 27–40, 2009.
- [311] G. G. Chowdhury, “Natural language processing,” *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2003.
- [312] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, “LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop,” 2015, *arXiv:1506.03365*. [Online]. Available: <https://arxiv.org/abs/1506.03365>
- [313] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, and A. Bolton, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, no. 7676, p. 354, 2017.



**FOUZIA ALTAF** received the M.Phil. and M.Sc. degrees from the University of the Punjab, Pakistan. She is currently pursuing the Ph.D. degree with the School of Science, Edith Cowan University (ECU), Australia. Her current research pertains to the areas of deep learning, medical image analysis, and computer vision. She was a recipient of the Research Training Program Scholarship at ECU.



**SYED M. S. ISLAM** received the B.Sc. degree in electrical and electronic engineering from the Islamic Institute of Technology, in 2000, the M.Sc. degree in computer engineering from the King Fahd University of Petroleum and Minerals, in 2005, and the Ph.D. degree (Hons.) in computer engineering from The University of Western Australia (UWA), in 2011. He was a Research Assistant Professor with UWA, from 2011 to 2015, a Research Fellow with Curtin University, from 2013 to 2015, and a Lecturer with UWA, from 2015 to 2016. Since 2016, he has been a Lecturer in computing and security sciences with the School of Science, Edith Cowan University, and an Adjunct Lecturer with the School of Computer Science and Software Engineering, UWA. He has published around 53 research articles and got nine public media releases. He obtained 14 competitive research grants for his research in the area of image processing, computer vision, and medical imaging. He has co-supervised to complete seven honors and postgraduate students and currently supervising one M.S. and seven Ph.D. students. He is serving as an Associate Editor for 13 international journals, a Technical Committee Member of 24 conferences, and a regular Reviewer for 26 journals. He is also serving on seven professional bodies, including the IEEE and the Australian Computer Society as a Senior Member.



**NAVEED AKHTAR** received the master's degree in computer science from Hochschule Bonn-Rhein-Sieg (HBRs), Germany, and the Ph.D. degree in computer vision from The University of Western Australia (UWA). He was a Research Fellow with UWA and The Australian National University. He is currently a Lecturer with UWA. His current research interests include medical imaging, adversarial machine learning, hyperspectral image analysis, and 3D point cloud analysis. His research in computer vision and pattern recognition is regularly published in the prestigious venues of the field, including the IEEE CVPR and the IEEE TPAMI. He has also served as a Reviewer for these and twelve other research venues in the broader field of machine learning. During his Ph.D. degree, he was a recipient of multiple scholarships and a winner of the Canon Extreme Imaging Competition, in 2015. He is currently serving as an Associate Editor for IEEE ACCESS.



**NAEEM KHALID JANJUA** is currently a Lecturer with the School of Science, Edith Cowan University, Perth. He has published an authored book, a book chapter, and various articles in international journals and refereed conference proceedings. His areas of active research are defeasible reasoning, argumentation, ontologies, data modeling, cloud computing, machine learning, and data mining. He works actively in the domain of business intelligence and web-based intelligent decision support systems. He is an Associate Editor of the *International Journal of Computer System Science and Engineering* (IJCSSE) and the *International Journal of Intelligent systems* (IJEIS).

• • •