Health Data Science
A SCIENCE PARTNER JOURNAL

Review Article
Advances in Deep Learning-Based Medical Image Analysis

Xiaoqing Liu,[1] Kunlun Gao,[1] Bo Liu ,[1] Chengwei Pan ,[1] Kongming Liang,[1] Lifeng Yan ,[1] Jiechao Ma,[1] Fujin He,[1] Shu Zhang,[1] Siyuan Pan ,[2] and Yizhou Yu[1,3]

[1]*DeepWise AI Lab, Beijing, China*
[2]*Shanghai Jiaotong University, Shanghai, China*
[3]*The University of Hong Kong, Hong Kong*

Correspondence should be addressed to Yizhou Yu; yizhouy@acm.org

Importance. With the booming growth of artificial intelligence (AI), especially the recent advancements of deep learning, utilizing advanced deep learning-based methods for medical image analysis has become an active research area both in medical industry and academia. This paper reviewed the recent progress of deep learning research in medical image analysis and clinical applications. It also discussed the existing problems in the field and provided possible solutions and future directions. *Highlights*. This paper reviewed the advancement of convolutional neural network-based techniques in clinical applications. More specifically, state-of-the-art clinical applications include four major human body systems: the nervous system, the cardiovascular system, the digestive system, and the skeletal system. Overall, according to the best available evidence, deep learning models performed well in medical image analysis, but what cannot be ignored are the algorithms derived from small-scale medical datasets impeding the clinical applicability. Future direction could include federated learning, benchmark dataset collection, and utilizing domain subject knowledge as priors. *Conclusion*. Recent advanced deep learning technologies have achieved great success in medical image analysis with high accuracy, efficiency, stability, and scalability. Technological advancements that can alleviate the high demands on high-quality large-scale datasets could be one of the future developments in this area.

1. Introduction

With rapid developments of artificial intelligence (AI) technology, the use of AI technology to mine clinical data has become a major trend in medical industry [1]. Utilizing advanced AI algorithms for medical image analysis, one of the critical parts of clinical diagnosis and decision-making, has become an active research area both in industry and academia [2, 3]. Recent applications of deep leaning in medical image analysis involve various computer vision-related tasks such as classification, detection, segmentation, and registration. Among them, classification, detection, and segmentation are fundamental and most widely used tasks.

Although there exist a number of reviews on deep learning methods on medical image analysis [4–13], most of them emphasize either on general deep learning techniques or on specific clinical applications. The most comprehensive review paper is the work of Litjens et al. published in 2017 [12]. Deep learning is such a quickly evolving research field; numerous state-of-the-art works have been proposed since

then. In this paper, we review the latest developments in the field of medical image analysis with comprehensive and representative clinical applications.

We briefly review the common medical imaging modalities as well as technologies for various specific tasks in medical image analysis including classification, detection, segmentation, and registration. We also give more detailed clinical applications with respect to different types of diseases and discuss the existing problems in the field and provide possible solutions and future research directions.

2. AI Technologies in Medical Image Analysis

Different medical imaging modalities have their unique characteristics and different responses to human body structure and organ tissue and can be used in different clinical purposes. The commonly used image modalities for diagnostic analysis in clinic include projection imaging (such as X-ray imaging), computed tomography (CT), ultrasound imaging, and magnetic resonance imaging (MRI). MRI sequences
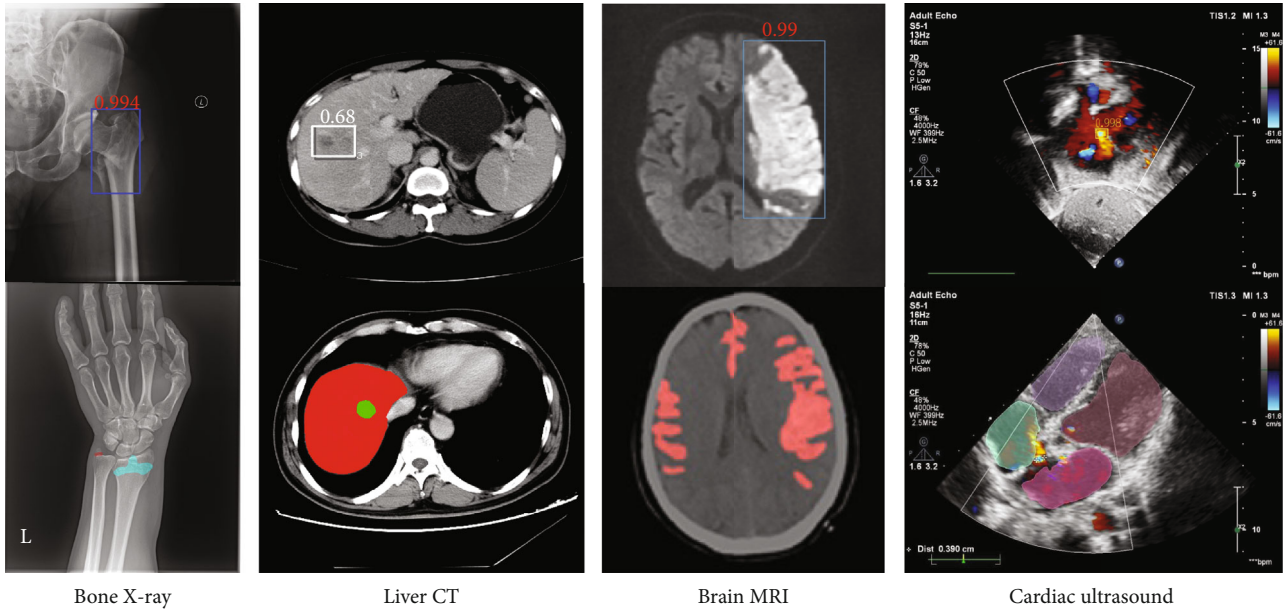
FIGURE 1: Examples of medical image modalities and their corresponding applications (original copy).

include T1, T1-w, T2, T2-w, diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC), and fluid attenuation inversion recovery (FLAIR). Figure 1 demonstrates a few examples of medical image modalities and their corresponding clinical applications.

2.1. Image Classification for Medical Image Analysis. As a fundamental task in computer vision, image classification plays an essential role in computer-aided diagnosis. A straightforward use of image classification for medical image analysis is to classify an input image or a series of images as either containing one (or a few) of predefined diseases or free of diseases (i.e., healthy case) [14, 15]. Typical clinical applications of image classification tasks include skin disease identification in dermatology [16, 17], eye disease recognition in ophthalmology (such as diabetic retinopathy [18, 19], glaucoma [20], and corneal diseases [21]). Classification of pathological images for various cancers such as breast cancer [22] and brain cancer [23] also belongs to this area.

Convolutional neural network (CNN) is the dominant classification framework for image analysis [24]. With the development of deep learning, the framework of CNN has continuously improved. AlexNet [25] was a pioneer convolutional neural network, which was composed of repeated convolutions, each followed by ReLU and max pooling operation with stride for downsampling. The proposed VGGNet [26] used 3×3 convolution kernels and 2×2 maximum pooling to simplify the structure of AlexNet and showed improved performance by simply increasing the number and depth of the network. Via combining and stacking 1×1, 3×3, and 5×5 convolution kernels and 3×3 pooling, the inception network [27] and its variants [28, 29] increased the width and the adaptability of the network. ResNet [30] and DenseNet [31] both used skip connections to relieve the gradient vanishing. SENet [32] proposed a squeeze-and-excitation module which enabled the model to

pay more attention to the most informative channel features. The family of EfficientNet [33] applied AUTOML and a compound scaling method to uniformly scale the width, depth, and resolution of the network in a principled way, resulting in improved accuracy and efficiency. Figure 2 demonstrates some of the commonly used CNN-based classification network architectures.

Besides the direct use for image classification, CNN-based networks can also be applied as the backbone models for other computer vision tasks, such as detection and segmentation.

To evaluate the algorithms of image classification, researchers use different evaluation metrics. Precision is the proportion of true positives in the identified images. The recall is the proportion of all positive samples in the test set that are correctly identified as positive samples. The accuracy rate is used to evaluate the global accuracy of a model. The F_1 score can be considered a harmonic average of the precision and the recall of the model, which takes both the precision and recall of the classification model into account. ROC (receiver operating characteristic) curve is usually used to evaluate the prediction effect of the binary classification model, and the kappa coefficient is a method to measure the accuracy of the model in multiclassification tasks.

$$
\begin{aligned}
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{n}, \\
F_1 &= 2 \bullet \frac{\text{Precision} \bullet \text{Recall}}{\text{Precision} + \text{Recall}}.
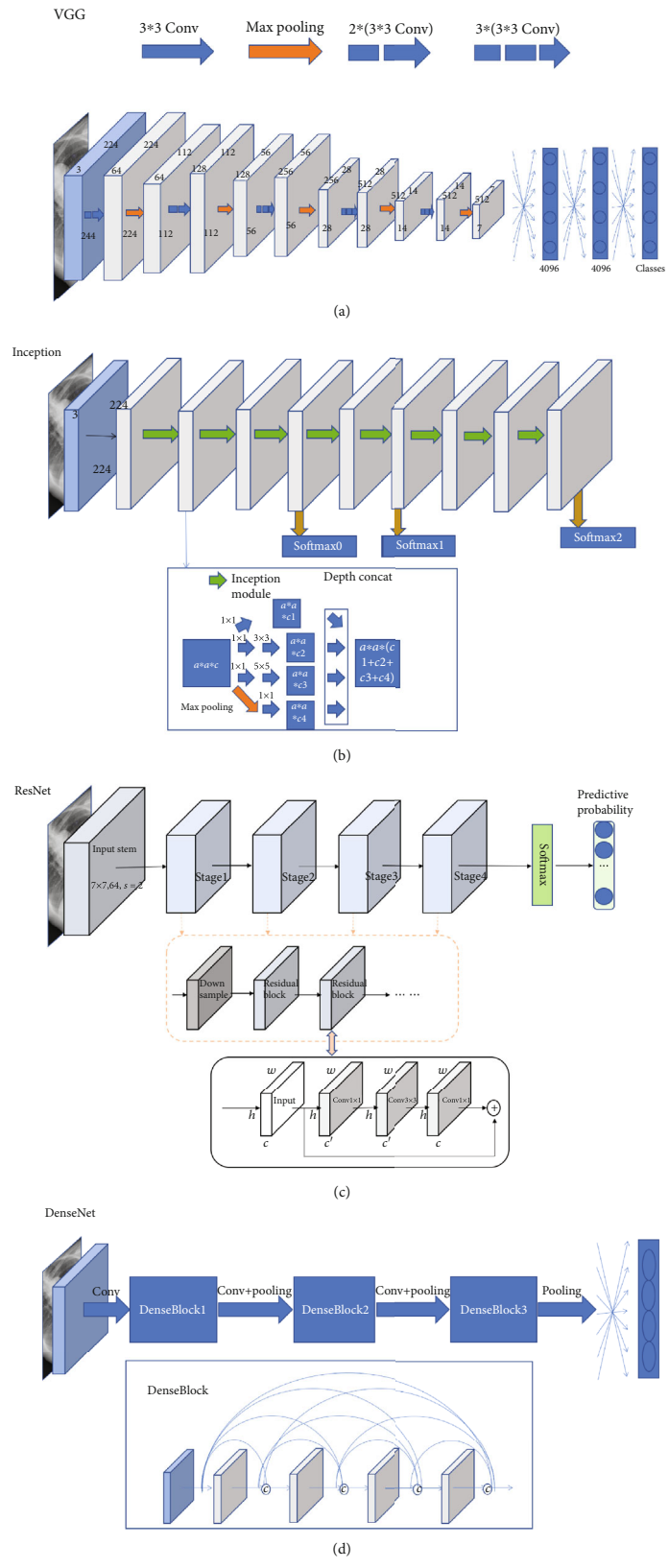\end{aligned}
\tag{1}
$$

FIGURE 2: Examples of CNN-based classification networks (original copy).

Here, we denote TP as true positives, FP as false positives, FN as false negatives, TN as true negatives, and n as the number of the testing samples.

2.2. Object Detection for Medical Image Analysis. Generally speaking, object detection algorithms include both identification and localization tasks. The identification task refers to

judging whether objects belonging to certain classes appear in regions of interest (ROIs) whereas the localization task refers to localizing the position of the object in the image. In medical image analysis, detection is commonly aimed at detecting the earliest signs of abnormality in patients. Exemplar clinical applications of detection tasks include lung nodule detection in chest CT or X-ray images [34, 35], lesion detection on CT images [36, 37], or mammograms [38].

Object detection algorithms can be categorized into two approaches, the anchor-based approach or anchor-free approach, where anchor-based algorithms can be further divided as single-stage algorithms or two/multistage algorithms. In general, single-stage algorithms are computationally efficient whereas two/multistage algorithms have better detection performance. The family of YOLO [39] and the single-shot multibox detector (SSD) [40] are two classic and widely used single-stage detectors with simple model architectures. As shown in Figures 3(a) and 3(b), both architectures are based on feed-forward convolutional networks producing a fixed number of bounding boxes and their corresponding scores for the presence of object instances of given classes in the boxes. A nonmaximum suppression step is applied to generate the final predictions. Different from YOLO which works on a single-scale feature map, the SSD utilizes multiscale feature maps, thereby producing better detection performance. Two-stage frameworks generate a set of ROIs and classify each of them through a network. The Faster-RCNN framework [41] and its descendant Mask-RCNN [42] are the most popular two-stage frameworks. As shown in Figure 3(c), the Faster/Mask-RCNN first generates object proposals through a region proposal network (RPN) and then classifies those generated proposals. The major difference between the Faster-RCNN and the Mask-RCNN is that the Mask-RCNN has an instance segmentation branch. Recently, there is a research trend on developing anchor-free algorithms. CornerNet [43] is one of the popular ones. As illustrated in Figure 3(d), CornerNet is a single convolutional neural network which eliminates the use of anchor boxes via utilizing paired key points where an object bounding box is indicated by the top-left corner and the bottom-right corner.

There are two main metrics to evaluate the performance of detection methods: the mean average precision (mAP) and the false positive per image (FP/I @ recall). mAP is used to calculate the average of all average precisions (APs) of all categories. FP/I @ recall rate is a measure of false positive (FP) of each image under a certain recall rate which takes into account the balance between false positives and the missing rate.

2.3. Segmentation for Medical Image Analysis. Image segmentation is a pixel labeling problem, which partitions an image into regions with similar properties. For medical image analysis, segmentation is aimed at determining the contour of an organ or anatomical structure in images. Segmentation tasks in clinical applications include segmenting a variety of organs, organ structures (such as the whole heart [44] and pancreas [45]), tumors, and lesions (such as the liver and liver tumor [46]) across different medical imaging modalities.

Since the fully convolutional neural network (FCN) [47] has been proposed, image segmentation has achieved great success. FCN was the first CNN which turned the classification task to dense segmentation task with in-network upsampling and a pixelwise loss. Through a skip architecture, it combined coarse, semantic, and local information to dense prediction. Medical image segmentation methods can be divided into two categories: the 2D methods and the 3D methods according to the input data dimension. The U-Net architecture [48] is the most popular FCN for medical image segmentation. As shown in Figure 4, U-Net consists of a contracting path (the downsample side) and an expansive path (the upsample side). The contracting path follows the typical CNN architecture. It consists of the repeated application of convolutions, each followed by ReLU and max pooling operation with stride for downsampling. At each downsampling step, it also doubles the number of feature channels. Each step in the expansive path is composed of feature map upsampling followed by deconvolution that halves the number of feature channels; a concatenation with the correspondingly cropped feature map from the contracting path is also applied. Variants of U-Net-based architectures have been proposed. Isensee et al. [49] proposed a general framework called nnU-Net (No new U-Net) for medical image segmentation, which applied a dataset fingerprint (representing the key properties of the dataset) and a pipeline fingerprint (representing the key design of the algorithms) to systematically optimize the segmentation task via formulating a set of heuristic rules from domain knowledge. The nnU-Net achieved state-of-the-art performance on 19 different datasets with 49 segmentation tasks across a variety of organs, organ structures, tumors, and lesions in a number of imaging modalities (such as CT, MRI).

Dice similarity coefficient and intersection over union (IOU) are the two major evaluation metrics to evaluate the performance of segmentation methods, and they are defined as follows:

$$
\begin{aligned}
\text{Dice} &= \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}, \\
\text{IOU} &= \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}},
\end{aligned}
\tag{2}
$$

where TP, FP, and FN denote true positive, false positive, and false negative, respectively.

2.4. Image Registration for Medical Image Analysis. Image registration, also known as image warping or image fusion, is a process of aligning two or more images. The goal of medical image registration is aimed at establishing optimal correspondence within images acquired at different times (for longitudinal studies), by different imaging modalities (such as CT, MRI), across different patients (for intersubject studies), or from distinct viewpoints. Image registration plays a crucial preprocessing step in many clinical applications including computer-aided intervention and treatment planning [50], image-guided/assisted surgery or simulation [51], and fusion of anatomical images (e.g.,
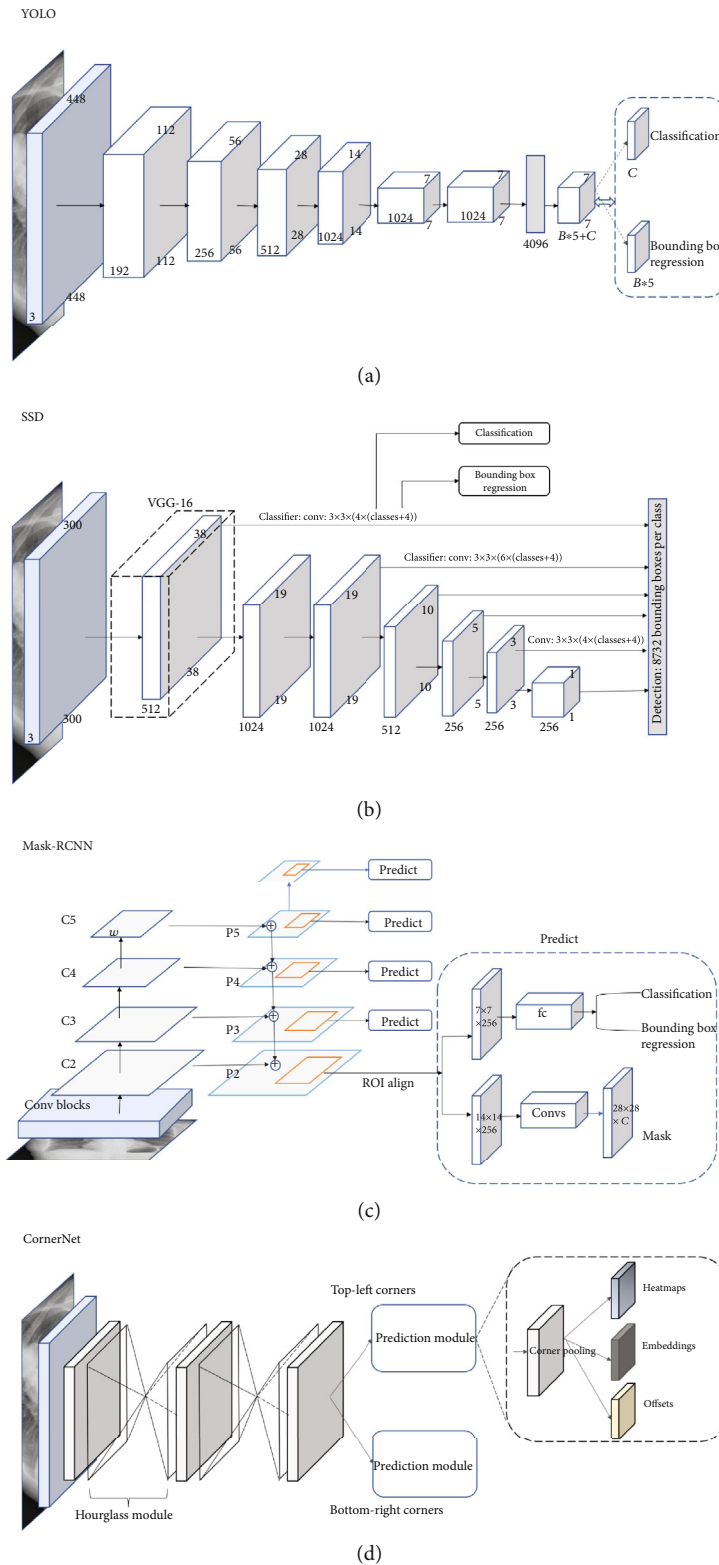
YOLO

SSD

Mask-RCNN

CornerNet

Figure 3: Examples of object detection frameworks (original copy).

CT or MRI images) with functional images (such as positron emission tomography, single-photon emission computed tomography, or functional MRI) for disease diagnosis and monitoring [52].

Depending on different points of view, image registration methodologies can be categorized differently. For instance, image registration methods can be classified as monomodal or multimodal based on imaging modalities involved. From
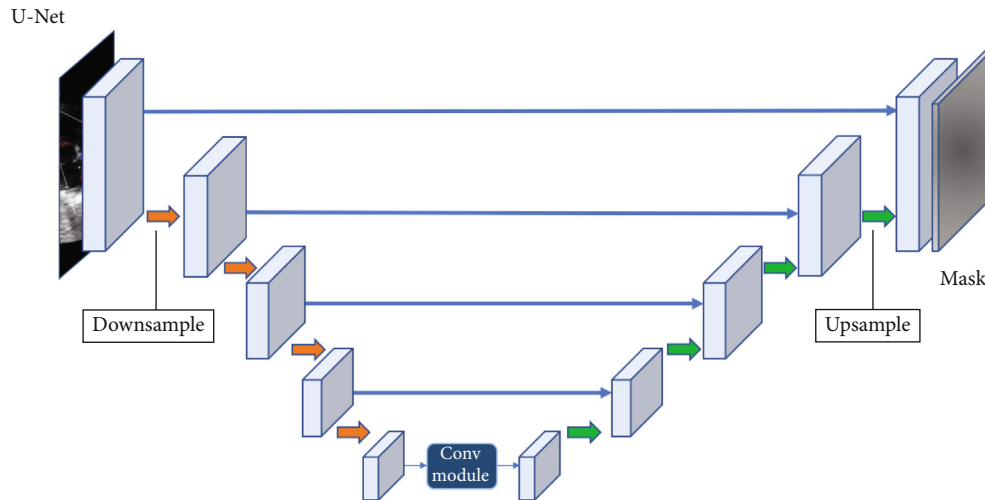
FIGURE 4: Examples of image segmentation frameworks (original copy).

the nature of geometric transformation, methods can also be categorized as rigid or nonrigid classes. By data dimensionality, registration methods can be classified as 2D/2D, 3D/3D, 2D/3D, etc., and from similarity measure point of view, registration can be categorized as feature-based or intensity-based groups. Previously, image registration has been extensively explored as an optimization problem whose aim is to search the best geometric transformation iteratively through optimizing a similarity measure such as sum of squared differences (SSD), mutual information (MI), and cross-correlation (CC). Ever since the beginning of the deep learning renaissance, various deep learning-based registration methods have been proposed and achieved the state-of-the-art performance [53].

Yang et al. [54] proposed a fully supervised deep learning method to align 2D/3D intersubject brain MR in a single step via a U-Net-like FCN. Jun et al. [55] also applied a CNN to perform deformable registration of abdominal MR images to compensate respiration deformation. Despite the success of supervised learning-based methods, the nature of acquisition of reliable ground truth remains significantly challenging. Weakly supervised and/or unsupervised methods can effectively alleviate the issue of lack of training datasets with ground truth. Li and Fan [56] trained an FCN to perform deformable 3D brain MR images using self-supervision. Inspired by the spatial transfer network (STN) [57], Kuang et al. [58] applied a STN-based CNN to perform deformable registration of MRI T1-W brain volumes.

Recently, Generative Adversarial Network- (GAN-) and Reinforcement Learning- (RL-) based methods have also motivated great attentions. Yan et al. [59] performed a rigid registration of 3D MR and ultrasound images. In their work, the generator was trained to estimate rigid transformation where the discriminator was used to distinguish between images that were aligned by ground-truth transformations or by predicted ones. Kreb et al. [60] applied a RL method to perform the nonrigid deformable registration of 2D/3D prostate MRI images where they utilized a low-resolution

deformation model for registration and a fuzzy action control to influence the action selection.

For performance evaluation, Dice coefficient and mean square error (MSE) are two major evaluation metrics. Target registration error (TRE) can also be applied if landmark correspondence can be acquired.

3. Clinical Applications

In this section, we review state-of-the-art clinical applications in four major systems of the human body involving the nervous system, the cardiovascular system, the digestive system, and the skeletal system. To be more specific, AI algorithms on medical image diagnostic analysis for the following representative diseases including brain diseases, cardiac diseases, and liver diseases, as well as orthopedic trauma, are discussed.

3.1. Brain. In this section, we discuss three most critical brain diseases, namely, stroke, intracranial hemorrhage, and intracranial aneurysm.

3.1.1. Stroke. Stroke is one of the leading causes of death and disability worldwide and imposes an enormous burden for health care systems [61]. Accurate and automatic segmentation of stroke lesions can provide insightful information for neurologists.

Recent studies have presented tremendous ability in stroke lesion segmentation. Chen et al. [62] used DWI images as input to segment acute ischemic lesions and achieved an average Dice score of 0.67. Clèrigues et al. [63] proposed a deep learning methodology for acute and subacute stroke lesion segmentation using multimodal MRI images, and the Dice scores of the two segmentation tasks were 0.84 and 0.59, respectively. Liu et al. [64] used a U-shaped network (Res-CNN) to automatically segment acute ischemic stroke lesions from multimodality MRIs, and the average Dice coefficient was 0.742. Zhao et al. [65] proposed a semisupervised learning method using the weakly labeled subjects to detect

the suspicious acute ischemic stroke lesions and achieved a mean Dice coefficient of 0.642. Compared to using MRI, a 2D patch-based deep learning approach was proposed to segment the acute stroke lesion core from CT perfusion images [66], and the average Dice coefficient was 0.49.

3.1.2. Intracranial Hemorrhage. Recent studies have also shown great promise in automated detection of intracranial hemorrhage and its subtypes. Chilamkurthy et al. [67] achieved an AUC of 0.92 for detecting intracranial hemorrhage based on a publicly available dataset called CQ500 consisting of 313,318 head CT scans from 20 centers. They use the original clinical radiology report and consensus of three independent radiologists as the gold standard to evaluate their method. Ye et al. [68] proposed a novel three-dimensional (3D) joint convolutional and recurrent neural network (CNN-RNN) for the detection of intracranial hemorrhage. They developed and evaluated their method on a total of 2,836 subjects (ICH/normal, 1,836/1,000) from three institutions. Their algorithm achieved an AUC of 0.94 for intraparenchymal, 0.93 for intraventricular, 0.96 for subdural, 0.94 for extradural, and 0.89 for subarachnoid for the subtype classification task. Ker et al. [69] proposed to apply an image thresholding in the preprocessing step to improve the classification F1 score from 0.919 to 0.952 for their 3D CNN-based acute brain hemorrhage diagnosis. Singh et al. [70] also proposed an image preprocessing method to improve the 3D CNN-based acute brain hemorrhage detection via normalizing 3D volumetric scans using intensity profile. Their experimental results demonstrated the best F1 scores of 0.96, 0.93, 0.98, and 0.99, respectively, for four types of acute brain hemorrhages (i.e., subarachnoid, intraparenchymal, subdural, and intraventricular) on the CQ500 dataset [67].

3.1.3. Intracranial Aneurysm. Intracranial aneurysm is a common life-threatening disease usually caused by trauma, vascular disease, or congenital development with a prevalence of 3.2% in the population [71]. Rupture of an intracranial aneurysm is a serious incident with high mortality and morbidity rates [72]. As such, the accurate detection of intracranial aneurysms is also important. Computed tomography angiography (CTA) and magnetic resonance angiography (MRA) are noninvasive methods and widely used for the diagnosis and presurgical planning of intracranial aneurysms [73]. Nakao et al. [74] used a CNN classifier to predict whether each voxel was inside or outside aneurysms by inputting MIP images generated from a volume of interest around the voxel. They detected 94.2% of aneurysms with 2.9 false positives per case. Stember et al. [75] employed a CNN based on U-Net architecture to detect aneurysms on MIP images and then to derive aneurysm size. Sichtermann et al. [76] established a system based on an open-source neural network named DeepMedic for the detection of intracranial aneurysms from 3D TOF-MRA data. Ueda et al. [77] adopted ResNet for the detection of aneurysms from MRA images and reached a sensitivity of 91% and 93% for the internal and external test datasets, respectively. Allison et al. [78] proposed a segmentation model called HeadXNet to

segment aneurysms on CTA images. Recently, Shi et al. [79] proposed a 3D patch-based deep learning model for detecting intracranial aneurysm in CTA images. The proposed model utilized both spatial and channel attentions within a residual-based encoder-decoder architecture. Experimental results on multicohorta studies proofed the clinical applicability.

3.2. Cardiac/Heart. Echocardiography, CT, and MRI are commonly used medical imaging modalities for noninvasive assessment of the function and structure of the cardiovascular system. Automatic analysis of images from the above modalities can help physicians study the structure and function of heart muscle, find the cause of a patient's heart failure, identify potential tissue damages, and so on.

3.2.1. Identification of Standard Scan Planes. Identification of standard scan planes is an important step in clinical echocardiogram interpretation since many cardiac diseases are diagnosed based on standard scan planes. Zhang et al. [80] built a fully automated, scalable, analysis pipeline for echocardiogram interpretation, including view identification, cardiac chamber segmentation, quantification of structure and function, and disease detection. They trained a 13-layer CNN on 14,035 echocardiograms spanning on a 10-year period for identification of 23 viewpoints and trained a cardiac chamber segmentation network across 5 common standard scan planes. Then, the segmentation output was used to quantify chamber volumes and LV mass, determine ejection fraction, and facilitate automated determination of longitudinal strain through speckle tracking. Howard et al. [81] trained a two-stream network on over 8,000 echocardiographic videos for 14 different scan plane identification, which contained a time-distributed network to get spatial feature and a temporal network to get optical flow feature of moving objects between frames. Experiments showed that the proposed method can halve the error rate for video scan plane classification, and the types of misclassification the method made were very similar to differences of opinion between human experts.

3.2.2. Segmentation of Cardiac Structures. Vigneault et al. [82] presented a novel deep CNN architecture called Ω-Net for fully automatic whole-heart segmentation. The network was trained end to end from scratch to segment five foreground classes (the four cardiac chambers plus the LV myocardium) in three views (SA, 4C, and 2C) with data acquired from both 1.5-T and 3-T magnets as part of a multicenter trial involving 10 institutions. Xiong et al. [83] developed a 16-layer CNN model called AtriaNet to automatically segment the left atrial (LA) epicardium and endocardium. AtriaNet consists of a multiscaled dual-pathway architecture with two different sizes of input patches centered on the same region that captures both the local arterial tissue and geometry and the global positional information of LA. Benchmarking experiments showed that AtriaNet had outperformed the state-of-the-art CNNs, with a Dice score of 0.940 and 0.942 for the LA epicardium and endocardium at the time. Moccia et al. [84] modified and trained the ENet, a fully

convolutional neural network, to provide scar-tissue segmentation in the left ventricle. Bai et al. [85] proposed an image sequence segmentation algorithm by combining a fully convolutional network with a recurrent neural network, which incorporated both spatial and temporal information into the segmentation task. The proposed method achieved an average Dice metric of 0.960 for the ascending aorta and 0.953 for the descending aorta. Morris et al. [86] developed a novel pipeline that paired MRI/CT data that were placed into separate image channels to train a 3D neural network using the entire 3D image for sensitive cardiac substructure segmentation. The paired MR/CT multichannel data inputs yielded robust segmentations on noncontrast CT inputs, and data augmentation and 3D Conditional Random Field (CRF) postprocessing improved deep learning contour agreement with ground truth.

3.2.3. Coronary Artery Segmentation.

Shen et al. [87] proposed a joint framework for coronary CTA segmentation based on deep learning and traditional-level set method. A 3D FCN was used to learn the 3D semantic features of coronary arteries. Moreover, an attention gate was added to the entire network, aiming to enhance the vessels and suppress irrelevant regions. The output of 3D FCN with the attention gate was optimized by the level set to smooth the boundary to better fit the ground-truth segmentation. The coronary CTA dataset used in this work consisted of 11,200 CTA images from 70 groups of patients, of which 20 groups of patients were used as a test set. The proposed algorithm provided significantly better segmentation results than vanilla 3D FCN intuitively and quantitatively. He et al. [88] developed a novel blood vessel centerline extraction framework utilizing a hybrid representation learning approach. The main idea was to use CNNs to learn local appearances of vessels in image crops while using another point-cloud network to learn the global geometry of vessels in the entire image. This combination resulted in an efficient, fully automatic, and template-free approach to centerline extraction from 3D images. The proposed approach was validated on CTA datasets and demonstrated its superior performance compared to both traditional and CNN-based baselines.

3.2.4. Coronary Artery Calcium and Plaque Detection.

Zhang et al. [89] established an end-to-end learning framework for artery-specific coronary calcification identification in noncontrast cardiac CT, which can directly yield accurate results based on given CT scans in the testing process. In this framework, the intraslice calcification features were collected by a 2D U-DenseNet, which was the combination of DenseNet and U-Net. While those lesions spanned multiple adjacent slices, authors performed 3D U-Net extraction to the interslice calcification features, and the joint semantic features of 2D and 3D modules were beneficial to artery-specific calcification identification. The proposed method was validated on 169 noncontrast cardiac CT exams collected from two centers by cross-validation and achieved a sensitivity of 0.905, a PPV of 0.966 for calcification number, a sensitivity of 0.933, a PPV of 0.960, and a F_1 score of 0.946 for calcification volume, respectively. Liu et al. [90] proposed a vessel-focused

3D convolutional network for automatic segmentation of artery plaque including three subtypes: calcified plaques, noncalcified plaques, and mixed calcified plaques. They first extracted the coronary arteries from the CT volumes and then reformed the artery segments into straightened volumes. Finally, they employed a 3D vessel-focused convolutional neural network for plaque segmentation. This proposed method was trained and tested on a dataset of multiphase CCTA volumes of 25 patients. The proposed method achieved Dice scores of 0.83, 0.73, and 0.68 for calcified plaques, noncalcified plaques, and mixed calcified plaques, respectively, on the test set, which showed a potential value for clinical application.

3.3. Liver.

CT and MRI are widely used for the early detection, diagnosis, and treatment of liver diseases. Automatic segmentation of the liver and/or liver lesion with CT or MRI is of great importance in radiotherapy planning, liver transplantation planning, and so on.

3.3.1. Liver Lesion Detection and Segmentation.

Vorontsov et al. used deep CNNs to detect and segment liver tumors [91]. For lesion sizes smaller than 10 mm ($n = 30$), 10–20 mm ($n = 35$), and larger than 20 mm ($n = 40$), the detection sensitivities of the method were 10%, 71%, and 85%; positive predictive values were 25%, 83%, and 94%; and dice similarity coefficients were 0.14, 0.53, and 0.68. Wang et al. proposed an attention network by using an extra network to gather information from continuous slices for lesion segmentation [92]. This method had a Dice per case score of 74.1% on LiTS test dataset. In order to improve the performance on small lesions, modified U-Net (mU-Net) is proposed by Seo et al. which obtained a Dice score of 89.72% on validation set for liver tumor segmentation [93]. An edge enhanced network was proposed by Tang et al. [94] for liver tumor segmentation with a Dice per case score of 74.8% on LiTS test dataset.

3.3.2. Liver Lesion Classification.

Unlike liver lesion segmentation or detection, there are few works about lesion classification, as there is no public dataset about lesion classification, and it is difficult to collect enough data. A liver tumor classification system trained with 1,210 patients and validated in 201 patients based on deep learning was proposed by Zhen et al. [95]. The system can distinguish malignant from benign liver tumors with an AUC score of 94.6% using only unenhanced images, and the performance can be improved a lot with clinical information.

3.3.3. Liver Fibrosis Staging.

Liver fibrosis staging is important for the prevention and treatment of chronic liver disease. Although the amount of the works based on deep learning for liver fibrosis staging is few, these methods have shown their capability for this task. Liu et al. proposed a method using CNNs and SVM to classify the capsules on ultrasound images to get the stage score, and this method had a classification AUC score of 97.03% [96]. Yasaka et al. proposed two deep CNNs models to obtain stage scores, respectively, from CT [97] and MRI [98] images, achieving AUC scores of 0.73-0.76 and 0.84-0.85, respectively. Choi et al. trained a

TABLE 1: Publicly available Benchmark datasets.

Dataset name	Organ/modalities	Image size	No. classes	No. of cases	Tasks	Resources
LIDC-IDRI	Lung/CT	$133 \times 512 \times 512$	3	1018	Lung nodules	[114]
LUNA	Lung/CT	$133 \times 512 \times 512$	1	888	Lung nodules	[115]
DDSM	Breast/mammography	—	3	2,500	Breast mass	[116]
DeepLesion	Diversity CT	—	3+	4427	Lung nodules, liver tumors, lymph nodes	[117]
LiTS	Liver/CT	$432 \times 512 \times 512$	2	131	Liver, liver tumors	[118]
Brain tumor	Brain/MRI	$138 \times 169 \times 138$	3	484	Edema, tumor, necrosis	[119]
Heart	Heart/MRI	$115 \times 320 \times 232$	1	20	Left ventricle	[119]
Prostate	Prostate/MRI	$20 \times 320 \times 319$	2	32	Peripheral and transition zone	[119]
Pancreas	Pancreas/CT	$93 \times 512 \times 512$	2	282	Pancreas, pancreas cancer	[119]
Spleen	Spleen/CT	$90 \times 512 \times 512$	1	41	Spleen	[119]
Colon	Colon/CT	$95 \times 512 \times 512$	1	126	Colon cancer	[119]

model based on deep learning using 7,491 patients and validated on 891 patients, and the AUC score on the validation dataset was 0.95-0.97 [99]. Recently, a model based on multimodal ultrasound images received an AUC score of 0.93-0.95 [100] which used transfer learning to improve the classification performance.

3.3.4. Other Liver Disease. Prediction of microvascular invasion (MVI) before surgery is valuable for liver cancer patients' treatment planning since MVI is an adverse prognostic factor for these patients [101]. Men et al. proposed 3D CNNs with LSTM to predict MVI on enhanced MRI images receiving an AUC score of 89% [102]. Jiang et al. [103] also reported a 3D CNN-based one with enhanced CT images achieving an AUC score of 90.6%.

3.4. Bone. Bone fracture, also called orthopedic trauma, is a relatively common disease. Bone fracture recognition in X-ray images has become a promising research direction since 2017 with the development of deep learning technology. In general, there are two main approaches for bone fracture recognition, namely, the classification-based approach and the object detection-based approach.

3.4.1. Classification-Based Approach. For the classification-based approach, researchers usually use the labels of "no fracture" and "fracture" for the whole image. The pioneer and dedicated work of the classification pipeline was from Olczak et al. [104]. By adopting the VGGNet as the backbone of the classification pipeline, they trained the model on 256,000 well-labeled images of the wrists, hands, and ankles for recognizing fractures. With a large amount of validating data, the model set a strong and credible baseline of the accuracy of 83%. Urakawa et al. [105] used the same network architecture as Olczak et al.'s in classifying intertrochanteric hip fractures on 3,346 radiographs. The results have shown a 95.5% accuracy whereas an accuracy of orthopedic surgeons was reported at 92.2%. Gale et al. [106] extracted 53,000 clinical X-rays to get an area under the ROC curve of 0.994 whereas Krogue et al. [107] labeled 3,034 images to get an area under the

curve of 0.973. They both applied DenseNet into the classification task on hip fracture radiographs.

3.4.2. Object Detection-Based Approach. The object detection-based approach is aimed at localizing the fracture locations in the images. Gan et al. [108] trained a Faster R-CNN model to locate the area of wrist fracture; then, they sent the ROI to an inception framework for classification. The AUC score achieved 0.96 overpassing radiologists' performance by 9% in accuracy on a set of 2,340 anteroposterior wrist radiographs. Thian et al. [109] employed the same Faster R-CNN architecture and also ran the model on wrist radiographs with a larger volume of the dataset of 7,356 images. The result had an indistinctive AUC score of 0.957. Still on wrist radiographs, using the idea of semantic segmentation, Lindsey et al. [110] adopted an extension of U-Net to predict a heat map probability of fractures for each image pixel. Even using 135,409 wrist radiographs, the article only reported an average clinician sensitivity of 91.5% and specificity of 93.9% aided with a trained model, which seemed to be inferior to the above research. Wu et al. [111] proposed an end-to-end multidomain facture detection network which treated each body part as a domain. The proposed network was composed of two subnetworks, namely, a domain classification network for predicting the domain type of an image and a fracture detection network for detecting fractures on X-ray images of different domains. By constructing feature enhancement modules and multifeature-enhanced r-CNN, the proposed network extracted more representative features for each domain. Experimental results on real-clinical data demonstrated the effectiveness with the best F-score on all the domains over existing Faster R-CNN-based state-of-the-art methods. Recently, Wu et al. [112] proposed a novel feature ambiguity mitigation model to improve the bone fracture detection on X-ray radiographs. A total of 9,040 radiographic images for various body parts including the hand, wrist, elbow, shoulder, pelvic, knee, ankle, and foot were studied. Experimental results demonstrated performance improvements in all body parts.

4. Challenges and Future Directions

Although deep learning models have achieved great success in medical image analysis, small-scale medical datasets are still the main bottleneck in this field. Inspired by the idea of transfer learning technique, one possible way is to do domain transfer which adapts a model trained on natural images to medical image applications or from one image modality to another. Another possible way is to apply federated learning [113] by which training can be performed among multiple data centers collaboratively. In addition, researchers have also begun to collect benchmark datasets for various medical image analysis purposes. Table 1 summarized examples of the publicly available datasets.

Class imbalance is another major problem of medical image analysis. A number of researches on novel loss function design, such as focal loss [120], grading loss [121], contrastive loss [122], and triplet loss [123], have been proposed to tackle this problem. Making use of domain subject knowledge is another direction. For instance, Jiménez-Sánchez et al. [124] proposed a curriculum learning method to classify proximal femoral fractures in X-ray images, whose core idea is to control the sampling weight of samples in the training process based on a priori knowledge. Chen et al. [125] also proposed a novel pelvic fracture detection framework based on bilaterally symmetric structure assumption.

5. Conclusion

The rise of advanced deep learning methods has enabled great success in medical image analysis with high accuracy, efficiency, stability, and scalability. In this paper, we reviewed the recent progress of CNN-based deep learning techniques in clinical applications including image classification, object detection, segmentation, and registration. More detailed image analysis-based diagnostic applications in four major systems of the human body involving the nervous system, the cardiovascular system, the digestive system, and the skeletal system were reviewed. To be more specific, state-of-the-art works for different diseases including brain diseases, cardiac diseases, and liver diseases, as well as orthopedic trauma, are discussed. This paper also described the existing problems in the field and provided possible solutions and future research directions.

Conflicts of Interest

The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. The authors have no conflicts of interest to declare.

Authors' Contributions

Y. Yu and X. Liu conceptualized, organized, and revised the manuscript. X. Liu contributed to all aspects of the preparation of the manuscript. K. Gao, B. Liu, C. Pan, K. Liang, L. Yan, J. Ma, F. He, S. Pan, and S. Zhang were involved in the writing of the manuscript. All authors con-tributed to this paper. Xiaoqing Liu, Kunlun Gao, Bo Liu, Chengwei Pan, Kongming Liang, and Lifeng Yan contributed equally to this work.

References

[1] H. T. Shen, X. Zhu, Z. Zhang et al., "Heterogeneous data fusion for predicting mild cognitive impairment conversion," *Information Fusion*, vol. 66, pp. 54–63, 2021.

[2] Y. Zhu, M. Kim, X. Zhu, D. Kaufer, and G. Wu, "Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data," *Medical Image Analysis*, vol. 67, p. 101825, 2021.

[3] X. Zhu, B. Song, F. Shi et al., "Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan," *Medical Image Analysis*, vol. 67, p. 101824, 2021.

[4] S. Mitra and B. Uma Shankar, "Medical image analysis for cancer management in natural computing framework," *Information Sciences*, vol. 306, pp. 111–131, 2015.

[5] E. Miranda, M. Aryuni, and E. Irwansyah, "A survey of medical image classification techniques," in *2016 International Conference on Information Management and Technology (ICIMTech)*, Bandung, Indonesia, 2016.

[6] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[7] K. Suzuki, "Survey of deep learning applications to medical image analysis," *Medical Imaging Technology*, vol. 35, pp. 212–226, 2017.

[8] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*, Academic Press, 2017.

[9] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[10] S. Liu, Y. Wang, X. Yang et al., "Deep learning in medical ultrasound analysis: a review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.

[11] A. Maier, C. Syben, and T. Lasser, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, pp. 86–101, 2019.

[12] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[13] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D deep learning on medical images: a review," *Sensors*, vol. 20, no. 18, article 5097, 2020.

[14] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, p. 113, 2019.

[15] C. Wang, F. Zhang, Y. Yu, and Y. Wang, "BR-GAN: bilateral residual generating adversarial network for mammogram classification," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020*, A. L.

Martel, Ed., vol. 12262 of Lecture Notes in Computer Science, Springer, Cham, 2020.

[16] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[17] H. Wu, H. Yin, H. Chen et al., "A deep learning, image based approach for automated diagnosis for inflammatory skin diseases," *Annals of Translational Medicine*, vol. 8, no. 9, p. 581, 2020.

[18] D. S. W. Ting, C. Y. L. Cheung, G. Lim et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic Populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[19] V. Gulshan, L. Peng, M. Coram 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.

[20] X. Bai, S. I. Niwas, W. Lin et al., "Learning ECOC code matrix for multiclass classification with application to glaucoma diagnosis," *Journal of Medical Systems*, vol. 40, no. 4, 2016.

[21] H. Gu, Y. Guo, L. Gu et al., "Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs," *Scientific Reports*, vol. 10, no. 1, p. 17851, 2020.

[22] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1868–1873, Banff, AB, Canada, 2017.

[23] J. Ker, Y. Bai, H. Y. Lee, J. Rao, and L. Wang, "Automated brain histology classification using machine learning," *Journal of Clinical Neuroscience*, vol. 66, pp. 239–245, 2019.

[24] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642–3649, Providence, RI, USA, 2012.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[26] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, Computer, International Conference on Learning Representations, San Diego, CA, USA, 2014.

[27] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, https://arxiv.org/abs/1512.00567.

[29] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2016, https://arxiv.org/abs/1602.07261.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, USA, 2018.

[33] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105–6114, Long Beach, California, USA, 2019.

[34] S.-C. B. Lo, S.-L. A. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995.

[35] J. Liu, G. Zhao, F. Yu, M. Zhang, Y. Wang, and Y. Yizhou, "Align, attend and locate: chest x-ray diagnosis via contrast induced attention network with limited supervision," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10632–10641, Seoul, Korea, 2019.

[36] Z. Li, S. Zhang, J. Zhang, K. Huang, Y. Wang, and Y. Yizhou, "MVP Net: multi-view FPN with position-aware attention for deep universal lesion detection," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, D. Shen, Ed., vol. 11769 of Lecture Notes in Computer Science, Springer, Cham, 2019.

[37] S. Zhang, J. Xu, Y.-C. Chen et al., "Revisiting 3D context modeling with supervised pre-training for universal lesion detection in CT slices," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020*, A. L. Martel, Ed., vol. 12264 of Lecture Notes in Computer Science, Springer, Cham, 2020.

[38] Y. Liu, F. Zhang, Q. Zhang, S. Wang, Y. Wang, and Y. Yizhou, "Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.

[39] 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*, pp. 779–788, 2016.

[40] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot Multi-Box detector," in *Computer Vision – ECCV 2016. ECCV 2016*, vol. 9905 of Lecture Notes in Computer Science, Springer, Cham.

[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[42] G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961–2969, 2017.

[43] H. Law, "CornerNet: detecting objects as paired keypoints," in *Computer Vision – ECCV 2018. ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., vol. 11218 of Lecture Notes in Computer Science, pp. 765–781, Springer, Cham, 2018.

[44] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, pp. 23421–23429, 2019.

[45] C. Fang, G. Li, C. Pan, Y. Li, and Y. Yizhou, "Globally guided progressive fusion network for 3D pancreas segmentation," in *Medical Image Computing and Computer Assisted*

Intervention – MICCAI 2019. MICCAI 2019, D. Shen, Ed., vol. 11765 of Lecture Notes in Computer Science, Springer, Cham, 2019.

[46] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[47] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE trans Pattern Anal Mach Intel*, vol. 39, no. 4, pp. 640–651, 2014.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9351 of Lecture Notes in Computer Science, Springer, Cham, 2015.

[49] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," https://arxiv.org/abs/1904.08128.

[50] M. Staring, U. A. van der Heide, S. Klein, M. A. Viergever, and J. Pluim, "Registration of cervical MRI using multifeature mutual information," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1412–1421, 2009.

[51] K. Miller, A. Wittek, G. Joldes et al., "Modelling brain deformations for computer-integrated neurosurgery," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 26, no. 1, pp. 117–138, 2010.

[52] Xishi Huang, Jing Ren, G. Guiraudon, D. Boughner, and T. M. Peters, "Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 11, pp. 1802–1814, 2009.

[53] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, no. 1-2, 2020.

[54] X. Yang, R. Kwitt, and M. Niethammer, "Fast predictive image registration," *Deep Learning and Data Labeling for Medical Applications.*, pp. 48–57, 2016.

[55] J. Lv, M. Yang, J. Zhang, and X. Wang, "Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study," *The British Journal of Radiology*, vol. 91, 2018.

[56] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1075–1078, Washington, DC, USA, 2018.

[57] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transfer networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 2017–2025, 2015.

[58] D. Kuang and T. Schmah, "FAIM-a ConvNet method for unsupervised 3D medical image registration," 2018, https://arxiv.org/abs/1811.09243.

[59] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, "Adversarial image registration with application for MR and TRUS image fusion," 2018, https://arxiv.org/abs/1804.11024.

[60] J. Kreb, T. Mansi, H. Delingette et al., "Robust non-rigid registration through agent-based action learning," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. MICCAI 2017*, M. Descoteaux, L. Maier-Hein,

A. Franz, P. Jannin, D. Collins, and S. Duchesne, Eds., vol. 10433 of Lecture Notes in Computer Science, pp. 344–352, Springer, Cham, 2017.

[61] M. Katan and A. Luft, "Global burden of stroke," *Seminars in Neurology*, vol. 38, no. 2, p. 208, 2018.

[62] L. Chen, P. Bentley, and D. Rueckert, "Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks," *Neuroimage Clin*, vol. 15, pp. 633–643, 2017.

[63] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "Acute and sub-acute stroke lesion segmentation from multimodal MRI," *Computer Methods and Programs in Biomedicine*, vol. 194, article 105521, 2020.

[64] L. Liu, S. Chen, F. Zhang, F. X. Wu, Y. Pan, and J. Wang, "Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6545–6558, 2020.

[65] B. Zhao, S. Ding, H. Wu et al., "Automatic acute ischemic stroke lesion segmentation using semi-supervised learning," 2019, https://arxiv.org/abs/1908.03735.

[66] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks," *Computers in Biology and Medicine*, vol. 115, article 103487, 2019.

[67] S. Chilamkurthy, "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.

[68] H. Ye, F. Gao, Y. Yin et al., "Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network," *European Radiology*, vol. 29, no. 11, pp. 6191–6201, 2019.

[69] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, and L. Wang, "Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans," *Sensors*, vol. 19, no. 9, p. 2167, 2019.

[70] S. Singh, L. Wang, S. Gupta, B. Gulyas, and P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sensors Journal*, p. 1, 2020.

[71] M. H. Vlak, A. Algra, R. Brandenburg, and G. J. E. Rinkel, "Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis," *Lancet Neurology*, vol. 10, no. 7, pp. 626–636, 2011.

[72] D. J. Nieuwkamp, L. E. Setz, A. Algra, F. H. H. Linn, N. K. de Rooij, and G. J. E. Rinkel, "Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis," *The Lancet Neurology*, vol. 8, no. 7, pp. 635–642, 2009.

[73] N. Turan, R. A. Heider, A. K. Roy et al., "Current perspectives in imaging modalities for the assessment of unruptured intracranial aneurysms: a comparative analysis and review," *World Neurosurgery*, vol. 113, pp. 280–292, 2018.

[74] T. Nakao, S. Hanaoka, Y. Nomura et al., "Deep neural network-based computer assisted detection of cerebral aneurysms in MR angiography," *Journal of Magnetic Resonance Imaging*, vol. 47, no. 4, pp. 948–953, 2018.

[75] J. N. Stember, P. Chang, D. M. Stember et al., "Convolutional neural networks for the detection and measurement of

cerebral aneurysms on magnetic resonance angiography," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 808–815, 2019.

[76] T. Sichtermann, A. Faron, R. Sijben, N. Teichert, J. Freiherr, and M. Wiesmann, "Deep learning–based detection of intracranial aneurysms in 3D TOF-MRA," *American Journal of Neuroradiology*, vol. 40, no. 1, pp. 25–32, 2019.

[77] D. Ueda, A. Yamamoto, M. Nishimori et al., "Deep learning for MR angiography: automated detection of cerebral aneurysms," *Radiology*, vol. 290, no. 1, pp. 187–194, 2019.

[78] A. Park, C. Chute, P. Rajpurkar et al., "Deep learning–assisted diagnosis of cerebral aneurysms using the HeadXNet model," *JAMA Network Open*, vol. 2, no. 6, article e195600, 2019.

[79] Z. Shi, C. Miao, U. J. Schoepf et al., "A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images," *Nature Communications*, vol. 11, no. 1, p. 6090, 2020.

[80] J. Zhang, S. Gajjala, P. Agrawal et al., "Fully automated echocardiogram interpretation in clinical practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018.

[81] J. P. Howard, J. Tan, M. J. Shun-Shin et al., "Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography," *Journal of Medical Artificial Intelligence*, vol. 3, 2020.

[82] D. M. Vigneault, W. Xie, C. Y. HodDavid, D. A. Bluemke, and J. A. Noble, "Ω-Net (Omega-Net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks," *Medical Image Analysis*, vol. 48, pp. 95–106, 2018.

[83] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Mecleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 515–524, 2019.

[84] S. Moccia, R. Banali, C. Martini et al., "Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 32, no. 2, pp. 187–195, 2019.

[85] W. Bai, H. Suzuki, C. Qin et al., "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018*, A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., vol. 11073 of Lecture Notes in Computer Science, Springer, Cham, 2019.

[86] E. D. Morris, A. I. Ghanem, M. Dong, M. V. Pantelic, E. M. Walker, and C. K. Glide-Hurst, "Cardiac substructure segmentation with deep learning for improved cardiac sparing," *Medical Physics*, vol. 74, no. 2, pp. 576–586, 2020.

[87] Y. Shen, Z. Fang, Y. Gao, N. Xiong, C. Zhong, and X. Tang, "Coronary arteries segmentation based on 3D FCN with attention gate and level set function," *IEEE Access*, vol. 7, 2019.

[88] J. He, C. Pan, C. Yang et al., "Learning hybrid representations for automatic 3D vessel centerline extraction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020*, A. L. Martel, Ed., vol. 12266 of Lecture Notes in Computer Science, Springer, Cham, 2020.

[89] W. Zhang, J. Zhang, X. Du, Y. Zhang, and S. Li, "An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT," *Computing*, vol. 101, no. 6, pp. 667–678, 2019.

[90] J. Liu, C. Jin, J. Feng, Y. Du, J. Lu, and J. Zhou, "A vessel-focused 3D convolutional network for automatic segmentation and classification of coronary artery plaques in cardiac CTA," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges. STACOM 2018*, M. Pop, Ed., vol. 11395 of Lecture Notes in Computer Science, Springer, Cham, 2018.

[91] E. Vorontsov, M. Cerny, P. Régnier et al., "Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases," *Radiology: Artificial Intelligence*, vol. 1, no. 2, article 180014, 2019.

[92] X. Wang, S. Han, Y. Chen, D. Gao, and N. Vasconcelos, "Volumetric attention for 3D medical image segmentation and detection," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, D. Shen, Ed., vol. 11769 of Lecture Notes in Computer Science, Springer, Cham, 2019.

[93] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1316–1325, 2020.

[94] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, and R. M. Summers, "E2Net: an edge enhanced network for accurate liver and tumor segmentation on CT scans," https://arxiv.org/abs/2007.09791.

[95] S.-h. Zhen, M. Cheng, Y.-b. Tao et al., "Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data," *Frontiers in Oncology*, vol. 10, p. 680, 2020.

[96] X. Liu, J. L. Song, S. H. Wang, J. W. Zhao, and Y. Q. Chen, "Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification," *Sensors*, vol. 17, p. 149, 2017.

[97] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, and S. Kiryu, "Deep learning for staging liver fibrosis on CT: a pilot study," *European Radiology*, vol. 28, no. 11, pp. 4578–4585, 2018.

[98] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, and S. Kiryu, "Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images," *Radiology*, vol. 287, no. 1, pp. 146–155, 2018.

[99] K. J. Choi, J. K. Jang, S. S. Lee et al., "Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver," *Radiology*, vol. 289, no. 3, pp. 688–697, 2018.

[100] L. Y. Xue, Z. Y. Jiang, T. T. Fu et al., "Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis," *European Radiology*, vol. 30, no. 5, pp. 2973–2983, 2020.

[101] Z. Tang, W. R. Liu, P. Y. Zhou et al., "Prognostic value and predication model of microvascular invasion in patients with intrahepatic cholangiocarcinoma," *Journal of Cancer*, vol. 10, no. 22, pp. 5575–5584, 2019.

[102] S. Men, H. Ju, L. Zhang, and W. Zhou, "Prediction of microvascular invasion of hepatocellar carcinoma with contrast-enhanced MR using 3D CNN And LSTM," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*,, pp. 810–813, Venice, Italy, 2019.

[103] Y.-Q. Jiang, S.-E. Cao, S. Cao et al., "Preoperative identification of microvascular invasion in hepatocellular carcinoma

by XGBoost and deep learning," *Journal of Cancer Research and Clinical Oncology*, vol. 147, pp. 821–833, 2021.

[104] J. Olczak, N. Fahlberg, A. Maki et al., "Artificial intelligence for analyzing orthopedic trauma radiographs," *Acta Orthopaedica*, vol. 88, no. 6, pp. 581–586, 2017.

[105] T. Urakawa, "Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network," *Skeletal Radiology*, vol. 48, no. 2, pp. 239–244, 2019.

[106] W. Gale, L. Oakden-Rayner, G. Carneiro, A. P. Bradley, and L. J. Palmer, "Detecting hip fractures with radiologist-level performance using deep neural networks," 2017, https://arxiv.org/abs/1711.06504.

[107] J. D. Krogue, "Automatic hip fracture identification and functional subclassification with deep learning. Radiology," *Artificial Intelligence*, vol. 2, no. 2, article e190023, 2020.

[108] K. Gan, D. Xu, Y. Lin et al., "Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments," *Acta Orthopaedica*, vol. 90, no. 4, pp. 394–400, 2019.

[109] Y. L. Thian, Y. Li, P. Jagmohan, D. Sia, V. E. Y. Chan, and R. T. Tan, "Convolutional neural networks for automated fracture detection and localization on wrist radiographs," *Radiology: Artificial Intelligence*, vol. 1, article e180001, 2019.

[110] R. Lindsey, A. Daluiski, S. Chopra et al., "Deep neural network improves fracture detection by clinicians," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 45, pp. 11591–11596, 2018.

[111] S. Wu, L. Yan, X. Liu, Y. Yu, and S. Zhang, "An end-to-end network for detecting multi-domain fractures on X-ray images," in *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, October 2020.

[112] H.-Z. Wu, L. F. Yan, X. Q. Liu et al., "The feature ambiguity mitigate operator model helps improve bone fracture detection on X-ray radiograph," *Scientific Reports*, vol. 11, no. 1, article 1589, 2021.

[113] P. Kairouz, H. McMahan, B. Avent et al., "Advances and open problems in Federated Learning," https://arxiv.org/abs/1912.04977.

[114] I. I. I. Armato, G. McLennan, L. Bidaut et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[115] A. A. A. Setio, A. Traverso, T. de Bel et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.

[116] K. Bowyer, D. Kopans, W. P. Kegelmeyer et al., "The digital database for screening mammography," in *Third international workshop on digital mammography*, vol. 58, p. 27, 1996.

[117] K. Yan, X. Wang, L. Lu, and R. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of Medical Imaging*, vol. 5, 2018.

[118] P. Bilic, P. F. Christ, E. Vorontsov et al., "The liver tumor segmentation benchmark (LiTS)," https://arxiv.org/abs/1901.04056.

[119] A. L. Simpson, M. Antonelli, S. Bakas et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019, https://arxiv.org/abs/1902.09063.

[120] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.

[121] M. Husseini, A. Sekuboyina, M. Loeffler, F. Navarro, B. H. Menze, and J. S. Kirschke, "Grading loss: a fracture grade-based metric loss for vertebral fracture detection," 2020, https://arxiv.org/abs/2008.07831.

[122] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, vol. 2, pp. 1735–1742, New York, NY, USA, 2006.

[123] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, Boston, MA, USA, 2015.

[124] A. Jiménez-Sánchez, D. Mateus, S. Kirchhoff et al., "Medical-based deep curriculum learning for improved fracture classification," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019*, D. Shen, Ed., vol. 11769 of Lecture Notes in Computer Science, Springer, Cham, 2019.

[125] H. Chen, Y. Wang, K. Zheng et al., "Anatomy-aware Siamese network: exploiting semantic asymmetry for accurate pelvic fracture detection in X-ray images," 2020, https://arxiv.org/abs/2007.01464.