MEDICAL IMAGE SEGMENTATION USING 3D CONVOLUTIONAL NEURAL NETWORKS: A REVIEW

A PREPRINT

S Niyas [a], **S J Pawan**[a], **M Anand Kumar**[b], **and Jeny Rajan** [a]

[a]Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal, India

[b]Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India

ABSTRACT

Computer-aided medical image analysis plays a significant role in assisting medical practitioners for expert clinical diagnosis and deciding the optimal treatment plan. At present, convolutional neural networks (CNN) are the preferred choice for medical image analysis. In addition, with the rapid advancements in three-dimensional (3D) imaging systems and the availability of excellent hardware and software support to process large volumes of data, 3D deep learning methods are gaining popularity in medical image analysis. Here, we present an extensive review of the recently evolved 3D deep learning methods in medical image segmentation. Furthermore, the research gaps and future directions in 3D medical image segmentation are discussed.

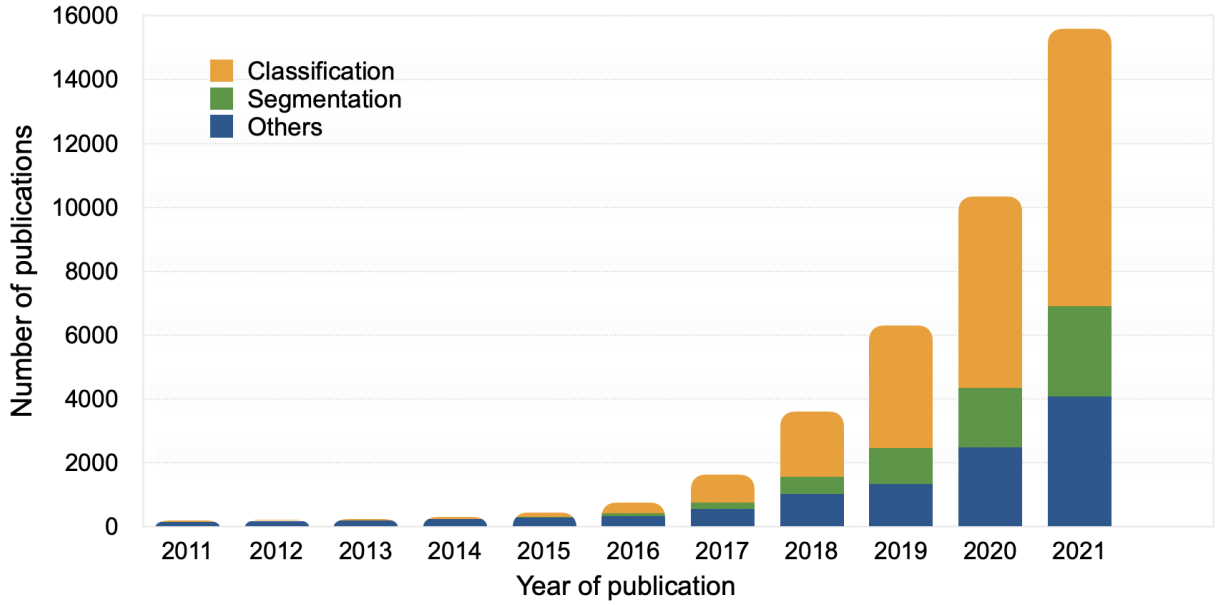
Keywords 3D deep learning · Convolutional Neural Networks · Medical image analysis

1 Introduction

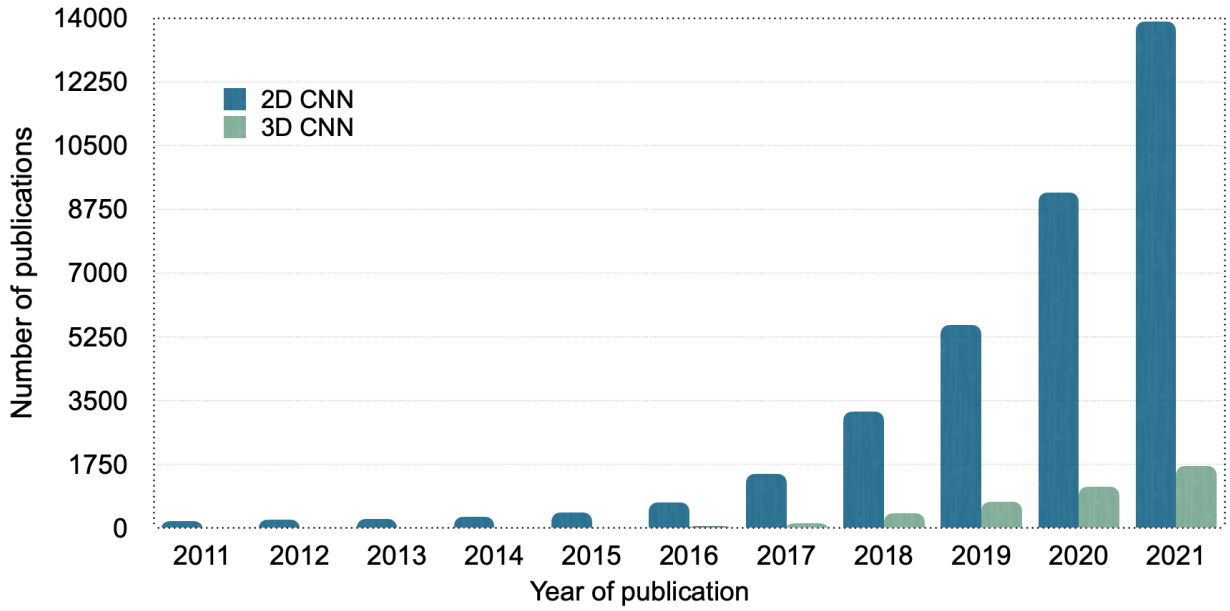
Artificial intelligence (AI) techniques have been increasingly used in modern healthcare devices to analyze medical images. Computer-aided diagnosis (CADx) refers to analyzing medical and radiologic data and extracting useful information using computer software to assist clinicians in making a rapid and error-free diagnosis. These diagnostic systems reduce the subjectivity in decision-making and the overall cost involved. In recent years, the performance of popular medical imaging modalities, such as X-ray, computed tomography (CT), ultrasonography (USG), and magnetic resonance imaging (MRI), has improved in terms of multiple factors such as acquisition time, image quality, and cost-effectiveness [1]. However, these techniques suffer from certain limitations such as high noise, motion artifacts, and non-uniform contrasts, leading to an unreliable clinical diagnosis. Furthermore, registration errors between adjacent frames in three-dimensional (3D) imaging methods can degrade the image quality.

Medical image processing encompasses problem-specific strategies using image processing algorithms. Some of its common applications include image registration, denoising, enhancement, compression, classification, and segmentation. Earlier, medical imaging data were usually processed using unsupervised machine learning techniques (such as thresholding and clustering) that do not require huge training data or extensive computation resources. However, the increase in the volume and complexity of medical image data has accelerated the development of algorithms for generalized feature extraction, which has increased the need for supervised machine learning algorithms.

Traditional supervised machine learning approaches are based on application-specific feature extraction techniques for analyzing image characteristics such as contrast variation, orientation, shape, and texture patterns. Moreover, the performance of supervised algorithms is highly dependent on several factors, such as image quality and complexity of feature patterns. Because handcrafted features are an integral part of such methods, the algorithm design requires



(a) Breakdown of papers based on the application.



(b) Breakdown of papers based on the deep learning approach (2D or 3D CNN).

Figure 1: Statistics of papers retrieved from PubMed on medical image analysis using deep learning techniques (from 2011-2021). The data for the year 2021 has been extrapolated from the papers listed till June 2021.

domain experts, making human intervention inevitable. These limitations of traditional machine learning algorithms have resulted in the development of adaptive learning approaches such as artificial neural networks (ANN) [2].

Studies on neural network-based decision-making are being conducted since the 1950s. Several challenges were faced initially, mainly due to the lack of ample data and adequate computational capabilities. However, factors such as data availability, advancements in parallel distributed computing, and the surge in the semiconductor industry accelerated further research on neural networks. Eventually, ANN emerged as the predominant method over traditional machine learning algorithms owing to its enhanced learning performance. With the increasing prominence of image data in the

digital data space, research on developing neural network models for images increased tremendously as evident from the introduction of convolutional neural networks (CNN). The first popular CNN method was published by LeCun et al. [3] for handwritten character recognition. Subsequently, several successful deep CNN models have been developed to solve various classification [4] and segmentation problems [5].

Medical image processing has also benefited from these developments periodically. For instance, several studies have been conducted on the use of deep learning-based medical image analysis, which in turn has renewed the research interest in this field. However, despite the significant growth in the research on deep learning-based medical image analysis during the last couple of years, statistics on classification outweigh that on segmentation approaches. The statistics of CNN-based medical image processing papers retrieved from PubMed are shown in Fig. 1a.

The literature reports numerous dedicated reviews on CNN-based medical image analysis. For instance, the review article by Shen et al. [6] provides a comprehensive overview of medical image processing using advanced machine learning, especially deep-learning techniques. Their study highlights the advantages of various CNN models over traditional machine learning methods in different medical applications. Similarly, Litjens et al. [7] thoroughly discussed the applications of deep learning in various fields of medical image analysis. In addition, several review papers [8–13] provide a basic understanding of CNN models and their applications in medical image analysis. However, most of these reviews have focused on analyzing deep learning approaches in an application-specific manner.

Since CNN models are designed to operate using image data, the dimensionality of the images plays a significant role in selecting the appropriate model. Generally, the majority of the medical image data are constructed using either two-dimensional (2D) or 3D imaging techniques. For example, digital X-rays, retinal fundus images, microscopic images in pathology, mammograms, etc., belong to the 2D image categories in the biomedical domain. Similarly, MRI, CT, and ultrasound (US) are widely used 3D medical images in clinical diagnosis. However, most of the published studies on 3D medical image analysis have used 2D CNN models. Fig. 1b substantiates our statement showing the breakdown of studies that have used 2D and 3D deep learning techniques in medical image analysis.

Segmentation of organs or anatomical structures is a crucial step in medical image processing. It is primarily used to detect abnormalities and estimate the exact extent of the organ or lesion region. Fig. 1 show the data reporting the lack of research in medical image segmentation and the inappropriate utilization of 3D medical data. In [14], Singh et al. presented an overview of various building blocks in 3D CNN architectures and several deep-learning approaches in volumetric medical image analysis. To the best of our knowledge, there are no comprehensive reviews in the literature focusing on medical image segmentation using 3D deep learning techniques. These research gaps motivated us to conduct an in-depth review of current deep learning trends in 3D medical image segmentation. We have also discussed the future perspectives in 3D medical image segmentation.

This review covers several reputed articles that mainly focused on the 3D deep-learning techniques for volumetric medical image segmentation. Papers reporting overlapping techniques have been excluded from this review to minimize redundancy. The primary contributions of this review article are to provide

1. an overview of 3D deep learning for medical image processing.
2. an in-depth review on various state-of-the-art 3D deep learning-based medical image segmentation.
3. a discussion on research gaps and future directions in 3D medical image segmentation.

The rest of this review is organized as follows. Section 2 presents a brief overview of 2D and 3D CNN models and a detailed survey on existing 3D deep learning models along with their contributions to medical image segmentation. Section 3 provides a critical discussion and an outlook for future research.

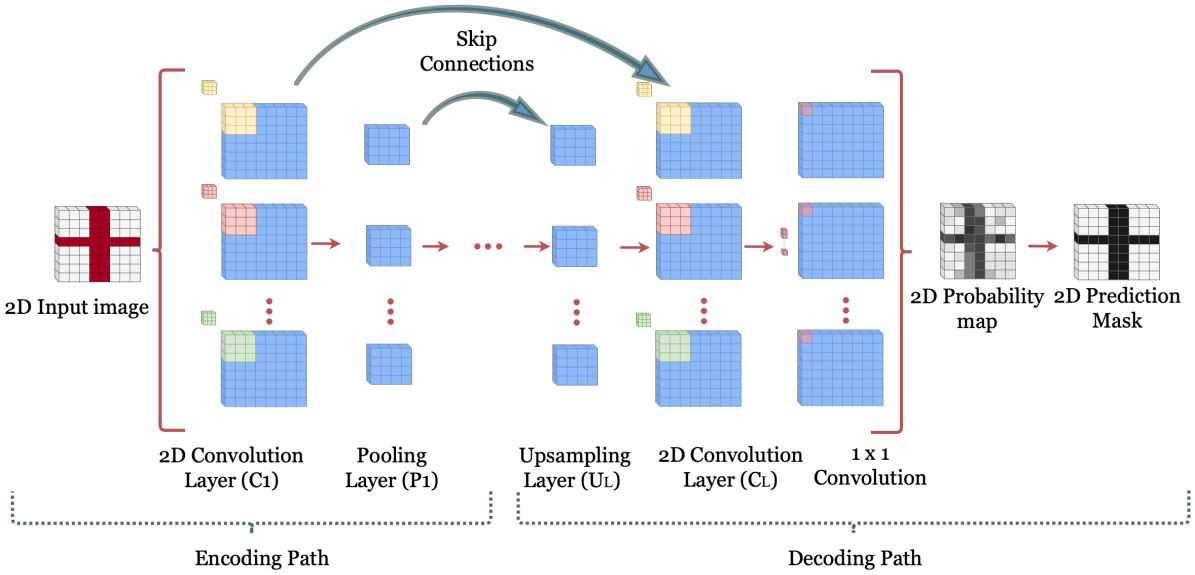
2 3D CNN in Medical Image Segmentation

2.1 Convolutional Neural Networks: An overview

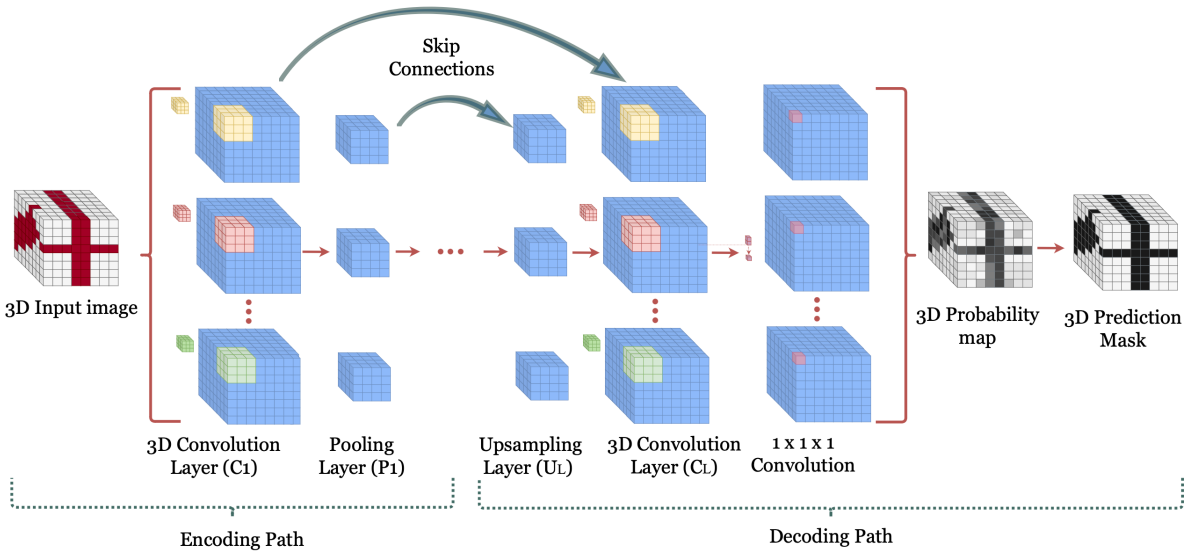
CNN is designed to learn spatial features using higher-dimensional data space (2D, 3D, etc.) and analyze the images adaptively without using any handcrafted features. CNN architecture generally consists of three building blocks: convolution layers, pooling layers, and fully connected dense layers [12]. A convolution layer uses small multi-dimensional kernels to extract spatial features from image locations. The pooling layer is designed to pass relevant features to the subsequent layers by downsampling the feature space. In addition, pooling reduces the number of trainable parameters in the subsequent dense layers and provides translation invariance to small shift errors.

A typical CNN model uses multiple convolutions and pooling layers for extracting multi-scale features; the complexity of image features advances as the model depth increases. The fully connected dense layer is designed to transform the

feature map from the previous layer to one-dimensional (1D) feature vectors. The final fully connected layer operates as a classification layer and provides the probabilities of the target classification task. Furthermore, a CNN model uses several additional modules such as activation, batch normalization, regularization, and dropout, to improve the learning process [14].



(a) A 2D CNN architecture for segmentation.



(b) A 3D CNN architecture for segmentation.

Figure 2: 2D and 3D CNN architectures for segmentation.

In image segmentation, every pixel needs to be classified simultaneously. CNN model for segmentation uses convolution and pooling layers similar to those in a classification network. However, the fully connected layers are usually

excluded in the segmentation model to retain the spatial relationship of pixels throughout the network. In addition, adequate upsampling layers are added to the final layers of the network to compensate for the pooling operation. Hence, the probability of each pixel is identified and finally, the segmentation mask is created by processing multiple overlapping patches and forward passing of each patch through the network. Although several CNN-based segmentation architectures have been proposed in the literature, the encoding–decoding-based fully convolutional neural networks (FCNN) architectures and their derivatives [15–17] are considered the best performing models for several segmentation problems. A general workflow of an encoding–decoding architecture for segmentation is depicted in Fig. 2a.

2.2 An Overview on 3D Medical Data

Medical images use different modalities based on the imaging principles. The characteristics of these images differ, in terms of spatial resolution, image intensity range, size of the image, and noise. In addition, they vary in terms of dimensionality and its way of representation. 2D data mostly use relatively simple Euclidean representation, which describes each data point using pixel intensity values. In medical images, these pixel values represent the state of the anatomical structure under consideration. For instance, the pixel intensity of X-ray images varies with the radiation absorption, whereas it depends on the acoustic pressure in ultrasound images or the radio frequency (RF) signal amplitude in MRI.

The representation of 3D data is more complex. Nevertheless, several standard representations are reported in the literature, such as projections, volumetric representations, point clouds, and graphs [18]. The majority of the 3D medical image data belong to the volumetric representation, where volumetric elements or *voxels* (analogous to *pixels* in a 2D image) are used to model 3D data by describing how the 3D object is distributed through the three perpendicular axes. While acquiring 3D medical data, the device scans the body parts in any of the three planes: axial (front to back), coronal (top to bottom), and sagittal (side to side). Normally, multiple 2D slices are acquired across the area under consideration and are stacked to produce 3D image volumes using adequate image registration techniques [19–21].

The 3D image visualization has provided a great opportunity for clinicians to evaluate the cross-section of anatomic structures. This has increased the understanding of the complex patterns and structural morphology, mostly in radiology. Currently, 3D imaging is commonly used in several modalities such as CT, MRI, USG, and PET. Although 3D imaging techniques have numerous advantages over 2D images, they have certain limitations as well. For example, compared to 2D imaging methods, they require significantly large storage space and are often expensive. However, with decreasing computational expenses, higher networking speeds, and the availability of powerful graphics processing units (GPU), visualization, and analysis of 3D medical images have become easy [22].

2.3 Applications of 3D CNN in Medical Image Segmentation

A 3D CNN-based segmentation model uses 3D images as input, and a similar-sized 3D prediction mask is expected as the output. The network structure is similar to that of a standard 2D CNN model, but the convolution and pooling layers use 3D extensions to process the volumetric data. Here, the convolution layers perform filtering with 3D kernels, and the 3D pooling layers subsample the data in all three dimensions to compress the size of the feature space. Hence the volumetric data is analyzed as cubic patches from layer to layer and can learn spatial features in all three dimensions. A 3D CNN architecture for segmentation is shown in Fig. 2b.

The segmentation architecture usually consists of an encoding and decoding path like the U-Net architecture proposed by Ronneberger et al. [15]. In the encoding path, 3D convolution layers and pooling layers create 4D feature space. The number of filters used in each convolution layer determines the extent of the fourth dimension. Hence, the intermediate convolution layers use a 4D convolution operation. Since the pooling reduces the feature space dimension in the encoding path, the decoding path uses a sufficient number of upsampling layers to gradually increase the feature space dimension similar to the input data. Normally, transpose convolution is used for upsampling the data in a learnable fashion. Skip connections are also used to preserve the fine features, by merging features between the corresponding encoding and decoding layers. A $(1 \times 1 \times 1)$ filter is used in the final layer to project the stacked features into a feature space with the same dimension as the input image. The probability map is then created using a non-linear activation function (commonly using the Softmax [23] activation function to limit the probability values between 0 and 1), followed by thresholding which creates the final prediction mask.

3D CNNs have been successfully used in several medical image segmentation tasks. In this study, we considered 3D CNN papers (for medical image segmentation) published between 2015 and 2021, and are classified into one of the following categories:

1. 3D extensions of 2D CNN architectures
2. 3D Patch-wise segmentation methods

3. 3D segmentation using 2D CNN models
4. 3D CNN with Semi-supervised learning

The subsequent sections performs a detailed review and provides in-depth comparison of state-of-the-art 3D medical image segmentation methods within each category.

2.3.1 3D extensions of 2D CNN architectures

The CNN models that use 3D equivalents of existing 2D segmentation architectures are reviewed under this category. Cizec et al. [24] proposed the first promising 3D CNN network for volumetric segmentation that learns from sparsely labeled 3D images. The proposed model extends the classical U-Net architecture [15] by substituting all 2D operations with their 3D equivalents. The study in [24] outlined two cases: a semi-automated model and a fully automated model. In both cases, the network learns from sparse annotated data and this helps to reduce the human effort in ground truth labeling without considerable degradation in the segmentation performance. The experiments were conducted on the Xenopus kidney dataset [25], and both the semi-automated and fully automated models showed considerable performance improvements over state-of-the-art 2D CNN architectures.

Milletari et al. [26] presented an advanced 3D U-Net architecture (V-Net) with residual blocks (instead of cascaded convolution blocks) and strided convolution (instead of max-pooling). The methodology use a Dice score based loss function to reduce the class imbalance between the voxel classes. This model was evaluated on the PROMISE2012 dataset [27], and the results were very close to the state-of-the-art 2D segmentation models. Bui et al. [28] proposed another 3D CNN architecture inspired by the 2D FCNN model proposed by Long et al. [29] for brain segmentation in infant brain MRI volumes [30]. The encoder path uses multiple coarse and fine dense 3D convolution layers to extract multi-scale features. Downsampling makes use of strided convolution to reduce the feature size and to increase the receptive field. Multiple convolution layers are used to extract four different scales of feature space, then upsampled and concatenated to generate the final probability map. However, direct merging of the upsampled features in the decoder path creates semantic gaps in the feature space, and may create undesired outcomes in the segmentation results.

In [31], Kayalibay et al. proposed another 3D encoder-decoder architecture with deep supervision [32]. In this model, the feature space created at different decoder levels in the network are merged using an element-wise summation of the features. Hence, the learning is based on the error between the ground truth and a combination of feature maps from different decoding levels. In this deep supervision approach, the learning process is directly dependent on the coarse features from the different decoder levels, and helps to speed-up the convergence. However, the element-wise summation may limit the learning process when the semantic gap between the different decoder levels is significant. Dou et al. [33] proposed a similar deep supervision-based 3D CNN model that can work on relatively small 3D datasets. The model uses an encoder-decoder FCNN with 3D convolutions. The outputs from each decoder layer are upsampled using 3D de-convolution and merged to get the final segmentation mask. The integration of deep supervision and the post-processing using CRF [34] also help to improve the overall segmentation performance.

In [35], Li et al. presented a 3D CNN model using dilated convolution [36] instead of max-pooling, and residual connections. For utilizing multi-scale features, the dilation factor of the 3D convolution kernel is steadily increased in the advanced layers. Hence, the spatial resolution of the feature space is kept constant throughout the network. The residual blocks merge the of features from different layers to reduce vanishing gradients problem and feature degradation. However, the absence of max-pooling reduces the translation in-variance in the feature space and accelerates the computation complexity.

Chen et al. [37] proposed a residual deep 3D CNN architecture for segmenting the brain region from 3D MRI volumes. The proposed architecture employs a voxel-wise residual network (VoxResNet). The architecture uses a 3D extension of 2D deep residual networks that extracts features from multiple scales using a series of convolution operations, and the features at multiple scales are finally fused to generate the segmentation mask. For training such a deep network with small training data, multi-modal contextual information is integrated into the network to take advantage of additional information from various modalities. Experiments were conducted using three imaging modalities: T1, T1-IR, and T2-FLAIR from the brain structure segmentation task [38].

In [39], Niyas et al. presented a 3D Residual U-Net architecture for segmenting focal cortical dysplasia (FCD) from 3D brain MRI. The method aims to retain the advantages of both 2D and 3D CNN methods by an effective design of input data slices and CNN architecture. The model proposes a shallow sliced stacking approach to generate large number of 3D samples from small datasets. The customized 3D U-Net architecture with residual connections in the encoder path helps in extracting multi-scale features with fewer trainable parameters.

Schlemper et al. [40] proposed an attention gate (AG) based 3D U-Net architecture for medical image segmentation which automatically learns to concentrate on various target structures. The features from the encoding path to the

decoding path are propagated through the skip connections with self-attention gating modules [41]. The AG module uses grid-based gating that allows attention coefficients to focus more on local neighborhoods, that helps to locate the relevant regions in the image. Performance improvements of the proposed network over U-Net are experimentally observed to be consistent across different imaging datasets [42], [43]. However, in an FCNN with encoder-decoder architecture, features in the lower depth are minimal and naive, while the features at advanced depths are of higher granularity. Hence, the attention modules at different depths may not help passing the salient features to the upsampling layers [44].

Wang et al. [45] proposed another 3D FCNN method that integrates recursive residual blocks and pyramid pooling to extract more complex features. The recursive residual blocks contain multiple residual connections that can minimize feature degradation problems. Pyramid pooling [46] generates fused feature maps at different decoding levels for obtaining both local and global information. It helps to eliminate the fixed size constraints of CNN without losing spatial information. The proposed architecture gave better multi-class segmentation performance while detecting white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) from 3D MR images from CANDI, IBSR18, and IBSR20 datasets [47, 48].

A 3D version of multi-scale U-Net segmentation was presented in [49] by Peng et al. The model uses multiple U-Net modules to extract long-distance spatial information at different scales. The U-Net blocks use Xception [50] modules instead of normal convolution to extract more complex features. Feature maps at different resolutions are up-sampled and fused to generate the segmentation mask. The authors claimed that the upsampled feature maps at different scales could extract and utilize appropriate features with faster learning. Traditional 3D convolutions are replaced with depth-wise separable convolutions, to reduce the computation complexity. However, the complex structure of the overall network and the deep convolution layers demand high memory and computation requirements.

The model cascade (MC) strategy is one of the popular schemes in CNN-based image segmentation that can alleviate the class imbalance problem by using a set of individual networks for coarse-to-fine segmentation. Despite its notable performance, it leads to undesired system complexity and overlook the correlation among the models. Zhou et al. [51] proposed a lightweight deep 3D CNN architecture: one-pass multi-task network (OM-Net) that can handle the native flaws in the MC approach. OM-Net combines discrete segmentation tasks into a one-pass deep model, to learn joint features and solve the class imbalance problem better than MC. The prediction results between tasks are correlated using a cross-task guided attention (CGA) block that can adaptively re-calibrate channel-wise feature responses. The model shows significant performance improvements in multiple public datasets [52, 53].

In the above-mentioned medical image segmentation approaches, the 3D CNN models mainly use a straight-forward extension of various 2D CNN building blocks. However, the volumetric data analysis brings a significant increase in the memory requirement and computation costs while training deep 3D CNN models compared to corresponding 2D versions. Several studies were also conducted in developing 3D segmentation models that works with limited training data and hardware resources. Such approaches are discussed in the following sections.

2.3.2 3D Patch-wise Segmentation

The resolution of input 3D data is the primary reason behind the need for large memory and higher computation complexity in training a 3D CNN segmentation model. Hence, patch-wise analysis becomes an interesting research topic in 3D medical image analysis, and such patch-wise 3D deep learning based techniques are discussed in this section.

Yu et al.[54] proposed a deep supervised 3D fractal network for whole heart and great vessel segmentation in MRI volumes. This method is an expansion of the FractalNet [55] and uses deep supervision using multiple auxiliary classifiers deployed at expanding layers to reduce the vanishing gradient problem. The methodology uses cropped 3D patches of size $(64 \times 64 \times 64)$ and reported good results in the HVSMR 2016 challenge dataset [56].

A deep dual pathway 3D CNN architecture was proposed by Kamnitsas et al. [57] for brain lesion segmentation. They employed a two-way network that simultaneously learns from multiple image scales to analyze features from different receptive fields. The approach uses 3D patches at two different scales and classifies the center voxels as any target classes using fully connected dense layers. A 3D fully connected conditional random field (CRF) is also utilized in the post-processing stage to reduce false alarms in the final segmentation mask. This method shows improved segmentation results over BRATS 2015 [52] and ISLES [58] datasets, compared with various 2D CNN methods. An advanced version of this model was also proposed in [59], which used multiple residual connections in the encoding path. The model reported better segmentation performance and was bench-marked at BRATS 2016 challenge [60].

In [61], Kamnitsas et al. presented an ensembles of multiple models and architectures (EMMA) for robust performance by aggregating predictions from multiple models. This method use an ensemble of a DeepMedic model [59], three 3D FCNN models [29], and two 3D U-Net models [15]. An ensembler computes the confidence maps, and finds the

average class confidence of the individual models. Finally, each voxel in the segmentation map was labeled with the class label with the highest confidence score. The EMMA approach won the first place in the BRATS 2017 challenge [52, 53, 62]. Nevertheless, the ensemble with seven 3D architectures requires higher computation complexity and training time.

In, [63], Chen et al. proposed a hierarchical 3D CNN architecture to segment Glioma from multi-modal brain MRI volumes. The model uses two distinct scales of 3D image patches to examine multi-scale features. The 3D CNN architecture uses a series of hierarchical dense convolutions without pooling layers. The model uses a patch-wise analysis that reduces the class imbalance issue and the computational cost involved in training large 3D datasets. This method was validated on the BRATS 2017 [53] dataset and had reported better segmentation performance compared to similar patch-wise 3D approaches.

The patch-based segmentation helps to reduce the hardware resources for training and generates a large number of data samples that favor adequate learning. However, the patch-wise analysis often fails to extract global features from the actual image volumes. This may limit the learning performance when the abnormality is region-specific.

2.3.3 3D segmentation using 2D CNN models

Since the straight-forward 3D CNN models are highly susceptible to large computation costs and require large datasets, several alternative techniques have been proposed that can use simulated 3D architectures with cost-effective approximations. Such prominent approximations of 3D medical image segmentation are discussed in this subsection.

Extraction of inter-slice features from a 3D volume is possible by analyzing three orthogonal planes (sagittal, coronal, and transverse planes) and by classifying the voxel at the intersection of three planes. This method was successfully applied by Ciompi et al. [64] for classifying lung modules. This model learn from 2D patches in three perpendicular planes centered at a given voxel, at three different scales. Fully connected layers combine the streams and perform the voxel classification. However, CNN-based systems with fully-connected layers are computationally less efficient for the segmentation task compare to FCNN architectures. A similar 3D segmentation approach by combining information from three orthogonal planes was discussed by Kitrungrotsakul et al. [65]. The proposed method (VesselNet) uses 2D DenseNet [66] network for classifying voxels in three different 2D planes. The features from the individual networks are then fused for getting the probability to predict the segmentation map. The approach is not an end-to-end segmentation model and hence misses most contextual and location-based information.

Mlynarski et al. [67] proposed a similar CNN architecture that combines the benefits of the short-range 3D context and the long-range 2D context. This architecture uses modality-specific sub-networks to focus on missing MR sequences. During training, three individual 2D U-Net models were used to create features from axial, coronal, and sagittal slices. The final 3D network was then trained using these 2D features along with 3D patches from the input volumes. The study also suggests considering the outputs from multiple 3D models to minimize the constraints of specific choices of neural network models. However, the use of multiple 2D and 3D models demands more hardware resources, and patch-wise analysis fails to learn region-specific features.

Another 3D CNN model was introduced by Heinrich et al. [68], which uses trainable 3D convolution kernels that learn both filter coefficients and spatial filter offsets in a continuous space, based on the principle of differentiable image interpolation. The proposed kernel: one binary extremely large and inflecting sparse kernel (OBELISK) works with fewer parameters and reduced memory consumption. The deformable convolutions in the OBELISK filter replace the continuously sampled spatial filter with a sparse sampling approach. This helps to extract information from wider spatial contexts and replace multiple small filter kernels at different scales. However, the computation complexity may be higher in OBELISK due to the unoptimized filter sampling.

Roth et al. [69] proposed an automated segmentation system for 3D CT pancreas volumes, based on a dual-stage cascaded method: a localization followed by a segmentation stage. 2D holistically nested convolutional networks (HNN) [70] on the three orthogonal directions, were used in the localization stage. The 2D HNN pixel probability maps are then merged to get a 3D bounding box of the foreground regions. In the second stage, the model focuses on semantic segmentation over the voxels in the bounding box. Two different HNN realizations are integrated in the segmentation stage that extract mid-level cues of deeply-learned boundary maps of the target organ. The authors also presented an advanced multi-class segmentation model [71] for segmenting the liver, spleen, and pancreas from CT volumes. The cascaded network helps to provide boundary preserving segmentation and reduces false detection in the 3D volumes.

Chen et al. [72] presented a separable 3D U-Net architecture that targets to extract spatial information from the image volumes with limited memory and computation cost. The model uses a U-Net architecture for the end-to-end training with separable 3D (S3D) convolution blocks. The S3D block is an arrangement of parallel 2D convolutions and exploits the advantages of the residual inception architecture. The model is evaluated on BRATS 2018 [73] data, and the results justify the improvement in the segmentation performance compared to a standard 2D or 3D U-Net architecture.

Another 3D CNN architecture proposed by Rickmann et al.[74] incorporates a compress-process-recalibrate (CPR) pipeline using 3D recalibration methods. The method uses a project & excite (PE) module that compress the intermediate high dimensional 4D feature maps into three 2D projection vectors. The convolution layers in the *processor* module learns from this 2D projections and the final recalibration stage generate the 4D tensors for the subsequent layers. This approach reduces the computation cost, without degrading the segmentation performance. The paper reported improved overall accuracy over different multi-class segmentation datasets [75–78].

The volumetric medical segmentation using the above-discussed approaches is highly useful in reducing the computation complexity and the need for large datasets. The results in the above discussed works prove its supremacy over other conventional 2D and 3D CNN approaches. However, the 3D approximations are highly dependent on the characteristics of the input data and the type of segmentation needed. This may limit the generic use of models across different medical image modalities, and the designing of such a universal model remains an open research challenge.

2.3.4 Semi-supervised Learning

The 3D CNN models discussed in Section 2.3.1-2.3.3 mostly use fully supervised deep learning algorithms that require thousands of annotated ground-truth 3D volumes. However, accurate marking of ground-truth images requires medical experts, and is a tedious process. Hence, the supervised learning algorithms are more expensive in terms of both time and cost. Consequently, research also commenced on semi-supervised learning approaches [79–81] that require only a few labeled image samples. 3D medical image segmentation using semi-supervised techniques are discussed in this section.

In the article [82], Mondal et al. proposed a 3D multi-modal medical image segmentation method based on generative adversarial networks (GANs) [83]. The method uses a semi-supervised training with a mix of labeled and unlabeled images. The proposed architecture uses several design considerations to modify the standard adversarial learning approaches to generate 3D volumes of multiple modalities. The 3D GANs generate fake samples and are used along with labeled and unlabeled 3D volumes, and a discriminator defines separate loss functions for these labeled, unlabeled and synthetic (fake) training samples. However, generating useful synthetic samples may not represent the distribution of the actual data and become very challenging while working with 3D medical image data.

Zhou et al. [84] proposed a straight-forward semi-supervised segmentation approach named deep multi-planar co-training (DMPCT), which uses parallel training to extract information from multiple planes. The multiplanar fusion generates reliable pseudo labeling to train deep segmentation networks. The DMPCT framework consists of a teacher network that uses fully labeled images for training. The trained model then creates the pseudo labels from the unlabelled training data with a multi-planar fusion module. Subsequently, the student model uses both labeled and pseudo labeled data for the final training process.

Yang et al. [85] proposed a similar semi-supervised segmentation technique to detect catheter from volumetric ultrasound images. The segmentation model uses a deep Q network (DQN) for localizing the target region. After the catheter localization, the method uses a twin-UNet model for the semantic segmentation of the catheter volume around the localized region by a patch-based strategy. This model uses a typical teacher network followed by a student network to train both labeled and unlabeled 3D patches based on a set of hybrid constraints.

In [86], Li et al. presented a shape-aware 3D segmentation for medical images to use extensive unlabeled data so as to enforce a geometric shape analysis on the segmentation output. The model uses a deep CNN architecture that predicts semantic segmentation and signed distance map (SDM) of object surfaces. The network uses an adversarial loss between the predicted SDMs of labeled and unlabeled data during training to leverage shape-aware features.The integration of adversarial loss, which uses a generative discrimination function, helps supervise learning with unlabelled data and extract generalized features.

In [87], Wang et al. proposed a tailored modern semi-supervised learning (SSL) method named as FocalMix for the detection of lesions from 3D medical images. The model is built on MixMatch [88] SSL framework, which uses a prediction for unlabeled images and MixUp augmentation. The proposed 3D CNN model uses a Soft-target Focal Loss along with an anchor level target prediction to improve lesion detection. The study also uses two adaptive augmentation methods: image-level MixUp and object-level MixUp, to generate the final training data. Most of the papers published under this category use a similar teacher-student network as in 2D CNN models. This limits the complete utilization of the inter-slice relationships from the volumetric data.

Some pre-trained 3D medical segmentation models have also been discussed in the literature. Chen et al. [89] propose a heterogeneous 3D network called Med3D by co-training multi-domain 3D datasets to develop multiple pre-trained 3D medical image segmentation models. The authors use datasets from various medical challenges to create a 3D segmentation data set (3DSeg-8) with different modalities, target organs and pathologies.The pre-trained models were experimented on several 3D medical datasets [90] using transfer learning and they observed with improved results

and faster convergence. The model can work well on similar image modalities with less training time and can be considered as a suitable option for small 3D image datasets. Zhou et al. [91] proposed a similar set of pre-trained 3D CNN networks for classification and segmentation tasks in 3D CT and MRI. The models are collectively known as Generic Autodidactic Models (Models Genesis), which uses learning by self-supervision. However, transfer learning may not be an appropriate solution in various scenarios due to the difference in image features across different image modalities and targeted abnormalities.

We summarise various medical image segmentation approaches using the 3D deep learning techniques in Table 1. The CNN architecture used, datasets, and remarks are also included in this table. A detailed discussion regarding the currently facing challenges in 3D medical image segmentation, possible solutions, and future directions are discussed in Section 3.

3 Discussion & Conclusion

3.1 Summary

In this study, we reviewed nearly 30 articles on 3D medical image segmentation using deep learning techniques. The articles were classified into different categories according to the different 3D deep learning approaches used to solve the segmentation problem. We believe that deep learning significantly contributes to medical image segmentation with 3D medical images. Although 3D deep learning methods are considerably less in use than 2D approaches, the exponential increase in the publications related to 3D deep learning techniques in the past couple of years in medical image analysis is remarkable. Nevertheless, the majority of the 3D CNN-based segmentation approaches employs the 3D extensions of successful 2D deep learning techniques. This bypass is advantageous for easy implementation; however, it introduces several optimization issues and restricts the analysis of 3D data to its full potential.

Numerous studies have focused on solving the native issues of 3D deep learning, such as the huge computation cost and memory requirements. The need for large volumetric datasets is another major challenge in 3D deep learning-based medical image segmentation. The key aspects of such successful alternative approaches in 3D deep learning-based segmentation were also reviewed in this study. The following subsections provide a detailed overview of the unique challenges faced in 3D medical image segmentation and the future trends.

3.2 Challenges in 3D Medical Image Segmentation

The application of deep learning methods in medical image analysis is associated with several challenges. The use of 3D deep learning techniques for volumetric medical image segmentation increases the complexities. The primary challenge faced by the researchers in implementing effective 3D CNN models is the need for large training data. The use of picture archiving and communication system (PACS)-based tools significantly improved the medical image management and storage standardization. However, huge amounts of archived medical data often become worthless due to the lack of proper annotation and the inconsistency in data attributed to the heterogeneity in imaging systems.

Most of the traditional 2D and 3D CNN models require task-specific labeled data, generally obtained from domain experts. Working with 3D medical data involves a significant time investment, particularly during slice by slice annotation. Similar to other supervised machine learning approaches, subjectivity in labeling is another challenge faced during 3D segmentation. Multiple domain experts might be required to decode the information corresponding to large 3D datasets. This may also introduce labeling noise in the training data and reduce the efficiency of volumetric segmentation.

In medical image analysis, generalized feature representation is often limited by the intra-class variance and inter-class similarity in large 3D datasets. Variations in image acquisition systems, contrast variations, and noise are the major reasons for overlapping features among different classes that critically influence the decision-making in unseen test data.

Another limiting factor in 3D medical image segmentation is class imbalance. The class imbalance may result in over-segmenting of the majority class of voxels due to its higher prior probability. Generally, 3D medical image segmentation is required for detecting abnormal or lesion regions from a scanned volumetric image. Hence, the ratio of voxels belonging to the abnormal and normal classes is extremely low and will seriously affect the training process. The class imbalance problem may be more critical in multi-class segmentation cases. Another challenge faced by the researchers while implementing 3D deep learning algorithms is training a 3D CNN model on a large volumetric dataset. This requires large memory and GPUs with extremely high computation capability.

Table 1: Overview of articles using 3D deep learning techniques for medical image segmentation.

Reference	Dataset	Remarks
3D extensions of 2D CNN architectures		
Cizec et al. [24]	Xenopus kidney dataset [25]	3D U-Net with sparsely annotated volumes.
Milletari et al. [26]	PROMISE2012 MRI dataset [27]	3D V-Net with residual blocks and strided convolution.
Bui et al. [28]	Infant brain MRI (iSeg [30])	D FCNN with dense blocks and strided convolution.
Kayalibay et al. [31]	Hand MRI [92], BRATS 2013 [93], and BRATS 2015 [52]	3D U-Net with deep supervision
Dou et al. [33]	CT Liver (SLiver07 [94]) & Heart MRI (HVSMR [56])	3D U-Net with deep supervision and CRF [34] based post-processing.
Li et al. [35]	Brain MRI (ADNI [95])	3D Res-Net with dilated convolution.
Chen et al. [37]	Multi-modal Brain MRI (MR-BrainS [38])	VoxResNet: a multi-modal dataset to learn from small datasets.
Niyas et al. [39]	Brain MRI (Private dataset)	3D Residual U-Net with shallow sliced stacking for generating 3D samples.
Schlemper et al. [40]	Abdominal CT datasets [42, 43]	Attention U-Net that can learn region specific features.
Wang et al. [45]	Brain MRI (CANDI [47], and IBSR [48])	3D U-Net with pyramid pooling [46] for fusing feature maps from different decoding levels.
Peng et al. [49]	Brain MRI (BRATS 2015 [52])	Multiple U-Net architectures with Xception [50] blocks.
Zhou et al. [51]	Brain MRI (BRATS 2015 [52], BRATS 2017[53], and BRATS 2018 [73])	OM-Net integrated with a cross-task guided attention (CGA) module.
3D Patch-wise Segmentation		
Yu et al.[54]	Heart MRI (HVSMR [56])	3D Fractal network with deep supervision.
Kamnitsas et al. [59]	Brain MRI (BRATS 2015 [52] and BRATS 2016 [60])	DeepMedic: Multi-scale 3D CNN with residual connections with 3D fully connected CRF.
Kamnitsas et al. [61]	Brain MRI (BRATS 2017 [53])	An ensemble of two DeepMedic models, three 3D FCNN models, and two 3D U-Net models.
Chen et al. [63]	Brain MRI (BRATS 2017 [53])	Multi-modal 3D CNN using 3D patches at two different scales.
3D segmentation using 2D CNN models		
Ciompi et al. [64]	CT Lung (MILD [96] dataset)	ConvNet: Multi-stream voxel classification in three orthogonal planes at three different scales.
Kitrungrotsakul et al. [65]	CT Liver (Private dataset)	VesselNet: Voxel classification in three orthogonal planes.
Mlynarski et al. [67]	Brain MRI (BRATS 2017 [53])	Used features from three orthogonal planes with 2D CNNs, and 3D image volumes.
Heinrich et al. [68]	CT datasets [76, 97]	OBELISK-Net that requires fewer trainable parameters.
Roth et al. [69]	Abdomen CT dataset [97]	A localization (using Holistically-nested convolutional networks (HNN) on three orthogonal planes) followed by a semantic segmentation stage.
Chen et al. [72]	Brain MRI (BRATS 2018 [73])	Separable 3D (S3D) U-Net: A parallel 2D convolutions with a residual inception architecture.
Rickmann et al.[74]	Multiple 3D medical datasets [75–78]	3D ConvNet with project & excite (PE) modules.
Semi-supervised Learning		
Mondal et al. [82]	Infant brain MRI (iSeg [30] & Multi-modal brain MRI (MR-BrainS [38])	The 3D GAN [83] that generates synthetic samples to train with labeled and unlabeled 3D volumes
Zhou et al. [84]	Abdomen CT (Private dataset)	Deep multi-planar co-training (DMPCT).
Yang et al. [85]	3D heart ultrasound (Private dataset)	A localization (using deep Q network (DQN)), followed by semantic segmentation (using a dual U-Net).
Li et al. [86]	3D heart MRI [98]	3D V-Net with signed distance map (SDM).
Wang et al. [87]	Thoracic CT scans [99, 100]	FocalMix: A model built on MixMatch [88] SSL framework.

3.3 Future Directions

Several of the above-discussed challenges cannot be eliminated from their source. In addition, it is impossible to always work with clean data. Hence, several studies are being conducted to design deep learning models that can circumvent such issues. Because the 3D medical image segmentation is primarily affected by the unavailability of labeled datasets, several recent studies have used semi-supervised deep learning over 3D medical images, which can use a mix of numerous unlabeled images with a few labeled images. Furthermore, such semi-supervised models can reduce the impact of label error and subjectivity to a certain extent. However, the successful 3D semi-supervised approaches are still in the premature phase and often use the 3D extension of 2D semi-supervised CNN models. Hence, developing sophisticated semi-supervised 3D CNN models for volumetric medical image segmentation is an open challenge for the research community.

Another possibility for improving the 3D segmentation performance is by reducing the huge class imbalance in 3D medical data. Several attempts [101–103] have been made to reduce class imbalance by using asymmetric loss functions to provide custom weights for different voxel classes. Although several alternative approaches, such as selective sampling and adaptive augmentation, are available to reduce the class imbalance problem, such strategies fail for highly imbalanced datasets. Hence, novel methods to eliminate class imbalance issues, especially in 3D medical data, are required. Future research in such techniques would be highly welcome. GANs-based data synthesis is another solution to resolve several of the current 3D (or 2D) medical image segmentation challenges. GANs-based analysis might be extremely relevant if it can generate real-like medical image volumes. Hence, such models for generating pseudo-medical data will be an interesting research topic in the future.

References

[1] Shouvik Chakraborty, Sankhadeep Chatterjee, Amira S Ashour, Kalyani Mali, and Nilanjan Dey. Intelligent computing in medical imaging: A study. In *Advancements in applied metaheuristic computing*, pages 143–163. IGI global, 2018.

[2] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Abubakar Malah Umar, Okafor Uchenwa Linus, Humaira Arshad, Abdullahi Aminu Kazaure, Usman Gana, and Muhammad Ubale Kiru. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 7:158820–158846, 2019.

[3] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[4] Anamika Dhillon and Gyanendra K Verma. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2):85–112, 2020.

[5] Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[6] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[8] Koichiro Yasaka, Hiroyuki Akai, Akira Kunimatsu, Shigeru Kiryu, and Osamu Abe. Deep learning with convolutional neural network in radiology. *Japanese journal of radiology*, 36(4):257–272, 2018.

[9] Berkman Sahiner, Aria Pezeshk, Lubomir M Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny H Cha, Ronald M Summers, and Maryellen L Giger. Deep learning in medical imaging and radiation therapy. *Medical physics*, 46(1):e1–e36, 2019.

[10] Geoff Currie, K Elizabeth Hawk, Eric Rohren, Alanna Vial, and Ran Klein. Machine learning and deep learning in medical imaging: intelligent imaging. *Journal of medical imaging and radiation sciences*, 50(4):477–487, 2019.

[11] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21, 2016.

[12] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.

[13] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017.

[14] Satya P Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan, and Balázs Gulyás. 3d deep learning on medical images: a review. *Sensors*, 20(18):5097, 2020.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[17] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[18] Eman Ahmed, Alexandre Saint, Abd El Rahman Shabayek, Kseniya Cherenkova, Rig Das, Gleb Gusev, Djamila Aouada, and Bjorn Ottersten. A survey on deep learning advances on different 3d data representations. *arXiv preprint arXiv:1808.01462*, 2018.

[19] Robert W Cox and Andrzej Jesmanowicz. Real-time 3d image registration for functional mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(6):1014–1018, 1999.

[20] Joseph V Hajnal and Derek LG Hill. *Medical image registration*. CRC press, 2001.

[21] Georgios Sakas. Trends in medical imaging: from 2d to 3d. *Computers & Graphics*, 26(4):577–587, 2002.

[22] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the Institute of Radio Engineers*, 109:820–838, 2021.

[23] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *2nd International Conference on Computational Sciences and Technology, (INCCST)*, pages 124–133, 2021.

[24] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[25] PD Nieuwkoop and J Faber. Normal table of xenopus laevis (daudin) garland. *New York*, 1994.

[26] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[27] Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Medical image analysis*, 18(2):359–373, 2014.

[28] Toan Duc Bui, Jitae Shin, and Taesup Moon. 3d densely convolutional networks for volumetric segmentation. *arXiv preprint arXiv:1709.03199*, 2017.

[29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[30] Li Wang, Dong Nie, Guannan Li, Élodie Puybareau, Jose Dolz, Qian Zhang, Fan Wang, Jing Xia, Zhengwang Wu, Jia-Wei Chen, et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: the iseg-2017 challenge. *IEEE transactions on medical imaging*, 38(9):2219–2230, 2019.

[31] Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*, 2017.

[32] Liwei Wang, Chen-Yu Lee, Zhuowen Tu, and Svetlana Lazebnik. Training deeper convolutional networks with deep supervision. *arXiv preprint arXiv:1505.02496*, 2015.

[33] Qi Dou, Lequan Yu, Hao Chen, Yueming Jin, Xin Yang, Jing Qin, and Pheng-Ann Heng. 3d deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis*, 41:40–54, 2017.

[34] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1529–1537, 2015.

[35] Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In *International conference on information processing in medical imaging*, pages 348–360. Springer, 2017.

[36] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7268–7277, 2018.

[37] Hao Chen, Qi Dou, Lequan Yu, Jing Qin, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, 170:446–455, 2018.

[38] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, et al. Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans. *Computational intelligence and neuroscience*, 2015, 2015.

[39] S Niyas, Vaisali S Chethana, Iwrin Show, T G Chandrika, S Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3d convolutional neural networks. *Biomedical Signal Processing and Control*, in press, 2021.

[40] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53:197–207, 2019.

[41] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[42] Roth, R. Holger, Amal Farag, E. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. Data from pancreas-ct. the cancer imaging archive. *IEEE Transactions on Image Processing*, 2016.

[43] H. Roth, Hirohisa Oda, Yuichiro Hayashi, Masahiro Oda, N. Shimizu, M. Fujiwara, K. Misawa, and K. Mori. Hierarchical 3d fully convolutional networks for multi-organ segmentation. *ArXiv*, abs/1704.06382, 2017.

[44] E. Thomas, S. Pawan, S. Kumar, Anmol Horo, S. Niyas, S. Vinayagamani, C. Kesavadas, and J. Rajan. Multi-res-attention unet : A cnn model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images. *IEEE journal of biomedical and health informatics*, PP, 2020.

[45] L. Wang, Cong Xie, and Nianyin Zeng. Rp-net: A 3d convolutional neural network for brain segmentation from magnetic resonance imaging. *IEEE Access*, 7:39670–39679, 2019.

[46] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:1904–1916, 2015.

[47] D. Kennedy, C. Haselgrove, S. Hodge, P. Rane, N. Makris, and J. Frazier. Candishare: A resource for pediatric neuroimaging data. *Neuroinformatics*, 10:319–322, 2011.

[48] T. Rohlfing, R. Brandt, R. Menzel, and C. Maurer. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21:1428–1442, 2004.

[49] Suting Peng, W. Chen, Jiawei Sun, and Boqiang Liu. Multi-scale 3d u-nets: An approach to automatic segmentation of brain tumor. *International Journal of Imaging Systems and Technology*, 30:17 – 5, 2020.

[50] F. Chollet. Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017.

[51] C. Zhou, Changxing Ding, X. Wang, Zhentai Lu, and D. Tao. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *IEEE Transactions on Image Processing*, 29:4516–4529, 2020.

[52] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[53] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The cancer imaging archive*, 286, 2017.

[54] Lequan Yu, Xin Yang, Jing Qin, and Pheng-Ann Heng. 3d fractalnet: dense volumetric segmentation for cardiovascular mri volumes. In *Reconstruction, segmentation, and analysis of medical images*, pages 103–110. Springer, 2016.

[55] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*, 2016.

[56] Danielle F Pace, Adrian V Dalca, Tal Geva, Andrew J Powell, Mehdi H Moghari, and Polina Golland. Interactive whole-heart segmentation in congenital heart disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 80–88. Springer, 2015.

[57] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[58] Hanna-Leena Halme, Antti Korvenoja, and Eero Salli. Isles (siss) challenge 2015: segmentation of stroke lesions using spatial normalization, random forest classification and contextual clustering. In *BrainLes 2015*, pages 211–221. Springer, 2015.

[59] Konstantinos Kamnitsas, Enzo Ferrante, Sarah Parisot, Christian Ledig, Aditya V Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Deepmedic for brain tumor segmentation. In *International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*, pages 138–149. Springer, 2016.

[60] Spyridon Bakas, Ke Zeng, Aristeidis Sotiras, Saima Rathore, Hamed Akbari, Bilwaj Gaonkar, Martin Rozycki, Sarthak Pati, and Christos Davatzikos. Glistrboost: combining multimodal mri segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In *BrainLes 2015*, pages 144–155. Springer, 2015.

[61] Konstantinos Kamnitsas, Wenjia Bai, Enzo Ferrante, Steven McDonagh, Matthew Sinclair, Nick Pawlowski, Martin Rajchl, Matthew Lee, Bernhard Kainz, Daniel Rueckert, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. In *International MICCAI Brainlesion Workshop*, pages 450–462. Springer, 2017.

[62] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.

[63] Lele Chen, Yue Wu, Adora M DSouza, Anas Z Abidin, Axel Wismüller, and Chenliang Xu. Mri tumor segmentation with densely connected 3d cnn. In *Medical Imaging 2018: Image Processing*, volume 10574, page 105741F. International Society for Optics and Photonics, 2018.

[64] F. Ciompi, K. Chung, S. J. Riel, A. A. A. Setio, P. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. Wille, A. Marchianó, U. Pastorino, M. Prokop, and B. Ginneken. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific Reports*, 7, 2017.

[65] Titinunt Kitrungrotsakul, Xian-Hua Han, Y. Iwamoto, Lanfen Lin, A. H. Foruzan, W. Xiong, and Yen-Wei Chen. Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 75:74–83, 2019.

[66] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[67] Pawel Mlynarski, H. Delingette, A. Criminisi, and N. Ayache. 3d convolutional neural networks for tumor segmentation using long-range 2d context. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 73:60–72, 2019.

[68] M. Heinrich, Ozan Oktay, and N. Bouteldja. Obelisk-net: Fewer layers to solve 3d multi-organ segmentation with sparse deformable convolutions. *Medical Image Analysis*, 54:1–9, 2019.

[69] Holger R Roth, Le Lu, Nathan Lay, Adam P Harrison, Amal Farag, Andrew Sohn, and Ronald M Summers. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis*, 45:94–107, 2018.

[70] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[71] Holger R Roth, Hirohisa Oda, Xiangrong Zhou, Natsuki Shimizu, Ying Yang, Yuichiro Hayashi, Masahiro Oda, Michitaka Fujiwara, Kazunari Misawa, and Kensaku Mori. An application of cascaded 3d fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, 66:90–99, 2018.

[72] W. Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and X. Qiao. S3d-unet: Separable 3d u-net for brain tumor segmentation. In *BrainLes@MICCAI*, 2018.

[73] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

[74] Anne-Marie Rickmann, Abhijit Guha Roy, Ignacio Sarasua, and Christian Wachinger. Recalibrating 3d convnets with project & excite. *IEEE transactions on medical imaging*, 39(7):2461–2471, 2020.

[75] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE transactions on medical imaging*, 38(2):540–549, 2018.

[76] Oscar Jimenez-del Toro, Henning Müller, Markus Krenn, Katharina Gruenberg, Abdel Aziz Taha, Marianne Winterstein, Ivan Eggel, Antonio Foncubierta-Rodríguez, Orcun Goksel, András Jakab, et al. Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: Visceral anatomy benchmarks. *IEEE transactions on medical imaging*, 35(11):2459–2475, 2016.

[77] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

[78] David N Kennedy, Christian Haselgrove, Steven M Hodge, Pallavi S Rane, Nikos Makris, and Jean A Frazier. Candishare: a resource for pediatric neuroimaging data, 2012.

[79] Xiaojin Zhu and Andrew B Goldberg. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130, 2009.

[80] Antti Rasmus, Harri Valpola, Mikko Honkala, Mathias Berglund, and Tapani Raiko. Semi-supervised learning with ladder networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, page 3546–3554, 2015.

[81] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 4080–4090, 2017.

[82] A. Mondal, J. Dolz, and Christian Desrosiers. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. *ArXiv*, abs/1810.12241, 2018.

[83] Ian J. Goodfellow, Jean Pouget-Abadie, M. Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, 2014.

[84] Yuyin Zhou, Yan Wang, Peng Tang, Song Bai, Wei Shen, Elliot Fishman, and Alan Yuille. Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 121–140. IEEE, 2019.

[85] Hongxu Yang, Caifeng Shan, Alexander F Kolen, et al. Deep q-network-driven catheter segmentation in 3d us by hybrid constrained semi-supervised learning and dual-unet. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 646–655. Springer, 2020.

[86] Shuailin Li, Chuyu Zhang, and Xuming He. Shape-aware semi-supervised 3d semantic segmentation for medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 552–561. Springer, 2020.

[87] Dong Wang, Yuan Zhang, Kexin Zhang, and Liwei Wang. Focalmix: Semi-supervised learning for 3d medical image detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3951–3960, 2020.

[88] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[89] S. Chen, K. Ma, and Y. Zheng. Med3d: Transfer learning for 3d medical image analysis. *ArXiv*, abs/1904.00625, 2019.

[90] S. Armato, G. McLennan, L. Bidaut, M. McNitt-Gray, C. Meyer, A. Reeves, B. Zhao, D. Aberle, C. Henschke, E. Hoffman, E. Kazerooni, H. MacMahon, Edwin J R Van Beeke, D. Yankelevitz, A. Biancardi, P. Bland, M. Brown, R. M. Engelmann, G. Laderach, D. Max, Richard C. Pais, D. Qing, R. Roberts, A. R. Smith, Adam Starkey, Poonam Batrah, P. Caligiuri, Ali O. Farooqi, G. Gladish, C. Jude, R. Munden, I. Petkovska, L. Quint, L. Schwartz, B. Sundaram, L. Dodd, C. Fenimore, D. Gur, N. Petrick, Jennifer Freymann, Justin Kirby,

B. Hughes, Alessi Vande Casteele, S. Gupte, Maha Sallamm, Michael D. Heath, Michael Kuhn, Ekta Dharaiya, Richard Burns, David Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, and B. Y. Croft. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38 2:915–31, 2011.

[91] Zongwei Zhou, Vatsal Sodha, Md Mahfuzur Rahman Siddiquee, Ruibin Feng, Nima Tajbakhsh, Michael B Gotway, and Jianming Liang. Models genesis: Generic autodidactic models for 3d medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 384–393. Springer, 2019.

[92] Agneta Gustus, Georg Stillfried, Judith Visser, Henrik Jörntell, and Patrick van der Smagt. Human hand modelling: kinematics, dynamics, applications. *Biological cybernetics*, 106(11):741–755, 2012.

[93] Michael Kistler, Serena Bonaretti, Marcel Pfahrer, Roman Niklaus, and Philippe Büchler. The virtual skeleton database: An open access repository for biomedical research and collaboration. *J Med Internet Res*, 15(11):e245, Nov 2013.

[94] Tobias Heimann, Bram Van Ginneken, Martin A Styner, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, György Bekes, et al. Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE transactions on medical imaging*, 28(8):1251–1265, 2009.

[95] Bradley T Wyman, Danielle J Harvey, Karen Crawford, Matt A Bernstein, Owen Carmichael, Patricia E Cole, Paul K Crane, Charles DeCarli, Nick C Fox, Jeffrey L Gunter, et al. Standardization of analysis sets for reporting results from adni mri data. *Alzheimer's & Dementia*, 9(3):332–337, 2013.

[96] Jesper H Pedersen, Haseem Ashraf, Asger Dirksen, Karen Bach, Hanne Hansen, Phillip Toennesen, Hanne Thorsen, John Brodersen, Birgit Guldhammer Skov, Martin Døssing, et al. The danish randomized lung cancer ct screening trial—overall design and results of the prevalence round. *Journal of Thoracic Oncology*, 4(5):608–614, 2009.

[97] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 556–564. Springer, 2015.

[98] Zhaohan Xiong, Qing Xia, Zhiqiang Hu, Ning Huang, Cheng Bian, Yefeng Zheng, Sulaiman Vesal, Nishant Ravikumar, Andreas Maier, Xin Yang, et al. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical Image Analysis*, 67:101832, 2021.

[99] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas van den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis*, 42:1–13, 2017.

[100] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

[101] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.

[102] Xiaoling Xia, Qinyang Lu, and Xin Gu. Exploring an easy way for imbalanced data sets in semantic image segmentation. In *Journal of Physics: Conference Series*, volume 1213, page 022003. IOP Publishing, 2019.

[103] Seyed Raein Hashemi, Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, Sanjay P Prabhu, Simon K Warfield, and Ali Gholipour. Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection. *IEEE Access*, 7:1721–1735, 2018.