BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 01/06 pp 1–18
Volume 1, Number 1, March 2018
DOI: 10.26599/BDMA.2018.9020001

Applications of Deep Learning to MRI Images: A Survey

Jin Liu, Yi Pan, Min Li, Ziyue Chen, Lu Tang, Chengqian Lu, and Jianxin Wang*

Abstract: Deep learning provides exciting solutions in many fields, such as image analysis, natural language processing, and expert system, and is seen as a key method for various future applications. On account of its non-invasive and good soft tissue contrast, in recent years, Magnetic Resonance Imaging (MRI) has been attracting increasing attention. With the development of deep learning, many innovative deep learning methods have been proposed to improve MRI image processing and analysis performance. The purpose of this article is to provide a comprehensive overview of deep learning-based MRI image processing and analysis. First, a brief introduction of deep learning and imaging modalities of MRI images is given. Then, common deep learning architectures are introduced. Next, deep learning applications of MRI images, such as image detection, image registration, image segmentation, and image classification are discussed. Subsequently, the advantages and weaknesses of several common tools are discussed, and several deep learning tools in the applications of MRI images are presented. Finally, an objective assessment of deep learning in MRI applications is presented, and future developments and trends with regard to deep learning for MRI images are addressed.

Key words: magnetic resonance imaging; deep learning; image detection; image registration; image segmentation; image classification

1 Introduction

Artificial intelligence[1–3] is not only a field of computer science that was created in the 1950s but also a thriving field with many practical applications and research hotspots. Artificial intelligence attempts to simulate human intelligence and produce a new intelligent machine that would be able to process information with human consciousness, behavior, and thinking. Its ultimate goal is to develop brain-like robots. Artificial intelligence has been applied to many fields, such as image analysis, natural language processing, robotics, and expert systems.

Machine learning[4–6] is the core of artificial intelligence and the fundamental approach toward designing intelligent computers. Machine learning involves a number of disciplines such as probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory. Machine learning mainly uses induction and synthesis to make computers acquire new knowledge by simulating human learning behavior and then reorganizes the existing knowledge to continually improve computer performance. Machine learning has also been widely applied in many fields, such as computer-aided disease diagnosis[7–9], bioinformatics[10–12], and computer vision[13–15]. The applications of machine learning span the entire field of artificial intelligence.

• Jin Liu, Min Li, Lu Tang, Chengqian Lu, and Jianxin Wang are with the School of Information Science and Engineering, Central South University, Changsha 410083, China. E-mail: {liujin06, limin, lutang, chengqlu, jxwang }@mail.csu.edu.cn.
• Yi Pan and Ziyue Chen are with the Department of Computer Science, Georgia State University, Atlanta, GA30302, USA. E-mail: yipan@gsu.edu; zchen2@student.gsu.edu.
∗ To whom correspondence should be addressed.
Manuscript received: 2017-07-26; accepted: 2017-11-01

With the deepening of artificial neural networks[16], the concept of deep learning[17, 18] has been proposed. Deep learning is not only an improvement in artificial neural networks, but also a new field in machine learning research[19–24]. The successful application of deep learning brings machine learning closer to artificial intelligence. The idea of the artificial neural networks arises from our understanding of the human brain, which comprises interconnections between neurons. The difference between artificial neural networks and the human brain is that any neuron in the human brain is connected to other neurons via a specific physical path, whereas neural networks contain discrete layers, connections, and data propagation directions. Since deep learning consists of more hidden layers in comparison to artificial neural networks, a more abstract high-level feature representation for different classes is formed by using multiple hidden layers to combine low-level features. Similar to artificial intelligence, deep learning also attempts to build and simulate the human brain to analyze the learning process of the neural network, which simulates the learning mechanism of the human brain when it attempts to understand unknown concepts. Deep learning has a good development momentum in data processing and analysis, and has been reviewed as one of the top 10 breakthrough technologies in the 2013 MIT technology review (`https://www.technologyreview.com/lists/technologies/2013/`). Thus far, deep learning has been widely used in the scientific[25, 26] and business community[27]. It is worth mentioning that Google launched the first generation of deep learning systems, namely, DistBelief[27], in 2011. By using the deep learning system, Google was able to scan thousands of their data center cores and build a larger neural network. The deep learning system has been widely deployed in Google's commercial products, such as Google Photos, Google Search, and Google Street View.

Feature representation plays an important role in medical image processing and analysis. As a technology, deep learning methods have two obvious advantages in feature representation, as follows:

- Deep learning can be used to automatically find features from a given dataset for each specific application. In general, traditional feature extraction methods are based on some prior knowledge to extract features in a particular application. Thus, these methods are semi-automatic learning methods.
- Deep learning can find new features that are suitable to specific applications, but have never been previously discovered by researchers. Traditional feature extraction methods are often limited by some a priori knowledge, which can only extract some features associated with a particular application.

Additionally, the two elements that affect the results of medical image processing and analysis, are image acquisition and image interpretation, as follows:

- Image acquisition: As we all know, the better the image quality, the better the results obtained in image processing and analysis. However, the quality of the image depends on image acquisition; therefore, the better the image acquisition, the higher the image quality. Magnetic Resonance Imaging (MRI) does not only have the characteristics of non-invasive and good soft tissue contrast, but also does not expose subjects to high ionizing radiation. Since MRI can provide a lot of invaluable information about tissue structures, such as shape, size, and localization, it is attracting more and more attention for clinical routine and computer-aided diagnosis[28–30]. Therefore, in this article, we focus on MRI images. MRI can be divided into structural and functional imaging. Structural imaging includes T1-weighted MRI (T1w), T2-weighted MRI (T2w), Diffusion Tensor Imaging (DTI), etc.; functional imaging includes resting state functional MRI (rs-fMRI), tasking state functional MRI (ts-fMRI), etc.
- Image interpretation: In clinical practice, most medical image interpretations are basically performed by clinicians to determine whether the subjects are abnormal. However, due to limitations with regard to the clinician's personal skills, subjectivity, energy, and other factors, the medical image interpretations by clinicians often differ significantly. To obtain accurate image interpretation results, it is imperative to develop an automatic image interpretation system that includes many functions, such as image detection, image registration, image segmentation, and image classification. To realize this system, many machine learning methods have been widely applied. However, due to the fact that deep learning architectures can obtain high-level latent

features, many researchers have applied deep learning architectures to the development of this automatic image interpretation system. Therefore, in this survey, we focus on deep learning.

In this survey, we provide a comprehensive review for the architectures of deep learning and their applications to MRI images based on the above analysis. First, we introduce some common architectures of deep learning. Then, we present several applications of deep learning in MRI images, such as image detection, image registration, image segmentation, and image classification. Subsequently, the advantages and weaknesses of several common tools are discussed, and several deep learning tools applied to MRI images are demonstrated. Finally, an objective assessment about deep learning in MRI applications is presented, and future developments and trends are addressed for deep learning by using MRI images.

2 Deep Learning Architectures

2.1 Artificial neural networks

Artificial neural networks[16] were proposed in the 1980s and have become a hot topic in the field of artificial intelligence. The essence of artificial neural networks is to learn the information processing process of the human brains neural network, and then develop a simpler model, which will form a specific network for a given network connection. Artificial neural networks are computational models consisting of a large number of nodes and connections. For a specific artificial neural network, each of its nodes represents a specific output function (denoted as an activation function), and each of its connections between two nodes represents the ability to transmit information between these two nodes (denoted as the weight). Therefore, an artificial neural network is often approximated as an algorithm, and the output of the algorithm depends mainly on the activation functions and the weights of the neural network. In general, the artificial neural networks can be divided into feedforward neural networks and feedback neural networks, according to the different connections of the networks, as follows:

- Feedforward neural networks: A feedforward neural network can be represented as a directed acyclic graph, without feedback in the network. The network implements the transformation of the information from the input space to the output space, and its information processing ability comes from the combination of many simple nonlinear functions. The topology structure of the network is relatively simple and easy to implement.
- Feedback neural networks: A feedback neural network can be represented as an undirected complete graph, with feedback in the network. The information processing of the network is the state transformation, which can use the dynamic system theory to deal with processing. The stability of the network is closely related to the associative memory function.

Over the last ten years, with the deepening of artificial neural network research, great progress has been made in multiple fields, such as pattern recognition, biology, intelligent robots, economy, and medicine. Furthermore, artificial neural networks have also been successfully applied to solving many practical problems, and exhibit good intelligence characteristics. In general, the characteristics of artificial neural networks are mainly reflected on three abilities; namely, the ability of self-learning, ability of associative memory, and ability of quickly finding an optimal solution.

2.2 Deep feedforward networks

Feedforward neural networks (also called deep feedforward networks) are the classical deep learning models[31, 32]. The purpose of training deep feedforward networks is to approximate the corresponding objective functions. A deep feedforward network can be defined as a mapping $y = f(x;\theta)$, which learns parameters θ to obtain the best function approximation. In general, a deep feedforward network includes an input layer, multiple hidden layers, and an output layer. Furthermore, the flow of information in the deep feedforward network flows only in one direction and never goes backward, as shown in Fig. 1. This is an example of a deep feedforward network with an input layer, three hidden layers, and an output layer.

As can be seen from Fig. 1, if given an input x, and three hidden functions f_1, f_2, and f_3, an output $f(x)$ can be obtain by training the deep feedforward network: $f(x) = f_3(f_2(f_1(x)))$. If y is the corresponding label of x, $f(x)$ should be very close to y. Thus, we can denote $f(x)$ as y. Such a chain structure is most common in deep feedforward networks. In general, the length of a chain structure can be called the depth of a deep feedforward network. Additionally, there has been a new professional terminology, namely, deep learning.

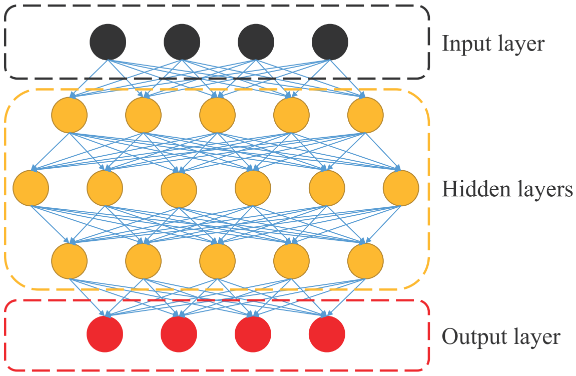


Fig. 1 An example of a deep feedforward network with an input layer, three hidden layers, and an output layer.

Therefore, the deep feedforward network is one of the most primitive deep learning architectures.

2.3 Stacked autoencoders

An autoencoder[33–35] is a simple deep feedforward network, which includes an input layer, a hidden layer, and an output layer. Meanwhile, according to different functions, an autoencoder can be divided into two parts: the encoder and decoder. The encoder (denoted as $f(x)$) is used to generate a reduced feature representation from an initial input x by a hidden layer h, and the decoder (denoted as $g(f(x))$) is used to reconstruct the initial input from the output of the encoder by minimizing the loss function:

$$L(x, g(f(x))) \tag{1}$$

By these two processes, high-dimensional data can be converted to low-dimensional data. Therefore, the autoencoder is very useful in classification and similar tasks.

The autoencoder has three common variants: the sparse autoencoder[36,37], denoising autoencoder[38–40], and contractive autoencoder[41,42], as follows:

- Sparse autoencoder: Unlike autoencoders, the sparse autoencoders add a sparse constraint $\Omega(h)$ to the hidden layer h. Thus, its reconstruction error can be evaluated by

$$L(x, g(f(x))) + \Omega(h) \tag{2}$$

- Denoising autoencoder: Unlike sparse autoencoders, which add a sparse constraint to the hidden layer, denoising autoencoders are aimed at minimizing the loss function:

$$L(x, g(f(\widetilde{x}))) \tag{3}$$

where $\widetilde{x}$ is based on x with some noise.

- Contractive autoencoder: Similar to sparse autoencoders, contractive autoencoders add the explicit regularizer $\Omega(h)$ to the hidden layer h, and minimize the explicit regularizer. The explicit regularizer is

$$\Omega(h)=\lambda\left\|\frac{\partial f(x)}{\partial x}\right\|_{\mathrm{F}}^{2} \tag{4}$$

where $\Omega(h)$ is the squared Frobenius norm[43] of the Jacobian partial derivative matrix of the encoder function $f(x)$, and λ is a free parameter.

A stacked autoencoder[31,44] is a neural network with multiple autoencoder layers as shown in Fig. 2. Furthermore, the input of the next layer comes from the output of the previous layer in the stacked autoencoders. An autoencoder usually consists of only three layers and does not have a deep learning architecture. However, the stacked autoencoder does have a deep learning architecture with a stacked number of autoencoders. It is worth mentioning that the training of the stacked autoencoder can only accomplish an action, layer-by-layer. For example, if we want to train a network with an $n \rightarrow m \rightarrow k$ architecture using a stacked autoencoder, we must first train the network $n \rightarrow m \rightarrow n$ to get the transformation $n \rightarrow m$, and then train the network $m \rightarrow k \rightarrow m$ to get the transformation $m \rightarrow k$, and finally stack the two transformations to form the stacked autoencoder (i.e., $n \rightarrow m \rightarrow k$). This process is also called layer-wise unsupervised pre-training[17].

2.4 Deep belief networks

The Boltzmann machine[45–48] is derived from statistical physics and is a modeling method based on energy functions that can describe the high order interaction between variables. Although the Boltzmann machine is relatively complex, it has a relatively complete physical interpretation and a strict mathematical statistics theory as its basis. The Boltzmann machine is a symmetric coupled random feedback binary unit neural network, which includes a visible layer and multiple hidden layers. The nodes of the Boltzmann machine can be divided into visible units and hidden units. In a Boltzmann machine, its visible and hidden units are used to represent the random neural network learning model, and its weights between two units in the model are used to represent the correlation between the

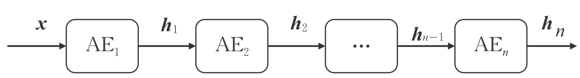


Fig. 2 An example of a stacked autoencoder with n autoencoders (i.e., AE_1, AE_2,.., AE_n).

corresponding two units.

A restricted Boltzmann machine[49, 50] is a special form of Boltzmann machine, which only includes a visible layer and a hidden layer. Unlike feedforward neural networks, the connections between the nodes of the hidden layer and the nodes of the visible layer in the restricted Boltzmann machines can be bi-directionally connected. Compared to Boltzmann machines, since the restricted Boltzmann machines only have one hidden layer, they have faster calculation speed and better flexibility. In general, restricted Boltzmann machines have two main functions: (1) Similar to autoencoders, restricted Boltzmann machines are used to reduce the dimension of data; (2) Restricted Boltzmann machines are used to obtain a weight matrix, which is used as the initial input of other neural networks.

Similar to stacked autoencoders, deep belief networks[51–54] are also neural networks with multiple restricted Boltzmann machine layers. Furthermore, in deep belief networks, the input of the next layer comes from the output of the previous layer. Deep belief networks adopt the hierarchical unsupervised greedy pre-training method[51] to pre-train each restricted Boltzmann machine in a hierarchical way. The obtained results were used as the initial input of the supervised learning probability model, whose learning performance improved greatly.

2.5 Convolutional neural networks

Convolutional neural networks[55–58] are also deep feedforward networks, and have been widely used in recognition tasks, such as document recognition[59], handwriting recognition[60], and image classification[61–64]. The only difference between the fully connected feedforward neural networks and the convolution neural networks is that the two adjacent layers of the two neural networks are connected in different ways. The former only has some nodes connected between the adjacent two layers, while the latter has all nodes connected between the adjacent two layers. The biggest problem of using a fully connected feedforward neural network is that there are too many parameters for the network. In general, increasing the parameters will not only lead to slower calculation speed, but will also lead to overfitting problems. To effectively reduce the number of parameters in the neural networks, more reasonable neural network architectures are required. Therefore, convolutional neural networks were proposed to achieve this goal.

Convolutional neural networks include two kernel layers, namely, the convolutional and pooling layers, as follows:

- **Convolutional Layer**: Only a small patch of the previous layer is used as the input of each node in the convolutional layer, and the size of the small patch is often 3×3 or 5×5. The convolutional layer attempts to analyze each small patch of the neural network in depth, which results in the higher abstraction of feature representation.
- **Pooling Layer**: There is often a pooling layer followed by the convolutional layer. The pooling layer can effectively reduce the size of the matrix from the previous convolutional layer; thus, it can reduce the number of parameters in the neural network. Therefore, the use of pooling layers can not only speed up the calculation, but can also prevent the problem of overfitting.

In general, there are two types of convolution neural network architectures, according to the different connection modes of the different convolutional layers. One is to connect different convolutional layers in series such as LeNet-5[59], AlexNet[61], and ZFNet[65], while the other one consists of connecting different convolutions in parallel, such as Inception and its follow-up versions[66–68].

3 Deep Learning Applications

In recent years, many deep learning methods have been proposed for application in the field of MRI image processing and analysis, such as image detection, image registration, image segmentation, and image classification. All of these can be formulated as feature representation problems, and can thus be solved effectively by using deep learning methods to find an effective set of features. In this section, we review the recent progress of applying deep learning architectures in the image detection, image registration, image segmentation, and image classification of MRI images.

3.1 Image detection

Image detection plays an important role in computer-aided detection routines. Its main purpose is to find the tissues of interest, and then measure and analyze whether these tissues produce lesions. Some deep learning methods have been proposed for performing MRI image detection as follows.

To perform organ detection from a given complex dataset with abnormalities, for which it is difficult to identify the labels of the samples in the dataset, Shin et al.[69] proposed a deep learning model with a stacked sparse autoencoder. In this study, the stacked sparse autoencoder model was generated by stacking several unsupervised feature learning layers, which were trained by using greedy methods. Subsequently, a pooling operation was applied to compress the features of gradually increased input regions, and to generate a part-based model to perform multiple organ detection in MRI images.

To achieve the automatic detection of lacunes of presumed vascular origin, Ghafoorian et al.[70] proposed an automated two-step deep convolutional network method. First, a fully convolutional network was applied to the detection of the initial candidates. Then, a 3D convolutional network was applied to reduce false positives. In this study, Ghafoorian et al.[70] suggested that location information plays an important role in the detection of candidate tissue. Therefore, to further improve detection performance, contextual information was generated by using multiple scale analysis and a combination of explicit location features to add into the convolutional network.

To detect cerebral microbleeds (CMBs) from MRI images, Dou et al.[71] proposed an automatic 3D convolutional network method. The 3D convolutional network was used to extract high-level features for CMBs via a data driven approach, which can effectively encode the spatial contextual information from MRI images. Since the 3D convolutional network adopted a traditional sliding window strategy, the computational cost of using the method to detect CMBs was relatively high. To further improve the performance of CMBs detection, a two-step cascaded 3D fully convolutional network framework was proposed. The 3D fully convolutional network was first used to rapidly retrieve potential candidates, and then used these potential candidates to further accurately distinguish CMBs from challenging mimics.

3.2 Image registration

Image registration is the process of matching and superimposing two or more images at different times, different sensors (such as imaging equipment) or different conditions (such as illumination, position, and angle)[72]. The general process of image registration is as follows:

- The features are obtained by the feature extraction of two images;
- The feature pairs are found by performing a similarity measure;
- The image space coordinate transformation parameters are obtained by matching feature pairs;
- The image space coordinate transformation parameters are used to perform image registration.

Image registration has been widely applied in medical image processing. Its main purpose is to combine various medical images, which display their information in the same image, and thereby provide multiple information for clinical diagnosis. Therefore, to achieve medical image registration, the building of accurate and effective correspondences between the two images is required. In general, the correspondences between two images can be represented by maximizing the similarity of the feature pairs.

Recently, as neuroimaging techniques developed, various new modalities have been emerging to make the diagnosis and treatment of diseases more accurate. Thus, image registration operations, which combine different modality data, are required. Many learning-based image registration methods have been proposed to help select the best related features, which are used to guide the corresponding detection between samples with large changes. However, for most of the existing learning-based image registration methods, there is a great limitation with regard to the fact that they need a lot of known correspondences during the training process. To address this limitation, Wu et al.[73] proposed an unsupervised deep learning framework to extract optimal image features for MRI image registration. In this study, first, a stacked convolutional independent subspace analysis network was developed to learn the hierarchical representations of patches from MRI brain images. The stacked convolutional independent subspace analysis network included two layers: (1) The first layer was used to extract the low-level feature representations of patches from MRI brain images; (2) The second layer was used to obtain the hierarchical representations. Then, the hierarchical representations were used to perform correspondence detection in the image registration process.

Later, in the same team, Wu et al.[74] also proposed an unsupervised deep learning framework to learn the hierarchical representations from MRI images for the purpose of image registration. The unsupervised deep learning framework contained a stacked autoencoder with a convolutional network. The 3D image patches were used as inputs for training the stacked autoencoder with a convolutional network. In this study, the stacked autoencoder mainly consisted of two networks; namely, the encoder and decoder networks, from which the former was used to learn the low-dimensional features from 3D image patches, and the latter was used to recover the 3D image patches from the learned low-dimensional features. However, if the inputs of the stacked autoencoder are very large, the computational cost of directly using the stacked autoencoder will be very high. In this study, a convolutional network was used to learn the translational invariant features, which can reduce the dimensionality of the original features, to reduce the computational cost of the stacked autoencoder.

3.3 Image segmentation

Automatic tissue segmentation in MRI images is of great importance in modern medical research and clinical routines. Many medical image segmentation challenges have been held to encourage the development of automatic segmentation techniques, such as Ischemic Stroke Lesion Segmentation (ISLES, `http://www.isles-challenge.org/`), Multimodal Brain Tumor Image Segmentation (BRATS, `http://braintumorsegmentation.org/`), MR Brain Image Segmentation (MRBrainS, `http://mrbrains13.isi.uu.nl/`), and cardiac MR Left Ventricle (LV) segmentation (`http://smial.sri.utoronto.ca/LV_Challenge/Home.html`). Many deep learning methods have also been proposed to perform the segmentation of various tissues in MRI images[75–78].

In MRI brain images, one of the most common image segmentations is the segmentation of Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF). To segment infant brain tissue images into GM, WM, and CSF, Zhang et al.[76] proposed the use of convolutional networks to achieve this goal by combining multi-modal MRI images, which are T1, T2, and Fractional Anisotropy (FA) images. In this study, four convolutional network architectures were designed according to different input patch sizes. These convolutional network architectures contained a different number of convolutional layers and resulting feature maps. To obtain the nonlinear mappings between the inputs and outputs of each convolutional network, the local response normalization scheme, the fully-connected layers, and the softmax layers were also applied to these convolutional networks. Moreover, to segment neonatal brain tissue images into Brain Stem (BS), cortical GM (cGM), myelinated WM (mWM), Basal Ganglia and Thalami (BGT), unmyelinated WM (uWM), ventricular CSF (vCSF), extracerebral CSF (eCSF), and cerebellum (CB), Moeskops et al.[77] also proposed a convolutional network to automatically segment these tissues. Similar to the convolutional networks previously proposed by Zhang et al.[76], the convolutional network also contained multiple convolutional layers and the resulting feature maps. Additionally, the fully-connected layers were also used in the convolutional network to represent each input patch size, and a single softmax output layer was used to connect these convolutional and fully-connected layers to perform the final segmentation..

Since most brain tumors can affect a patient's health, and even shorten their life expectancy, automatic and reliable segmentation techniques for removing brain tumors are required. However, most brain tumors have large spatial and structural variability, which makes them difficult to segment. Thus, automatic and reliable segmentation has become a challenging problem. To address the problem, many deep learning-based brain tumor segmentation methods have been proposed[79–84]. For example, Pereira et al.[83] used a convolutional network with small convolutional kernels to segment gliomas, which are the most common and aggressive brain tumors in MRI images. They believed that by using smaller kernels more convolutional layers could be stacked, and that the same results with larger kernels could be obtained. Additionally, to further improve the segmentation performance, both intensity normalization and volumetric constrains were used in the convolutional network. Later, Havaei et al.[84] also presented a fully automatic brain tumor segmentation method with a convolutional network. Unlike most traditional convolutional networks, the convolutional network in this study included three new components; namely, a two-pathway architecture, cascaded architecture, and two-phase training. The two-pathway architecture was used to obtain global contextual and local features, respectively. The cascaded architecture contained input concatenation, local pathway concatenation, and pre-output concatenation, and was used to exploit the output efficiency of a convolutional network. The two-phase training procedure was used to deal with the imbalance labels of brain tumors in MRI images.

The measurement of cardiac ventricle, including the LV and Right Ventricle (RV), plays an important role in the clinical assessment of cardiac structures and functions, such as ventricular volume, wall thickness, and ejection fraction. Therefore, the accurate and automatic segmentation of cardiac ventricle is also necessary. Recently, many deep learning methods have been proposed to segment the cardiac ventricle[85–87]. For example, to segment the LV from MRI images, Avendi et al.[86] proposed a methodology, which combined deep learning architectures and deformable models to perform this task. The method mainly included three steps: (1) Using convolutional networks to estimate the location of the LV from MRI images; (2) Using stacked autoencoders to infer the shape of the LV; (3) The inferred LV shape was used to perform the final segmentation of the LV by a deformable model. Later, Ngo et al.[87] proposed a combination method with deep learning architecture and a level set algorithm for the automated segmentation of the LV from MRI images. The method also includes three steps: (1) A deep belief network to estimate the location of the LV from MRI images; (2) Another deep belief network to delineate the endocardial and epicardial borders; (3) The estimated location of the LV and the delineation of the endocardial and epicardial borders were incorporated into the distance regularized level set method to perform the final segmentation of the LV.

In addition to the segmentation of the above tissues, the deep learning segmentation methods in the MRI images were also applied to other tissues, such as multiple sclerosis[88, 89], prostate[90], striatum[91], tibial cartilage[92], abdominal adipose tissues[93], and anterior visual pathway[94].

3.4 Image classification

Image classification plays an important role in automatic disease diagnosis and cognitive recognition, such as the classification of different severity diseases and the recognition of different brain activities. Many deep learning methods have also been proposed for performing image classification tasks in MRI images[95–97].

3.4.1 Alzheimer's disease classification

The automatic diagnosis of Alzheimer's Disease (AD), especially in its early stage, plays an important role in human health. Since AD is a neurodegenerative disease, it has a long incubation period. Therefore, it is necessary to analyze the symptoms of AD at different stages. Currently, many researchers have proposed using image classification to perform AD diagnosis. Moreover, many deep learning methods have been proposed to perform severity classification for different AD patients by using MRI images[95, 98–106].

To diagnose AD and its prodromal stage, namely, Mild Cognitive Impairment (MCI), Suk et al.[95] proposed a deep learning method for finding high-level latent and shared features from two imaging modalities: MRI images and Positron Emission Tomography (PET) images. In this study, a statistical significance test was first applied to obtain discriminative patches between classes. A multi-modal deep Boltzmann machine was built to find high-level latent and shared features from the paired patches. In the multi-modal deep Boltzmann machine, a Gaussian Restricted Boltzmann Machine was trained to transform the paired patches into binary vectors. The binary vectors were used as inputs to the multi-modal deep Boltzmann machine. After finding high-level latent and shared features by using the paired patches and trained multi-modal deep Boltzmann machine, an image-level classifier was developed to perform the final classification. The construction of the classifier mainly included three steps: (1) patch-level classifier learning; (2) mega-patch construction; (3) ensemble learning. Later, in the same team, Suk et al.[104] also proposed a deep learning method for AD classification to improve the previous classification performance. In this study, a stacked autoencoder was first developed to find high-level latent features from low-level features extracted from three data sources, namely MRI images, PET images, and CSF. Then, a sparse representation learning method was applied to select the most discrimination features from high-level latent features and two clinical scores. Finally, a multi-kernel support vector machine was applied to combine the selected multi-modal features to perform the final classification.

In another team, the deep learning methods around AD classification were also presented. For example, Liu et al.[99] designed a deep learning method, which combined a stacked sparse autoencoder and a softmax regression layer to diagnose AD and MCI. The stacked sparse autoencoder was trained to obtain high-level latent features from two imaging modalities, namely, MRI images and PET images. The softmax regression layer was used to obtain the probability of each subject to classify all experimental subjects. Subsequently, Liu et al.[100] also proposed a multi-phase feature representation learning framework to perform AD classification to improve the previous classification performance. Similar to Ref. [99], the purpose of the first phase was to obtain high-level latent features by using a stacked autoencoder from two imaging modalities; namely MRI images and PET images. The second phase consisted of using linear regression to optimize the obtained high-level latent features by adding some low-dimensional biomarkers. The third phase consisted of classifying all experimental subjects by using a softmax regression layer that was similar to Ref. [99].

3.4.2 Schizophrenia classification

Schizophrenia (SCZ) is a complex psychiatric disorder characterized by cognitive deterioration, aberrant sensory perception, and disturbed thinking[107,108]. Patients with SCZ may seem to lose their grasp to reality. Their families and society at large are also impacted by SCZ. Many patients with SCZ have difficulty in performing tasks or caring for themselves; therefore, they rely on others for help. Approximately 8 out of 1000 individuals will have an SCZ episode in their lifetime. Therefore, automatic diagnosis of SCZ is also necessary. Recently, many deep learning methods have been proposed to perform SCZ image classification[97,109–111].

Pinaya et al.[97] trained a deep neural network that combined a deep belief network and a softmax layer to extract high-level latent features from MRI images for the purpose of diagnosing patients with SCZ from health controls. The deep neural network was trained by using two steps: (1) pre-trained by a deep belief network; (2) supervised fine-tuning by a softmax layer. The pre-trained network was used to find high-level latent features from MRI images. The softmax layer was used to refine the pre-trained network by supervised fine-tuning, and to perform the final classification.

Later, with regard to the problem of SCZ classification, Kim et al.[110] also presented a deep neural network with multiple hidden layers and a softmax layer to obtain high-level latent features from low-level features extracted from MRI images. To further improve the accuracy of SCZ classification, both L_1-norm regularization and a stacked autoencoder were incorporate into the deep neural network. The L_1-norm regularization was used to control the weight sparsity in each hidden layer. The stacked autoencoder was used to pre-train the deep neural network weights for initialization.

3.4.3 Brain activity classification

In general, different external stimuli correspond to different brain activities, and different brain activities exhibit different functional brain images[112,113]. Therefore, image classification plays an important role in identifying different brain activities. Recently, many deep learning methods were proposed to perform image classification of different brain activities[96,114–116].

To identify different brain activities including emotions, social, motor, working memory, gambling, relational and language activities, Koyamada et al.[96] trained a feedforward deep neural network from fMRI images to implement this task. The feedforward deep neural network included multiple hidden layers and a maxsoft layer. Similarly, these hidden layers were used to obtain high-level latent features, and the softmax layer was used to calculate the probability of each subject in a class. Additionally, minibatch stochastic gradient descent, dropout[117], and principal sensitivity analysis[118] were incorporated into the feedforward deep neural network to improve the final classification performance.

Recently, Jang et al.[116] employed fully connected feed-forward deep neural networks with multiple hidden layers to classify different sensorimotor tasks including auditory attention, right-hand clenching, visual stimulus, and left-hand clenching. In this study, a deep belief network with a restricted Boltzmann machine was pre-trained and used to initialize the weights of fully connected feedforward deep neural networks. Then, a back-propagation algorithm was used to fine-tune the deep belief network to control the weight-sparsity levels across hidden layers.

In addition to the above three classifications, the deep learning classification methods of MRI images has also been applied to other classification fields, such as classification of Attention Deficit Hyperactivity Disorder (ADHD)[119,120], age prediction[121–123], stroke diagnosis[124], emotional response prediction[125], and discrimination of cerebellar ataxia types[126].

4 Deep Learning Tools

4.1 General deep learning tools

Deep learning is a complex technology. To achieve the abovementioned deep learning architectures, researchers need to spend a lot of time and energy.

Fortunately, in recent years, many deep learning tools have been developed as shown in Table 1. These tools are convenient for researchers; thus, they promote the application of deep learning architectures. Some common and widely used deep learning tools are shown in Table 1 and are briefly introduced as follows.

Caffe is not only the first major industry-level deep learning tool, but also the most popular tool in the field of computer vision. Caffe is an open deep learning framework created by Yangqing Jia[127]. The advantages and weaknesses of Caffe are as follows:

- **Advantages:**
  - ▷ Fast running;
  - ▷ Specializes in image processing;
  - ▷ Fine-tunes existing networks directly;
  - ▷ Trains models directly without writing any code;
  - ▷ Supports Python as the Application Program Interface (API);
- **Weaknesses:**
  - ▷ Layer-based network structure; its scalability is not good, and requires writing code for new layers;
  - ▷ Too much extension and dependence, which results in increased inflating.

Torch is a scientific computing framework, and supports many machine learning algorithms. The main development languages of this framework are C and Lua. Several large technology companies, such as Facebook and Twitter have adopted this framework. The advantages and weaknesses of Torch are as follows:

- **Advantages:**
  - ▷ Fast running and good flexibility;
  - ▷ Optimizes basic computing units and it is easy to write new layers and run on GPU;
  - ▷ Includes many common computational models based on Lua;
  - ▷ Includes many pre-trained models;
- **Weaknesses:**
  - ▷ Lua has a steep learning curve;
  - ▷ Layer-based network structure; its scalability is not good and requires writing code for new layers;
  - ▷ Does not support Python as API.

Theano was developed by the Montreal Institute of Technology (MIT) in 2008. Python is the main development language of Theano. Theano derives a lot of Python packages with deep learning, such as Pylearn2 and Keras. Theano is the first architecture to describe the model using symbolic tensor graphs. The advantages and weaknesses of Theano are as follows:

- **Advantages:**
  - ▷ Good flexibility and suitable for academic research;
  - ▷ Good support for recursive network and language modeling;
  - ▷ Many high-level deep learning packages such as Keras and Pylearn2;
  - ▷ Good portability;
- **Weaknesses:**
  - ▷ Slow compilation;
  - ▷ Difficult to modify codes for developer;
  - ▷ Less pre-trained models.

To be compatible with traditional machine learning and deep neural networks, TensorFlow was created by Google to replace Theano. TensorFlow is also an open

Table 1 Some common and widely used deep learning tools.

Name	Link	Reference
DeepLearnToolbox	`https://github.com/rasmusbergpalm/DeepLearnToolbox`	[128]
Caffe	`http://caffe.berkeleyvision.org/`	[127]
Torch	`http://torch.ch/`	[129]
Theano	`http://deeplearning.net/software/theano`	[130]
Pylearn2	`http://deeplearning.net/software/pylearn2/`	[131]
Keras	`https://github.com/EderSantana/keras`	[132]
TensorFlow	`https://www.tensorflow.org/`	[133]
CNTK	`https://www.microsoft.com/en-us/research/product/cognitive-toolkit/`	[134]
MXNet	`https://github.com/dmlc/mxnet`	[135]
Chainer	`http://chainer.org/`	[136]
Deeplearning4j	`https://deeplearning4j.org/`	[137]
SINGA	`http://www.comp.nus.edu.sg/~dbsystem/singa/`	[138]
MatConvNet	`http://www.vlfeat.org/matconvnet/`	[139]
maxDNN	`https://github.com/eBay/maxDNN`	[140]

source software library, which uses data flow graphs to implement numerical computation. Each node in the data flow graphs represents a mathematical operation, and each edge in the data flow graphs represents the relationship between two multidimensional data arrays. TensorFlow can run on multiple platforms, such as one or more CPUs (or GPUs), mobile devices, and servers. The advantages and weaknesses of TensorFlow are as follows:

- **Advantages:**
  - ▷ High quality meta frameworks;
  - ▷ Supports multiple GPUs;
  - ▷ Faster compilation than Theano;
  - ▷ Rapid development for new networks;
  - ▷ Supports distributed training;
  - ▷ Good portability;
- **Weaknesses:**
  - ▷ Slower running, and need to lager memory;
  - ▷ Less pre-trained models;
  - ▷ Does not support dynamic input of convolutional operation, and does not support convolution of time series.

4.2 Deep learning tools applied to MRI images

In recent years, based on the abovementioned general deep learning tools, some deep learning tools applied in MRI images have also been developed, as shown in Table 2. They are briefly introduced as follows.

- BrainNet: This tool was developed based on TensorFlow, and aims to train deep neural networks to segment GM and WM from brain MRI images.
- LiviaNET: This tool was also developed based on Theano, and aims to train 3D fully convolutional neural networks by using MRI images to segment subcortical brain structures.
- DIGITS: This tool was also developed to rapidly train accurate deep neural networks for image segmentation, classification, and tissue detection tasks. For example, DIGITS is used to perform Alzheimers disease prediction by using MRI images and obtain good results[142].
- resnet_cnn_mri_adni: This tool was developed to train residual and plain convolutional neural networks by performing AD classification of MRI images.
- mrbrain: This tool was developed to train convolutional neural networks by using MRI images to predict the age of humans.
- DeepMedic: This tool was developed based on Theano, and aims to train multi-scale 3D convolutional neural networks for brain lesion segmentation from MRI images. Moreover, this tool has shown excellent performance in brain lesion segmentation tasks, and was the winner of the ISLES 2015 competition.

5 Conclusion and Outlook

In summary, the aim of this survey was to provide valuable insights for researchers, with regard to applying deep learning architectures in the field of MRI-based research. To our knowledge, we are the first to review MRI-based deep learning applications. As can be seen in Section 3, deep learning architectures have been widely applied in MRI image processing and analysis, in areas such as image detection, image registration, image segmentation and image classification, and can obtain better results.

Although many researchers have paid more attention to MRI-based deep learning applications and have obtained some relatively good results, there are many problems and challenges that need to be solved urgently due to various limitations. In particular, the two main problems and challenges are as follows:

- Limited dataset size and class imbalance: It is known that the larger the dataset, the better the results of deep learning. However, since MRI image acquisition processes are usually complex and expensive, the size of an MRI image dataset is limited in many applications. Furthermore, for privacy considerations, many MRI images

Table 2 Some deep learning tools applied in MRI images.

Name	Link	Reference
BrainNet	`https://github.com/kaspermarstal/BrainNet`	[67]
LiviaNET	`https://github.com/josedolz/LiviaNET`	[141]
DIGITS	`https://developer.nvidia.com/digits`	[142]
resnet_cnn_mri_adni	`https://github.com/neuro-ml/resnet_cnn_mri_adni`	[143]
mrbrain	`https://github.com/lanpa/mrbrain`	[144]
DeepMedic	`https://github.com/Kamnitsask/deepmedic`	[145]

(especially disease-related MRI images) are rarely obtained. Therefore, the size of the dataset with MRI images is often small. In addition, if the training of MRI images has class imbalance during the training process, it is often very difficult to obtain suitable deep neural networks. Presently, the two main strategies to improve the above problems are as follows:

- ▷ Sampling: includes oversampling and undersampling, which is widely used to enlarge and rebalance the size of the dataset with MRI images. Oversampling is applied to generate new MRI images from existing MRI images, and undersampling is applied to selecting some MRI images from existing MRI images.
- ▷ Pre-training: When the size of the dataset with MRI images is limited, an unsupervised pre-training operation can help to prevent overfitting and generate more regularized results. Therefore, pre-training is also widely applied to deal with limited dataset size and class imbalance. In general, the pre-training operation is followed by a fine-tuning operation.

- Choosing a suitable deep learning architecture and its corresponding hyperparameters for a particular application: since, thus far, the advantages and weaknesses of each deep learning architecture are only roughly understood by most researchers, finding a way toward choosing a suitable deep learning architecture remains an unsolved problem with regard to obtaining good results for a particular application. Even if a deep learning architecture for a particular application is determined, finding a way toward setting the hyperparameters of the architecture also remains an unsolved problem. Presently, most researchers are based on experimental experience to address the abovementioned problems. Therefore, finding a way to choose the most suitable deep learning architecture and its corresponding hyperparameters for a particular application is not only an urgent problem to be solved, but also a great challenge to be addressed in future work.

With the continuous advancement of medical big data, the size of the MRI image dataset will no longer be a problem. Moreover, as the understanding of deep learning architectures expands, choosing a suitable deep learning architecture and its corresponding hyperparameters for a particular application will also become achievable. It is reasonable to expect that deep learning applications in MRI images will attain even more remarkable achievements in the near future.

Acknowledgment

The authors would like to express their gratitude for the support from the National Natural Science Foundation of China (Nos. 61232001, 61420106009, and 61622213).

References

[1] R. A. Brooks, Intelligence without representation, *Artif. Intell.*, vol. 47, nos. 1–3, pp. 139–159, 1991.

[2] N. J. Nilsson, *Principles of Artificial Intelligence*. Los Altos, CA, USA: Morgan Kaufmann, 2014.

[3] P. R. Cohen and E. A. Feigenbaum, *The Handbook of Artificial Intelligence*. Los Altos, CA, USA: Butterworth-Heinemann, 2014.

[4] P. Baldi and S. Brunak, *Bioinformatics: The Machine Learning Approach*. Cambridge, MT, USA: MIT Press, 2001.

[5] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[6] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine Learning: An Artificial Intelligence Approach*. Berlin, Germany: Springer Science & Business Media, 2013.

[7] J. Liu, M. Li, W. Lan, F. X. Wu, Y. Pan, and J. X. Wang, Classification of Alzheimer's disease using whole brain hierarchical network, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, doi: 10.1109/TCBB.2016.2635144.

[8] J. Liu, M. Li, Y. Pan, F. X. Wu, X. G. Chen, and J. X. Wang, Classification of schizophrenia based on individual hierarchical brain networks constructed from structural MRI images, *IEEE Trans. NanoBiosci.*, vol. 16, no. 7, pp. 600–608, 2017.

[9] J. Liu, J. X. Wang, B. Hu, F. X. Wu, and Y. Pan, Alzheimer's disease classification based on individual hierarchical networks constructed with 3-D texture features, *IEEE Trans. NanoBiosci.*, vol. 16, no. 6, pp. 428–437, 2017.

[10] W. Lan, M. Li, K. J. Zhao, J. Liu, F. X. Wu, Y. Pan, and J. X. Wang, LDAP: A web server for LncRNA-disease association prediction, *Bioinformatics*, vol. 33, no. 3, pp. 458–460, 2017.

[11] W. Lan, J. X. Wang, M. Li, J. Liu, Y. H. Li, F. X. Wu, and Y. Pan, Predicting drug–target interaction using positive-unlabeled learning, *Neurocomputing*, vol. 206, pp. 50–57, 2016.

[12] W. Lan, J. X. Wang, M. Li, J. Liu, F. X. Wu, and Y. Pan, Predicting microRNA-disease associations based on improved microRNA and disease similarities, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, doi: 10.1109/TCBB.2016.2586190.

[13] G. R. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2008.

[14] A. Vedaldi and B. Fulkerson, Vlfeat: An open and portable library of computer vision algorithms, in *Proc.18th ACM Int. Conf. Multimedia*, Firenze, Italy, 2010, pp. 1469–1472.

[15] R. Cipolla, S. Battiato, and G. M. Farinella, *Machine Learning for Computer Vision*. Springer, 2013.

[16] B. Yegnanarayana, *Artificial Neural Networks*. New Delhi, India: PHI Learning Pvt. Ltd., 2009.

[17] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, Greedy layer-wise training of deep networks, in *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2007, p. 153.

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

[19] D. Erhan, Y. Bengio, A. Courville, P. A. Manzagol, P. Vincent, and S. Bengio, Why does unsupervised pre-training help deep learning?, *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, 2010.

[20] X. Glorot, A. Bordes, and Y. Bengio, Domain adaptation for large-scale sentiment classification: A deep learning approach, in *Proc. 28th Int. Conf. Machine Learning*, Bellevue, WA, USA, 2011, pp. 513–520.

[21] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, On optimization methods for deep learning, in *Proc. 28th Int. Conf. Machine Learning*, Bellevue, WA, USA, 2011, pp. 265–272.

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

[23] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] Z. Yu, N. Yu, Y. Pan, and T. R. Li, A novel deep learning network architecture with cross-layer neurons, in *Proc. 2016 IEEE Int. Conf. Big Data and Cloud Computing* (*BDCloud*), *Social Computing and Networking* (*SocialCom*), *Sustainable Computing and Communications* (*SustainCom*)(*BDCloud-SocialCom-SustainCom*), Atlanta, GA, USA, 2016, pp. 111–117.

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

[26] N. Yu, Z. Yu, and Y. Pan, A deep learning method for lincRNA identification using auto-encoder algorithm, in *Proc. 6th Int. Conf. Computational Advances in Bio and Medical Sciences* (*ICCABS*), Atlanta, GA, USA, 2016, 2016, p. 1.

[27] J. Dean, G. S. Corrado, R. Monga, K. Chen, M. Devin, Q. V. Le, M. Z. Mao, M. Ranzato, A. Senior, P. Tucker, et al., Large scale distributed deep networks, in *Proc. 25th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012, pp. 1223–1231.

[28] J. Liu, M. Li, J. X. Wang, F. X. Wu, T. M. Liu, and Y. Pan, A survey of MRI-based brain tumor segmentation methods, *Tsinghua Sci. Technol.*, vol. 19, no. 6, pp. 578–595, 2014.

[29] J. Liu, J. X. Wang, Z. J. Tang, B. Hu, F. X. Wu, and Y. Pan, Improving Alzheimeres disease classification by combining multiple measures, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, doi: 10.1109/TCBB.2017.2731849.

[30] J. Liu, M. Li, Y. Pan, W. Lan, R. Q. Zheng, F. X. Wu, and J. X. Wang, Complex brain network analysis and its applications to brain disorders: A survey, *Complexity*, vol. 2017, p. 8362741, 2017.

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

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[33] H. Bourlard and Y. Kamp, Auto-association by multilayer perceptrons and singular value decomposition, *Biol. Cybern.*, vol. 59, nos. 4&5, pp. 291–294, 1988.

[34] G. E. Hinton and R. S. Zemel, Autoencoders, minimum description length, and Helmholtz free energy, in *Proc. 6th Int. Conf. Neural Information Processing Systems*, Denver, CO, USA, 1994, p. 3.

[35] G. Alain and Y. Bengio, What regularized auto-encoders learn from the data-generating distribution, *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3563–3593, 2014.

[36] M. A. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, Efficient learning of sparse representations with an energy-based model, in *Proc. 20th Annual Conf. Neural Information Processing Systems*, Vancouver, Canada, 2006, pp. 1137–1144.

[37] M. A. Ranzato, Y. L. Boureau, and Y. LeCun, Sparse feature learning for deep belief networks, in *Proc. 20th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2008, pp. 1185–1192.

[38] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, Extracting and composing robust features with denoising autoencoders, in *Proc. 25th Int. Conf. Machine Learning*, Helsinki, Finland, 2008, pp. 1096–1103.

[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, pp. 3371–3408, 2010.

[40] Y. Bengio, L. Yao, G. Alain, and P. Vincent, Generalized denoising auto-encoders as generative models, in *Proc. 26th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2013, pp. 899–907.

[41] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, Contractive auto-encoders: Explicit invariance during feature extraction, in *Proc. 28th Int. Conf. Machine Learning* (*ICML-11*), Bellevue, WA, USA, 2011, pp. 833–840.

[42] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, Higher order contractive auto-encoder, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Athens, Greece, 2011, pp. 645–660.

[43] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins University Press, 2012.

[44] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in *Proc. 21th Int. Conf. Artificial Neural Networks*, Espoo, Finland, 2011, pp. 52–59.

[45] S. E. Fahlman, G. E. Hinton, and T. J. Sejnowski, Massively parallel architectures for AI: Netl, thistle, and Boltzmann machines, in *Proc. 3rd AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 1983.

[46] G. E. Hinton, T. J. Sejnowski, and D. H. Ackley, Boltzmann machines: Constraint satisfaction networks that learn, Technical Report CMU-CS-84-119, Carnegie-Mellon University, Pittsburgh, PA, USA, 1984.

[47] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, *Cogn. Sci.*, vol. 9, no. 1, pp. 147–169, 1985.

[48] G. E. Hinton and T. J. Sejnowski, Learning and relearning in Boltzmann machines, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland, eds. Cambridge, MA, USA: MIT Press, 1986, pp. 282–317.

[49] G. E. Hinton, A practical guide to training restricted Boltzmann machines, in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K. R. Müller, eds. Springer, 2012, pp. 599–619.

[50] R. D. Hjelm, V. D. Calhoun, R. Salakhutdinov, E. A. Allen, T. Adali, and S. M. Plis, Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks, *NeuroImage*, vol. 96, pp. 245–260, 2014.

[51] G. E. Hinton, S. Osindero, and Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[52] G. E. Hinton, Learning multiple layers of representation, *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, 2007.

[53] H. Lee, C. Ekanadham, and A. Y. Ng, Sparse deep belief net model for visual area V2, in *Proc. 21st Conf. Neural Information Processing Systems*, Vancouver, Canada, 2007, pp. 873–880.

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

[55] Y. LeCun and Y. Bengio, Convolutional networks for images, speech, and time-series, in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, ed. Cambridge, MA, USA: MIT Press, 1995,.

[56] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in *Proc. 26th Annual Int. Conf. Machine Learning*, Montreal, Canada, 2009, pp. 609–616.

[57] Y. LeCun, K. Kavukcuoglu, and C. Farabet, Convolutional networks and applications in vision. in *Proc. 2010 IEEE Int. Symposium on Circuits and Systems (ISCAS)*, Paris, France, 2010, pp. 253–256.

[58] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, Striving for simplicity: The all convolutional net, arXiv preprint arXiv: 1412.6806, 2014.

[59] 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, 1998.

[60] P. Y. Simard, D. Steinkraus, and J. C. Platt, Best practices for convolutional neural networks applied to visual document analysis. in *Proc. 7th Int. Conf. Document Analysis and Recognition*, Edinburgh, UK, 2003, pp. 958–962.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proc. 25th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[62] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv: 1409.1556, 2014.

[63] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, Deep residual learning for image recognition, arXiv preprint arXiv: 1512.03385, 2015.

[64] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, Identity mappings in deep residual networks, arXiv preprint arXiv: 1603.05027, 2016.

[65] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, in *Proc. 13th European Conf. Computer Vision*, Zurich, Switzerland, 2014, pp. 818–833.

[66] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 1–9.

[67] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, arXiv preprint arXiv: 1602.07261, 2016.

[68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision, in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[69] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, 2013.

[70] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F. E. de Leeuw, B. van Ginneken, E. Marchiori, et al., Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin, arXiv preprint arXiv: 1610.07442, 2016.

[71] Q. Dou, H. Chen, L. Q. Yu, L. Zhao, J. Qin, D. F. Wang, V. C. Mok, L. Shi, and P. A. Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.

[72] B. Zitová and J. Flusser, Image registration methods: A survey, *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[73] G. R. Wu, M. Kim, Q. Wang, Y. Z. Gao, S. Liao, and D. G. Shen, Unsupervised deep feature learning for deformable registration of MR brain images, in *Proc. 16th Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, Nagoya, Japan, 2013, pp. 649–656.

[74] G. R. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. G. Shen, Scalable high-performance image registration framework by unsupervised deep feature representations learning, *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, 2016.

[75] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation, in *Proc. 28th* textit Int. *Conf. Neural Information Processing Systems*, Montreal, Canada, 2015, pp. 2998–3006.

[76] W. L. Zhang, R. J. Li, H. T. Deng, L. Wang, W. L. Lin, S. W. Ji, and D. G. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, *NeuroImage*, vol. 108, pp. 214–224, 2015.

[77] 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. Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[78] 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, 2016.

[79] D. Zikic, Y. Ioannou, A. Criminisi, and M. Brown, Segmentation of brain tumor tissues with convolutional neural networks, *Proceedings MICCAI Workshop on Multimodal Brain Tumor Segmentation Challenge*, Boston, MA, USA, 2014, pp. 36–39.

[80] G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, Multi-modal brain tumor segmentation using deep convolutional neural networks, in *Proc. MICCAI BraTS* (*Brain Tumor Segmentation*) *Challenge*, 2014, pp. 31–35.

[81] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, An ensemble of 2D convolutional neural networks for tumor segmentation, in *Proc. 19th Scandinavian Conf. Image Analysis*, Copenhagen, Denmark, 2015, pp. 201–211.

[82] P. Dvorak and B. H. Menze, Structured prediction with convolutional neural networks for multimodal brain tumor segmentation, in *Proc. Multimodal Brain Tumor Image Segmentation Challenge*, Munich, Germany, 2015, pp. 13–24.

[83] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, Brain tumor segmentation using convolutional neural networks in mri images, *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[84] 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, 2017.

[85] X. T. Zhen, Z. J. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation, *Med. Image Anal.*, vol. 30, pp. 120–129, 2016.

[86] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI, *Med. Image Anal.*, vol. 30, pp. 108–119, 2016.

[87] T. A. Ngo, Z. Lu, and G. Carneiro, Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance, *Med. Image Anal.*, vol. 35, pp. 159–171, 2017.

[88] Y. Yoo, T. Brosch, A. Traboulsee, D. K. B. Li, and R. Tam, Deep learning of image features from unlabeled data for multiple sclerosis lesion segmentation, in *Proc. 5th Int. Workshop on Machine Learning in Medical Imaging*, Boston, MA, USA, 2014, pp. 117–124.

[89] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, and R. Tam, Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation, *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.

[90] Y. R. Guo, Y. Z. Gao, and D. G. Shen, Deformable mr prostate segmentation via deep feature learning and sparse patch matching, *IEEE Trans. Med. Imaging*, vol. 35, no. 4, pp. 1077–1089, 2016.

[91] [91] H. Choi and K. H. Jin, Fast and robust segmentation of the striatum using deep convolutional neural networks,*J. Neurosci. Methods*, vol. 274, pp. 146–153, 2016.

[92] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in *Proc. 16th Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, Nagoya, Japan, 2013, pp. 246–253.

[93] F. Jiang, H. T. Li, X. H. Hou, B. Sheng, R. M. Shen, X. Y. Liu, W. P. Jia, P. Li, and R. G. Fang, Abdominal adipose tissues extraction using multi-scale deep neural network, *Neurocomputing*, vol. 229, pp. 23–33, 2017.

[94] A. Mansoor, J. J. Cerrolaza, R. Idrees, E. Biggs, M. A. Alsharid, R. A. Avery, and M. G. Linguraru, Deep learning guided partitioned shape model for anterior visual pathway segmentation, *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1856–1865, 2016.

[95] H. I. Suk, S. W. Lee, and D. G. Shen, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, *NeuroImage*, vol. 101, pp. 569–582, 2014.

[96] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii, Deep learning of fMRI big data: A novel approach to subject-transfer decoding, arXiv preprint arXiv: 1502.00093, 2015.

[97] W. H. Pinaya, A. Gadelha, O. M. Doyle, C. Noto, A. Zugman, Q. Cordeiro, A. P. Jackowski, R. A. Bressan, and J. R. Sato, Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia, *Sci. Rep.*, vol. 6, p. 38897, 2016.

[98] F. Y. Liu and C. H. Shen, Learning deep convolutional features for MRI based Alzheimer's disease classification, arXiv preprint arXiv: 1404.3366, 2014.

[99] S. Q. Liu, S. D. Liu, W. D. Cai, S. Pujol, R. Kikinis, and D. G. Feng, Early diagnosis of Alzheimer's disease with deep learning, in *Proc. 2014 IEEE 11th Int. Symposium on Biomedical Imaging* (*ISBI*), Beijing, China, 2014, pp. 1015–1018.

[100] S. Q. Liu, S. D. Liu, W. D. Cai, S. Pujol, R. Kikinis, and D. D. Feng, Multi-phase feature representation learning for neurodegenerative disease diagnosis, in *Proc. 1st Australasian Conf. Artificial Life and Computational Intelligence*, Newcastle, Australia, 2015, pp. 350–359.

[101] F. Li, L. Tran, K. H. Thung, S. W. Ji, D. G. Shen, and J. Li, A robust deep model for improved classification of ad/mci patients, *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1610–1616, 2015.

[102] A. Payan and G. Montana, Predicting alzheimer's disease: A neuroimaging study with 3D convolutional neural networks, arXiv preprint arXiv: 1502.02506, 2015.

[103] V. K. Ithapu, V. Singh, O. C. Okonkwo, R. J. Chappell, N. M. Dowling, and S. C. Johnson, Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment, *Alzheimers Dement.*, vol. 11, no. 12, pp. 1489–1499, 2015.

[104] H. I. Suk, S. W. Lee, and D. G. Shen, Latent feature representation with stacked auto-encoder for AD/MCI diagnosis, *Brain Struct. Funct.*, vol. 220, no. 2, pp. 841–859, 2015.

[105] S. Sarraf, J. Anderson, and G. Tofighi, DeepAD: Alzheimer's disease classification via deep convolutional neural networks using mri and fMRI, bioRxiv, doi: 10.1101/070441.

[106] E. Hosseini-Asl, R. Keynton, and A. El-Baz, Alzheimer's disease diagnostics by adaptation of 3D convolutional network, in *Proc. 2016 IEEE Int. Conf. Image Processing* (*ICIP*), Phoenix, AZ, USA, 2016, pp. 126–130.

[107] R. Tandon, H. A. Nasrallah, and M. S. Keshavan, Schizophrenia, "just the facts" 4. clinical features and conceptualization, *Schizophr. Res.*, vol. 110, nos. 1–3, pp. 1–23, 2009.

[108] M. S. Keshavan, H. A. Nasrallah, and R. Tandon, Schizophrenia, "just the facts" 6. moving ahead with the schizophrenia concept: from the elephant to the mouse, *Schizophr. Res.*, vol. 127, nos. 1–3, pp. 3–13, 2011.

[109] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, and V. D. Calhoun, Deep learning for neuroimaging: A validation study, arXiv preprint arXiv: 1312.5847, 2013.

[110] J. Kim, V. D. Calhoun, E. Shim, and J. H. Lee, Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia, *NeuroImage*, vol. 124, pp. 127–146, 2016.

[111] J. Qi and J. Tejedor, Deep multi-view representation learning for multi-modal features of the schizophrenia and schizo-affective disorder, in *Proc. 2016 IEEE Int. Conf. Acoustics*, *Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 952–956.

[112] K. Shibata, T. Watanabe, Y. Sasaki, and M. Kawato, Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation, *Science*, vol. 334, no. 6061, pp. 1413–1415, 2011.

[113] T. Horikawa, M. Tamaki, Y. Miyawaki, and Y. Kamitani, Neural decoding of visual imagery during sleep, *Science*, vol. 340, no. 6132, pp. 639–642, 2013.

[114] Y. Hatakeyama, S. Yoshida, H. Kataoka, and Y. Okuhara, Multi-voxel pattern analysis of fMRI based on deep learning methods, in *Soft Computing in Big Data Processing*, K. Lee, S. J. Park, and J. H. Lee, eds. Springer, 2014, pp. 29–38.

[115] O. Firat, L. Oztekin, and F. T. Y. Vural, Deep learning for brain decoding, in *Proc. 2014 IEEE Int. Conf. Image Processing* (*ICIP*), Paris, France, 2014, pp. 2784–2788.

[116] H. Jang, S. M. Plis, V. D. Calhoun, and J. H. Lee, Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks, *NeuroImage*, vol. 145, pp. 314–328, 2017.

[117] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv: 1207.0580, 2012.

[118] S. Koyamada, M. Koyama, K. Nakae, and S. Ishii, Principal sensitivity analysis, in *Proc. 19th Pacific-Asia Conf. Knowledge Discovery and Data Mining*, Ho Chi Minh City, Vietnam, 2015, pp. 621–632.

[119] D. P. Kuang and L. H. He, Classification on ADHD with deep learning, in *Proc. 2014 Int. Conf. Cloud Computing and Big Data* (*CCBD*), Wuhan, China, 2014, pp. 27–32.

[120] X. B. Han, Y. F. Zhong, L. F. He, P. S. Yu, and L. P. Zhang, The unsupervised hierarchical convolutional sparse auto-encoder for neuroimaging data classification, in *Proc. 8th Int. Conf. Brain Informatics and Health*, London, UK: 2015, pp. 156–166.

[121] H. Jang and J. H. Lee, Deep neural network for age prediction using resting-state fMRI data, in *The Organization for Human Brain Mapping (OHBM) 2016 Annual Meeting*, 2016.

[122] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, BrainNetCNN: Convolutional neural networks for brain networks towards predicting neurodevelopment, *NeuroImage*, vol. 146, pp. 1038–1049, 2016.

[123] J. H. Cole, R. P. K. Poudel, D. Tsagkrasoulis, M. W. A. Caan, C. Steves, T. D. Spector, and G. Montana, Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker, arXiv preprint arXiv: 1612.02572, 2016.

[124] T. Schmah, G. E. Hinton, R. S. Zemel, S. L. Small, and S. Strother, Generative versus discriminative training of rbms for classification of fMRI images, in *Proc. 21st Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2008, pp. 1409–1416.

[125] H. C. Kim and J. H. Lee, Deep learning to predict the emotional response using functional MRI data, in *The Organization for Human Brain Mapping (OHBM) 2016 Annual Meeting*, 2016.

[126] Z. Yang, S. H. Zhong, A. Carass, S. H. Ying, and J. L. Prince, Deep learning for cerebellar ataxia classification and functional score regression, in *Proc*. *5th Int*. *Workshop on Machine Learning in Medical Imaging*, Boston, MA, USA, 2014, pp. 68–76.

[127] 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*, Orlando, FL, USA, 2014, pp. 675–678.

[128] R. B. Palm, Prediction as a candidate for learning deep hierarchical models of data, Master's thesis, Technical University of Denmark, Denmark, 2012.

[129] R. Collobert, K. Kavukcuoglu, and C. Farabet, Torch7: A matlab-like environment for machine learning, in *BigLearn*, *NIPS Workshop*, *No. EPFL-CONF-192376*, 2011.

[130] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, et al., Theano: A python framework for fast computation of mathematical expressions, arXiv preprint arXiv: 1605.02688, 2016.

[131] I. J. Goodfellow, D. Warde-Farley, P. Lamblin, V. Dumoulin, M. Mirza, R. Pascanu, J. Bergstra, F. Bastien, and Y. Bengio, Pylearn2: A machine learning research library, arXiv preprint arXiv: 1308.4214, 2013.

[132] F. Chollet, Keras: Theano-based deep learning library, Code, https://github.com/fchollet.Documentation:http://keras.io, 2015.

[133] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. F. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., TensorFlow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv: 1603.04467, 2016.

[134] D. Yu, A. Eversole, M. Seltzer, K. S. Yao, O. Kuchaiev, Y. Zhang, F. Seide, Z. H. Huang, B. Guenter, H. M. Wang, et al., An introduction to computational networks and the computational network toolkit, Technical report, Tech. Rep. MSR-TR-2014-112, 2014.

[135] T. Q. Chen, M. Li, Y. T. Li, M. Lin, N. Y. Wang, M. J. Wang, T. J. Xiao, B. Xu, C. Y. Zhang, and Z. Zhang, MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems, arXiv preprint arXiv: 1512.01274, 2015.

[136] S. Tokui, K. Oono, S. Hido, and J. Clayton, Chainer: A next-generation open source framework for deep learning, in *Proc*. *Workshop on Machine Learning Systems* (*LearningSys*) *in the Twenty-Ninth Annual Conf*. *Neural Information Processing Systems* (*NIPS*), 2015.

[137] Deeplearning4j Development Team, Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0, http://deeplearning4j.org, 2017.

[138] B. C. Ooi, K. L. Tan, S. Wang, W. Wang, Q. C. Cai, G. Chen, J. Y. Gao, Z. J. Luo, A. K. H. Tung, Y. Wang, et al., SINGA: A distributed deep learning platform, in *Proc. 23rd ACM Int*. *Conf*. *Multimedia*, Brisbane, Australia, 2015, pp. 685–688.

[139] A. Vedaldi and K. Lenc, MatConvNet: Convolutional neural networks for MATLAB, in *Proc*. *23rd ACM Int*. *Conf*. *Multimedia*, Brisbane, Australia, 2015, pp. 689–692.

[140] A. Lavin, MaxDNN: An efficient convolution kernel for deep learning with maxwell GPUs, arXiv preprint arXiv: 1501.06633, 2015.

[141] J. Dolz, C. Desrosiers, and I. B. Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study, arXiv preprint arXiv: 1612.03925, 2016.

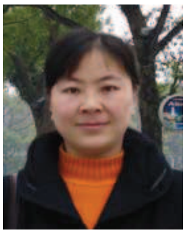
[142] S. Sarraf and G. Tofighi, Deep learning-based pipeline to recognize alzheimer's disease using fMRI data, in *Future Technologies Conf*. (*FTC*), San Francisco, CA, USA, 2016, pp. 816–820.

[143] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, arXiv preprint arXiv: 1701.06643, 2017.

[144] T. W. Huang, H. T. Chen, R. Fujimoto, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, Age estimation from brain MRI images using deep learning, in *Proc*. *14th Int*. *Symposium on Biomedical Imaging* (*ISBI 2017*), Melbourne, Australia, 2017, pp. 849–852.

[145] 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, 2017.

Jin Liu is a PhD candidate in School of Information Science and Engineering, Central South University, Changsha, China. He received the MS degree from University of Chinese Academy of Sciences in 2013. His current research interests include medical image analysis, machine learning, and pattern recognition.

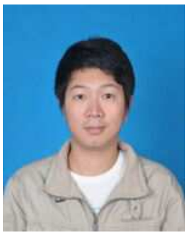


Min Li received the PhD degree in computer science from Central South University, China, in 2008. She is currently a Professor at the School of Information Science and Engineering, Central South University, China. Her main research interests include bioinformatics and systems biology.

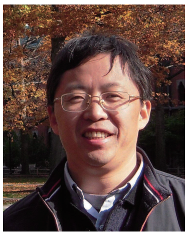
Yi Pan is a Regents' Professor of Computer Science and an Interim Associate Dean and Chair of Biology at Georgia State University, USA. Dr. Pan joined Georgia State University in 2000 and was promoted to full professor in 2004, named a Distinguished University Professor in 2013 and designated a Regents' Professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. He served as the Chair of Computer Science Department from 2005—2013. He is also a visiting Changjiang Chair Professor at Central South University, China. Dr. Pan received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, USA, in 1991. His profile has been featured as a distinguished alumnus in both Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. Dr. Pan's research interests include parallel and cloud computing, wireless networks, and bioinformatics. Dr. Pan has published more than 330 papers including over 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 8300 times. Dr. Pan has served as an editor-in-chief or editorial board member for 15 journals including 7 IEEE Transactions. He is the recipient of many awards including IEEE Transactions Best Paper Award, 4 other international conference or journal Best Paper Awards, 4 IBM Faculty Awards, 2 JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.

Ziyue Chen received the MS degree in both computer science and chemistry from Georgia State University, USA, in 2017. She is currently an engineer in McKinsey & Company. Her research interests include bioinformatics and machine learning.

Lu Tang is an undergraduate in School of Information Science and Engineering, Central South University, China. Her current research interests include medical image analysis and machine learning.



Chengqian Lu is a PhD candidate in School of Information Science and Engineering, Central South University, Changsha, China. He received the MS degree from Yunan University, China, in 2011. His current research interests include non-coding RNA, matrix completion, and machine learning.



Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the chair of and a professor in Department of Computer Science, Central South University, China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics and computer network. He is a senior member of the IEEE.