

A Survey on Domain Knowledge Powered Deep Learning for Medical Image Analysis

Xiaozheng Xie, Jianwei Niu, *Senior Member, IEEE*, Xuefeng Liu, Zhengsu Chen, and Shaojie Tang, *Member, IEEE*

Abstract—Although deep learning models like CNNs have achieved a great success in medical image analysis, small-sized medical datasets remain to be the major bottleneck in this area. To address this problem, researchers start looking for external information beyond the current available medical datasets. Traditional approaches generally leverage the information from natural images. More recent works utilize the domain knowledge from medical doctors, by letting networks either resemble how they are trained, mimic their diagnostic patterns, or focus on the features or areas they particular pay attention to. In this survey, we summarize the current progress on introducing medical domain knowledge in deep learning models for various tasks like disease diagnosis, lesion, organ and abnormality detection, lesion and organ segmentation. For each type of task, we systematically categorize different kinds of medical domain knowledge that have been utilized and the corresponding integrating methods. We end with a summary of challenges, open problems, and directions for future research.

Index Terms—medical image analysis, medical domain knowledge, deep neural networks.

1 INTRODUCTION

RECENT years have witnessed a tremendous progress in computer-aided detection/diagnosis (CAD) in medical imaging and diagnostic radiology, primarily thanks to the advancement of deep learning techniques. Having achieved great success in computer vision tasks, various deep learning models, mainly convolutional neural networks (CNNs), soon be applied to CAD. Among the applications are the early detection and diagnosis of breast cancer, lung cancer, glaucoma and skin cancer [1], [2], [3], [4], [5].

However, the size of medical datasets remains to be the major bottleneck to obtain a satisfactory deep-learning model for CAD. It is a generally accepted notion that bigger datasets result in better deep learning models [6]. In traditional computer vision tasks, there are many large-scale and well-annotated datasets like ImageNet [7] (with more than 14M labeled images from 20k categories) and COCO [8] (with more than 200k annotated images from 80 categories). In contrast, Table 1 lists some popular publicly available medical datasets and we can see that most of them are much smaller. For example, among the datasets for classification task, only two, namely the NIH and DeepLesion, contain more than 10K labeled medical images. Some datasets only have a few thousands or even hundreds of medical images.

The lack of medical datasets is represented in three aspects. Firstly, one common situation is the amount of medical images in the datasets is small. This problem is mainly due to the high cost associated with the data

collection. Medical images are collected from computerized tomography (CT), magnetic resonance imaging (MRI) scans, positron emission tomography (PET), all of which are expensive and labor-intensive to collect. Secondly, it is also quite common that although there are a large amount of medical images, only a small portion is annotated. The annotations of the collected images, including the classification labels (e.g. benign or malignant), the segmentation annotations of lesion areas, etc., require efforts from experienced radiologists. Thirdly, some diseases are rare in nature, and it is quite difficult to collect enough positive cases, caused unbalanced datasets.

The direct consequence of the lack of well annotated medical data is that the trained deep learning models can easily suffer from the overfitting problem [9]: they perform very well on training datasets, but fail when dealing with new data from the problem domain. Correspondingly, many works of medical image analysis adopt techniques designed for addressing overfitting in traditional computer vision tasks, like reducing the complexity of the network [10], [11], [12], [13], [14], [15], adopting some regularization techniques [16], [17], or using data augmentation strategies [18], [19] [20], [21] [22].

However, in essence, neither decreasing the model complexity nor leveraging data augmentation techniques introduces any new information into deep learning models. We argue that introducing more information beyond the given medical datasets should be a more promising approach to address the problem of small-sized medical datasets.

The idea of introducing external information to improve the performance of deep learning models for CAD is not new. For example, it is a routine that a deep learning model is firstly trained on some natural image datasets like ImageNet, and then be fine-tuned on target medical datasets [23], [24]. This process, called as transfer learning [25], implicitly introduces information from natural images.

- X. Xie, J. Niu, X. Liu and Z. Chen are with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China. E-mails: {xiexzheng,niujianwei,liu_xuefeng,danczs}@buaa.edu.cn
- J. Niu is also with the Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC) and Hangzhou Innovation Institute of Beihang University.
- S. Tang is in Jindal School of Management, The University of Texas at Dallas. E-mails: tangshaojie@gmail.com

TABLE 1
Examples of some popular databases in Medical Domain.

Name	Purpose	Type	Imaging	Images
ANDI	Classification	Brain	Multiple	800 patients
OpenNeuro	Classification	Brain	Multiple	4718 patients
ABIDE	Classification	Brain	MRI	539 patients and 573 typical controls
DDSM	Classification	Breast	Mammography	2500 patients
MIAS	Classification	Breast	Mammography	322 digitised films
ACDC	Classification	Cardiac	MRI	150 patients
ChestX-ray14	Detection	Chest	X-ray	112120 images from 20805 patients
LIDC-IDRI	Detection	Lung	CT, X-ray	1018 patients
LUNA16	Detection	Lung	CT	888 images with 1084 rumors
MURA	Detection	Musculoskeletal	X-ray	40895 images from 14982 patients
OASIS	Segmentation	Brain	MRI, PET	1098 patients
BRATS2018	Segmentation	Brain	MRI	542 images
DRIVE	Segmentation	Eye	SLO	40 images
STARE	Segmentation	Eye	SLO	400 images
MICCAI Gland Challenge	Segmentation	Glands	Histopathology	165 images
PRETERM	Segmentation	Left ventricular	MRI	234 patients
INBreast	Classification/Detection	Breast	Mammography	410 images
NIH	Classification/Detection	Chest	X-ray	112820 images from 30805 patients
DeepLesion	Classification/Detection	Multiple	CT	32735 images from 4427 patients
Prostate	Classification/Detection	Prostate	MRI	60 patients
Cardiac MRI	Classification/Segmentation	Cardiac	MRI	7980 images from 33 cases
JSRT	Detection/Segmentation	Lung	X-Ray	154 and 93 images with and without a nodule
ISIC2018	Classification/Detection/Segmentation	Skin	Dermoscopic	13000 images

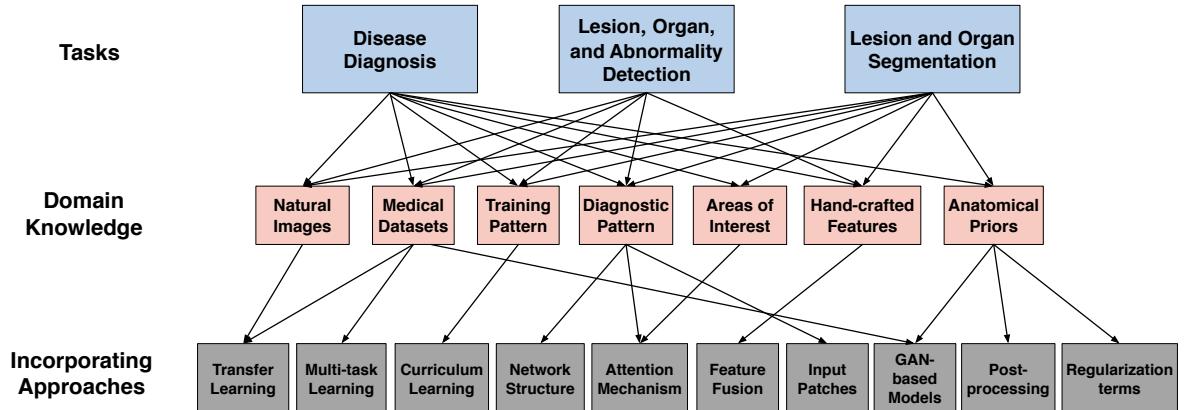


Fig. 1. Researches with different information categories and integrating methods in disease diagnosis, detection and lesion segmentation.

Besides natural images, multi-modal medical datasets or medical images from different but related diseases can also be utilized to improve the performance of deep learning models [26], [27].

Moreover, as experienced medical doctors (e.g. radiologists, ophthalmologists and dermatologists) can generally give fairly accurate results, it is not surprising that ‘the domain knowledge of medical doctors’, like the way they browse images, the particular areas they usually focus on, and the features they give special attentions to, etc., can be potentially very informative and help the deep learning models to better accomplish the designated tasks.

These types of knowledge are accumulated, summarized, and validated by a large number of practitioners over many years based on a huge amount of cases and hence are potentially far more informative than the given medical datasets.

For a given task for a certain disease, if we have identified a type of domain knowledge to be introduced, then how to incorporate it into the deep learning model is another issue requiring a careful design. In the last few years, various approaches have been designed to incorporate different types of domain knowledge into networks. For example, to incorporate the important

features identified by radiologists for lesion classification (benign or malignant), a simple approach is to combine the hand-crafted features with the ones extracted from deep learning models and then feed them into a classifier [28]. In some works, network architectures are revised to simulate the pattern of radiologists when they read images [29]. Attention mechanism, which allows a network pay more attention to a certain region of an image, is a powerful technique to incorporate radiologists' knowledge about the areas they usually focus on a medical image [30]. In addition, multi-task learning and meta learning are also widely utilized to introduce medical domain knowledge into deep learning models [31] [32].

This survey serves as a demonstration that, for almost all tasks of medical image analysis including disease diagnosis, detection, lesion segmentation etc., *identifying one or more types of appropriate domain knowledge related to the designated task with a carefully designed integrating approach will generally improve the performance of deep learning models.* Correspondingly, we organize the existing works in the following three aspects: *the types of given tasks, the types of domain knowledge that are introduced, and the ways of introducing the domain knowledge.*

Although there are a number of reviews on deep learning for medical image analysis, including [33], [34], [35], [36], [37], [38], [39], [40], [41] and [42], they all describe the existing works from the application point view, i.e. how deep learning techniques are applied to various medical applications. To the best of our knowledge, there is no review that gives systematic introduction on *how medical domain knowledge can help deep learning models.* This aspect, we believe, is the unique feature that distinguishes deep learning models for CAD from those for general computer vision tasks.

Fig. 1 gives the overview on how we organize the related works. At the top level, they are classified according to the three tasks of medical imaging analysis: (1) disease diagnosis, (2) lesion, organ and abnormality detection, and (3) lesion and organ segmentation. These three tasks cover most of the medical image analysis. Then the research works of each task are further classified according to the types of extra knowledge that have been incorporated. At the bottom level, they are further categorized according to different approaches to integrate the domain knowledge.

This survey contains over 270 papers, most of them recent, on a wide variety of applications of deep learning techniques for medical image analysis. Additionally, most of the corresponding works are from the conference proceedings for MICCAI, EMBC, ISBI and some journals such as TMI, Medical Imaging, JBHI and so on. The key words we use includes 'domain knowledge', 'expert knowledge', 'priors', 'deep learning', 'CNN', 'medical image processing' and so on. The papers without using deep learning, and those using deep learning but without incorporating any domain knowledge are not incorporated.

Summarizing, with this survey we aim to:

- systematically summarize and classify different types of domain knowledge in medical areas that are utilized to improve the performance of deep learning models in various applications.

- systematically summarize and classify different ways of introducing medical domain knowledge into deep learning models.
- give the outlook of challenges, open problems and future directions in integrating medical domain knowledge into deep learning models.

The remaining of the survey is organized as follows. Section 2, 3 and 4 introduces the related works for the major three tasks for medical image analysis. Besides these three major tasks, other tasks in medical image analysis are described in Section 5. In each section, we first introduce general architectures of deep learning models for the task, and then categorize the related works according to the types of the domain knowledge that been integrated. Various incorporating methods for each type of domain knowledge are also described. At last, Section 6 discusses the research challenges, open problems, and gives the outlook of future directions.

2 DISEASE DIAGNOSIS

2.1 Disease Diagnosis and the Corresponding General Structures of Deep Learning Models

Disease diagnosis refers to the task of determining the type and condition of possible diseases based on the images. Traditional disease diagnosis is made by radiologists based on their experience. To reduce the operator dependency and improve the diagnostic accuracy, CAD systems based on machine learning techniques have been developed in the last few decades [43], [44], [45], [46]. In essence, disease diagnosis is a classification task, and an image can be classified as normal or diseased, benign or malignant, or different levels of severity. These CAD systems generally first extract some features from images and then feed them into a classifier to give final conclusion [44], [45]. However, which features to be selected are generally determined by radiologists.

More recently, deep learning techniques, especially CNNs, have gained a great success in various computer vision tasks, mainly thanks to their capability to automatically extract discriminative features. Different types of CNNs can be directly applied to disease diagnosis.

Fig. 2 shows the structure of a typical CNN that used for disease diagnosis in chest X-ray image. The CNN employs alternating convolutional and pooling layers, and contains trainable filter banks per layer. Each individual filter in a filter bank is able to generate a feature map. This process is alternated and the CNN can learn increasingly more and more abstract features that will later be used by the fully connected layers to accomplish the classification task.

Different types of CNN architectures, from AlexNet [47], GoogLeNet [48], VGGNet [49], ResNet [50] to DenseNet [51], have achieved a great success in the diagnosis of various diseases. For example, the AlexNet is utilized in [52] for the diagnosis of diabetic retinopathy (DR) and achieves 97.93% classification accuracy on the standard KAGGLE fundus datasets. GoogLeNet, ResNet, and VGGNet models are used in the diagnosis of canine ulcerative keratitis [53], and most of them achieve accuracies of over 90% when classifying superficial and deep corneal ulcers. DenseNet

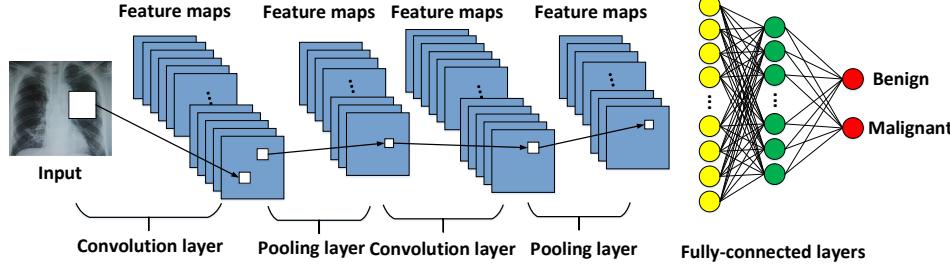


Fig. 2. A typical architecture of CNN for medical disease diagnosis.

is adopted to diagnose lung nodules on chest X-ray radiograph [54]. The experimental results show that more than 99% of lung nodules can be detected. Among various CNN architectures, it is found that VGGNet and ResNet are more effective for many medical diagnostic tasks [29], [55], [56], [57].

However, the above works generally directly apply general CNNs to medical image analysis or with only slight modifications (e.g. by modifying the number of kernels, channels or the sizes of filters), and no medical knowledge is incorporated.

In the following sections, we give a systematic review on the researches that utilize medical domain knowledge for the disease diagnosis. The knowledge sources, the types of knowledge and the incorporating methods of these research works are summarized in Table 2.

2.2 Incorporating Knowledge from Natural Images or Other Medical Datasets

For many computer vision tasks, it is a de-facto standard practice to pre-train the model on a very large dataset (e.g. ImageNet), and then use the pre-trained network either as an initialization or a fixed feature extractor for the given task. This also applies to medical image diagnosis. Despite the disparity between natural images and medical ones, it has been demonstrated that CNNs comprehensively trained on the large scale well-annotated natural image datasets can still be helpful for disease diagnosis tasks [59], [60]. Intrinsically speaking, this transfer learning process introduces knowledge from natural images into the network for medical image diagnosis.

The networks pre-trained on natural images can be leveraged via two different ways [33]. The first is to utilize them as fixed feature extractors, and the second is to utilize them as an initialization which will then be fine-tuned on target medical datasets. These two approaches are respectively illustrated in Fig. 3(a) and Fig. 3(b).

In particular, the first strategy takes a pre-trained network, removes its last fully-connected layer, and then treats the rest of the network as a fixed feature extractor. The extracted features are then fed into a linear classifier (e.g. support vector machine (SVM)) which is trained on the target medical datasets. Applications in this category include mammography mass lesion classification [61], [64], chest pathology identification [62], glaucoma identification [63], skin cancer classification [58] and ECG arrhythmia classification [65].

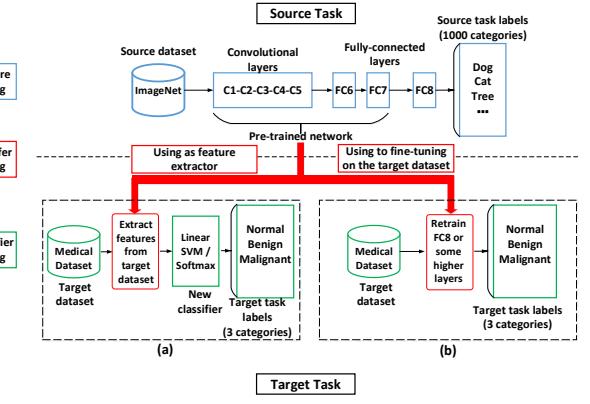


Fig. 3. Two strategies to utilize the pre-trained network on natural images. (a) as a feature extractor and (b) as an initialization which will be fine-tuned on the target dataset.

Although there is a clear gap between the natural images (e.g. ImageNet) and medical datasets, the above approaches also show success on diseases diagnosis. This phenomenon can be attributed to the fact that a network pre-trained on natural images, especially in the earlier layers, contain more generic features (e.g. edge detectors or color blob detectors) that should be useful to many tasks, including disease diagnosis [107].

The second strategy is to fine-tune the weights of the pre-trained network on the medical datasets. It is possible to fine-tune the weights of all layers of the network, or to keep some of the earlier layers fixed and only fine-tune some higher-level portion of the network. Applications in this category include the classification of skin cancer [4], interstitial lung diseases [66], [71], breast cancer [68] and thorax diseases [69].

Besides the information from natural images, using images from other medical datasets is also quite popular.

Medical datasets containing images of the same or similar modality with the target ones have similar distribution and therefore can be helpful. In [72], to classify malignant and benign breast masses in digitized screen-film mammograms (SFMs), a multi-task transfer learning DCNN is proposed to incorporate the information from digital mammograms (DMs). The multi-task transfer learning DCNN is found to have significantly higher performance compared to the single-task transfer learning DCNN which only utilize SFMs.

In addition, even medical images with different modalities can provide complementary information. For

TABLE 2
The list of researches in disease diagnosis and the knowledge they incorporated.

Knowledge Source	Knowledge Type	Incorporating Method	References
natural domain	natural images	transfer learning	[4], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71]
medical dataset	multi-modal images	transfer learning	[26], [72], [73], [74], [75], [76]
	datasets from other diseases	multi-task learning	[77], [78]
	training pattern	curriculum learning	[38], [79], [80], [81], [82]
	diagnostic pattern	network structure design	[29], [83]
	areas doctors focus on	attention mechanism (as labels, part of the feature maps)	[30], [57], [84]
		decision-level feature fusion	[64], [85], [86], [87], [88], [89]
medical doctors	features doctors give specifical attention	feature-level feature fusion input-level fusion as labels of CNNs	[28], [68], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99]
	other related information (extra category labels, extra radiological reports)	multi-task learning network structure design	[56], [100], [101], [102], [103], [31], [71], [104]
			[55], [105], [106]

example, [26] shows that using a model firstly pre-trained on a mammography dataset can obtain better results than directly trained on the target dataset with digital breast tomosynthesis (DBT) images. Another example is in prostate cancer classification, where the radiofrequency ultrasound images are first used to pretrain the DCNN, then the model is fine-tuned on B-mode ultrasound images [73]. Other examples include [74], [75], [76].

Images in medical datasets with different diseases can also have a similar morphological structure or distribution and therefore can be helpful to each other. For example in [77], a multi-task deep learning (MTDL) method is proposed. MTDL can simultaneously utilize multiple cancer datasets so that hidden representations can provide more information to small-scale cancer datasets and enhance the classification performance. Another example is a cross-disease attention network (CANet) proposed in [78]. CANet characterizes and leverages the relationship between diabetic retinopathy (DR) and diabetic macular edema (DME) in fundus images using a special designed disease-dependent attention module. Experimental results on two public datasets show that CANet outperforms other methods on diagnosing both of the two diseases.

Summarizing, when incorporating knowledge from natural images or extra medical datasets, two kinds of approaches are widely adopted: the transfer learning and the multi-task learning. For the former, a deep learning model is firstly trained on the selected external dataset, and then be fine-tuned on the target one. For the latter, the external dataset and the target one are utilized respectively as the auxiliary and the main tasks of a multi-task network. The intrinsic network sharing in the multi-task network enables the information sharing of information from the two datasets. We can also find from related works that an external dataset is particularly useful when (1) it has a large size, and (2) the distribution of the contained images is close to that of the target one.

2.3 Incorporating the Knowledge from Medical Doctors

Some experienced medical doctors can give fairly accurate conclusion on the given medical images, mainly thanks to the training they have received and the expertise

they have accumulated over many years on many cases. They generally follow some certain patterns or take some procedures when reading medical images. Incorporating these knowledge from medical doctors can improve the diagnostic performance of deep learning models.

The types of domain knowledge of medical doctors that have been utilized in deep learning models can be summarized into five categories:

- 1) the training pattern,
- 2) the general diagnostic patterns they view images,
- 3) the areas on which they usually focus,
- 4) the features (e.g. characteristics, structures, shapes) they give special attentions to, and
- 5) other related information for diagnosis.

The related works for each category will be described in the following subsections.

2.3.1 Training Pattern of Medical Doctors

The training process of medical students has a character: they are trained by tasks with increasing difficulty. For example, a student is first given some easier tasks, like deciding whether an image contains lesions, and then is required to accomplish more challenging tasks like determining whether the lesions are benign or malignant. Finally, more challenging tasks will be given like determining the subtypes of the lesions.

The above pattern can be introduced in the training process of deep neural networks via a technique called as curriculum learning [108]. Specifically, a curriculum determines a sequence of training samples ranked in ascending order of learning difficulty. Curriculum learning has been an active research topic in computer vision and has been recently utilized for medical image diagnosis.

For example, a teacher-student curriculum learning strategy is proposed for breast screening classification from DCE-MRI [38]. A deep learning model is trained on simpler tasks before introducing the hard problem of malignancy detection. This strategy shows the better performance when compared with state-of-the-art methods.

Similarly, a CNN based attention-guided curriculum learning framework is presented in [79], which leverages the severity-level attributes mined from radiology reports.

Images in order of difficulty (grouped by different severity-levels) are fed to CNN to boost the learning gradually.

In [80], the curriculum learning is adopted to support the classification of proximal femur fracture from X-ray images. The approach assigns a degree of difficulty to each training sample. By first learning ‘easy’ examples and then ‘hard’ ones, the model can reach a better performance even with fewer data. Other examples of using curriculum learning for disease diagnosis can be found in [81] and [82].

2.3.2 General Diagnostic Pattern of Medical Doctors

Experienced medical doctors generally follow some patterns when they read medical images. These patterns can be incorporated into deep learning models, generally in their architecture design.

For example, radiologists generally take the following three-staged approach when they read chest X-ray images: first browsing the whole image, then concentrating on the local lesion areas, and finally combining the global and local information to make decisions. This pattern is incorporated in the architecture design of the network for thorax disease classification [29] (see Fig. 4). The proposed network has three branches, one is used to view the whole image, the second for viewing the local areas, and the third one for combining the global and local information together. The network yields state-of-the-art accuracy on the ChestX-ray14 dataset.

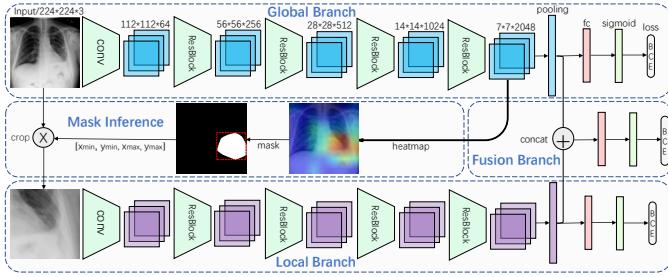


Fig. 4. An example of leveraging the diagnostic pattern of radiologists for thorax disease diagnosis [29].

For the diagnosis of skin lesions, experienced dermatologists generally first locate lesions, then identify dermoscopic features from the lesion areas, and finally make diagnosis based on the features. This pattern is mimicked in the design of the network for the diagnosis of skin lesions [83]. The proposed network, called as DermaKNet, is composed of several subnetworks, each one devoted to a specific task: lesion-skin segmentation, detection of dermoscopic features, and global lesion diagnosis. The DermaKNet achieves higher performance compared to the traditional CNN models.

2.3.3 The Areas Medical Doctors Usually Focus on

When experienced medical doctors read an image, they generally focus on a few specific areas, as these areas are more informative than other places for the purpose of disease diagnosis. Therefore, the information about where medical doctors focus may be useful for deep learning models to give better results.

The knowledge above is generally represented as ‘attention maps’, which are annotations given by medical doctors indicating the areas they focus on when reading images. For example, for glaucoma diagnosis, a CNN named AG-CNN explicitly incorporates the ‘attention maps’ [30]. The attention maps are labeled by ophthalmologists for images in the dataset indicating where they focus when reading images (shown in Fig. 5). To incorporate the attention maps, an attention prediction subnet in AG-CNN is designed, and the attention prediction loss measuring the difference between the generated and ground truth attention maps (provided by ophthalmologists) is utilized to supervise the training process. Experimental results show that AG-CNN significantly advances state-of-the-art glaucoma detection methods.

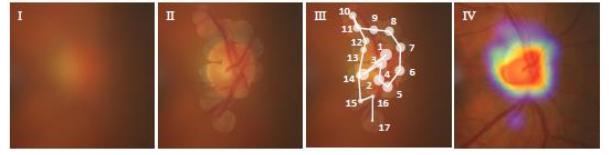


Fig. 5. Example of the generation process of attention maps of ophthalmologists when they read images in glaucoma detection [30]. I: The original blurred fundus image. II: Fixations of ophthalmologists with cleared regions. III: The order of clearing the blurred regions. IV: The generated attention map based on the captured fixations.

Another example in this category is the lesion-aware CNN (LACNN) for the classification of retinal optical coherence tomography (OCT) images [84]. The LACNN simulates the pattern of ophthalmologists’ diagnosis by focusing on local lesion-related regions. The ‘attention maps’ are represented as the annotated OCT images delineating the lesion regions using bounding polygons. To incorporate the information, the LACNN proposes a lesion-attention module to enhance the features from local lesion-related regions while still preserving the meaningful structures in global OCT images. The experimental results on two clinically acquired OCT datasets demonstrate the effectiveness of introducing the attention maps for retinal OCT image classification, with 8.3% performance gain when compared with the baseline method.

Furthermore, an Attention Branch Network (ABN) is proposed to incorporate the knowledge given by the radiologists in diabetic retinopathy [57]. ABN introduces a branch structure which generates attention maps that highlight the attention regions of the network. It allows the attention maps to be modified with semantic segmentation labels of disease regions. Experimental results on the disease grade recognition of retina images show that ABN achieves 93.73% classification accuracy and its interpretability is clearer than conventional approaches.

2.3.4 Features on Which Medical Doctors Give Special Attentions to

In the last decades, many guidelines and rules have gradually developed in various medical fields. They generally pointed out some important features for diagnosis. As these features are designated by medical doctors, they are called as ‘*hand-crafted features*’. For

example, the popular ABCD rule [109] is widely adopted by dermatologists to classify melanocytic tumors. The ABCD rule points out four distinguishing features, namely asymmetry, border, color and differential structures, to determine whether a melanocytic skin lesion under the investigation is benign or malignant.

Another example is in the field of breast cancer diagnosis. Radiologists use the BI-RADS (Breast Imaging Reporting and Data System) score [110] to place abnormal findings into different categories, with score 1 indicating healthy and score 6 the breast cancer. More importantly, BI-RADS indicates some features, including margin, shape, micro-calcification, and echo pattern, that closely related to BI-RADS scores. For example, lesions with smooth, thin and regular margins are more likely to be benign ones, while lesions with irregular and thick margins are highly suspected to be malignant. Other features that can help to classify benign and malignant breast tumors are shown in Table 3.

TABLE 3
Features in the BI-RADS guideline to classify benign and malignant breast tumors in ultrasound images [111].

	Benign	Malignant
Margin	smooth, thin, regular	irregular, thick
Shape	round or oval	irregular
Microcalcification	no	yes
Echo pattern	clear	unclear
Acoustic attenuation	not obvious	obvious
Side acoustic shadow	obvious	not obvious

Similarly, as pointed out in [71], for the benign-malignant risk assessment of lung nodules, six high-level nodule features, including calcification, sphericity, margin, lobulation, spiculation and texture, have shown a tightly connection with malignancy scores (see Fig. 6).

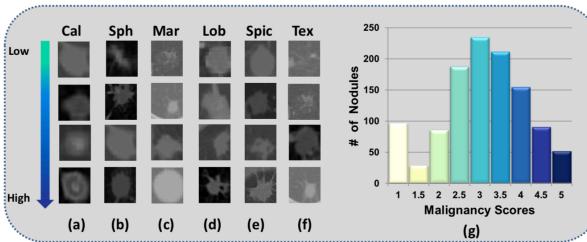


Fig. 6. Lung nodule attributes with different malignancy scores [71].

These hand-crafted features have been widely used in many traditional CAD systems. These systems generally first extract these features from medical images, and then feed them into some classifiers like SVM or Random Forest [112], [113]. For example, for the lung nodule classification on CT images, many CAD systems utilize features including the size, shape, morphology, and texture from the suspected lesion areas [56], [114], [115], [116], [117]. Similarly, in the CAD systems for the diagnosis of breast ultrasound images, features such as intensity, texture and shape are selected [118], [119], [120], [121].

When using deep learning models like CNNs which have the ability to automatically extract representative features, there are four approaches to combine ‘hand-crafted features’ with features extracted from CNNs.

- the decision-level fusion: The two types of features are fed into two classifiers respectively, and the decisions from two classifiers are combined.
- the feature-level fusion: the two types of features are directly combined via techniques like concatenation.
- the input-level fusion: the hand-crafted features are represented as image patches which are taken as inputs to the CNNs.
- as labels of CNNs: the hand-crafted features are annotated firstly and then utilized as labels for CNNs during training process.

Decision-level fusion: The structure of this approach is illustrated in Fig. 7. In this approach, the hand-crafted features and the features extracted from CNNs are fed into two classifiers respectively. Then the classification results from both classifiers are combined using some decision fusion techniques to give final classification results.

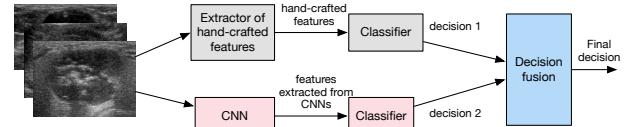


Fig. 7. Decision-level fusion: the decisions from two classifiers, one based on hand-crafted features, and the other on the CNNs, are combined.

For example, a skin lesion classification model proposed in [85] combines the results from two SVM classifiers. The first one uses hand-crafted features (i.e. RSurf features and Local Binary Patterns) as input and the second one employs features derived from a CNN. Both of the classifiers predict the category for each tested image with a classification score. These scores are subsequently used to determine the final classification result.

Similarly, a mammographic tumor classification method also combines features in decision-level fusion [64]. After individually performing classification with CNN features and analytically extracted features (e.g. contrast, texture, margin spiculation), the method adopts the soft voting to combine the outputs from both individual classifiers. The experimental results show that the performance of the ensemble classifier was significantly better than the individual ones.

Another example is the fuse-TSD algorithm proposed in [86] for lung nodule classification. TSD incorporates the texture, shape with deep model-learned information at the decision level by the ensemble classifiers. Experimental results show fuse-TSD can have the higher AUC in distinguishing benign and malignant lung nodules. Other examples that utilize this approach include the breast cancer diagnosis [87], the skin lesion classification [88] and the classification of cardiac CT slices [89].

Feature-level fusion: In this approach, hand-crafted features and features extracted from CNNs are

concatenated, and the combined features are fed into a classifier for diagnosis. The structure of this approach is illustrated in Fig. 8.

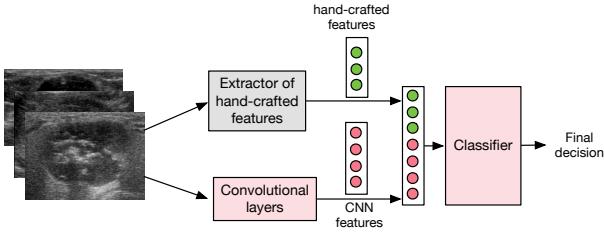


Fig. 8. Feature level fusion: the hand-crafted features are combined with the features extracted from CNNs as the new feature vectors.

For example, a combined-feature based classification approach, called as CFBC, is proposed for lung nodule classification [28]. In CFBC, the hand-crafted features (including the texture and shape descriptors) and the features learned by a nine-layer CNN are combined and fed into a back-propagation neural network. Experimental results on a publicly available dataset show that compared with a purely CNN model, incorporating hand-crafted features improves the accuracy, sensitivity and specificity by 3.87%, 6.41% and 3.21%, respectively.

Another example in this category is the glaucoma diagnosis method proposed in [90]. Concretely, a multi-branch neural network model is designed in which deep features extracted by CNNs and the hand-crafted features (including cup-to-disc ratio, peripapillary atrophy (PPA), etc.), are concatenated and reformed as a vector in a fully connected layer. Results show that the model outperforms classical CNNs.

Furthermore, in the breast cancer histology image classification, two hand-crafted features, namely the parameter-free threshold adjacency statistics (PFTAS) and gray-level co-occurrence matrix (GLCM), are fused with the five groups of deep learning features extracted from five different deep models [68]. Results show that after incorporating hand-crafted features, the accuracy of the deep learning model can be significantly improved.

Other examples of employing the above feature-level fusion can also be found in the chest pathology identification [91], skin lesion classification [92], [93], lung nodule classification [94], [95], [96], [99], histology image classification [97] and ECG classification [98].

Input-level fusion: In this approach, hand-crafted features are firstly represented as patches which highlight the corresponding features. Then these patches are taken as inputs to CNNs to make the final conclusion. Using this approach, the CNNs are expected to pay more attention to the hand-crafted features. It should be noted that generally speaking, one CNN is required for each type of hand-crafted feature, and information obtained from these CNNs need to be combined in some manner to make a final decision. The structure of this approach is illustrated in Fig. 9.

For example in [100], three types of hand-crafted features, namely the contrast information of the initial

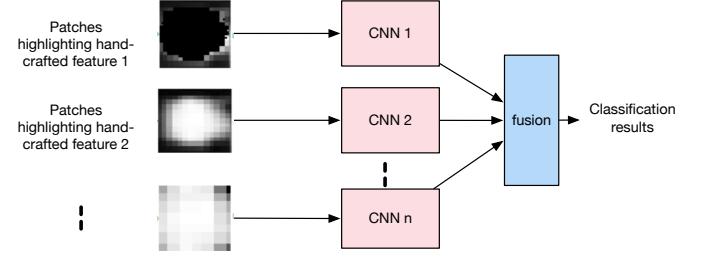


Fig. 9. Input-level fusion: the hand-crafted features are represented as image patches which are taken as inputs to the CNNs.

nodule candidates (INCs) and the outer environments, histogram of oriented gradients (HOG) feature, and LBP feature are transformed into the corresponding patches and are taken as inputs of multiple CNNs. The results show that this approach outperforms both conventional CNN-based approaches and traditional machine-learning approaches based on hand-crafted features.

Another example is the MV-KBC algorithm proposed for the lung nodule classification [56]. The MV-KBC first extracts patches corresponding to three features: the overall appearance (OA), voxel values (HVV) and heterogeneity in shapes (HS). Each type of patches are feed into a ResNet. The outputs of these ResNets are combined to generate the final classification results. An example of using OA and HS features can also be found in [101].

Moreover, [102] proposes the dual-path semi-supervised conditional generative adversarial networks (DScGAN) for the thyroid classification. Specifically, the features from the ultrasound B-mode images and the ultrasound elastography images are first extracted as the OB patches (indicating the region of interest (ROI) in B-mode images), OS patches (characterizing the shape of nodules) and OE patches (indicating the elasticity ROI). Then these patches are utilized as the input of the DScGAN. Using the information from these patches is demonstrated to be effective to improve the classification performance.

Furthermore, another work in this category can be found in [103] for the diagnosis of breast ultrasound images. To incorporate the shape information of tumors, besides original images, extra three types of images are also fed into the network, including images of tumor regions, segmentation maps and the images concatenated by the aforementioned three types of images.

As labels of CNNs

In this approach, besides the original classification labels of images, medical doctors also provide labels for some hand-crafted features. This extra information is generally incorporated into deep learning models via the multi-task learning structure.

For example, in [31], the nodules in lung images were quantitatively rated by radiologists with regard to 9 hand-crafted features (e.g. spiculation, texture and margin). The multi-task learning (MTL) framework is proposed to incorporate the above information into the main task of lung nodule classification.

In addition, for the benign-malignant risk assessment of lung nodules in low-dose CT scans [71], the binary labels about the presence of six high-level nodule attributes,

namely the calcification, sphericity, margin, lobulation, spiculation and texture, are utilized. Six CNNs are designed and each aims at learning one attribute. The automatically learned discriminative features by CNNs for these attributes are fused in a MTL framework to obtain the final risk assessment scores.

Similarly in [104], each glioma nuclear image is exclusively labeled in terms of shapes and attributes for the centermost nuclei of the image. These features are then learned by a multi-task CNN. Experiments demonstrate the proposed method outperforms the baseline CNN.

2.3.5 Other Types of Information Related to Diagnosis

In this section, we summarize other types of information that can help the deep learning models to improve the diagnostic performance. These types of information include extra category labels and clinical reports.

Extra category labels

For medical images, besides a classification label (i.e. normal, malignant or benign), radiologists may give some extra categorical labels. For example, in ultrasonic diagnosis of breast cancer, an image usually has a BI-RADS label which classifies the image into 0~6 [111] (category 0 suggests examination again, categories 1 and 2 prone to be a benign lesion, category 3 suggests probably benign findings, categories 4 and 5 are suspected to be malignant, category 6 definitely suggests malignant). These labels also contain information about the condition of lesions. In [55], the BI-RADS label for each medical image is incorporated in a multi-task learning structure as an auxiliary task. Experimental results show that the incorporating of these BI-RADS labels can improve the diagnostic performance of major classification task.

Extra radiological reports

A radiological report is a summary of all the clinical findings and impressions determined during examination of a radiography study. It usually contains rich information and reflects the findings and descriptions of radiologists. Incorporating radiological reports into CNNs designed for disease diagnosis can usually be beneficial. As medical reports are generally handled by recurrent neural networks (RNNs), to incorporate information from medical reports, generally hybrid networks containing both CNNs and RNNs are needed.

For example, a Text-Image embedding network (TieNet) is designed to classify the common thorax disease in chest X-rays [105]. TieNet has an end-to-end CNN-RNN architecture enabling it to integrate information of radiological reports. The classification results are significantly improved (about 6% increase on average in AUCs) compared to the baseline CNN purely based on medical images.

In addition, using semantic information from diagnostic reports is explored in [106] for understanding pathological bladder cancer images. A dual-attention model is designed to facilitate the high-level interaction of semantic information and visual information. Experiments demonstrate that incorporating information from diagnostic reports significantly improves the performance over the baseline method.

3 LESION, ORGAN AND ABNORMALITY DETECTION

3.1 Object Detection in Medical Images and the Corresponding Structures of Deep Learning Models

The task of detecting objects in medical images (e.g. lesions, abnormalities and organs) is important. In many conditions, lesion detection is a key part of disease diagnosis. Similarly, organ detection is an essential preprocessing step for image registration, organ segmentation and lesion detection. Detecting abnormalities in medical images, like cerebral microbleeds in brain MRI images and hard exudates in retinal images is also required in many applications.

The detection of these objects is one of the most labor-intensive tasks for medical doctors, and therefore, there has been a long research tradition to design CAD systems to accomplish the task [122] [123]. More recently, deep learning models have been applied to detect objects in medical images. Among the applications are pulmonary nodule detection in CT images [13], breast tumor detection in ultrasound images [124], [125], retinal diseases detection in retinal fundus images [126], [127] and so on.

According to structures, existing object detection models for medical images can be classified into the following three categories.

In the first category, original images are firstly cropped into small patches based on features like colors or textures. Patches are then classified using CNNs as target or non-target areas. Areas with the same labels are finally combined to obtain target candidates. Examples in this category can be found in [12], [13], [128], [129], [130].

Approaches in the second category generally adopt the two-stage detectors like Faster R-CNN [131] and Mask R-CNN [132]. These detectors have been widely utilized in the field of computer vision. They generally consist of a region proposal network (RPN) that hypothesizes candidate object locations and a detection network that refines region proposals. Examples in this category include colitis detection in CT images [133], intervertebral disc detection in X-ray images [134] and the detection of architectural distortions in mammograms [135].

Approaches in the third category adopt the one-stage objector like YOLO (You Only Look Once) [136], SSD (Single Shot MultiBox Detector) [137] and RetinaNet [138]. These networks skip the region proposal stage and run detection directly by considering the probability that the object appears at each point in image. Compared with the two-stage models, models in this approach are generally faster and simpler. Examples in this category can be found in the breast tumor detection in ultrasound images [124] and mammograms [125], and pulmonary lung nodule detection in CT images [139].

The above networks are mainly designed for detecting objects in 2D images. To detect objects in volumetric 3D images like CT and MRI, one straightforward way is to employ conventional 2D CNNs based on a single slice and process the slices sequentially [140], [141]. This solution disregards the contextual information along the third dimension, so its performance would be degraded. Alternatively, some models aggregate adjacent slices or orthogonal planes (i.e., axial, coronal and sagittal) to

enhance complementary spatial information [12], [142], [143], [144]. In particular, RNN and Long Short-Term Memory (LSTM) are used to incorporate temporal information in some detection tasks on videos [145], [146]. More recently, 3D CNNs are designed and applied to medical images domain [51], [147].

In the following sections, we will introduce the related works which incorporate external knowledge into deep learning models for object detection in medical images. The summary of these works are listed in Table 4.

3.2 Incorporating Knowledge from Natural Images or Other Medical Datasets

Similar to disease diagnosis, it is quite popular to pre-train a large natural image dataset (generally ImageNet). Examples can be found in the lymph node detection [14], breast tumor detection [24], the polyp and pulmonary embolism detection [148] and colorectal polyps detection [149], [150].

In addition, utilizing another medical dataset is also quite popular. For example, [151] develops a strategy to detect breast masses from digital tomosynthesis by fine-tuning the model pre-trained on mammography datasets. In addition, PET images are used to help the lesion detection in CT scans of liver [152].

3.3 Incorporating the Knowledge from Medical Doctors

In this section, we summarize the existing works on incorporating the knowledge of medical doctors for detecting objects in medical images. The types of domain knowledge from medical doctors are mainly focused on the following four categories:

- 1) the training patterns,
- 2) the general detection patterns they view images,
- 3) the features (e.g. locations, structures, shapes) they give special attentions to, and
- 4) other related information for detection.

3.3.1 Training Patterns of Medical Doctors

The training pattern of medical doctors, which is generally characterized as giving tasks with increasing difficulty, is also adopted for object detection in medical images. Similar to the disease diagnosis, most related works utilize the curriculum learning to introduce this pattern. For example, an attention-guided curriculum learning (AGCL) framework is presented to locate the lesion in chest X-rays [79]. During this process, images in order of difficulty (grouped by different severity-levels) are fed into CNN gradually, and the disease heatmaps generated from the CNN are used to locate the lesion areas.

Another work is called as CASED proposed for lung nodule detection in chest CT [153]. CASED adopts a curriculum adaptive sampling technique to address the problem of extreme data imbalance. In particular, CASED lets the network to first learn how to distinguish nodules from their immediate surroundings, and then adds a greater proportion of difficult-to-classify global context, until uniformly samples from the empirical data distribution. In this way, CASED tops the LUNA16 challenge leader-board with a score of 88.35%.

3.3.2 General Detection Pattern of Medical Doctors

When experienced medical doctors are locating possible lesions in medical images, they also have particular patterns, and these patterns can be incorporated into deep learning models for object detection. To detect objects in medical images, experienced medical doctors generally have the following patterns:

- they usually take into account images collected under different settings (e.g. brightness and contrast),
- they often compare bilateral images, and
- they generally read adjacent slices.

For example, to locate possible lesions during visual inspection of the CT images, radiologists would combine images collected under different settings (e.g. brightness and contrast). To imitate the above process, a multi-view feature pyramid network (FPN) is proposed in [154], where multi-view features are extracted from images rendered with varied brightness and contrast. The multi-view information is then combined using a position-aware attention module. Experiments show that the proposed model achieves an absolute gain of 5.65% over the previous state-of-the-art method on the NIH DeepLesion dataset.

In addition, the bilateral information is widely adopted by radiologists. For example, in standard mammographic screening, images are captured from both two breasts, and experienced radiologists generally compare bilateral mammogram images to find masses. To incorporate this pattern, a contrasted bilateral network (CBN) is proposed in [155], where the bilateral images are coarsely aligned first and then fed into a pair of networks to extract features for the following detection steps (shown in Fig. 10). Experimental results demonstrate the effectiveness of the bilateral information.

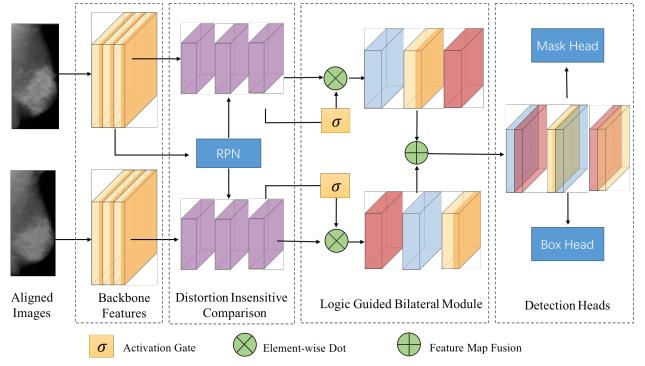


Fig. 10. Mammogram masses detection by integrating the bilateral information [155].

Similarly, to detect acute stroke signs in non-contrast CT images, experienced neuroradiologists routinely compare the appearance and Hounsfield Unit intensities of the left and right hemispheres, and then find the regions most commonly affected in stroke episodes. This pattern is mimicked by [156] for the detection of dense vessels and ischaemia. The experimental results show that introducing the pattern greatly improves the performance for detecting ischaemia. Another example of integrating the bilateral

TABLE 4

The list of researches of lesion, organ and abnormality detection and the knowledge they incorporated.

Knowledge Source	Knowledge Type	Incorporating Method	References
natural domain	natural images	transfer learning	[14], [24], [148], [149], [150]
medical dataset	multi-modal images	transfer learning	[151], [152]
	training pattern	curriculum learning	[79], [153]
		using multi-view images as input	[154]
	diagnostic pattern	comparing bilateral images	[155], [156], [157]
		analyzing adjacent slices	[139]
medical doctors	features doctors give specific attention	feature-level feature fusion	[158], [159], [160]
	other related information (classification labels, radiological reports)	multi-task learning training pattern design	[79], [161]

feature comparison into a CNN is the thrombus detection system proposed in [157].

When looking for small nodules in medical images, radiologists often observe each CT slice together with adjacent slices, similar to detecting an object in a video. This workflow is imitated in [139] to detect pulmonary lung nodule in CT images, where the state-of-the-art object detector, SSD, is applied in this process. The method obtains state-of-the-art result with FROC score 0.892 in LUNA16 dataset.

3.3.3 Features on Which Medical Doctors Give Special Attentions to

Similar to disease diagnosis, medical doctors also propose many ‘hand-crafted’ features to help them to find the target objects (e.g. nodules or lesions) in medical images. Incorporating these features can also be beneficial to deep learning models.

For example in [158], to detect mammographic lesions, many types of hand-crafted features (e.g. contrast features, geometrical features, location features) are firstly extracted, and then concatenated with those learned from a CNN (see Fig. 11). The results show that these hand-crafted features, especially the location and context features (which contain information not available to the CNN), can complement the network generating a higher specificity over the CNN.

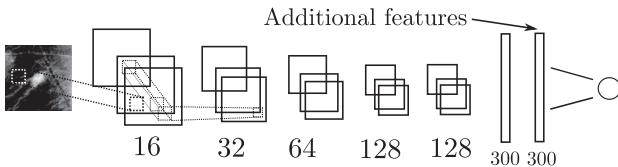


Fig. 11. Introducing hand-crafted features into a deep neural network for the detection of mammographic lesions.

Similarly in [159], a deep learning model based on Faster R-CNN is presented to detect abnormalities in the esophagus from endoscopic images. In particular, to enhance texture details, the proposed detection system incorporates the Gabor handcrafted features with the CNN features through concatenation in the detection stage. The results on two datasets (Kvasir and MICCAI 2015) show that the model is able to surpass the state-of-the-art performance.

Another example can be found in [160] for the detection of lung nodules, where the 88 handcrafted features

including intensity, shape, and texture are extracted and combined with features extracted by a CNN and then feed into a classifier. Experimental results demonstrate the effectiveness of the combination of handcrafted features and CNN features.

3.3.4 Other Types of Information Related to Detection

Similar with that in disease diagnosis, there are also other information (extra labels, radiological reports) can be integrated into the lesion detection process.

In [161], information of the classification labels is incorporated to help the lesion localization in chest X-rays and mammograms. In particular, a framework named as self-transfer learning (STL) is proposed, which jointly optimizes both classification and localization networks to help the localization network focus on correct lesions. Results show that STL can achieve significantly better localization performance compared to previous weakly supervised localization approaches.

Another example can be found [79]. To locate thoracic diseases on chest radiographs, the difficulty of each sample, represented as the severity level of the thoracic disease, is extracted from radiology reports firstly. Then the curriculum learning technique is adopted, in which the training samples are presented to the network in order of increasing difficulties. Experimental evaluations on the ChestXray14 database validate the effectiveness on significant performance improvement over baseline methods.

4 LESION AND ORGAN SEGMENTATION

4.1 Object Segmentation in Medical Images and the Corresponding Structures of Deep Learning Models

Medical image segmentation devotes to identifying pixels of lesions or organs from the background, and is generally regarded as a prerequisite step for the lesion assessment and disease diagnosis. Different from traditional segmentation systems which are generally based on edge detection filters and mathematical methods, segmentation methods based on deep learning models have become the dominant technique in recent years and have been widely used for the segmentation of lesions such as brain tumors [162], [163], breast tumors [164], and organs such as livers [165] and pancreas [166]. The deep learning models utilized for medical image segmentation are generally based on the CNN, the FCN, the U-Net and the GAN.

As the medical image segmentation can be seen as the pixel-level classification problem, the CNNs perform well in disease diagnosis can also be used for medical image segmentation. For these CNN-based methods, the original medical images are cropped into small patches. Then these patches are used to train a CNN-based classification network. At last, the classification results of these patches are combined as the final segmentation results. The examples in this category can be found in brain tumor segmentation [162], [167].

As an extension of the classical CNN, the fully convolutional network (FCN) is a popular pixel-based segmentation network structure [168]. FCN involves up-sampling layers to make the size of output match that of the input image. By combining the coarse abstractions from deep layers with fine details from shallow layers, FCN has been proven to perform well in various medical image segmentation tasks [169], [170], [171]. In addition, some variants of FCN, such as the cascaded FCN [172], parallel FCN [173], focal FCN [174], multi-stream FCN [175] and recurrent FCN [176] [177] are also widely used for segmenting medical images.

The third category of network structure for medical image segmentation is the U-Net [178] and its variants. U-Net builds upon FCN structure, mainly consists of a series of convolutional and deconvolutional layers, and with the short connections between the layers of equal resolution. By providing the high-resolution features to the corresponding deconvolutional layers, U-Net and its variants like UNet++ [179] and recurrent residual U-Net [180] perform well in many medical image segmentation tasks [181], [182], [183]. In addition, the 3D U-Net proposed in [184] and V-Net proposed in [185] are also widely utilized for the segmentation in 3D medical images [186], [187].

The fourth category for the medical image segmentation is the GAN-based models [188], [189]. In these methods, the generator is used to predict the mask of target based on some encoder-decoder structures (such as FCN or U-Net). The discriminator serves a shape regulator that helps the generator to obtain satisfactory segmentation results. Applications of GANs in medical image segmentation include brain segmentation [190], [191], myocardium and blood pool segmentation [192], splenomegaly segmentation [193], skin segmentation [194], vessel segmentation [195] and anomaly segmentation [196] in retinal fundus image, breast mass segmentation [197].

In the following sections, we introduce researches that incorporate domain knowledge into deep learning models for segmentation. The summary of these works are listed in Table 5.

4.2 Incorporating Knowledge From Natural Images or Other Medical Datasets

4.2.1 Knowledge from Natural Images or Videos

It is also quite common that a deep learning model for segmentation is firstly trained on a large-scale natural image dataset (e.g. ImageNet) and then fine-tuned on the target one. Using the above transfer learning strategy to introduce knowledge from natural images has demonstrated to achieve a better performance in medical image

segmentation. Examples can be found in intima-media boundary segmentation [148] and the segmentation of prenatal ultrasound images [198]. Besides ImageNet, [171] adopts the off-the-shelf DeepLab model trained on the PASCAL VOC dataset for anatomical structure segmentation in ultrasound images. This pre-trained model is also used in the deep contour-aware network (DCAN) which designed for the gland segmentation in histopathological images [199].

Besides using models pre-trained on ‘static’ datasets like ImageNet and PASCAL VOC, many deep neural networks, especially those designed for the segmentation of 3D medial images, leverage models pre-trained on large-scale video datasets. For example, in the automatic segmentation of proximal femur in 3D MRI [186], the C3D pre-trained model is adopted as the encoder of the proposed 3D U-Net. The C3D model is trained on the Sports-1M dataset, which is the largest video classification benchmark with 1.1 million sports videos in 487 categories [235].

4.2.2 Knowledge from Other Medical Datasets with the Different Modalities

In addition to natural images, using knowledge from external medical datasets with different modalities and with different diseases is also quite popular.

Transfer learning

For example, [200] investigates the transferability of the acquired knowledge of a CNN model initially trained for WM hyper-intensity segmentation on legacy low-resolution data when applied to new data from the same scanner but with higher image resolution. Likewise, the datasets of other MRI scanners and protocols are used in [201] to help the multi sclerosis segmentation process via transfer learning.

Another example is the MRI semantic segmentation in [202]. It is found that different segmentation networks for MRI semantic segmentation, if appropriately trained, share the similar distribution in the kernels, in contrast to noisy kernels from models trained on small datasets. Therefore, prior distribution of kernels in a network, even trained on a dataset in a different domain, should improve segmentation quality on the target segmentation problem. To leverage this prior distribution in kernels, a new transfer learning method is proposed based on the Deep Weight Prior (DWP) [236]. In this method, a Variational Autoencoder (VAE) is adopted to approximate the distribution of the kernels of the segmentation network trained on an external medical dataset, and then be utilized to solve the target problem. Experimental results on the BRATS2018 database demonstrate that the proposed method outperforms the traditional transfer learning method.

Multi-task learning and multi-modal learning

In [203], the multi-task learning is adopted, where the data of brain MRI, breast MRI and cardiac CT angiography (CTA) are used simultaneously as multiple tasks. On the other hand, [204] adopts a multi-modal learning structure for organ segmentation. A dual-stream encoder-decoder architecture is proposed to learn modality-independent, and thus, generalisable and robust features shared among medical datasets with different modalities (MRI and CT images). Experimental results prove the effectiveness of this multi-modal learning structure.

TABLE 5
A short list of researches of lesion, organ segmentation and the knowledge they incorporated.

Knowledge Source	Knowledge Type	Incorporating Method	References
medical dataset	natural domain	transfer learning	[148], [171], [186], [198], [199]
	multi-modal images	transfer learning	[200], [201], [202]
		multi-task/multi-modal learning	[203], [204]
		modality transformation/synthesis using GAN-based models	[191], [205], [206], [207], [208], [209], [210]
	datasets from other diseases	transfer learning	[211], [212], [213]
		disease domain transformation using GAN-based models	[214]
medical doctors	training pattern	curriculum learning	[215], [216]
	diagnostic pattern	using different views as input	[217], [218]
		attention mechanism	[219]
	anatomical priors	incorporated in the post-processing stage	[220], [221], [222]
		incorporated in the loss function	[32], [223], [224], [225], [226], [227], [228]
	other hand-crafted features	learned by using generative models	[218], [229], [230], [231]
		feature-level fusion	[232], [233]
		input-level fusion	[234]

Using GAN-based models

Besides using the multi-task learning and multi-modal learning, many works leverage GANs to achieve the domain transformation among datasets with different modalities. For example, in the left/right lung segmentation process, a model called SeUDA (unsupervised domain adaptation) is proposed [205], which leverages the semantic-aware GAN to transfer the knowledge from one chest dataset to another. In particular, target images are first mapped towards the source data space via the constraint of a semantic-aware GAN loss. Then the segmentation results are obtained from the segmentation DNN learned on the source domain. Experimental results show that the segmentation performance of SeUDA is highly competitive.

Similarly, for the segmentation of lung tumors in MR images, [207] adopts GANs to transform information from CT images. This solution first synthesizes a reasonably large number of MR images from a CT dataset, then combines the synthesized MR images with a fraction of real MR ones with corresponding labels and train the segmentation network. Experiments achieve reasonably accurate cancer segmentation from limited MRI datasets. Other examples of using GANs can also be found in the cardiac segmentation [206], [208], [209], liver segmentation [210], lung segmentation [211], brain MRI segmentation [191], [212], cardiac substructure and abdominal multi-organ segmentation [213].

4.2.3 Knowledge from Other Medical Datasets with Different Diseases

There are a few works that utilize the datasets of other diseases. For instance, [214] first builds a union dataset (3DSeg-8) by aggregating 8 different 3D medical segmentation datasets, and designs the Med3D network to co-train based on 3DSeg-8. Then the pre-trained models obtained from Med3D are transferred into lung and liver segmentation tasks. Experiments show that this method not only improves the accuracy, but also accelerates the training convergence speed.

Moreover, the annotated retinal images are used to help the cardiac vessel segmentation without annotations in [27]. In particular, a shape-consistent generative adversarial network (SC-GAN) is used to generate the synthetic images

and the corresponding labels. Then the synthetic images are used to train the segmentor. Experiments demonstrate the supreme accuracy of coronary artery segmentation.

4.3 Incorporating Knowledge From Medical Doctors

The domain knowledge of medical doctors is also widely adopted in the design of deep learning models for segmentation tasks in medical images. The types of domain knowledge from medical doctors utilized in deep segmentation models can be divided into four categories:

- 1) the training pattern,
- 2) the general diagnostic patterns they view images,
- 3) the anatomical priors (e.g. shape, location, topology) of lesions or organs, and
- 4) other hand-crafted features they give special attentions to.

4.3.1 Training Pattern of Medical Doctors

Many research works for segmenting medical images also mimic the training pattern of medical doctors, which is represented as giving tasks with increasing difficulties. They generally utilize the curriculum learning technique or its derivative methods like self-paced learning (SPL) [237].

For example in [215], for the segmentation of multi-organ CT images, each annotated medical image is divided into small patches. During the training process, patches producing large error by the network will be selected with a higher probability. In this manner, the network will focus sampling on difficult regions, resulting in improved performance.

Besides the curriculum learning, SPL, which is also inspired by the learning process of humans that gradually incorporates easy-to-hard samples into training [237], is also used in medical image segmentation. For example, the SPL is combined with the active learning for the pulmonary nodule segmentation in 3D images [216]. The system achieves the state-of-the-art segmentation performance.

4.3.2 General Diagnostic Pattern of Medical Doctors

In the lesion or organ segmentation tasks, some specific patterns that medical doctors adopted are incorporated into the network.

For example, during visual inspection of CT images, radiologists often change window widths and window centers to help to make decision on uncertain nodules. This pattern is mimicked in [217]. In particular, image patches of different window widths and window centers are stacked together as the input of the deep learning model to gain rich information. The evaluation implemented on the public LIDC-IDRI dataset indicates that the proposed method achieves promising performance on lung nodule segmentation compared with the state-of-the-art methods.

In addition, experienced clinicians generally assess the cardiac morphology and function from multiple standard views, using both long-axis (LA) and short-axis (SA) images to form an understanding of the cardiac anatomy. Inspired by the above observation, a cardiac MR segmentation method is proposed which takes three LA and one SA views as the input [218]. In particular, the features are firstly extracted using a multi-view autoencoder (MAE) structure, and then are feed in a segmentation network. Experimental results show that this method has a superior segmentation accuracy over state-of-the-art methods.

Furthermore, expert manual segmentation usually relies on the boundaries of anatomical structures of interest. For instance, a radiologist segmenting a liver from CT images would usually trace liver edges first, and then deduce the internal segmentation mask. Correspondingly, boundary-aware CNNs are proposed in [219] for medical image segmentation. The networks are designed to account for organ boundary information, both by providing a special network edge branch and edge-aware loss terms. The effectiveness of these boundary aware segmentation networks are tested on BraTS 2018 dataset for the task of brain tumor segmentation.

4.3.3 Anatomical Priors of Lesions or Organs

In comparison to non-medical images, medical images have many anatomical priors such as the shape, position and topology of organs or lesions. Experienced medical doctors greatly rely on these anatomical priors when they are doing segmentation tasks on these images. Incorporating the anatomical prior knowledge in deep learning models has been demonstrated to be an effective way for accurate medical image segmentation. Generally speaking, there are three different approaches to incorporate these anatomical priors into deep learning models.

Incorporating anatomical priors in the post processing stage

The first approach is to incorporate the anatomical priors in the post processing stage. The result of a deep segmentation network is often blurry and post-processing is generally needed to refine the segmentation result.

For example, according to pathology, most of breast cancer cases begin in glandular tissues and are located inside the mammary layer [238]. This position feature is utilized by [220] in its post-processing stage where a fully connected conditional random field (CRF) model is employed. In particular, the position of tumors and their relative locations with mammary layer are added as a new term in CRF energy function to obtain better segmentation results.

Besides, some researches first learn the anatomical priors, and then incorporate them in the post-processing stage to help produce anatomically plausible segmentation results [221], [222]. For instance, the latent representation of anatomically correct cardiac shape is first learned by using adversarial variational autoencoder (aVAE), then be used to convert erroneous segmentation maps into anatomically plausible ones [221]. Experiments manifest that aVAE is able to accommodate any cardiac segmentation method, and convert its anatomically implausible results to plausible ones without affecting its overall geometric and clinical metrics.

Another example in [222] introduces the post-processing step based on denoising autoencoders (DAE) for lung segmentation. In particular, the DAE is trained using only segmentation masks, then the learned representations of anatomical shape and topological constraints are imposed on the original segmentation results (as shown in Fig. 12). By applying the Post-DAE on the resulting masks from arbitrary segmentation methods, the lung anatomical segmentation of X-ray images shows plausible results.

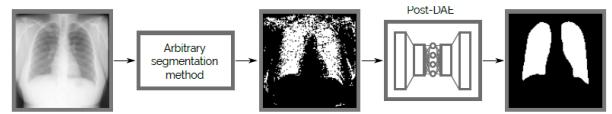


Fig. 12. The sample of integrating the shape prior in the post-process stage [222].

Incorporating anatomical priors as regularization terms in the loss function

The second approach is incorporating anatomical priors as regularization terms in the objective function of deep segmentation networks. For example, for the segmentation of cardiac MR images, [32] proposes a network called as SRSCN. SRSCN comprises a shape reconstruction neural network (SRNN) and a spatial constraint network (SCN). SRNN aims to maintain a realistic shape of the resulting segmentation. In addition, as the shapes and appearances of the heart in the basal and apical slices can vary significantly, the SCN is adopted to incorporate the spatial information of the 2D slices. The loss of the SRSCN comes from three parts: the segmentation loss, the shape reconstruction (SR) loss for shape regularization, and the spatial constraint (SC) loss to assist segmentation. The results using images from 45 patients demonstrate the effectiveness of the SR and SC regularization terms, and show the superiority of segmentation performance of the SRSCN over the conventional schemes.

Another example in this category is the one designed for skin lesion segmentation [223]. In this work, the star shape prior is encoded as a new loss term in a FCN to improve its segmentation of skin lesions from their surrounding healthy skin. In this manner, the non-star shape segments in FCN prediction maps are penalized to guarantee a global structure in segmentation results. The experimental results on the ISBI 2017 skin segmentation challenge dataset demonstrate the advantage of regularizing FCN parameters by the star shape prior.

Similarly in [224], a FCN-based network for histology gland segmentation is proposed. In this network, the

geometric priors (boundary smoothness) and topological priors (containment or exclusion) of lumen in epithelium and stroma are devised as additional regularization terms and added into the loss function of the FCN network. The results on the segmentation of histology glands from a dataset of 165 images demonstrate the advantage of the loss terms. More examples in this category can be found in kidney segmentation [225], liver segmentation [226] and cardiac segmentation [227], [228].

Learning anatomical priors via generative models

In the third approach, the anatomical priors (especially the shape prior) are learned by some generative models first and then incorporated into segmentation networks.

For example, in the cardiac MR segmentation process, a shape multi-view autoencoder (MAE) is proposed to learn shape priors from MR images of multiple standard views [218]. The information encoded in the latent space of the trained shape MAE is incorporated into multi-view U-Net (MV U-Net) in the fuse block to guide the segmentation process. The detail structures of shape MAE and MV U-Net are shown in Fig. 13.

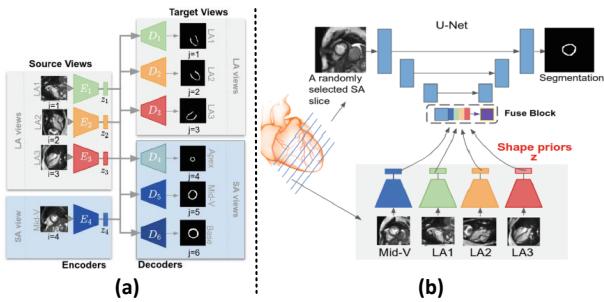


Fig. 13. The sample of using shape prior in the cardiac segmentation [218]. (a) The shape MAE used to learn shape priors of multiple standard views and (b) MV U-Net incorporates the shape priors learned from shape MAE.

Another example is shown in [229], where the shape constrained network (SCN) is proposed to incorporate the shape prior into the eye segmentation network. More specifically, the prior information is first learned by a VAE-GAN, and then the pre-trained encoder and discriminator are leveraged to regularize the training process. Other examples can also be found in brain geometry segmentation in MRI [230], 3D fine renal artery segmentation [231].

4.3.4 Other Hand-crafted Features

Besides anatomical priors, some hand-crafted features are also utilized for segmentation tasks. Generally speaking, there are two ways to incorporate the hand-crafted features into deep learning models: the feature-level fusion and the input-level fusion.

In the feature-level fusion, the hand-crafted features and the features learned by the deep models are concatenated. For example, for the gland segmentation in histopathology images [232], two handcrafted features, namely invariant LBP features as well as $H\&E$ components, are firstly calculated from images. Then these features are concatenated with the features generated from the last convolutional layer of the network for predicting the

segmentation results. In the brain structure segmentation [233], the spatial atlas prior is first represented as a vector and then concatenated with the deep features.

For the input-level fusion, the hand-crafted features are transformed into the input patches. Then the original image patches and the feature-transformed patches are fed into a deep segmentation network. For example in [234], for automatic brain tumor segmentation in MRI images, three handcrafted features (i.e. mean intensity, LBP and HOG) are firstly extracted. Based on these features, a SVM is employed to generate confidence surface modality (CSM) patches. Then the CSM patches and the original patches from MRI images are fed into a segmentation network. This method achieves good performance on BRATS2015 dataset.

5 OTHER MEDICAL APPLICATIONS

In the previous sections, we focus on three major applications in medical image analysis, namely, (1) disease diagnosis, (2) lesion, organ and abnormality detection, and (3) lesion or organ segmentation. In this section, we briefly introduce the works on incorporating medical domain knowledge in other related applications.

5.1 Medical Image Reconstruction

The objective of medical image reconstruction is reconstructing a diagnostic image from a certain number of measurements, e.g. X-ray projections in CT or the spatial frequency information (k-space data) in MRI. Deep learning based methods have been widely applied in this field [239], [240], [241].

It is also quite common that external information is incorporated into deep learning models for medical image reconstruction. First of all, the knowledge from natural images and medical datasets is helpful for the image reconstruction. For example, the pre-trained VGG-net model is incorporated into the optimization framework to ensure perceptual similarity in MRI reconstruction [242]. In addition, domain shift between the real data and the in silico data can also be tackled by using transfer learning network for optical tomography image reconstruction [243].

Besides, some other methods also incorporate some hand-crafted features in the medical image reconstruction process. For example, in [242], a network model called as DAGAN is proposed for the reconstruction of compressed sensing magnetic resonance imaging (CS-MRI). In the DAGAN, to better preserve texture and edges in the reconstruction process, the adversarial loss is coupled with a content loss. In addition, the frequency-domain information is incorporated to enforce similarity in both the image and frequency domains. Experimental results show that the DAGAN method provides superior reconstruction with preserved perceptual image details.

In [243], a new image reconstruction method is proposed to solve the limited-angle and limited sources breast cancer diffuse optical tomography (DOT) image reconstruction problem in a strong scattering medium. By adaptively focusing on important features and filtering irrelevant and noisy ones using the Fuzzy Jaccard loss, the network is able to reduce false positive reconstructed pixels and reconstruct more accurate images.

Similarly, to recover MRI images of the target contrast, [244] proposes a GAN-based method. The method simultaneously leverages the relatively low-spatial-frequency information available in the collected evidence for the target contrast and the relatively high-spatial frequency information available in the source contrast. Demonstrations on brain MRI datasets indicate the proposed method outperforms state-of-the-art reconstruction methods, with enhanced recovery of high-frequency tissue structure, and improved reliability against feature leakage or loss.

5.2 Medical Image Retrieval

The hospitals having diagnostic and investigative imaging facilities are producing large amount of imaging data. Therefore, the development of medical image retrieval, especially the content based image retrieval (CBIR) systems can be of great help to aid clinicians in browsing these large datasets. Intrinsically, CBIR is a computer vision technique which is based on the image features like color, texture and shape or any other features being derived from the image itself. Therefore, the performance of the CBIR system mainly depends on these selected features [245]. Deep learning methods have been applied to CBIR and have achieved high performance due to their superior capability for extracting features automatically.

It is also quite common that these deep learning models for CBIR utilize external information beyond the given medical datasets. Some methods adopt the transfer learning to utilize the knowledge from natural images or external medical datasets [246], [247], [248]. For example, the VGG model pre-trained based on ImageNet is used in brain tumor retrieval process [246], where a block-wise fine-tuning strategy is proposed to enhance the retrieval performance on the T1-weighted CE-MRI dataset. Another example can be found in x-ray image retrieval process [247], where a model pre-trained on the large augmented dataset is fine-tuned on the target dataset to extract general features.

Besides, as features play an important role in the similarly analysis in CBIR, some methods fuse prior features with deep features. In particular, in the chest radiograph image retrieval process, the decision values of binary features and texture features are combined with the deep features in the form of decision-level fusion [249]. Similarly, the metadata such as patients' age and gender is combined with the image-based features extracted from deep CNN for X-ray chest pathology image retrieval [250]. Furthermore, the features extracted from saliency areas can also be injected into the features extracted from the whole image for the high retrieval accuracy [248].

5.3 Medical Report Generation

A medical report generally includes the information of findings (e.g. medical observations of both normal and abnormal features), impressions or conclusions indicating the most prominent medical observation and other related information. Report-writing can be error-prone for unexperienced physicians, time-consuming and tedious for experienced physicians. To address these issues, many techniques, especially deep learning models for image

captioning have been successfully applied for automatic generation of medical reports [251], [252].

It is also found that incorporating external knowledge can help deep learning models to generate better medical reports. For example, to better extract the rich high-level features of images in the target datasets, many methods use the pre-trained models either based on natural image datasets [253], [254], [255] or some publicly available large medical datasets [256].

Some methods try to incorporate specific or general patterns that doctors adopt when writing reports. For example, radiologists generally write reports using certain templates. Therefore, some templates are used during the sentence generation process [255], [257]. Furthermore, as the explanation given by doctors is fairly simple and phrase changing does not change their meaning, [258] presents a model-agnostic model to learn the short text description to explain this decision process.

In addition, radiologists follow some procedures when writing reports: they generally first check a patient's images for abnormal findings, then write reports by following certain templates, and adjust statements in the templates for each individual case when necessary [259]. This process is mimicked in [257], which first transfers the visual features of medical images into an abnormality graph, then retrieves text templates based on the abnormalities and their attributes for chest X-ray images.

In [260], a pre-constructed graph embedding module (modeled with a graph CNN) on multiple disease findings is utilized to assist the generation of reports. The incorporation of knowledge graph allows for dedicated feature learning for each disease finding and the relationship modeling between them. Experiments on the publicly accessible dataset (IU-RR) demonstrate the superior performance of the method integrated with the proposed graph module.

6 RESEARCH CHALLENGES, OPEN PROBLEMS AND DIRECTIONS

In this section, we summarize the research challenges, open problems and future directions in embedding medical domain knowledge in deep learning models.

6.1 The Medical Domain Knowledge Itself

Although using medical domain knowledge in deep learning models is quite popular, there are many difficulties, challenges and open problems related to the identification, representation and evaluation of medical domain knowledge. They are listed as follows.

The subjective and individual-dependent domain knowledge

The medical domain knowledge to be selected is generally based on the experience of medical doctors. However, identifying experience from medical doctors is not an easy task. The experience of humans, including medical doctors, are intrinsically subjective and fuzzy. Even experienced medical doctors may not be able to give accurate and objective descriptions on what kinds of experiences they have leveraged to finish a given task. Furthermore, experiences of medical doctors are dependent

on many factors include the trainings they have received, the patients they met and therefore can vary significantly or even contradictory to each other. How to identify the common domain knowledge accepted by the community is not a trivial task.

The difficulty in the representation of domain knowledge

Second, the representation of the domain knowledge is also a challenging task. The original domain knowledge of medical doctors is generally in the form of descriptive sentences like ‘we will focus more on the margin areas of a tumor to determine whether it is benign or malignant’, or ‘we often compare bilateral images to make decision’. How to transform the knowledge into appropriate representations that can be incorporated into deep learning models need a careful design.

In addition, some types of knowledge is relatively simple to be represented but some are quite difficult, if not impossible. We utilize features in Table 3 as an example. We can see that the margin is an important feature for radiologists to distinguish between benign and malignant tumors in the breast ultrasound images. This knowledge can be represented as hand-crafted features: we can utilize some techniques to extract the features representing the smoothness or regularity of the tumor margin and add these features into a deep learning model. Another way to represent margin information is simply to let a deep learning network pay more attention on the margin areas of tumors using some attention mechanisms. However, the representation of other features, including the acoustic attenuation and the echo pattern is much more difficult, as these features do not reflect the characteristics of local areas. How to find a generic approach to efficiently and accurately represent a given type of domain knowledge is still an open problem.

The overlap of the domain knowledge and the knowledge learned by deep learning models

Even we have correctly identified a type of domain knowledge and incorporated it into a deep learning model in an appropriate representation, it is not guaranteed that the domain knowledge can help to improve its performance. One major reason is that *the domain knowledge may have been learned by the deep learning model based on the training data*. Intuitively speaking, only the knowledge that is not easily learned by a deep learning model based on a relatively small-sized dataset can help the model to improve the performance. Given a type of medical domain knowledge, whether it can be easily learned by neural networks from datasets is still an open problem. Some experiences of medical doctors can be biased and may even degrade the performance of deep learning models after being added. How to evaluate the extent to which the medical domain knowledge is helpful to deep learning models is important, as the answer to this problem will give us important clue on the selection of the medical domain knowledge.

The dilemma of the specificity and generality of medical domain knowledge

A few types of medical domain knowledge, like the training pattern of medical doctors and some diagnostic patterns, are relatively generic. However, most types of the domain knowledge of medical doctors are dependent on

diseases and imaging types (e.g. CT, MRI, and ultrasound). For example, radiologists will adopt different rules to diagnose brain tumors and lung nodules. Even for the same disease (e.g. breast cancer), radiologists adopt different patterns when reading ultrasound images and MR images. This greatly lowers down the generality of medical domain knowledge. In addition, there is usually a dilemma of the specificity and generality of medical domain knowledge: the generic domain knowledge contain less information, while the specific domain knowledge, although can be informative, is only limited in for a certain disease with a certain imaging type.

6.2 The Incorporating Methods

The mechanism to adjust the amount of medical domain knowledge to be incorporated

Although various incorporating methods have been designed, how to adjust the amount of medical domain knowledge to be injected into deep learning models has not been studied. Generally speaking, incorporating domain knowledge has a price to pay. For example, if we let the network pay too much attention to margin areas as medical doctors usually do, the network may overlook other areas that may contain more useful information for the task. To the extreme case, the network will act exactly like a medical doctor and lost its ability to learn from data, which violates the principle of ‘using medical domain knowledge to assist deep learning models’.

Based on the above discussion, we can see that it would be beneficial if we can design a mechanism to control the amount of medical domain knowledge to be incorporated into deep learning models. In addition, the mechanism should be adaptive, as the optimal answer depends on many factors, including the domain knowledge itself, the deep learning models and the size of the given datasets. How to design a mechanism which optimally control the amount of medical domain knowledge to be incorporated into deep learning models deserves further study.

The transfer learning and multi-task learning designed specifically for medical image analysis

In essence, medical domain knowledge is just a special type of extra information to be incorporated into deep learning models. In general computer vision tasks, there have many strategies to incorporate extra information into deep learning models, and two popular ones are the transfer learning and the multi-task learning. Not surprisingly, these two techniques have been extensively adopted to introduce medical domain knowledge into deep learning models. However, medical domain has many distinct features from general computer vision fields, in terms of the distribution of images, and the tasks to be accomplished. How to find the transfer learning or the multi-task learning specifically designed for medical field is an interesting topic deserves further investigation.

Incorporating multi-modal medical domain knowledge simultaneously

Most of the existing works only incorporate a single type of medical domain knowledge, or a few types of medical domain knowledge of the same modality (e.g. a number of hand-crafted features). However, when reading images,

experienced medical doctors usually combines different experience in different stages. For example, in the diagnosis of breast cancer, experienced radiologists generally take the three-staged approach: first browsing the whole image, then concentrating on the local lesion areas, and finally combining the global and local information to make decisions. When scrutinizing local lesion areas, they focus on some characteristics such as shape, texture, calcification, edge and so on. In addition, they will also compare historical medical images and the previous radiology reports of the patient to track the change of the lesions. In summary, medical doctors comprehensively utilize many types of domain knowledge before making a decision. Similarly, how to design some architectures which is able to incorporate multi-modal domain knowledge is still an open problem.

Incorporating method design and evaluation

Generally speaking, most of the existing works are the case-by-case study, or are application-dependent. Some types of domain knowledge and the incorporating methods, although effective on some certain medical datasets, are not proven to work on other datasets. This greatly reduce the utility of the identified domain knowledge and the integration methods. When facing with different sizes of medical datasets or deep models with different network structures, the integrating methods of domain knowledge may also different. How to design an effective and useful integrating method is essential for medical image analysis.

Besides, different from the general computer vision fields, where many large-scale datasets can be used to test the effectiveness of a neural network model. From this aspect, we believe that medical field may also need some large medical datasets serving as benchmarks, which are able to give a good evaluation on a certain given medical domain knowledge and an integrating method.

6.3 Other Research Directions

Besides the open problems mentioned above, there are several directions that we feel need further investigation in the future.

Domain adaptation

Domain adaptation is developed to transfer the information from a source domain to a target one. Via techniques like adversarial learning [261], domain adaptation is able to narrow the domain shift between the source domain and the target one in input space [262], feature space [263], [264] and output space [265], [266], [267]. Domain adaptation can be naturally adopted to transfer knowledge of one medical dataset to another, even when they have different imaging modes or belong to different diseases.

In addition, unsupervised domain adaptation (UDA) is a promising avenue to enhance the performance of deep neural networks on the target domain, using labels only from the source domain. This is especially useful for medical field, as annotating the medical images is quite labor-intensive and the lack of annotations is quite common in medical datasets. Some examples have demonstrated the effectiveness of UDA [268], but further depth study needs to be implemented in the future.

The knowledge graph

We believe the knowledge graph is another interesting direction in this field that deserves further study. According to different relationships in graphs, we can establish three types of knowledge graphs. The first knowledge graph reflects the relationship among different kinds of medical domain knowledge with respect to a certain disease. This knowledge graph can help us identify a few key types of knowledge that may help to improve deep learning models better. The second type of knowledge graph reflects the relationship among different diseases. This knowledge graph can help us to find out the potential domain knowledge from other related diseases. The third type of knowledge graph describes the relationship among medical datasets. These datasets can belong to different diseases and in different imaging modes (e.g. CT, MRI, ultrasound). This type of knowledge graph will help to identify the external datasets that may help to improve the performance of the current deep learning model.

We believe that the knowledge graph, with the character of embedding different types of knowledge, is a generic and flexible approach to incorporate multi-modal medical domain knowledge.

The generative models

The generative models, like GAN and AE, have shown great promise to be applied to incorporate medical domain knowledge into deep learning models, especially for segmentation tasks. GAN has shown its capability to leverage information from extra datasets with different imaging modes (e.g. using a MRI dataset to help segmenting CT images [205], [207]). In addition, GAN is able to learn important features contained in medical images in a weakly or fully unsupervised manner and therefore is quite suitable for medical image analysis.

AE-based models have already achieved a great success in extracting features, especially the shape feature in objects like organs or lesions in a fully unsupervised manner [218], [229]. The features learning by AE can also be easily integrated into the training process of networks.

Network architecture search (NAS)

At last, we have mentioned in the previous section that one open problem is to find appropriate network architectures to incorporate medical domain knowledge. We believe one approach to address this problem is the technique of NAS. NAS has demonstrated its capability to automatically find a good network architecture in many computer vision tasks. We believe it is also a generic approach to find an optimal structure to integrate any given medical domain knowledge.

For instance, when some hand-crafted features is used as the domain knowledge, with the help of NAS, a network structure can be identified with the special connections between domain knowledge features and deep features. In addition, instead of designing the feature fusion method (feature-level fusion, decision-level fusion or input-level fusion) for these two kinds of features, the integrating phase and integrating intensity of these two kinds of features can also be determined during the searching process.

REFERENCES

- [1] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee, "Joint weakly and semi-supervised deep learning for localization and

- classification of masses in breast ultrasound images," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 762–774, 2019.
- [2] X. Xu, Q. Lu, L. Yang, S. X. Hu, D. Z. Chen, Y. Hu, and Y. Shi, "Quantization of fully convolutional networks for accurate biomedical image segmentation," *computer vision and pattern recognition*, pp. 8300–8308, 2018.
- [3] Z. Zhou, J. Y. Shin, L. Zhang, S. R. Gurudu, M. B. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," *computer vision and pattern recognition*, pp. 4761–4772, 2017.
- [4] A. Esteva, B. Kuprel, R. A. Novoa, J. M. 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, pp. 115–118, 2017.
- [5] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification," *workshop on applications of computer vision*, pp. 673–681, 2018.
- [6] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 8–12, 2009.
- [7] "Imagenet," <http://www.image-net.org/about-overview>.
- [8] "Coco," <http://mscoco.org/>.
- [9] H. Chen, X. J. Qi, J. Z. Cheng, and P. A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Thirtieth AAAI conference on artificial intelligence*, 2016.
- [10] J. Cheng, D. Ni, Y. Chou, J. Qin, C. M. Tiu, Y. Chang, C. Huang, D. Shen, and C. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans." *Scientific Reports*, vol. 6, no. 1, pp. 24 454–24 454, 2016.
- [11] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative *et al.*, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.
- [12] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1170–1181, 2015.
- [13] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. 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 transactions on medical imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.
- [14] 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 transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [15] 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," *Medical image analysis*, vol. 26, no. 1, pp. 195–202, 2015.
- [16] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th international conference on control automation robotics & vision (ICARCV)*. IEEE, 2014, pp. 844–848.
- [17] B. J. Erickson, P. Korfatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.
- [18] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.
- [19] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," in *Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, p. 105741M.
- [20] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [22] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.
- [23] B. Huynh, K. Drukker, and M. Giger, "Mo-de-207b-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks," *Medical physics*, vol. 43, no. 6Part30, pp. 3705–3705, 2016.
- [24] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.
- [25] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [26] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 686–696, 2018.
- [27] F. Yu, J. Zhao, Y. Gong, Z. Wang, Y. Li, F. Yang, B. Dong, Q. Li, and L. Zhang, "Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 714–722.
- [28] Y. Xie, J. Zhang, S. Liu, W. Cai, and Y. Xia, "Lung nodule classification by jointly using visual descriptors and deep features," in *Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging*. Springer, 2016, pp. 116–125.
- [29] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv preprint arXiv:1801.09927*, 2018.
- [30] L. Li, M. Xu, X. Wang, L. Jiang, and H. Liu, "Attention based glaucoma detection: A large-scale database and cnn model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 571–10 580.
- [31] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, and J.-Z. Cheng, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in ct images," *IEEE transactions on medical imaging*, vol. 36, no. 3, pp. 802–814, 2016.
- [32] Q. Yue, X. Luo, Q. Ye, L. Xu, and X. Zhuang, "Cardiac segmentation from lge mri using deep neural network incorporating shape and spatial priors," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 559–567.
- [33] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [34] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [35] K. Suzuki, "Survey of deep learning applications to medical image analysis," *Med Imaging Technol*, vol. 35, no. 4, pp. 212–226, 2017.
- [36] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, p. 112821, 2019.
- [37] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, "Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on mri," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 939–954, 2019.
- [38] G. Maicas, A. P. Bradley, J. C. Nascimento, I. Reid, and G. Carneiro, "Training medical image analysis systems like radiologists," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 546–554.
- [39] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: a review," *Engineering*, 2019.
- [40] T. G. Debelee, F. Schwenker, A. Ibenthal, and D. Yohannes,

- "Survey of deep learning in breast cancer image analysis," *Evolving Systems*, pp. 1–21, 2019.
- [41] L. M. Pehrson, M. B. Nielsen, and C. Ammitzbøl Lauridsen, "Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the lidec-idi database: A systematic review," *Diagnostics*, vol. 9, no. 1, p. 29, 2019.
- [42] F. Altaf, S. M. Islam, N. Akhtar, and N. K. Janjua, "Going deep in medical image analysis: Concepts, methods, challenges, and future directions," *IEEE Access*, vol. 7, pp. 99 540–99 572, 2019.
- [43] T. Hirai, Y. Korogi, H. Arimura, S. Katsuragawa, M. Kitajima, M. Yamura, Y. Yamashita, and K. Doi, "Intracranial aneurysms at mr angiography: effect of computer-aided diagnosis on radiologists' detection performance," *Radiology*, vol. 237, no. 2, pp. 605–610, 2005.
- [44] F. Li, H. Arimura, K. Suzuki, J. Shiraishi, Q. Li, H. Abe, R. Engelmann, S. Sone, H. MacMahon, and K. Doi, "Computer-aided detection of peripheral lung cancers missed at ct: Roc analyses without and with localization," *Radiology*, vol. 237, no. 2, pp. 684–690, 2005.
- [45] H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz *et al.*, "Computer-aided diagnosis in chest radiography: results of large-scale observer tests at the 1996–2001 rsna scientific assemblies," *Radiographics*, vol. 23, no. 1, pp. 255–265, 2003.
- [46] N. Bhooshan, M. Giger, M. Medved, H. Li, A. Wood, Y. Yuan, L. Lan, A. Marquez, G. Karczmar, and G. Newstead, "Potential of computer-aided diagnosis of high spectral and spatial resolution (hiss) mri in the classification of breast lesions," *Journal of Magnetic Resonance Imaging*, vol. 39, no. 1, pp. 59–67, 2014.
- [47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *neural information processing systems*, vol. 141, no. 5, pp. 1097–1105, 2012.
- [48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *computer vision and pattern recognition*, pp. 1–9, 2015.
- [49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *international conference on learning representations*, 2015.
- [50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *computer vision and pattern recognition*, pp. 770–778, 2016.
- [51] G. Huang, Z. Liu, L. V. Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *computer vision and pattern recognition*, pp. 2261–2269, 2017.
- [52] R. F. Mansour, "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy," *Biomedical engineering letters*, vol. 8, no. 1, pp. 41–57, 2018.
- [53] J. Y. Kim, H. E. Lee, Y. H. Choi, S. J. Lee, and J. S. Jeon, "Cnn-based diagnosis models for canine ulcerative keratitis," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019.
- [54] X. Li, L. Shen, X. Xie, S. Huang, Z. Xie, X. Hong, and J. Yu, "Multi-resolution convolutional networks for chest x-ray radiograph based lung nodule detection," *Artificial intelligence in medicine*, p. 101744, 2019.
- [55] J. Liu, W. Li, N. Zhao, K. Cao, Y. Yin, Q. Song, H. Chen, and X. Gong, "Integrate domain knowledge in training cnn for ultrasonography breast cancer diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 868–875.
- [56] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. J. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 991–1004, 2019.
- [57] M. Mitsuhasha, H. Fukui, Y. Sakashita, T. Ogata, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Embedding human knowledge in deep neural network via attention map," *arXiv preprint arXiv:1905.03540*, 2019.
- [58] F. P. dos Santos and M. A. Ponti, "Robust feature spaces from pre-trained deep network layers for skin lesion classification," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018, pp. 189–196.
- [59] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, and V. Vaidya, "Understanding the mechanisms of deep transfer learning for medical images," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 188–196.
- [60] G. Wimmer, S. Hegenbart, A. Vécsei, and A. Uhl, "Convolutional neural network architectures for the automated diagnosis of celiac disease," in *International Workshop on Computer-Assisted and Robotic Endoscopy*. Springer, 2016, pp. 104–113.
- [61] J. Arevalo, F. A. Gonzalez, R. Ramospollan, J. L. Oliveira, and M. A. G. Lopez, "Convolutional neural networks for mammography mass lesion classification," *vol. 2015*, pp. 797–800, 2015.
- [62] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414. International Society for Optics and Photonics, 2015, p. 94140V.
- [63] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *12th International Symposium on Medical Information Processing and Analysis*, vol. 10160. International Society for Optics and Photonics, 2017, p. 101600U.
- [64] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.
- [65] M. Salem, S. Taheri, and J.-S. Yuan, "Ecg arrhythmia classification using transfer learning from 2-dimensional deep cnn features," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2018, pp. 1–4.
- [66] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers *et al.*, "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018.
- [67] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 76–84, 2016.
- [68] H. Cao, S. Bernard, L. Heutte, and R. Sabourin, "Improve the performance of transfer learning without fine-tuning using dissimilarity-based multi-view learning for breast cancer histology images," pp. 779–787, 2018.
- [69] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.
- [70] W. Li, J. Li, K. V. Sarma, K. C. Ho, S. Shen, B. S. Knudsen, A. Gertych, and C. W. Arnold, "Path r-cnn for prostate cancer diagnosis and gleason grading of histological images," *IEEE transactions on medical imaging*, vol. 38, no. 4, pp. 945–954, 2018.
- [71] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d cnn-based multi-task learning," *international conference information processing*, pp. 249–260, 2017.
- [72] 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," *Physics in Medicine & Biology*, vol. 62, no. 23, p. 8894, 2017.
- [73] S. Azizi, P. Mousavi, P. Yan, A. Tahmasebi, J. T. Kwak, S. Xu, B. Turkbey, P. Choyke, P. Pinto, B. Wood *et al.*, "Transfer learning from rf to b-mode temporal enhanced ultrasound features for prostate cancer detection," *International journal of computer assisted radiology and surgery*, vol. 12, no. 7, pp. 1111–1121, 2017.
- [74] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. Richter, and K. Cha, "Cross-domain and multi-task transfer learning of deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis," in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575. International Society for Optics and Photonics, 2018, p. 105750Q.
- [75] X. Li, G. Qin, Q. He, L. Sun, H. Zeng, Z. He, W. Chen, X. Zhen, and L. Zhou, "Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification," *European radiology*, vol. 30, no. 2, pp. 778–788, 2020.
- [76] O. Hadad, R. Bakalo, R. Ben-Ari, S. Hashoul, and G. Amit, "Classification of breast lesions using cross-modal deep

- learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 109–112.
- [77] Q. Liao, Y. Ding, Z. L. Jiang, X. Wang, C. Zhang, and Q. Zhang, "Multi-task deep convolutional neural network for cancer diagnosis," *Neurocomputing*, vol. 348, pp. 66–73, 2019.
- [78] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "Canet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE transactions on medical imaging*, 2019.
- [79] Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers, "Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 249–258.
- [80] A. Jiménez-Sánchez, D. Mateus, S. Kirchhoff, C. Kirchhoff, P. Biberthaler, N. Navab, M. A. G. Ballester, and G. Piella, "Medical-based deep curriculum learning for improved fracture classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 694–702.
- [81] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale cnn and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 169–177.
- [82] C. Haarburger, M. Baumgartner, D. Truhn, M. Broeckmann, H. Schneider, S. Schrading, C. Kuhl, and D. Merhof, "Multi scale curriculum cnn for context-aware breast mri malignancy classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 495–503.
- [83] I. González-Díaz, "Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 547–559, 2018.
- [84] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, and Z. Liu, "Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1959–1970, Aug 2019.
- [85] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2016, pp. 1–6.
- [86] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct," *Information Fusion*, vol. 42, pp. 102–110, 2018.
- [87] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Medical physics*, vol. 44, no. 10, pp. 5162–5171, 2017.
- [88] V. Singh and I. Nwogu, "Analyzing skin lesions in dermoscopy images using convolutional neural networks," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 4035–4040.
- [89] M. Moradi, Y. Gur, H. Wang, P. Prasanna, and T. Syeda-Mahmood, "A hybrid learning approach for semantic labeling of cardiac ct slices and recognition of body position," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1418–1421.
- [90] Y. Chai, H. Liu, and J. Xu, "Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models," *Knowledge-Based Systems*, vol. 161, pp. 147–156, 2018.
- [91] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 294–297.
- [92] J. Hagerty, J. Stanley, H. Almubarak, N. Lama, R. Kasmi, P. Guo, R. Drugge, H. Rabinovitz, M. Olivero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images," *IEEE journal of biomedical and health informatics*, 2019.
- [93] B. M. Devassy, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "The impact of replacing complex hand-crafted features with standard features for melanoma classification using both hand-crafted and deep features," in *Proceedings of SAI Intelligent Systems Conference*. Springer, 2018, pp. 150–159.
- [94] C. Wang, A. Elazab, J. Wu, and Q. Hu, "Lung nodule classification using deep feature fusion in chest radiography," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 10–18, 2017.
- [95] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma," *Tomography*, vol. 2, no. 4, p. 388, 2016.
- [96] R. Paul, S. Hawkins, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Predicting malignant nodules by fusing deep features with classical radiomics features," *Journal of Medical Imaging*, vol. 5, no. 1, p. 011021, 2018.
- [97] Z. Lai and H. Deng, "Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron?" *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [98] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "Encase: an ensemble classifier for ecg classification using expert features and deep neural networks," 09 2017.
- [99] M. Buty, Z. Xu, M. Gao, U. Bagci, A. Wu, and D. J. Mollura, "Characterization of lung nodule malignancy using hybrid shape and appearance features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 662–670.
- [100] J. Tan, Y. Huo, Z. Liang, and L. Li, "Expert knowledge-infused deep learning for automatic lung nodule detection," *Journal of X-ray science and technology*, no. Preprint, pp. 1–20, 2019.
- [101] J. Shi, J. Wu, P. Lv, and J. Guo, "Breastnet: Entropy-regularized transferable multi-task learning for classification with limited breast data."
- [102] W. Yang, J. Zhao, Y. Qiang, X. Yang, Y. Dong, Q. Du, G. Shi, and M. B. Zia, "Dscgans: Integrate domain knowledge in training dual-path semi-supervised conditional generative adversarial networks and s3vm for ultrasonography thyroid nodules classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 558–566.
- [103] W. K. Moon, Y.-W. Lee, H.-H. Ke, S. H. Lee, C.-S. Huang, and R.-F. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Computer Methods and Programs in Biomedicine*, p. 105361, 2020.
- [104] V. Murthy, L. Hou, D. Samaras, T. M. Kurc, and J. H. Saltz, "Center-focusing multi-task cnn with injected features for classification of glioma nuclear images," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 834–841.
- [105] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9049–9058.
- [106] Z. Zhang, P. Chen, M. Sapkota, and L. Yang, "Tandemnet: Distilling knowledge from medical images using diagnostic reports as optional semantic references," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 320–328.
- [107] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
- [108] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 41–48.
- [109] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.
- [110] C. Dorsi, L. Bassett, S. Feig et al., "Breast imaging reporting and data system (bi-rads)," *Breast imaging atlas, 4th edn. American College of Radiology*, Reston, 1998.
- [111] B. W. e. a. Mendelson EB, Bohm-Vlez M, "Acr bi-rads ultrasound. in: Acr bi-rads atlas, breast imaging reporting and data system." Reston, VA, *American College of Radiology*, 2013.
- [112] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

- [113] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [114] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *Journal of digital imaging*, vol. 28, no. 1, pp. 99–115, 2015.
- [115] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung ct images," *Journal of digital imaging*, vol. 29, no. 4, pp. 466–475, 2016.
- [116] M. Alilou, M. Orooji, and A. Madabhushi, "Intra-perinodular textural transition (ipris): A 3d descriptor for nodule diagnosis on lung ct," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 647–655.
- [117] M. P. Paing, K. Hamamoto, S. Tungjitusolmun, and C. Pintaviroo, "Automatic detection and staging of lung tumors using locational features and double-staged classifications," *Applied Sciences*, vol. 9, no. 11, p. 2329, 2019.
- [118] P. Tsui, Y. Liao, C. Chang, W. Kuo, K. Chang, and C. Yeh, "Classification of benign and malignant breast tumors by 2-d analysis based on contour description and scatterer characterization," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 513–522, 2010.
- [119] A. Madabhushi and D. N. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 155–169, 2003.
- [120] L. Sellami, O. B. Sassi, k. Chtourou, and A. B. Hamida, "Breast cancer ultrasound images' sequence exploration using bi-rads features' extraction: Towards an advanced clinical aided tool for precise lesion characterization," *IEEE Transactions on NanoBioscience*, vol. 14, no. 7, pp. 740–745, Oct 2015.
- [121] S.-M. Hsu, W.-H. Kuo, F.-C. Kuo, and Y.-Y. Liao, "Breast tumor classification using different features of quantitative ultrasound parametric images," *International journal of computer assisted radiology and surgery*, vol. 14, no. 4, pp. 623–633, 2019.
- [122] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, and E. B. Mendelson, "Computerized lesion detection on breast ultrasound," *Medical Physics*, vol. 29, no. 7, pp. 1438–1446, 2002.
- [123] M.-C. Yang, C.-S. Huang, J.-H. Chen, and R.-F. Chang, "Whole breast lesion detection using naive bayes classifier for portable ultrasound," *Ultrasound in medicine & biology*, vol. 38, no. 11, pp. 1870–1880, 2012.
- [124] Z. Cao, L. Duan, G. Yang, T. Yue, Q. Chen, H. Fu, and Y. Xu, "Breast tumor detection in ultrasound images using deep learning," in *International Workshop on Patch-based Techniques in Medical Imaging*. Springer, 2017, pp. 121–128.
- [125] R. Platania, S. Shams, S. Yang, J. Zhang, K. Lee, and S.-J. Park, "Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid)," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2017, pp. 536–543.
- [126] 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," *Investigative ophthalmology & visual science*, vol. 57, no. 13, pp. 5200–5206, 2016.
- [127] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [128] Z. Xue, S. Jaeger, S. Antani, L. R. Long, A. Karargyris, J. Siegelman, L. R. Folio, and G. R. Thoma, "Localizing tuberculosis in chest radiographs with deep learning," in *Medical imaging 2018: imaging informatics for healthcare, research, and applications*, vol. 10579. International Society for Optics and Photonics, 2018, p. 105790U.
- [129] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 559–567.
- [130] M. J. Van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.
- [131] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," vol. 2015, pp. 91–99, 2015.
- [132] K. He, G. Gkioxari, P. Dollar, and R. B. Girshick, "Mask r-cnn," *international conference on computer vision*, pp. 2980–2988, 2017.
- [133] J. Liu, D. Wang, L. Lu, Z. Wei, L. Kim, E. B. Turkbey, B. Sahiner, N. A. Petrick, and R. M. Summers, "Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks," *Medical physics*, vol. 44, no. 9, pp. 4630–4642, 2017.
- [134] R. Sa, W. Owens, R. Wiegand, M. Studin, D. Capoferro, K. Barooah, A. Greaux, R. Ratray, A. Hutton, J. Cintineo *et al.*, "Intervertebral disc detection in x-ray images using faster r-cnn," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 564–567.
- [135] R. Ben-Ari, A. Akselrod-Ballin, L. Karlinsky, and S. Hashoul, "Domain specific convolutional neural nets for detection of architectural distortion in mammograms," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 552–556.
- [136] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [137] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [138] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [139] N. Li, H. Liu, B. Qiu, W. Guo, S. Zhao, K. Li, and J. He, "Detection and attention: Diagnosing pulmonary lung cancer from ct by imitating physicians," *arXiv preprint arXiv:1712.05114*, 2017.
- [140] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in pet/ct images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical physics*, vol. 43, no. 6Part1, pp. 2821–2827, 2016.
- [141] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Isgum, "2d image classification for 3d anatomy localization: employing deep convolutional neural networks," in *Medical Imaging 2016: Image Processing*, vol. 9784. International Society for Optics and Photonics, 2016, p. 97841Y.
- [142] 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 *International conference on medical image computing and computer-assisted intervention*. Springer, 2013, pp. 246–253.
- [143] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2014, pp. 520–527.
- [144] H. R. Roth, Y. Wang, J. Yao, L. Lu, J. E. Burns, and R. M. Summers, "Deep convolutional networks for automated detection of posterior-element fractures on spine ct," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 97850P.
- [145] H. Chen, Q. Dou, D. Ni, J.-Z. Cheng, J. Qin, S. Li, and P.-A. Heng, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 507–514.
- [146] B. Kong, Y. Zhan, M. Shin, T. Denny, and S. Zhang, "Recognizing end-diastole and end-systole frames via deep temporal regression network," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 264–272.
- [147] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3d convolutional neural network for automatic detection of lung nodules in chest ct," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 1013409.

- [148] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [149] J. J. Näppi, T. Hironaka, D. Regge, and H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in ct colonography," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 97852B.
- [150] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. Lau, and C. C. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level cnn features from nonmedical domain," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 41–47, 2016.
- [151] 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," in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575. International Society for Optics and Photonics, 2018, p. 1057525.
- [152] A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, "Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection," *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.
- [153] A. Jesson, N. Guizard, S. H. Ghalehghe, D. Goblot, F. Soudan, and N. Chapados, "Cased: curriculum adaptive sampling for extreme data imbalance," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 639–646.
- [154] Z. Li, S. Zhang, J. Zhang, K. Huang, Y. Wang, and Y. Yu, "Mvp-net: Multi-view fpn with position-aware attention for deep universal lesion detection," in *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 13–21.
- [155] Y. Liu, Z. Zhou, S. Zhang, L. Luo, Q. Zhang, F. Zhang, X. Li, Y. Wang, and Y. Yu, "From unilateral to bilateral learning: Detecting mammogram masses with contrasted bilateral network," in *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 477–485.
- [156] A. Lisowska, A. O'Neil, V. Dilys, M. Daykin, E. Beveridge, K. Muir, S. McLaughlin, and I. Poole, "Context-aware convolutional neural networks for stroke sign detection in non-contrast ct scans," in *Annual Conference on Medical Image Understanding and Analysis*. Springer, 2017, pp. 494–505.
- [157] A. Lisowska, E. Beveridge, K. Muir, and I. Poole, "Thrombus detection in ct brain scans using a convolutional neural network." in *BIOIMAGING*, 2017, pp. 24–33.
- [158] 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," *Medical image analysis*, vol. 35, pp. 303–312, 2017.
- [159] N. Ghatwary, X. Ye, and M. Zolgharni, "Esophageal abnormality detection using densenet based faster r-cnn with gabor features," *IEEE Access*, vol. 7, pp. 84374–84385, 2019.
- [160] L. Fu, J. Ma, Y. Ren, Y. S. Han, and J. Zhao, "Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 101340A.
- [161] S. Hwang and H.-E. Kim, "Self-transfer learning for weakly supervised lesion localization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 239–246.
- [162] 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," *Medical image analysis*, vol. 35, pp. 18–31, 2017.
- [163] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [164] J. Zhang, A. Saha, Z. Zhu, and M. A. Mazurowski, "Hierarchical convolutional neural networks for segmentation of breast tumors in mri with application to radiogenomics," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 435–447, 2018.
- [165] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic *et al.*, "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.
- [166] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in ct imaging," in *Medical Imaging 2015: Image Processing*, vol. 9413. International Society for Optics and Photonics, 2015, p. 94131G.
- [167] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, "An ensemble of 2d convolutional neural networks for tumor segmentation," in *Scandinavian Conference on Image Analysis*. Springer, 2015, pp. 201–211.
- [168] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *computer vision and pattern recognition*, pp. 3431–3440, 2015.
- [169] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 728–736.
- [170] P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen, and D. Kong, "Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets," *International journal of computer assisted radiology and surgery*, vol. 12, no. 3, pp. 399–411, 2017.
- [171] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, and S. K. Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 487–495.
- [172] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi *et al.*, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 415–423.
- [173] K. Kamnitsas, C. Ledig, V. F. 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," *Medical image analysis*, vol. 36, pp. 61–78, 2017.
- [174] X.-Y. Zhou, M. Shen, C. Riga, G.-Z. Yang, and S.-L. Lee, "Focal fcn: towards small object segmentation with limited training data," *arXiv preprint arXiv:1711.01506*, 2017.
- [175] G. Zeng and G. Zheng, "Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 136–140.
- [176] X. Yang, L. Yu, L. Wu, Y. Wang, D. Ni, J. Qin, and P.-A. Heng, "Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [177] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in ct and mri images via recurrent neural contextual learning and direct loss function," *arXiv preprint arXiv:1707.04912*, 2017.
- [178] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [179] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Uinet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.
- [180] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.
- [181] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey *et al.*, "A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," *Frontiers in neuroscience*, vol. 13, 2019.

- [182] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual u-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.
- [183] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovy, and S. Stirenko, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*. Springer, 2018, pp. 638–647.
- [184] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.
- [185] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.
- [186] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, and G. Zheng, "3d u-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3d mr images," in *International workshop on machine learning in medical imaging*. Springer, 2017, pp. 274–282.
- [187] Z. Zhong, Y. Kim, L. Zhou, K. Plichta, B. Allen, J. Buatti, and X. Wu, "3d fully convolutional networks for co-segmentation of tumors on pet-ct images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 228–231.
- [188] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 507–515.
- [189] M. Zhao, L. Wang, J. Chen, D. Nie, Y. Cong, S. Ahmad, A. Ho, P. Yuan, S. H. Fung, H. H. Deng *et al.*, "Craniomaxillofacial bony structures segmentation from mri with deep-supervision adversarial learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 720–727.
- [190] M. Rezaei, K. Harmuth, W. Gierke, T. Kellermeier, M. Fischer, H. Yang, and C. Meinel, "A conditional adversarial network for semantic segmentation of brain tumor," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 241–252.
- [191] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert *et al.*, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 597–609.
- [192] M. Rezaei, H. Yang, and C. Meinel, "Whole heart and great vessel segmentation with context-aware of generative adversarial networks," in *Bildverarbeitung für die Medizin 2018*. Springer, 2018, pp. 353–358.
- [193] Y. Huo, Z. Xu, S. Bao, C. Bermudez, A. J. Plassard, J. Liu, Y. Yao, A. Assad, R. G. Abramson, and B. A. Landman, "Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks," in *Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, p. 1057409.
- [194] S. Izadi, Z. Mirikhraj, J. Kawahara, and G. Hamarneh, "Generative adversarial networks to segment skin lesions," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 881–884.
- [195] A. Lahiri, K. Ayush, P. Kumar 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 42–48.
- [196] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 146–157.
- [197] W. Zhu, X. Xiang, T. D. Tran, G. D. Hager, and X. Xie, "Adversarial deep structured nets for mass segmentation from mammograms," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 847–850.
- [198] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, and D. Ni, "Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 663–666.
- [199] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "Dcan: Deep contour-aware networks for object instance segmentation from histology images," *Medical image analysis*, vol. 36, pp. 135–146, 2017.
- [200] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken *et al.*, "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 516–524.
- [201] S. Valverde, M. Salem, M. Cabezas, D. Pareto, J. C. Vilanova, L. Ramíortorrenta, A. Rovira, J. Salví, A. Oliver, and X. Lladó, "One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks," *NeuroImage: Clinical*, vol. 21, p. 101638, 2019.
- [202] A. Kuzina, E. Egorov, and E. Burnaev, "Bayesian generative models for knowledge transfer in mri semantic segmentation problems," *Frontiers in neuroscience*, vol. 13, p. 844, 2019.
- [203] P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Isgum, "Deep learning for multi-task medical image segmentation in multiple modalities," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 478–486.
- [204] V. V. Valindria, N. Pawlowski, M. Rajchl, I. Lavdas, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, "Multi-modal learning from unpaired images: Application to multi-organ segmentation in ct and mri," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 547–556.
- [205] C. Chen, Q. Dou, H. Chen, and P.-A. Heng, "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 143–151.
- [206] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9242–9251.
- [207] J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, G. S. Mageras, J. O. Deasy, and H. Veeraraghavan, "Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 777–785.
- [208] W. Yan, Y. Wang, S. Gu, L. Huang, F. Yan, L. Xia, and Q. Tao, "The domain shift problem of medical image segmentation and vendor-adaptation by unet-gan," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 623–631.
- [209] C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 865–872.
- [210] J. Yang, N. C. Dvornek, F. Zhang, J. Chapiro, M. Lin, and J. S. Duncan, "Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 255–263.
- [211] J. Jue, H. Jason, T. Neelam, R. Andreas, B. L. Sean, D. O. Joseph, and V. Harini, "Integrating cross-modality hallucinated mri with ct to aid mediastinal lung tumor segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 221–229.
- [212] C. S. Perone, P. Ballester, R. C. Barros, and J. Cohen-Adad, "Unsupervised domain adaptation for medical imaging segmentation with self-ensembling," *NeuroImage*, vol. 194, pp. 1–11, 2019.
- [213] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Transactions on Medical Imaging*, 2020.

- [214] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv: Computer Vision and Pattern Recognition*, 2019.
- [215] L. Berger, H. Eoin, M. J. Cardoso, and S. Ourselin, "An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation," in *Annual Conference on Medical Image Understanding and Analysis*. Springer, 2018, pp. 277–286.
- [216] W. Wang, Y. Lu, B. Wu, T. Chen, D. Z. Chen, and J. Wu, "Deep active self-paced learning for accurate pulmonary nodule segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 723–731.
- [217] B. Wu, Z. Zhou, J. Wang, and Y. Wang, "Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1109–1113.
- [218] C. Chen, C. Biffi, G. Tarroni, S. E. Petersen, W. Bai, and D. Rueckert, "Learning shape priors for robust cardiac mr segmentation from multi-view images," in *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 523–531.
- [219] A. Hatamizadeh, D. Terzopoulos, and A. Myronenko, "End-to-end boundary aware networks for medical image segmentation," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 187–194.
- [220] K. Huang, H.-D. Cheng, Y. Zhang, B. Zhang, P. Xing, and C. Ning, "Medical knowledge constrained semantic breast ultrasound image segmentation," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1193–1198.
- [221] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac mri segmentation with strong anatomical guarantees," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 632–640.
- [222] A. J. Larrazabal, C. E. Martinez, and E. Ferrante, "Anatomical priors for image segmentation via post-processing with denoising autoencoders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 585–593.
- [223] Z. Mirikhraj and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 737–745.
- [224] A. BenTaieb and G. Hamarneh, "Topology aware fully convolutional networks for histology gland segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 460–468.
- [225] H. Ravishankar, R. Venkataramani, S. Thiruvenkadam, P. Sudhakar, and V. Vaideya, "Learning and incorporating shape models for semantic segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 203–211.
- [226] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y. Chen, R. Tong, and J. Wu, "Semi-supervised segmentation of liver using adversarial learning with deep atlas prior," in *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 148–156.
- [227] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. ORegan *et al.*, "Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation," *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2017.
- [228] C. Zotti, Z. Luo, A. Lalande, and P. Jodoin, "Convolutional neural network with shape prior applied to cardiac mri segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1119–1128, 2019.
- [229] B. Luo, J. Shen, S. Cheng, Y. Wang, and M. Pantic, "Shape constrained network for eye segmentation in the wild," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 1952–1960.
- [230] A. V. Dalca, J. Guttag, and M. R. Sabuncu, "Anatomical priors in convolutional networks for unsupervised biomedical segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9290–9299.
- [231] Y. He, G. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J. Dillenseger, P. Shao, S. Zhang *et al.*, "Dpa-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep prior anatomy," in *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 139–147.
- [232] S. Rezaei, A. Emami, H. Zarrabi, S. Rafiei, K. Najarian, N. Karimi, S. Samavi, and S. R. Soroushmehr, "Gland segmentation in histopathology images using deep networks and handcrafted features," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 1031–1034.
- [233] K. Kushibar, S. Valverde, S. González-Villà, J. Bernal, M. Cabezas, A. Oliver, and X. Lladó, "Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features," *Medical image analysis*, vol. 48, pp. 177–186, 2018.
- [234] H. Khan, P. M. Shah, M. A. Shah, S. ul Islam, and J. J. Rodrigues, "Cascading handcrafted features and convolutional neural network for iot-enabled brain tumor segmentation," *Computer Communications*, 2020.
- [235] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.
- [236] A. Atanov, A. Ashukha, K. Struminsky, D. Vetrov, and M. Welling, "The deep weight prior," 2018.
- [237] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in *Advances in Neural Information Processing Systems*, 2010, pp. 1189–1197.
- [238] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. Sharma, "Various types and management of breast cancer: an overview," *Journal of advanced pharmaceutical technology & research*, vol. 1, no. 2, p. 109, 2010.
- [239] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic mr image reconstruction," *IEEE transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2017.
- [240] A. Kofler, M. Haltmeier, C. Kolbitsch, M. Kachelrieß, and M. Dewey, "A u-nets cascade for sparse view computed tomography," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 91–99.
- [241] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic mr image reconstruction," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 280–290, 2018.
- [242] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo *et al.*, "Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.
- [243] H. B. Yedder, M. Shokoufi, B. Cardoen, F. Golnaraghi, and G. Hamarneh, "Limited-angle diffuse optical tomography image reconstruction using deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 66–74.
- [244] S. U. H. Dar, M. Yurt, M. Shahdloo, M. E. Ildiz, and T. Çukur, "Synergistic reconstruction and synthesis via generative adversarial networks for accelerated multi-contrast mri," *arXiv preprint arXiv:1805.10704*, 2018.
- [245] Y. Liu, D. Zhang, G. Lu, and W. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.
- [246] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Content-based brain tumor retrieval for mr images using transfer learning," *IEEE Access*, vol. 7, pp. 17 809–17 822, 2019.
- [247] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, "A sequential search-space shrinking using cnn transfer learning and a radon projection pool for medical image retrieval," *Expert Systems with Applications*, vol. 100, pp. 224–233, 2018.
- [248] J. Ahmad, M. Sajjad, I. Mehmood, and S. W. Baik, "Sinc: Saliency-injected neural codes for representation and efficient retrieval of medical radiographs," *PloS one*, vol. 12, no. 8, 2017.
- [249] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2015, pp. 2940–2943.
- [250] ———, "Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval,"

- in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 978510.
- [251] *On the automatic generation of medical imaging reports*, 2018.
- [252] G. Liu, T. H. Hsu, M. B. A. McDermott, W. Boag, W. Weng, P. Szolovits, and M. Ghassemi, "Clinically accurate chest x-ray report generation." *arXiv: Computer Vision and Pattern Recognition*, 2019.
- [253] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a very large-scale radiology database," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1090–1099.
- [254] Y. Xue, T. Xu, L. R. Long, Z. Xue, S. Antani, G. R. Thoma, and X. Huang, "Multimodal recurrent model with attention for automated radiology report generation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 457–466.
- [255] Y. Li, X. Liang, Z. Hu, and E. P. Xing, "Hybrid retrieval-generation reinforced agent for medical image report generation," in *Advances in neural information processing systems*, 2018, pp. 1530–1540.
- [256] J. Yuan, H. Liao, R. Luo, and J. Luo, "Automatic radiology report generation based on multi-view image fusion and medical concept enrichment," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 721–729.
- [257] C. Y. Li, X. Liang, Z. Hu, and E. P. Xing, "Knowledge-driven encode, retrieve, paraphrase for medical image report generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6666–6673.
- [258] W. Gale, L. Oakden-Rayner, G. Carneiro, A. P. Bradley, and L. J. Palmer, "Producing radiologist-quality reports for interpretable artificial intelligence," *arXiv preprint arXiv:1806.00340*, 2018.
- [259] Y. Hong and C. E. Kahn, "Content analysis of reporting templates and free-text radiology reports," *Journal of digital imaging*, vol. 26, no. 5, pp. 843–849, 2013.
- [260] Y. Zhang, X. Wang, Z. Xu, Q. Yu, A. Yuille, and D. Xu, "When radiology report generation meets knowledge graph," *arXiv preprint arXiv:2002.08277*, 2020.
- [261] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [262] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," *arXiv preprint arXiv:1711.03213*, 2017.
- [263] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.
- [264] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 136–144.
- [265] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2507–2516.
- [266] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7472–7481.
- [267] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3723–3732.
- [268] Y. Zhang, Y. Wei, P. Zhao, S. Niu, Q. Wu, M. Tan, and J. Huang, "Collaborative unsupervised domain adaptation for medical image diagnosis," *arXiv preprint arXiv:1911.07293*, 2019.