Intelligent Oncology 1 (2025) 277–289

Contents lists available at ScienceDirect

Intelligent Oncology

journal homepage: www.keaipublishing.com/en/intonc/

Review

Deep learning in abdominal organ segmentation: A review

Peijun Hu[a], Jingsong Li[a,b], Yan Wang[c,*], Dexing Kong[d,*]

[a] Research Center for Scientific Data Hub, Zhejiang Lab, Hangzhou Zhejiang 311100, China
[b] College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou Zhejiang 310027, China
[c] School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China
[d] School of Mathematical Sciences, Zhejiang University, Hangzhou Zhejiang 310058, China

ARTICLE INFO

Keywords:
Abdominal organ segmentation
Deep learning
Prior knowledge
Supervised learning
Limited-supervised learning

ABSTRACT

Abdominal organ segmentation is an essential and fundamental medical procedure with many clinical and research applications. There is extensive variability in the size, location, and shape of the abdominal organs among individuals, and neighboring organs and structures exhibit similar textures and levels of intensity, which contribute to the difficulties encountered when developing robust, accurate, and automated segmentation approaches. In the past decade, deep learning (DL)-based methods have shown promising results based on a large amount of labeled data. However, acquiring large-scale images with high-quality annotations is both difficult and impractical in clinical practice. Furthermore, the images obtained from multicenters contain domain shift, which degenerates the model's performance on any new and unseen dataset. There have been extensive efforts to develop limited-supervised segmentation methods that can tackle this imperfect data problem. At the same time, prior knowledge from the medical domain may improve model performance while also constraining the results to an anatomically plausible range. In this paper, we provide a review of DL-based methods for abdominal organ segmentation that covers supervised and limited-supervised segmentation techniques, as well as the utilization of prior knowledge of abdominal organ and strategies in DL models. We present a categorized methodological overview of these approaches and summarize the relevant benchmarks and evaluation metrics used in this research area. Finally, we discuss the challenges and potential trends that may emerge in abdominal segmentation. Accordingly, this review systematically synthesizes advancements in DL for abdominal organ segmentation, provides relevant references for researchers in this field, and promotes the transformation of DL techniques into more precise clinical tools for this domain.

1. Introduction

The abdomen is a vital part of the human body and consists of those organs that are most important to human life, such as the liver, kidney, spleen, and pancreas. Medical imaging is a technique that provides a view of the internal parts of the human body, which allows for proper diagnosis, surgery planning, image-guided interventions, and cancer delineation and staging via image processing techniques. Different imaging technologies can be used to perform a medical image analysis, including different types of light (endoscopy, optical coherence tomography [OCT]), radioactive pharmaceuticals (nuclear medicine: positron emission tomography [PET], single photon emission computed tomography [SPECT]), magnets (magnetic resonance imaging [MRI]), sound (ultrasound), and X-rays (computed tomography, [CT]). Among the above-mentioned modalities, CT and MRI are the ones most often used to analyze abdominal data, and segmentation of the obtained images is the most fundamental task for performing a subsequent analysis and clinical application. The regions of interest (ROI) obtained by segmentation are used for image-guided interventions, radiological diagnosis, and radiotherapy.[1] In radiotherapy, the precise delineation of the target organ and others at risk is critical to ensure accurate irradiation.[2] The comprehensive analysis and quantitative measurement of both the shapes and volumes of the target organs serve as indicators of disorders.[3] In clinical practice, organs or structures are often manually delineated slice by slice, which is tedious, time-consuming, and labor-intensive. Table 1 presents the annotation time, annotation consistency, and main challenges associated with several typical organs and anatomical regions; in this case, annotation consistency is evaluated using the intersection-over-union metric. In some cases, the consistent and robust delineation of ROIs remains challenging due to

* Corresponding authors.
E-mail addresses: wycq2006@sina.com (Y. Wang), dkong@zju.edu.cn (D. Kong).

https://doi.org/10.1016/j.intonc.2025.06.007
Received 21 April 2025; Received in revised form 29 June 2025; Accepted 29 June 2025
Available online 10 July 2025
2950-2616/© 2025 Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Table 1
Annotation time, annotation consistency, and challenges associated with several typical organs.

Organ & imaging modality	Annotation time (min/case)	Annotation consistency (IoU)	Main annotation challenges
Brain MRI	15–30	0.75–0.85	Blurred gray matter/white matter boundary, difficult to locate small lesions
Lung CT	15–30	0.65–0.78	Unclear boundaries of ground-glass nodules, vascular interference
Liver CT	20–40	0.70–0.82	High tumor heterogeneity, complex portal vein segmentation
Heart MRI	30–50	0.68–0.80	Motion artifacts, thin ventricular wall structure
Prostate MRI	25–30	0.72–0.83	Difficult to distinguish central/peripheral zone boundary, multiple tumor foci

Abbreviations: 1. IoU: Intersection over Union 2. MRI: magnetic resonance imaging 3. CT: computed tomography.

the ill-defined boundary and is prone to intra- and inter-observer variabilities. In the past few decades, extensive efforts have been advanced towards developing robust and automatic segmentation methods and algorithms to aid in the process of contouring organ boundaries.

Image segmentation tasks can be divided into semantic segmentation and instance segmentation. Semantic segmentation is a pixel-level classification task that assigns each pixel to a particular category, while instance segmentation not only requires pixel-level segmentation as a first step, but also distinguishes instances in each category. Most abdominal image segmentation refers to semantic segmentation, through which each organ and tissue type is treated as belonging to a different category. Semantic segmentation does not distinguish between instances, as in the case of liver tumor segmentation.

Early abdominal organ segmentation process can be dived into image-based methods and learning-based methods. The most frequently used image-based methods for abdominal organ segmentation include thresholding,[4] region growing,[5] level set-based methods,[6] and graph cut-based methods,[7,8] which mainly rely on low-level image information, such as intensity and gradience. As a result, these methods are vulnerable to the challenges of leakage and under-segmentation caused by similar intensities or appearances around the boundaries and inhomogeneous regions inside organs. Learning-based methods for abdominal organ segmentation mainly consists of statistical shape models[3,9,10] and atlas-based methods.[11–13] Both types rely on prior knowledge, shape model, or probability maps to capture anatomical information about the target organs and have shown good accuracy and robustness in multi-organ segmentation. However, with limited training data, it is difficult to construct a typical shape prior to the procedure that can cover the significant variability of deformable organs, such as the liver. For the atlas-based method, it is critical to perform atlas selection and label fusion before segmentation, and the approach relies heavily on the registration accuracy of the abdominal images. To alleviate the initialization effect and improve localization accuracy, landmark detection and random forest regression[14,15] have also been applied in abdominal organ segmentation. Nevertheless, reliable anatomical landmarks are difficult to determine for deformable organs, and the handcrafted image feature extraction and selection processes are crucial to the determination of the results.

Compared to the above-mentioned traditional methods, deep learning (DL) models have exhibited state-of-the-art performance in the past decade. With the powerful capacity for feature representation and training on large amounts of data, DL models, such as convolutional neural networks (CNN), have realized automated segmentation and achieved reasonable accuracy, despite the extensive variability in the size, shape, location, and texture of the abdominal organs. However, a critical barrier to its widespread adoption is the domain shift, which refers to the discrepancy in data distribution between the training (source domain) and deployment (target domain) environments. This shift often leads to degraded model performance when applied to unseen data, thus limiting its clinical reliability. Domain shift arises from multiple technical and biological sources. For instance, technical sources include variations in scanner hardware (e.g., 1.5 T vs. 3.0 T MRI) and acquisition protocols (e.g., T1-weighted vs. T2-weighted sequences), while biological sources involve differences in patient anatomy (e.g., age, pathology). Notably, these sources do not affect all organs uniformly. For example, the pancreas, a slender organ (mean thickness: 1–3 cm) adjacent to gas-filled intestines, is highly sensitive to scanner-induced artifacts (e.g., magnetic susceptibility artifacts in 3.0 T MRI), which blur its boundaries. In contrast, the liver, a large and homogeneous soft-tissue organ, exhibits greater robustness to such artifacts due to its size and uniform tissue composition. Despite these critical differences, existing studies often treat domain shift as a generic problem, neglecting organ-specific vulnerabilities.

Several comprehensive surveys of DL in general medical image segmentation[16–21] and in abdominal organ segmentation in particular have been conducted.[22–31] For instance, Wang et al.[17] reviewed the evolution and development of DL model application in medical image segmentation in terms of the categories of supervised learning and weakly supervised learning. Taghanaki et al.[18] categorized the leading DL-based image segmentation methods into six main groups: deep architectural, data synthesis-based, loss function-based, sequenced, weakly-supervised, and multi-task methods, providing a comprehensive review of the contributions in each of these groups. In a related vein, Tajbakhsh et al.[19] considered solutions for the two major dataset limitations, which consist of scarce annotations and weak annotations, in medical image segmentation. Hesamian et al.[20] looked at network structures, training techniques and challenges in medical image segmentation. Peng et al.[21] surveyed medical image segmentation with limited supervision, including situations with a limited amount of annotations, sparse annotations, and inaccurate annotations. While each one of these works attended to some aspect of medical image segmentation, several surveys have focused in particular on single abdominal organ segmentation[22–27] and multi-organ segmentation.[28–31] For instance, Fu et al.[30] classified DL-based multi-organ segmentation methods into pixel-wise classification and end-to-end segmentation methods. The pixel-wise classification methods were further dived into auto-encoder and CNN approaches, whereas the end-to-end segmentation methods were divided into fully convolutional network (FCN), region-based FCN, generative adversarial networks (GAN), and synthetic image-guided segmentation. Likewise, Li et al.[31] provided an overview of traditional and DL-based abdominal multi-organ segmentation, in which the DL methods were divided into supervised and semi-supervised methods. As an alternative approach, Senkyire et al.[29] surveyed supervised and semi-supervised segmentation methods, and then summarized their applications in single abdominal organ and multi-organ cases.

Despite the proliferation of systematic reviews of medical image segmentation with DL, research that has focused specifically on abdominal organs is relatively scant, and a systematic classification for existing methods is lacking. In clinical practice, abdominal organ segmentation is vulnerable to the challenges of insufficient or inaccurate labeling. In more recent years, substantial effort has been put into developing segmentation models and algorithms with imperfect annotated medical image data based on limited-supervised learning methods. Concurrently, prior knowledge of this medical domain can

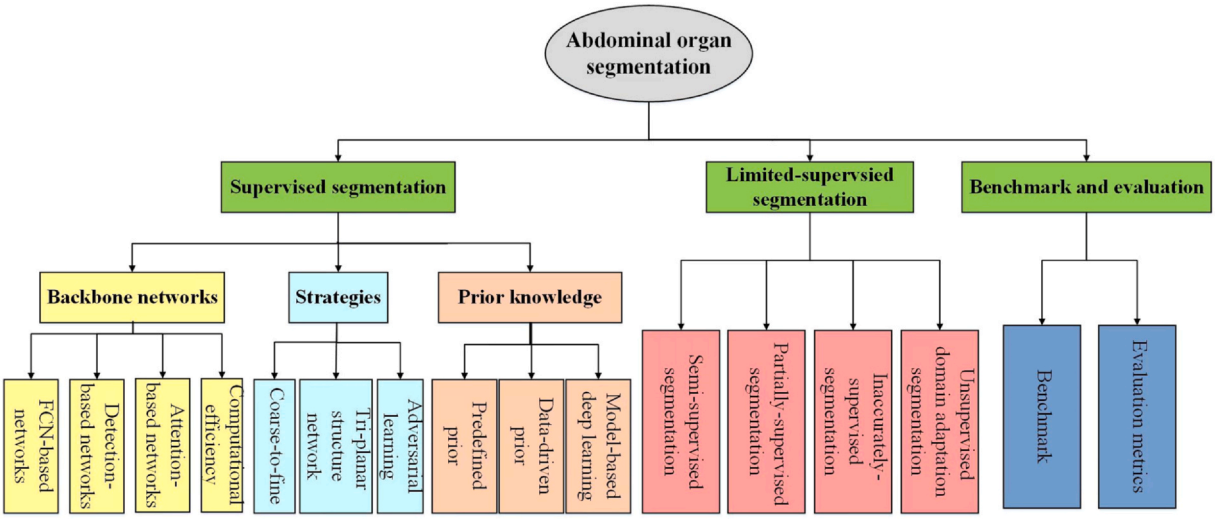


Figure 1. Overview of the main structure of this article.

improve model performance and constrain the results within the anatomically plausible range. Overall, how to represent and incorporate prior knowledge into DL model has attracted a significant amount of attention. In this paper, we review DL methods with prior knowledge of abdominal organ segmentation. We include single and multi-organ segmentation methods and categorize current DL methods into supervised and limited-supervised learning. In the section on supervised segmentation methods, we first introduce basic deep neural network (DNN) architectures with state-of-the-art strategies; we then investigate prior knowledge usage in abdomen images. The limited-supervised methods are classified into semi-supervised, partially-supervised, inaccurately-supervised, and unsupervised domain adaptation segmentation according to the available data annotations. In addition, we provide a summary of publicly available benchmarks and evaluation metrics related to abdominal organ segmentation. An overview of the main structure of this survey is provided in Figure 1.

The studies cited in this review have been collected from various databases, including Google Scholar, PubMed, and Web of Science. Keywords included but were not limited to “organ segmentation,” “deep learning,” “abdominal organ segmentation,” “semi-supervised segmentation,” and “domain adaptation.” Most of these works were published between 2012 and 2022. In this paper, we also included methods related to “multi-atlas,” “active shape,” and “active contour” models that have been combined with DL methods. Papers that did not focus on abdominal organ segmentation and those not published in English were not included.

The rest of this paper is organized as follows. Section 2 reviews supervised segmentation in abdominal organs, including backbone network architecture and strategies, and examines the prior knowledge used with DL. Section 3 focuses on limited-supervised segmentation approaches, including semi-supervised, partially-supervised, inaccurate-supervised, and unsupervised domain adaptation segmentation. Section 4 summarizes the public available benchmarks and related evaluation metrics on abdominal organ segmentation. Section 5 presents a discussion and possibilities for future work, and Section 6 draws a conclusion.

2. Supervised segmentation

In medical image segmentation, deep supervised methods are widely used and achieve high segmentation accuracy. With annotations of organs, DNN can fully learn the specific features related to certain targets. Within the scope of abdominal organ segmentation, which consists of individual and multiple organ segmentation, prior knowledge and anatomical information are often included to improve model performance. In this section, we first review the backbone network architecture types deployed in medical image segmentation. The widely-used networks for image segmentation can be divided into FCN-based models, detection-based models, and attention-based models. Then, we discuss the strategies that are frequently combined with these networks. In addition, we consider the prior knowledge incorporated into DL models for abdominal organ segmentation.

2.1. Backbone networks

Medical image segmentation based on DL can be categorized into pixel-wise classification and end-to-end region segmentation. Early DL-based methods performed segmentation by using sliding image patches that covered the entire image and predicted a label for each pixel.[32] CNNs are a commonly-used type of network backbone architecture for pixel-wise segmentation. A typical CNN consists of convolutional layers, activation functions, pooling layers, batch normalization layers, and dropout layers. Image segmentation is viewed as a classification task for each pixel, and the image batch located at the center of the target pixel is fed into the network to generate a class prediction for this pixel. Later, end-to-end networks in the encoder-decoder architecture are applied in image segmentation after the introduction of FCNs.[33] In contrast to CNN, FCN replaces the last fully connected layer of CNN with convolutional layers, which transform the feature maps into a prediction map that is the same size as the input image. The encoder in the FCN gradually reduces the feature maps and captures high-level features, and the decoder gradually recovers the spatial information and fine boundary. In order to localize the results, high resolution features from the contracting path are combined with the upsampled output. The prediction map then provides a one-to-one classification for each pixel in the input image.

FCN-based networks. Based on the FCN, improved networks with a similar encoder-decoder architecture and other special structures have been proposed for medical image segmentation, such as U-Net,[34] V-Net,[35] SegNet,[36] DenseNet,[37] and DeepLab.[38,39] The U-Net[34] modified and extended the FCN such that it could work with very few training images and displayed more precise segmentations. The upsampling part of U-Net has a large number of feature channels, which allow the network to propagate contextual information to higher resolution layers. The expansive path and contracting path are almost symmetric, yielding a U-shaped architecture. The segmentation performance of U-Net across different organs is listed in Table 2, which highlights the model’s adaptability to organ-specific characteristics. In contrast, SegNet[36] follows an FCN architecture that consists of an encoder network, a decoder network, and pixel-wise classification layers. The

Table 2
Segmentation performance of U-Net across different organs.

Organ	Dataset	Dice score	HD_{95}(mm)
Liver	LiTS (CT)	0.94 ± 0.02	2.1 ± 0.5
Pancreas	MSD-Pancreas (CT)	0.72 ± 0.04	4.2 ± 1.1
Kidney	KiTS19 (CT)	0.92 ± 0.03	2.0 ± 0.6
Spleen	MSD-Spleen (CT)	0.87 ± 0.02	2.8 ± 0.7

Abbreviations: 1. CT: computed tomography 2. MSD: Medical Segmentation Decathlon 3. HD95: 95% Hausdorff distance.

decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. The upsampled maps, which are sparse, are then convolved with trainable filters to produce dense feature maps. As another alternative, DenseNet[37] uses densely-connected convolutional layers, which can alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. Atrous spatial pyramid pooling with multiple dilated rates was introduced into CNNs to form the DeepLab[38,39] model, which can generate a larger field of view without increasing model parameters. DeepLabv3[39] relied on multi-proportional atrous convolution to capture multi-scale information in cascade or in parallel.

Detection-based networks. Some methods perform image segmentation within an object detection framework. The region-based convolutional neural network (R-CNN)[40] uses region proposals alongside the CNN framework for target detection and is further expended for semantic segmentation. According to the flowchart, the input image is obtained, and the candidate region is extracted. The candidate regions are entered into CNN separately, and the output of the CNN is then entered into the support vector machine for category determination. Faster R-CNN[41] introduces a region proposal network that shares full-image convolutional features with the detection network, which achieves rapid and precise detection. Based on Faster R-CNN, Mask R-CNN[42] was proposed for object detection and segmentation by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. As an example of application, Shu et al.[43] used a region proposal network and a prebackground classification subnetwork in Mask R-CNN for multi-organ segmentation.

Attention-based networks. The attention mechanism adaptively weighs features according to the input and has played an increasingly important role in computer vision tasks, including organ segmentation. Some studies have introduced the attention mechanism into CNN models to improve segmentation performance. For instance, Attention U-Net[44] incorporated a self-attention gating module into the CNN to suppress irrelevant regions and highlight salient features for the pancreas segmentation task. The variance-aware attention U-Net[45] for multi-organ CT segmentation was then proposed, in which variance uncertainty was embedded into an attention architecture. In a similar vein, Tang et al.[46] designed a combined 2D and 3D multi-organ segmentation framework, in which the self-attention mechanism was used to control spatial contextual information extracted from a low-resolution 3D model to guide high-resolution 2D segmentation. A multiscale attention mechanism with a U-shaped network[47] has also been proposed for liver segmentation. In another model, the guided self-attention mechanism[48] was used to integrate local features with corresponding global dependencies and adaptively highlight interdependent channel maps. This approach was applied in abdominal organs, cardiovascular structures, and brain tumors. Jin et al.[49] also developed a 3D hybrid attention-aware segmentation method, in which attention residual modules were integrated into the U-Net so that the attention-aware features could change adaptively.

In more recent years, transformers[50] designed for sequence-to-sequence prediction have emerged as alternative architectures that employ dispense convolution operators that rely solely on attention mechanisms. These transformers can model global contexts to solve the limitations of CNN-based approaches in modeling long-range relations. The vision transformer[51] was developed based on image patches, rather than pixels, as the units of information in an image. Taking 2D image patches with positional embeddings as inputs and pre-training on a large dataset, vision transformer achieved comparable performance with CNN-based methods. With the TransUNet,[52] a hybrid CNN-transformer architecture was utilized for medical image segmentation, in which the transformer encoded tokenized image patches from a CNN feature map as the input sequence for extracting global contexts. The TransUNet, evaluated in multi-organ segmentation and cardiac segmentation tasks, showed superior performance over the U-Net and Attention U-Net. A U-Net-like pure transformer called Swin-Unet[53] was then developed for medical image segmentation. The tokenized image patches were fed into the transformer-based U-shaped encoder-decoder architecture with skip-connections for local-global semantic feature learning. Then, a symmetric Swin transformer-based decoder with a patch expanding layer was built to perform the upsampling operation. The Swin-Unet was validated on a multi-organ segmentation task, achieving Dice similarity coefficient (DSC) 0.85 on BTCV (MICCAI 2015 multi-atlas beyond the Cranial Vault). However, in leveraging global attention to model long-range dependencies, transformers then underperform on small organs like the pancreas. The pancreas, with its small size (mean volume about 60 ml) and blurred boundaries, has local details (e.g., acinar structures) easily diluted by global attention, leading to imprecise boundaries. Therefore, some hybrid models have been developed to balance the global context with local features. For instance, Xie et al.[54] proposed CoTr to efficiently bridge the CNN and transformers; in this case, the CNN was constructed to extract feature representations, and a deformable transformer was built to model the long-range dependency on the extracted feature maps. The deformable transformer payed attention only to a small set of key positions by introducing the deformable self-attention mechanism. The results in multi-organ segmentation based on the BTCV dataset showed an improvement in accuracy over other CNN-based, transformer-based, and hybrid methods.

Computational efficiency. Computational efficiency is critical for clinical translation. For 3D abdominal CT (512 × 512 × 100), traditional CNNs (e.g., U-Net) take about 20 h to train (GPU: NVIDIA A100) with 5 s/inference; transformers (e.g., Swin-Unet) require about 40 h for training and 8 s/inference due to self-attention complexity; and hybrid models (e.g., CoTr) balance CNN local feature extraction with transformer global modeling, exhibiting a training time of 25 h and 6 s/inference between the two. Small organ segmentation (e.g., pancreas) further reduces efficiency by about 30% due to higher input resolution demands.

2.2. Strategies

Based on the backbone networks, strategies such as coarse-to-fine segmentation,[55] deep supervision, and adversarial learning have been suggested for application in specific segmentation tasks. While a 3D network for volumetric data can obtain inter-slice information and usually secure superior accuracy, it requires high memory and more computation time. Thus, the volumetric scan image is often sliced into sub-volumes and fed into networks for training and inference. For example, Gibson et al.[56] presented a DL-based segmentation algorithm based on Dense V-Networks and applied it on eight abdominal organs, which demonstrated the superiority of the registration free method over multi-atlas label fusion. The method was evaluated on 3T MRI images for organ-at-risk segmentation. In a related vein, Zhang et al.[57] utilized a cascaded V-Net multi-class network for multi-organ localization and a single-class network for individual organ segmentation. In liver segmentation, a 3D U-Net variant with residual blocks was proposed to automatically segment the abdominal organs and tissue structures.[58] To overcome challenges in 3D network modeling, a patch-based network with random spatial initialization and statistical fusion on overlapping ROIs was developed.[59]

An alternative approach to overcome the high memory problem in 3D models is to introduce 2.5D images composed of axial, coronal, and sagittal planes of volumetric data with a tri-planar structure network[60] or recurrent networks.[61–63] For example, Li et al.[64] proposed a multi-dimensional cascaded net for multi-organ segmentation, in which a 3D net and a 2.5D net were used to process downsampled CT volume and original image slices, respectively. In pancreas segmentation, a two-stage coarse-to-fine strategy and multi-scale feature representation[65] were used to overcome the small target and class imbalance problem. Likewise, Hu et al.[66] developed a two-stage coarse-to-fine segmentation method in which the coarse segmentation was converted into saliency information based on geodesic distance transformation to guide the fine segmentation. In an earlier study, Cai et al.[62] used a recurrent neural network (RNN) process to slice images and maintain spatial inconsistency across adjacent image slices in the pancreas segmentation task.

Adversarial learning[67] has attracted a lot of attention since its introduction, which is also widely used in image segmentation. The GAN architecture consists of a generation subnetwork (G) and a discriminant subnetwork (D). The basic idea is to train the networks iteratively in a dynamic game process, letting G trick D by generating a realistic image with random noise, while D tries to distinguish the fake images generated by G from real images. The idea of adversarial learning can be applied in image segmentation to decrease the gap between the predicted segmentation and the ground truth. In abdomen segmentation, Li et al.[68] used an adversarial U-Net with a multi-level pyramidal pooling module to achieve a better gradient information flow for improving pancreas segmentation performance. Some studies have considered image synthesis and image segmentation within a GAN architecture to improve data usage and model performance on cross-modality images. For example, Cai et al.[69] proposed a cycle- and shape-consistent generative adversarial network to perform cross-modality organ translation and segmentation. The method realized image translation between CT and MRI with two generators supervised by a cycle-consistent loss and a combination of an adversarial loss and a shape-consistency loss, which can provide augmented training with dual-domain data. The segmentation was performed by training dual-modality segmentation networks for different domain data. In this case, the discriminators were trained to distinguish between real and synthetic data.

2.3. Prior knowledge

Although DL-based segmentation techniques have achieved remarkably high performance, they are prone to generating problematic artifacts, such as fragmented structures and topological inconsistencies, since the basic neural network models do not take local geometric features, such as smoothness and anatomical shape, into consideration, especially when the model is trained on an insufficiently labeled dataset. Prior information about region, shape, and anatomy are commonly used in traditional segmentation models, which can constrain the model and help generate anatomically plausible results. In recent years, prior knowledge has been combined with DL models through pre-processing, post-processing, or a unified learning framework to improve segmentation accuracy and anatomical consistency. Prior knowledge can be divided into the categories of predefined prior, specific prior learned from target data, or statistically prior learned from training data. In this section, we review DL methods combined with prior knowledge and categorize them following the source of the latter. It should be noted that a similar study[70] reviewed DL with anatomy information for medical image segmentation.

2.3.1. Predefined prior

Predefined priors are often properties associated with the region, shape, and topology of objects, which are commonly taken as objects of regularization in traditional models, such as the active contour model (ACM)/level set, Markov random field (MRF)/conditional random field (CRF), and graph cut. The boundaries of organs are assumed to be smooth or have the shortest length among all the evolution curves. Another type of predefined priors consists of assumptions about objects with regular shapes, such as an elliptical shape,[71] star shape,[72] or convexity.[73] In addition, regional relationships, such as inclusion, exclusion, and neighboring, have been used in the case of abdominal organs.

To improve the boundary smoothness of a segmented region through CNN, a large category of methods have relied on MRFs/CRFs or ACM/level sets after the CNN as a form of post-processing. Given the initial rough segmentation by CNNs, MRFs/CRFs or ACM/level set models interpret the segmentation as the initialization and output refined results. In liver segmentation, Cai et al.[61] used two networks for the initial segmentation of the pancreas, which was then fed into CRF for fine segmentation. Christ et al.[74] cascaded two FCNs for liver and tumor segmentation, respectively, and then used a dense 3D CRF to refine the segmentations. In a similar vein, Dou et al.[75] proposed a 3D deeply supervised network, which was followed by a CRF to refine the segmentation results. Lu et al.[76] first localized the liver through a 3D CNN, after which a graph cut algorithm was applied to refine the contours of the liver. In an earlier study, Hu et al.[77] combined 3D CNN and surface evolution for liver segmentation, wherein prior information about the liver was calculated from the initial liver region predicted by the CNN, and then the prior was incorporated into an improved surface evolution model. Likewise, Hoogi et al.[78] used a CNN to initialize and adaptively estimate the parameters of ACM for liver lesion segmentation, Alirr[79] used an FCN followed by a localized level set to segment the liver and tumor. A similar approach[80] was proposed for pancreas segmentation.

Although obtaining prior knowledge through post-processing is easy to implement and takes advantage of the automated localization of CNNs and the regularization of traditional models, the two-stage processing approach is prone to issues associated with image-based models, such as boundary leakage and under-segmentation. Some studies have begun integrating CRF models directly into neural networks in an end-to-end training way. In such cases, each iteration of the CRF step is formulated as a stack of CNN layers, and multiple iterations can be then implemented by repeatedly executing this stack as an RNN. For example, Xu et al.[81] combined 3D FCN with CRF-RNN to perform bladder segmentation.

Except for the smoothness property of organ boundaries, predefined priors of general shapes, such as an elliptical shape,[71] star shape,[72] convexity,[73] and volume preservation,[82] have also been used.[83] The encoding methods for anatomical knowledge primarily include graph convolutional networks (GCNs) and statistical shape models. Graph convolutional networks model anatomical structures through nodes (organs) and edges (spatial relationships), enabling explicit constraints on positional relationships between organs. They perform exceptionally well in multi-organ segmentation. Statistical shape models, on the other hand, learn the statistical distribution of organ shapes via principal component analysis, making them suitable for precise segmentation of regular-shaped organs (e.g., the liver). However, they lack generalization capability for irregular organs (e.g., the pancreas). In summary, GCNs are more suitable for multi-organ spatial constraints, while statistical shape models excel at single-organ shape refinement.

In addition, the inter-relationships between organs and information about size can be interpreted as high-level prior and incorporated into the loss function of DL models.[84] The star shape prior was encoded into the loss function of FCN for end-to-end skin lesion segmentation.[85] The morphology prior, defined in terms of the general shape and size of the pancreas from multiphase CT images, was integrated into a U-Net model for multi-phase pancreas segmentation.[86] Shi et al.[87] proposed exclusion loss in the case of organs not overlapping in multi-organ segmentation. In cardiac segmentation, Kervadec et al.[88] introduced linear constraints of target-region size into the loss function of CNNs. The area of the target organ was estimated from the predicted probability map output by the CNN, and a penalty loss function about organ

size was added into the loss function. Despite the value of this work, other studies using these kinds of priors in abdominal organ segmentation are limited. One potential reason is that the applicable conditions for these priors remain unclear. For instance, in the case of the spleen, statistical shape priors were shown to improve the Dice score from 88.2% to 91.5% (with the 95% Hausdorff distance decreasing from 4.1 mm to 3.4 mm); conversely, for the colon, statistical shape priors excessively constrained shape variability, leading to a 2.3% reduction in the Dice score.

2.3.2. Data-driven prior

Some studies have shown how specific priors can be learned in data-driven way, with prior information being utilized through the loss function or the two-stage pre-/post-processing method. One class of methods constructs shape priors with statistical approaches. Other studies have taken advantage of the shape representation provided by the active shape model (ASM) and applied them in the post-processing of initial CNN segmentation. For example, Tabrizi et al.[89] used a fuzzy-ASM model to produce final segmentation based on the initial bounding box of a kidney provided by a CNN in 3D ultrasound. Ma et al.[90] integrated the segmentation from a DNN and statistical shape model into a Bayesian model for pancreas segmentation.

Another cluster of methods has shown how shape priors can be incorporated into DL models within a union framework or as a form of pre-/post-processing. For example, Mohagheghi et al.[91] demonstrated the learning of global information regarding 3D liver shapes in a low-dimensional latent space with a denoising auto-encoder, and the learned knowledge was defined as a loss function combined with the Dice loss in the model. Ravishankar et al.[92] proposed a ultrasound kidney segmentation approach, wherein a U-Net and a convolutional auto-encoder were trained end-to-end. The auto-encoder model was pre-trained on noisy augmented shapes and was then plugged into a cascading architecture to update the segmentation network through a custom loss function. In another study, Huang et al.[93] designed a deep atlas prior based on the probabilistic approach and incorporated the deep atlas prior loss and other likelihood losses into the Bayesian loss. The proposed method was applied in liver and spleen segmentation. In an earlier study, Zhao et al.[94] registered an image to the template for localization and obtained probabilistic atlases, which were taken as knowledge to aid in the FCN-based pancreas segmentation. More recently, Lee et al.[95] showed how voxel-wise relations from the same class could be enforced and projected to the same point with contrastive learning in a Siamese network, achieving a 2% improvement of the DSC value in multi-organ segmentation. Finally, Guo et al.[96] incorporated spatial regional and boundary information learned by a CNN model into ASM to get fine segmentation.

2.3.3. Model-based DL

Traditional models can easily incorporate domain knowledge, but they may not be optimal for specific tasks and have difficulty with parameter tuning. In contrast, DL models can be trained effectively and efficiently, but they have difficulties incorporating domain knowledge. As a result, some researchers have combined model-based and data-based DL approaches to capitalize on the advantages of both; these approaches have been referred to as model-based DL methods.[97,98] For instance, the inference system was implemented using a DNN similar to DL methods, but the architecture was tailored specifically for a scenario of interest based on a suitable model-based method. For example, in compressive sensing magnetic resonance imaging, the algorithm of alternating direction method of multipliers was unfolded into a deep network for efficient training and parameter learning.[99] However, despite its appeal, this type of approach is rarely found in image segmentation.

An alternative method is to represent a model-based algorithm with neural building blocks (i.e., the sub-networks of a DNN carry out the specific computations of the different building blocks constituting the model-based algorithm). In a multi-atlas segmentation framework, Ding et al.[100] designed a bi-directional registration network and a similarity estimation network for multi-modality image registration and label fusion, respectively, which was validated for liver segmentation. Yang et al.[101,102] proposed a deep Fusion-Net consisting of a feature extraction subnet and non-local patch-based label fusion subnet, which solved the important steps of atlas selection and label fusion in the multi-atlas segmentation. The method was then applied to cardiac MR images. Likewise, DeepSSM[103] uses a CNN that simultaneously localizes the biological structure of interest, establishes correspondences, and projects these points onto a low-dimensional shape representation in the form of principal component analysis loadings within a point distribution model.

3. Limited-supervised segmentation

Despite its outstanding performance, supervised segmentation requires large amounts of reliable annotated data and faces the challenge of domain shift when the trained model is applied to new unseen datasets. However, the annotation of medical images requires clinical expertise from doctors or radiologists to ensure the reliability of the training data, which restricts the size of the training set for medical image segmentation. Limited-supervised segmentation with imperfect training data can solve this problem to some extent. Limited-labeling can consist of (1) a small subset of labeled data together with large-scale unlabeled images; (2) partial annotations, including sparse labeling; and/or (3) inaccurate annotations including noisy labels. The extreme case of limited-supervised segmentation would consist of segmentation with no labeled data in the target image domain. Cross-modality or cross-domain segmentation uses annotated data from the source domain to train models for target domain data in a weakly-supervised way. In this section, we divide limited-supervised segmentation of abdomen images into semi-supervised, partially-supervised, inaccurately-supervised, and unsupervised domain adaptation segmentation, which correspond to insufficient labeling, partial labeling, inaccurate labeling, and unlabeled data in the target domain, respectively.

3.1. Semi-supervised organ segmentation

The semi-supervised segmentation approach is applied in those cases in which only a small subset of training data has annotations, and most data are unlabeled. The aim of semi-supervised segmentation is to effectively utilize a large quantity of unlabeled data and less manually annotated data to improve model performance. Suppose we denote $X=\{X_L,X_U\}$ as the entire dataset, including a small labeled subset $X_L=\{x_i\}_{i=1}^{L}$ with labels $Y_L=\{y_1,y_2,\ldots,y_L\}$ and a large-scale unlabeled dataset $X_U=\{x_i\}_{i=1}^{U}$, $L \ll U$. We can then assume that the labels are K classes, and the basic semi-supervised problem can be formulated as

$$\min_{\theta} \sum_{x\in X_L, y\in Y_L} L_s(x, y, \theta) + \alpha \sum_{x\in X_U} L_u(x, \theta) + \beta \sum_{x\in X} \mathscr{R}(x, \theta) \tag{1}$$

where L_s denotes the supervised loss, L_u denotes the unsupervised loss, and $\mathscr{R}$ denotes the regularization (e.g., the consistency loss or designed regularization). The unsupervised loss and regularization terms are not strictly distinguishable from each other. The parameter θ denotes the model parameters, whereas α and β denote the trade-off coefficients. Semi-supervised learning[104] can then be categorized into a generative method, a consistency regularization approach, a graph-based method, or self-training/pseudo-labeling. More recently, few-shot learning has been adopted to generalize a model to new-class datasets with only a few annotated data. Here, we provide an overview of the semi-supervised segmentation methods applied to abdominal organs. A more sufficient and detailed survey of semi-supervised learning methods can be found in related studies.[21,104]

Generative adversarial learning in semi-supervised segmentation involves training two subnetworks (i.e., the generator and the

discriminator) as an ordinary GAN model does. The discriminator is trained to identify whether a sample is a true label or generated label, and the generator is a semantic segmentation network that generates a segmentation probability map that is not distinguished by the discriminator. The discriminator can also discover trustworthy regions in predicted labels for unannotated data.[105] For example, Huang et al.[93] developed a semi-supervised liver segmentation model with adversarial learning, in which the discriminator network was trained to distinguish between the segmentation maps and the ground truths and thus able to choose a reliable mask for unannotated samples. In addition to the incorporation of prior knowledge of liver shape, a deep atlas prior loss was defined and combined with a Bayesian loss.

Consistency regularization assumes that a model should generate consistent predictions for similar inputs with perturbation. In this vein, Fang et al.[106] proposed a difference minimization network (DMNet) for semi-supervised kidney segmentation. To use unlabeled data, the DMNet adopted two decoder branches and minimized the difference between the soft masks generated by the two decoders. In this manner, each decoder can learn under the supervision of the other decoder and can thus be improved at the same time. In another study, Huang et al.[107] proposed co-train weight-averaged models for learning a unified multi-organ segmentation network from few-organ datasets. The two networks were trained collaboratively and were allowed to teach one another on unlabeled organs. In addition, the weighted-average models were adopted to alleviate the noisy teaching supervision and produce more reliable soft labels.

Graph-based semi-supervised learning relies on the basic assumption that a graph can be extracted from the raw dataset, in which each node represents a training sample, and each edge denotes some similarity measurement of the node pair. This method uses specific criteria (e.g., node similarity metrics or topological features) to propagate the information of a label on the graph, and its performance is dependent on the construction of the graph. As an example, Soberanis-Mukul et al.[108] utilized a semi-supervised GCN learning strategy to refine segmentation by CNN. The uncertainty analysis of the CNN was used to define a partial-labeled graph, and the refinement problem was formulated as a semi-supervised graph learning problem. This method was validated on a National Institutes of Health (NIH)-pancreas and Medical Segmentation Decathlon (MSD)-kidney dataset, showing better refinement performance than the CRF.

Self-training,[109] also known as pseudo-labeling, often relies on a teacher-student paradigm in which a teacher model with labeled data is learned, and pseudo-labels are generated on the unlabeled data, which are then combined with labeled data to train the student model. The loss function of pseudo-labeling for a neural network can be expressed simply as

$$\mathscr{L} = \frac{1}{n}\sum_{m=1}^{n}\sum_{i=1}^{K}\mathscr{R}(y_i^m, f_i^m) + \alpha\frac{1}{n'}\sum_{m=1}^{n'}\sum_{i=1}^{K}\mathscr{R}(y_i^{'m}, f_i^{'m}) \tag{2}$$

where n and n' are the numbers of labeled data and unlabeled data in a mini-batch, f_i^m and $f_i^{'m}$ correspond to the predicted output in the labeled data and unlabeled data, and y_i^m and $y_i^{'m}$ are the ground truth and pseudo label for the labeled data and unlabeled data, respectively. Parameter α is the coefficient balancing the supervised and unsupervised loss terms. The student model is trained on noisy data with the pseudo label. Some studies have selected the pseudo label to refer to the highest confidence in the ground truth of the unlabeled data. Lee et al.[110] proposed a semi-supervised method, in which a discriminator was trained on an unlabeled dataset with a quality assurance (QA) score, and the predicted QA was included in the loss and back propagated to increase the model's performance from QA-only data. Raju et al.[111] proposed co-heterogenous and adaptive segmentation that fused appearance-based semi-supervision, mask-based adversarial domain adaptation, and pseudo-labeling. This framework can train a model on a small labeled cohort of single phase data and adapt it to any unlabeled cohort of heterogenous multi-phase data, with possibly new clinical scenarios and pathologies. Likewise, Huang et al.[112] designed a semi-supervised network that used anatomy priors and unlabeled data in a self-training and multi-task learning manner. The anatomy prior was learned from the probabilistic atlas, and an image reconstruction task aided the network training.

Co-training is an extension of self-training and is also applied in medical image segmentation. Co-training exploits multi-view data descriptions to learn from a limited number of labeled examples and a large amount of unlabeled data. This method is based on the assumption that each sample has two different and complementary views, and each view is sufficient for training a good classifier. Then, the classifiers are applied to predict each view of the unlabeled data, and the most confident candidates are labeled for the other model. The procedure is iteratively repeated until the unlabeled data are exhausted, or some condition is reached. In this vein, Xia et al.[113] proposed an uncertainty-aware multi-view co-training framework, in which co-training was applied on unlabeled data by enforcing the consistency of multi-view segmentation results with an uncertainty estimation. The method was evaluated on an NIH pancreas segmentation dataset and a multi-organ segmentation dataset. Likewise, Zhou et al.[114] developed Multi-Planar CoTraining, through which they trained different models on the sagittal, coronal, and axial planes of a volumetric image in the teacher-student paradigm; in this case, it was found that multi-planar fusion can help improve pseudo labels and train a better student model.

Few-shot learning is a machine learning technique that addresses situations in which an existing model needs to generalize to an unknown semantic class with only a few labeled examples by utilizing previously acquired knowledge.[115] In few-shot segmentation, a segmentator learns via a training set $\mathscr{D}_{tr}$ consisting of annotated training classes $\mathscr{C}_{tr}$ and segments a testing image from the set $\mathscr{D}_{te}$ consisting of unseen classes $\mathscr{C}_{te}$ given only one or few reference labeled images, while there are no overlaps between $\mathscr{C}_{tr}$ and $\mathscr{C}_{te}$. For example, suppose that the training set consists of labels $C_{tr}=\{liver,kidney,spleen\}$ and the model needs to segment new classes $C_{te}=\{pancreas\}$ in the testing set. Let N be the number of semantic classes in C_{te}, and K be the number of examples for each semantic class in C_{te}; the few-shot learning problem can be also referred to as N-way K-shot learning.

In recent years, several studies has adopted few-shot learning for abdominal organ segmentation with scarcely annotated data. In a few-shot segmentation framework, Roy et al.[116] applied channel squeeze & spatial excitation blocks to facilitate efficient interaction between the conditioner and the segmenter arm, which allowed for no reliance on a pre-trained model. The method was validated on organ segmentation from the VISCERAL dataset.[117] Sun et al.[118] used a global correlation network with discriminative embedding, which utilized the geometrical and morphological patterns that appear in medical images. The method was evaluated on abdominal MR and CT images for multi-organ segmentation. With distance metric learning, Cui et al.[119] proposed a low-shot (one- and few-) image segmentation framework that learned multimodal representation for each category and performed dense predictions based on cosine distances between the pixels' deep embeddings and the category representations. In a related study, Makarevich et al.[120] presented Meta-MedSeg based on gradient-based meta-learning for volumetric medical data segmentation. The learning in this case consisted of two steps of meta-training and fine-tuning. In each round of meta-training, a set of tasks consisting of images and their corresponding segmentation maps were sampled and fed to the U-Net model, and the learned weights were aggregated. The final model was used as an initialization of the fine-tuning step. This method was evaluated on four organ segmentation tasks, including spleen segmentation.

3.2. Partially-supervised segmentation

The insufficient labeling scenario in semi-supervised segmentation can be extended to partial-labeling, which means that the annotation is

sparse or partial. Partially-labeled data in an abdominal organ refers to cases in which the abdominal images contain only annotated data for the target organ in specific tasks (e.g., the pancreas), while leaving the rest of the organs in the background. When these labeled datasets for different research objects are combined together to perform multi-organ segmentation, the annotations are supposed to be partial. Another requirement is to alleviate the burden of 3D volumetric labeling, as only 2D slices of the axial axis of each 3D image are annotated, which can be referred as sparse labeling.

For partially-labeled datasets, Zhang et al.[121] took nnU-Net as a backbone and introduced a conditioning strategy by feeding auxiliary information into the decoder architecture as an additional input layer. In another study, Petit et al.[122] identified pixels with the correct label and trained the FCN model with a multi-label loss. In addition, the iterative confidence self-training approach was used to relabel the missing pixel labels. This method was performed similarly to a model trained on all labels, even for large missing label proportions. Petit et al.[122] solved the problem of partial-labeling by using a strategy to identify pixels with correct labels and trained FCNs with a multi-label loss adapted to this context. Then, an iterative confidence self-training approach was designed to relabel the missing pixel labels relying on an uncertainty measure learned by a confidence network. The method was validated on the TCIA and LiTS datasets for multiple organ/structure segmentation. The Prior-Aware Neural Network (PaNN)[123] was proposed to handle such background ambiguity via the incorporation of prior knowledge on organ size distributions. In this case, the authors used a small fully-labeled dataset to estimate the prior, which was incorporated into PaNN via partially-labeled datasets. Jiang et al.[124] studied partially supervised multi-organ medical image segmentation, aiming to develop a unified semantic segmentation model by utilizing multiple partially labeled datasets, wherein each dataset provides labels for one organ category. Using 2D axial sparse labeling of 3D abdominal organs, Bitarafan et al.[125] proposed a self-training framework that consists of two steps: assigning pseudo annotations of unlabeled voxels and updating 3D segmentation network by employing both the labeled and pseudo labeled voxels. To improve pseudo-labeling accuracy, the propagation of labels (or pseudo-labels) between adjacent slices and the 3D processing of voxels were performed with a 2D registration method.

3.3. Inaccurately-supervised segmentation

Segmentation with inaccurate or imprecise annotations refers to the scenario in which true labels are corrupted with noises, or the annotation is not pixel-wise dense but rather is composed of bounding-box annotations without a precise object boundary. The label noise may come from inter- and intra-observer variability or erroneous annotations by non-experts or computers, and the imprecise annotations may stem from the clinical practice. Training models on inaccurate data may result in erroneous predictions in clinical tasks. However, correcting all the annotations is expensive and non-practical. Given the importance of a clean label to DL models, and the fact that label noise or corruption is common in practice, there have been many efforts to resolve the issue of noisy label learning.[126] However, it is still understudied in the medical image analysis community, especially with regard to abdominal organ segmentation.[127]

The methods used to handle label noise can be categoried into label cleaning and pre-processing, network architectures, label-noise-robust loss functions, data re-weighting, data and label consistency, and training procedures.[128] Specifically, Min et al.[128] developed a label-noise-robust method that jointly trained two separate models and updated the models only on data samples in which the predictions of the two models were not consistent. In addition, different attention modules were inserted at various depths in the networks to use the gradient information in different feature maps, which helped identify any down-weight samples with noisy labels. The method was validated on the segmentation of cardiac and glioma MR images. In abdominal segmentation, Wang et al.[129] developed an annotation-efficient deep-learning framework to deal with imperfect training datasets, including limited annotations, a lack of target domain annotations, and noisy annotations. The framework realized cross-model co-optimization learning by training two networks in parallel and conducting cross-model information exchange. With the exchanged information, self-label filtering and the correction of inexpensive noisy labels proceeded with cascaded local and global steps progressively in an elaborately designed schedule according to an observed small loss criterion. Abdomen MR images from the CHAOS challenge were used to evaluate the framework. Wang et al.[130] first used the GAN to generate pseudo labels for training images under the implicit high-level shape constraint. Then, a discriminator-guided generator channel calibration module that employed the discriminator's feedback was built to calibrate the generator for better pseudo labels. To learn from the pseudo labels that were noisy, a noise-robust iterative learning method was further introduced using noise-weighted Dice loss. To combat the noisy label in 3D image segmentation, Shi et al.[131] proposed a learning architecture containing three components. First, a slice-level label-quality awareness module was introduced to automatically generate label quality scores for slices in a set. Then, a shape-awareness regularization loss based on distance transform maps was proposed to introduce prior shape information. Third, a re-weighting strategy was used to weaken the negative effects caused by noisy labels. The proposed learning architecture was validated on abdominal liver and multi-organ segmentation tasks.

3.4. Unsupervised domain adaptation segmentation

In clinical applications that contain multiple modalities or datasets, with different imaging qualities being required for specific tasks, training models on the easier "annotated dataset" and then transferring the model to the "difficult dataset" is a solution to overcome the labeling problem. However, there may be a domain shift between the images from different medical institutes and those with different modalities. To mitigate the performance degradation by domain shift, some studies have focused on cross-modality or cross-domain segmentation. For example, in low field 3D stomach MRI segmentation,[132] CycleGAN[133] was used to construct and learn the mapping relationship between high and low field intensity MRI and to overcome domain shift. Then, the images generated by the high field intensity MRI through the CycleGAN were taken as extended data to aid the 3D Res-UNet segmentation network and the low filed MRI. In the work of Cone-beam CT-guided adaptive pancreatic radiotherapy,[134] a pre-trained CycleGAN was applied to generate synthetic CT images given by Cone-beam CT images. Then, a mask-scoring R-CNN was applied on the synthetic CT for final segmentation.

Single modalities (e.g., CT) are susceptible to scanning parameters. To define objects in multimodal images for clinical diagnosis, Yuan et al.[135] developed a GAN based unified framework to perform modality translation and multimodal segmentation in a single network, and the method was evaluated on abdominal organ segmentation. Chen et al.[136] described an unsupervised domain adaptation framework that conducted the synergistic alignment of domains from both image and feature perspectives; this approach was applied in cross-modality adaptation between MRI and CT images. Yang et al.[137] improved cross-device generalization ability through multimodal fusion in the diagnosis of diabetic retinopathy. Similar methods can be transferred to abdominal organ segmentation to enhance the model's robustness against differences in scanning protocols by combining CT-MRI features. In addition, adaptive DL architectures (e.g., self-supervised learning) can adapt to new data distributions by dynamically adjusting the model,[138] which is an alternative approach to handling variability in abdominal scans.

In conclusion, we present the performance comparison of several typical DL algorithms mentioned above in Table 3.

Table 3
The performance comparison of five typical deep learning algorithms.

Method type	Representative algorithm	Core advantages	Limitations
Supervised-CNN	U-Net	Strong local feature extraction, high efficiency	Weak global dependency modeling
Supervised-transformer	Swin-Unet	Superior global context modeling	High computational cost, small organ detail loss
Semi-supervised	Pseudo-labeling	Utilizing unlabeled data to improve performance	Pseudo-label noise affects models
Unsupervised domain adaptation	CycleGAN Transfer	Adapting to cross-modal data	Performance degrades with large modality gaps

4. Benchmark and evaluation

This section outlines the abdominal organ benchmarks and evaluation metrics used to measure the accuracy of automated segmentations.

4.1. Benchmark

To provide a fair comparison and promote progress in developing abdominal segmentation methodologies, many related abdomen datasets with annotated organs have been utilized. There are several public datasets and benchmarks for abdominal organ segmentation. For the liver, there are MICCAI-Sliver07 (MICCAI liver segmentation challenge),[139] LiTS (Liver Tumor Segmentation Benchmark),[140] 3DIRCADB (3Dircadb dataset from Research Institute Against Digestive Cancer)[141] and Medical Segmentation Decathlon-liver.[142] For the kidney, there is KiTS19 (The Kidney Tumor Segmentation Challenge).[143] For the pancreas, there are NIH-pancreas from[60] the National Institutes of Health Clinical Center and MSD-pancreas.[142] For the spleen, there is MSD-spleen.[142] These datasets are mostly in the CT or MRI modalities. For abdominal multi-organ segmentation, the datasets from VISCERAL,[117] BTCV,[144] CHAOS (Combined (CT-MR) Healthy Abdominal Organ Segmentation challenge),[145] and AbdomenCT-1k have been used.[146] The popular publicly available abdominal organ segmentation benchmarks in the past few decades are summarized in Table 4.

The MICCAI-Sliver07 dataset[139] consists of 30 CT scans, among which 20 scans are for training with ground truth and 10 scans are for testing with hidden ground truth. Datasets in this database are pathological and made up of tumors, metastases, and cysts of varied sizes. All the volumes of the MICCAI-Sliver07 dataset have an in-plane resolution of 512×512 pixels. The inner-slice pixel spacing ranges from 0.576 mm to 0.812 mm, and the slice thickness ranges from 0.7 mm to 5.0 mm. The number of slices in every case spans from 64 to 394.

The LiTS[140] includes 131 CT volumes with liver and liver tumor annotations, along with 70 testing cases whose labels are hidden. The tumor types have variable tumor contrast levels (hyper-/hypo-intense), abnormalities in tissues (metastasectomy), sizes, and amounts of lesions. In this benchmark, the best liver segmentation algorithm achieved a Dice score of 0.96 (MICCAI), whereas for tumor segmentation, the best algorithm was evaluated at 0.67 (ISBI) and 0.70 (MICCAI).

The 3DIRCADB dataset[141] is composed of 3D CT-scans from 10 women and 10 men with hepatic tumors in 75% of cases. The size and number of tumor lesions and their contrast enhancement differentiates the CT volumes substantially. Each image slice of the CT volume has a resolution of 512×512 width and height, and the number of slices lies between 74 and 260.

The CHAOS[145] contains 80 subjects, of whom 40 patients have a single CT scan and 40 have MRI scans. All the CT and MRI datasets include healthy abdominal organs that have no pathological abnormalities (such as tumors and metastasis). CHAOS includes five tasks: liver segmentation from CT-MRI, liver segmentation from CT, liver segmentation from MRI, four organ segmentation from CT-MRI, and four organ segmentation from MRI.

The KiTS19[143] includes 300 patients with CT scans. Among them, 210 training cases have kidney and kidney tumor annotations, and 90 testing cases have hidden annotations. On the KiTS19 challenge, the best team achieved a kidney DSC of 0.974 and a tumor DSC of 0.851 on the 90 testing cases.

The NIH-pancreas dataset[60] is the most-widely used benchmark for pancreas segmentation. It consists of 80 subjects with CT scans, among whom 17 subjects are healthy kidney donors scanned prior to nephrectomy and the remaining 65 patients neither had major abdominal pathologies nor pancreatic cancer lesions.

The MSD[142] provides 10 datasets with 17 target ROI to be segmented. Related abdominal organs in the MSD include the liver, pancreas, and spleen. The MSD-liver consists of 201 contrast-enhanced CT images from patients with primary cancers and metastatic liver disease, which contain annotations of the liver and liver tumors. The numbers of training and testing data were 131 and 70, respectively. The MSD-pancreas has 421 portal-venous phase CT scans of patients undergoing resection of pancreatic masses. The pancreatic parenchyma and pancreatic mass were annotated. The number of training and testing data were 282 and 139, respectively. The MSD-spleen contained 61 portal venous phase CT scans from patients undergoing chemotherapy treatment for liver metastases, and the spleen was labeled as ROI. In this dataset, the numbers for training and testing data were 41 and 20 respectively.

Table 4
List of the popular abdominal organ segmentation benchmark datasets.

Dataset name(abbr.)	Target	Modality	No. of Training/Test	Source and year
MICCAI-Sliver07[139]	Liver	CT	20/10	MICCAI 2007
3DIRCADB[141]	Liver	CT	20	–
LiTS[140]	Liver and liver tumor	CT	131/70	ISBI & MICCAI 2017
KiTS19[143]	Kidney and kidney tumor	CT	210/90	MICCAI 2019
NIH-pancreas[60]	Pancreas	CT	80	TCIA 2015
CHAOS[145]	Liver and multi-organ	CT/MRI	20/20	ISBI 2019
MSD-pancreas[142]	Pancreas and tumor	CT	282/139	MICCAI 2018
MSD-spleen[142]	Spleen	CT	41/20	MICCAI 2018
MSD-liver[142]	Liver	CT	131/70	MICCAI 2018
BTCV[144]	13 organs or structures	CT	20/10	MICCAI2015
VISCERAL[117]	20 anatomical structures	CT	80/40	ISBI & ECIR 2015
AbdomenCT-1K[146]	Liver, kidney, spleen, pancreas	CT	1112	MICCAI 2021

The BTCV[144] from the Vanderbilt University Medical Center focuses on multi-atlas segmentation with abdominal and cervix CT scans and was later widely-used for multi-organ segmentation evaluation. The benchmark has annotations of 13 organs or structures: spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, portal vein and splenic vein, pancreas, right adrenal gland, and left adrenal gland. The VISCERAL anatomy benchmark[117] provides a cloud-based infrastructure for multi-modal medical image analysis, including the abdominal organs with CT and MRI. Each modality has 30 clinical scans, with 20 for training and 10 for testing.

The AbdomenCT-1k[146] provides the largest dataset that consists of 1,112 annotated CT scans from 12 medical centers. It was based on previous benchmark datasets with more organ annotations, including LiTS,[140] MSD,[142] KiTS,[143] and NIH-pancreas.[60] The labeled organs include the liver, spleen, kidney, and pancreas. In addition, it established four organ segmentation benchmarks for fully-supervised, semi-supervised, weakly-supervised, and continual learning.

4.2. Evaluation metrics

To fully analyze the performance of different abdominal segmentation methods, various evaluation metrics have been used in previous studies, including volume-based metrics, such as the DSC, the Jaccard index (JAC), relative absolute volume difference (RAVD), and surface distance-based metrics, such as the HD_{95}, average symmetric surface distance (ASSD), and maximum symmetric surface distance (MSSD). The DSC measures the relative volumetric overlap between segmentations. It is mostly used to measure overall segmentation performance in recent years. There is no overlap between the segmented region and ground truth when the value of DSC is 0, and it is a perfect segmentation when DSC = 1. The JAC is a similar metric that is also known as intersection-over-union, which calculates the area of the overlap between the ground truth and the predicted segmentation. RAVD non-symmetrically calculates the volume difference between the predicted volume and ground truth. It reflects whether a method performs over- or under-segmentation compared to the ground truth.

The left three metrics measure the surface error by computing the symmetric distance from predicted segmentation to ground truth segmentation surface. HD_{95} quantifies the maximum 95 percentile distance, ASSD quantifies the mean surface distance, and MSSD reflects the maximum surface distance between two volumes. The detailed definitions are as follows.

- DSC is defined as

$$DSC = 2 \times \frac{|X \bigcap Y|}{|X| + |Y|}. \tag{3}$$

- JAC is defined as

$$JAC = \frac{|X \bigcap Y|}{|X \bigcup Y|} \tag{4}$$

- RAVD is defined as

$$RAVD = \left(abs \frac{|Y| - |X|}{|X|} \times 100\% \right). \tag{5}$$

- The HD_{95} is defined as

$$HD_{95}(X, Y) = max_{k95\%}[d(X, Y), d(Y, X)]. \tag{6}$$

- The ASSD is defined as

$$ASSD(X, Y) = \frac{1}{|\Omega_X| + |\Omega_Y|} \left(\sum_{x \in \Omega_X} d(x, Y) + \sum_{y \in \Omega_Y} d(y, X) \right). \tag{7}$$

- The MSSD is defined as

$$MSSD(X, Y) = \max\{ \max_{x \in \Omega_X} d(x, Y), \max_{y \in \Omega_Y} d(y, X) \}. \tag{8}$$

where X and Y refer to ground truth and algorithm segmentation, respectively. Ω_X and Ω_Y denote the set of pixels at the surface of segmentation. $d(x, Y) = min_{y \in \Omega_Y} ||x - y||_2$, $d(X, Y)$ is the total distances between X and Y.

5. Discussion and future work

By reviewing DL-based abdominal organ segmentation literature conducted in the recent decade, we found that abdominal organ segmentation is a highly active research area, and state-of-the-art DL and computer vision methods have been quickly applied to this area. DL-based methods have shown outstanding performance based on high-quality and large-scale annotated data. However, there are two main challenges that obstruct the clinical usage of abdominal organ segmentation algorithms. First, acquiring a large scale of annotated medical images for every medical image segmentation task is difficult and impractical, especially for high-quality annotations. Thus, the model's performance degenerates on a new dataset, especially on those with a domain shift. Second, the segmentation predicted by DL-based models usually lacks region, shape, or topology regularization, which may not be anatomically plausible. To solve the annotated data problem, many models and algorithms focus on training DL-based models with imperfectly-labeled training data. To improve the regularization and robustness of segmentation results, multiple methods try to combine model-based approaches with DL-based methods. From the reviewed literature, we can see that the following topics are currently understudied.

Annotation data for segmentation. Reducing annotation workload is a core challenge in advancing the clinical application of segmentation models. Future implementation can be achieved through the following strategies: (1) Weakly supervised learning by utilizing sparse annotations (e.g., labeling only key slices) combined with self-supervised pre-training (e.g., contrastive learning) to reduce full-annotation requirements; (2) Active learning by selecting the most informative samples for annotation via uncertainty sampling, which can improve annotation efficiency by 5–10 times; and (3) Cross-modal transfer by training models on annotation-rich modalities (e.g., CT) and transferring them to annotation-scarce modalities (e.g., MRI) to lower annotation costs. Preliminary experiments show that combining weakly supervised learning with active learning can reduce the number of annotations required for pancreatic segmentation from 80 fully annotated cases to 8, with the DSC decreasing by only 2%.

DL models with prior knowledge. The prior knowledge in the reviewed works is mainly about boundary smoothness and statistical shape priors, and these priors are often combined with DL models through two-stage pre-/post-processing or loss function. However, two-stage methods are prone to have accumulated errors, or they may be restricted by limitations of DL models or traditional models. Learning specific shape priors and incorporating them into a union end-to-end framework has exhibited promising practicability for clinical use. Previous works have used auto-encoders to represent shape priors in

low-dimensional latent space and restrict the object shapes through customer loss function.[91,92] However, these studies mainly center on large organs, including the liver and kidney. Modeling small organs and structures with irregular shapes requires further study.

The other trend of using prior knowledge is to represent high-level priors regarding object size, topology, shape, and inter-region constraints by loss functions.[84] The design of prior loss should consider the differentiability and computational complexity of the loss term. The variational and optimization-based segmentation approaches can be considered an ideal representation of how to model prior information about object shape and appearance. However, there are limited studies using high-level prior losses, especially for abdominal organs that have large size variability and irregular shapes. Future research is warranted to address multiple organ segmentation with priors about shape, size, topology, and inter-regional relationships.

Model-based DL models for abdominal organ segmentation. Model-based DL[98] aims to tackle the limitations associated with model-based methods and black-box DL, and these methods combine principled mathematical models with data-driven systems to benefit from the advantages of both approaches. The strategies of designing model-based DL can be divided into model-aided networks and DNN-aided inference systems. Model-aided networks mean that a specific DNN tailored for the problem is designed by following the operations of suitable model-based methods. DNN-aided inference uses conventional model-based methods, but its specific parts of the model-based algorithm are augmented with DL tools, allowing the resulting system to implement the algorithm while learning to overcome partial or mismatched domain knowledge from data. By incorporating domain knowledge, model-based DL requires less data and has better interpretability. Thus, these methods are promising to solve the prior knowledge usage and data labeling problem in abdominal organ segmentation. However, the study of model-based DL in abdomen segmentation remains relatively scant.

Limited-supervised methods for imperfect annotated data. Due to the difficulties of requiring data and precise annotations for medical data, it is challenging to design robust models with imperfect training data. The imperfect labeling can be insufficient labeling, noisy labeling, partial labeling, while imperfect data refer to diverse data with domain shifts or a limited amount of data. It is worth developing limited-supervised learning methods and taking advantage of domain knowledge to solve the data limitation issue for better applications in clinical usage.

6. Conclusions

In this review, we have surveyed DL-based methods in abdominal organ segmentation, as well as the function of prior knowledge combined with DL models, public datasets, and evaluation metrics. According to the available data annotations, we divided DL methods into supervised segmentation and limited-supervised segmentation. For supervised segmentation, we identified widely-used network backbones and strategies in medical image segmentation and looked into prior knowledge that was used in these models. For limited-supervised segmentation, we provided a taxonomy of limited-supervised learning, which consists of semi-supervised segmentation, partially-supervised segmentation, inaccurately-supervised segmentation, and unsupervised domain adaptation segmentation. We also summarized current benchmarks and public datasets for abdominal single/multi-organ segmentation and the evaluation metrics used in most studies. As a conclusion, we discussed the future directions of research into abdominal organ segmentation.

CRediT authorship contribution statement

Peijun Hu: Writing – original draft. **Jingsong Li:** Supervision. **Yan Wang:** Writing – review & editing. **Dexing Kong:** Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dexing Kong reports financial support, article publishing charges, travel, and writing assistance were provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (Grant No.: 12090020, 12090025, 12101571), the Natural Science Foundation of Zhejiang Province (Grant No.: LQ20H180001), the Major Scientific Project of Zhejiang Lab (Grant No.: 2022ND0AC01) and the Natural Science Foundation of Chongqing (Grant No.: CSTB2023NSCQ-LZX0054).

References

1. Chen X, Sun S, Bai N, et al. A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. *Radio Oncol.* 2021;160:175–184.
2. Molitoris JK, Diwanji T, Snider III JW, et al. Advances in the use of motion management and image guidance in radiation therapy treatment for lung cancer. *J Thorac Dis.* 2018;10(Suppl 21):S2437–S2450.
3. Linguraru MG, Pura JA, Pamulapati V, Summers RM. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Med Image Anal.* 2012;16(4):904–914.
4. Seo KS, Kim HB, Park T, Kim PK, Park JA. *Automatic liver segmentation of contrast enhanced CT images based on histogram processing. Advances in Natural Computation: First International Conference, ICNC 2005, Changsha, China, August 27-29, 2005, Proceedings, Part I 1*. Springer; 2005:1027–1030.
5. Ruskó L, Bekes G, Németh G, Fidrich M. Fully automatic liver segmentation for contrast-enhanced CT images. *MICCAI Wshp 3D Segm Clin A Gd Chall.* 2007;2(7):143–150.
6. Shehata M, Khalifa F, Soliman A, et al. *A level set-based framework for 3D kidney segmentation from diffusion MR images. 2015 IEEE International Conference on Image Processing (ICIP)*. Springer; 2015:4441–4445.
7. Peng J, Hu P, Lu F, Peng Z, Kong D, Zhang H. 3D liver segmentation using multiple region appearances and graph cuts. *Med Phys.* 2015;42(12):6840–6852.
8. Linguraru MG, Pura JA, Chowdhury AS, Summers RM. *Multi-organ segmentation from multi-phase abdominal CT via 4D graphs using enhancement, shape and location optimization. International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010:89–96.
9. Saito A, Nawano S, Shimizu A. Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs. *Med Image Anal.* 2016;28:46–65.
10. Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y. Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. *Med Image Anal.* 2015;26(1):1–18.
11. Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE Trans Med Imaging.* 2013;32(9):1723–1730.
12. Xu Z, Burke RP, Lee CP, et al. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Med Image Anal.* 2015;24(1):18–27.
13. Tong T, Wolz R, Wang Z, et al. Discriminative dictionary learning for abdominal multi-organ segmentation. *Med Image Anal.* 2015;23(1):92–104.
14. Gauriau R, Cuingnet R, Lesage D, Bloch I. Multi-organ localization with cascaded global-to-local regression and shape prior. *Med Image Anal.* 2015;23(1):70–83.
15. Criminisi A, Robertson D, Konukoglu E, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med Image Anal.* 2013;17(8):1293–1303.
16. Chi W, Ma L, Wu J, Chen M, Lu W, Gu X. Deep learning-based medical image segmentation with limited labels. *Phys Med Biol.* 2020;65(23):10.1088/1361-6560/abc363.
17. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: a survey. *IET Image Process.* 2022;16(5):1243–1267.
18. Taghanaki AbhishekSAsgari, Cohen, Cohen-Adad JP, Hamarneh J, Deep G. semantic segmentation of natural and medical images: a review. *Artif Intell Rev.* 2021;54(1):137–178.
19. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. *Med Image Anal.* 2020;63:101693.
20. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging.* 2019;32(4):582–596.

21. Peng J, Wang Y. Medical image segmentation with limited supervision: a review of deep network models. *IEEE Access.* 2021;9:36827–36851.
22. Yao X, Song Y, Liu Z. Advances on pancreas segmentation: a review. *Multimed Tools Appl.* 2020;79(9):6799–6821.
23. Torres HR, Queiros S, Morais P, Oliveira B, Fonseca JC, Vilaca JL. Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: a systematic review. *Comput Methods Prog Biomed.* 2018;157:49–67.
24. Bao R, Darzi E, He S, et al. Foundation AI model for medical image segmentation. *arxiv preprint arxiv:*2411.02745. Published online 2024.
25. Moghbel M, Mashohor S, Mahmud R, Saripan M. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artif Intell Rev.* 2018;50(4):497–537.
26. Kumar H, DeSouza SV, Petrov MS. Automated pancreas segmentation from computed tomography and magnetic resonance images: a systematic review. *Comput Methods Prog Biomed.* 2019;178:319–328.
27. Summers RM. Progress in fully automated abdominal CT interpretation. *AJR Am J Roentgenol.* 2016;207(1):67–69.
28. Rehman A, Khan FG. A deep learning based review on abdominal images. *Multimed Tools Appl.* 2021;80(20):30321–30352.
29. Senkyire IB, Liu Z. Supervised and semi-supervised methods for abdominal organ segmentation: a review. *Int J Autom Comput.* 2021;18(6):887–914.
30. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. A review of deep learning based methods for medical image multi-organ segmentation. *Phys Med.* 2021;85:107–122.
31. Li Q, Song H, Chen L, Meng X, Yang J, Zhang L. An overview of abdominal multi-organ segmentation. *Curr Bioinform.* 2020;15(8):866–877.
32. Zhang W, Li R, Deng H, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage.* 2015;108:214–224.
33. Long J, Shelhamer E, Darrell T. *Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015; 2015:3431–3440.
34. Ronneberger O, Fischer P, Brox T. *U-net: convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention.* Springer; 2015:234–241.
35. Milletari F, Navab N, Ahmadi SA. *V-Net: fully convolutional neural networks for volumetric medical image segmentation. 4th Int Conf 3D Vis (3DV).* Springer; 2016:565–571.
36. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(12):2481–2495.
37. Huang G, Liu Z, MaatenVan L, Der,Weinberger KQ. *Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017; 2017:4700–4708.
38. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell.* 2017;40(4):834–848.
39. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:*1706.05587. Published online 2017.
40. Girshick R, Donahue J, Darrell T, Malik J. *Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014; 2014:580–587.
41. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell.* 2016;39(6):1137–1149.
42. He K, Gkioxari G, Dollár P, Girshick R. *Mask R-CNN. Proceedings of the IEEE international conference on computer vision.* 2017; 2017:2961–2969.
43. Shu JH, Nian FD, Yu MH, Li X. An improved mask R-CNN model for multiorgan segmentation. *Math Probl Eng.* 2020;2020(1):8351725.
44. Oktay O, Schlemper J, Folgoc LL. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:*1804.03999. Published online 2018.
45. Lin H, Li Z, Yang Z, Wang Y. Variance-aware attention U-Net for multi-organ segmentation. *Med Phys.* 2021;48(12):7864–7876.
46. Tang H, Liu X, Han K, et al. *Spatial context-aware self-attention model for multi-organ segmentation. Proceedings of the IEEE/CVF winter conference on applications of computer vision.* 2021; 2021:939–949.
47. Wu J, Zhou S, Zuo S, et al. U-Net combined with multi-scale attention mechanism for in CT images. *BMC Med Inf Decis Mak.* 2021;21(1):1–12.
48. Sinha A, Dolz J. Multi-scale self-guided attention for medical image segmentation. *IEEE J Biomed Health Inf.* 2020;25(1):121–130.
49. Jin Q, Meng Z, Sun C, Cui H, Su R. RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans. *Front Bioeng Biotechnol.* 2020;8:605132.
50. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30.
51. Dosovitskiy A, Beyer L, Kolesnikov A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:*2010.11929. Published online 2020.
52. Chen J, Lu Y, Yu Q. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:*2102.04306. Published online 2021.
53. Cao H, Wang Y, Chen J. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:*2105.05537. Published online 2021.
54. Xie Y, Zhang J, Shen C, Xia Y. *Cotr: Efficiently bridging CNN and transformer for 3D medical image segmentation. Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27-October 1, 2021, Proceedings, Part III 24.* Springer; 2021:171–180.
55. Zhu Z, Xia Y, Shen W, Fishman E, Yuille A. *A 3D coarse-to-fine framework for volumetric medical image segmentation. 2018 International conference on 3D vision (3DV).* IEEE; 2018:682–690.
56. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense V-Networks. *IEEE Trans Med Imaging.* 2018;37(8):1822–1834.
57. Zhang L, Zhang J, Shen P, et al. Block level skip connections across cascaded V-Net for multi-organ segmentation. *IEEE Trans Med Imaging.* 2020;39(9):2782–2793.
58. Weston AD, Korfiatis P, Philbrick KA, et al. Complete abdomen and pelvis segmentation using U-Net variant architecture. *Med Phys.* 2020;47(11):5609–5618.
59. Tang Y, Gao R, Lee HH, et al. High-resolution 3D abdominal segmentation with random patch network fusion. *Med Image Anal.* 2021;69:101894.
60. Roth HR, Lu L, Farag A, et al. *Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. International conference on medical image computing and computer-assisted intervention.* Springer; 2015:556–564.
61. Cai J, Lu L, Zhang Z, Xing F, Yang L, Yin Q. *Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20.* Springer; 2017:674–682.
62. Cai J, Lu L, Xing F, Yang L. *Pancreas segmentation in CT and MRI via task-specific network design and recurrent neural contextual learning. Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics.* Springer; 2019:3–21.
63. Yu Q, Xie L, Wang Y, Zhou Y, Fishman EK, Yuille AL. *Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.* 2018; 2018:8280–8289.
64. Li C, Mao Y, Guo Y, Li J, Wang Y. Multi-Dimensional Cascaded Net with Uncertain Probability Reduction for Abdominal Multi-Organ Segmentation in CT Sequences. *Comput Methods Programs Biomed.* 2022;221:106887.
65. Fu M, Wu W, Hong X, et al. Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images. *BMC Syst Biol.* 2018;12(4):119–127.
66. Hu P, Li X, Tian Y, et al. Automatic pancreas segmentation in CT images with distance-based saliency-aware DenseASPP network. *IEEE J Biomed Health Inf.* 2020;25(5):1601–1611.
67. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst.* 2014;27.
68. Li M, Lian F, Wang C, Guo S. Accurate pancreas segmentation using multi-level pyramidal pooling residual U-Net with adversarial mechanism. *BMC Med Imaging.* 2021;21(1):1–8.
69. Cai J, Zhang Z, Cui L, Zheng Y, Yang L. Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network. *Med Image Anal.* 2019;52:174–184.
70. Liu L, Wolterink JM, Brune C, Veldhuis RNJ. Anatomy-aided deep learning for medical image segmentation: A review. *Phys Med Biol.* 2021;66(11):11TR01.
71. Slabaugh G, Unal G. *Graph cuts segmentation using an elliptical shape prior.* IEEE; 2005:1222.
72. Veksler O. *Star shape prior for graph-cut image segmentation. Computer Vision-ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III 10.* Springer Berlin Heidelberg; 2008:454–467.
73. Gorelick L, Veksler O, Boykov Y, Nieuwenhuis C. Convexity shape prior for binary segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2016;39(2):258–271.
74. Christ PF, Elshaer MEA, Ettlinger F, et al. *Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19.* Springer; 2016:415–423.
75. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA. *3D deeply supervised network for automatic liver segmentation from CT volumes. International conference on medical image computing and computer-assisted intervention.* Cham: Springer International Publishing; 2016:149–157.
76. Lu F, Wu F, Hu P, Peng Z, Kong D. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. *Int J Comput Assist Radio Surg.* 2017;12(2):171–182.
77. Hu P, Wu F, Peng J, Liang P, Kong D. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. *Phys Med Biol.* 2016;61(24):8676.
78. Hoogi A, Subramaniam A, Veerapaneni R, Rubin DL. Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis. *IEEE Trans Med Imaging.* 2016;36(3):781–791.
79. Alirr OI. Deep learning and level set approach for liver and tumor segmentation from CT scans. *J Appl Clin Med Phys.* 2020;21(10):200–209.
80. Gong Z, Zhu Z, Zhang G, Zhao D, Guo W. *Convolutional neural networks based level set framework for pancreas segmentation from CT images. Proceedings of the Third International Symposium on Image Computing and Digital Medicine.* 2019; 2019:27–30.
81. Xu X, Zhou F, Liu B. Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN. *Int J Comput Assist Radio Surg.* 2018;13(7):967–975.
82. Li H, Liu J, Cui L, Huang H, Tai X. Volume preserving image segmentation with entropy regularized optimal transport and its applications in deep learning. *J Vis Commun Image Represent.* 2020;71:102845.
83. Liu J, Wang X, Tai XC. Deep convolutional neural networks with spatial regularization, volume and star-shape priors for image segmentation. *J Math Imaging Vis.* 2022;64(6):625–645.

84. Jurdi R, El,Petitjean C, Honeine P, Cheplygina V, Abdallah F. High-level prior-based loss functions for medical image segmentation: a survey. *Comput Vis Image Under.* 2021;210:103248.
85. Mirikharaji Z, Hamarneh G. *Star shape prior in fully convolutional networks for skin lesion segmentation. Int Conf Med Image Comput Comput Assist Interv*. Springer; 2018:737–745.
86. Qureshi TA, Lynch C, Azab L, et al. Morphology-guided deep learning framework for segmentation of pancreas in computed tomography images. *J Med Imaging.* 2022;9(2):24002.
87. Shi G, Xiao L, Chen Y, Zhou SK. Marginal loss and exclusion loss for partially supervised multi-organ segmentation. *Med Image Anal.* 2021;70:101979.
88. Kervadec H, Dolz J, Tang M, Granger E, Boykov Y, Ayed IB. Constrained-CNN losses for weakly supervised segmentation. *Med Image Anal.* 2019;54:88–99.
89. Tabrizi PR, Mansoor A, Cerrolaza JJ, Jago J, Linguraru MG. *Automatic kidney segmentation in 3D pediatric ultrasound images using deep neural networks and weighted fuzzy active shape model. Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE; 2018:1170–1173.
90. Ma J, Lin F, Wesarg S, Erdt M. *A novel Bayesian model incorporating deep neural network and statistical shape model for pancreas segmentation. Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*. Springer; 2018:480–487.
91. Mohagheghi S, Foruzan AH. Incorporating prior shape knowledge via data-driven loss model to improve 3D liver segmentation in deep CNNs. *Int J Comput Assist Radio Surg.* 2020;15(2):249–257.
92. Ravishankar H, Venkataramani R, Thiruvenkadam S, Sudhakar P, Vaidya V. *Learning and incorporating shape models for semantic segmentation. International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer; 2017:203–211.
93. Huang H, Zheng H, Lin L, et al. Medical image segmentation with deep atlas prior. *IEEE Trans Med Imaging.* 2021;40(12):3519–3530.
94. Zhao Y, Li H, Wan S, et al. Knowledge-aided convolutional neural network for small organ segmentation. *IEEE J Biomed Health Inf.* 2019;23(4):1363–1373.
95. Lee CE, Chung M, Shin YG. Voxel-level siamese representation learning for abdominal multi-organ segmentation. *Comput Methods Prog Biomed.* 2022;213:106547.
96. Guo X, Schwartz LH, Zhao B. Automatic liver segmentation by integrating fully convolutional networks into active contour models. *Med Phys.* 2019;46(10):4455–4469.
97. Xu Z, Sun J. Model-driven deep-learning. *Natl Sci Rev.* 2018;5(1):22–24.
98. Shlezinger N, Whang J, Eldar YC, Dimakis AG. Model-based deep learning. *arXiv preprint arXiv:*2012.08405. Published online 2020.
99. Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. *Adv Neural Inf Process Syst.* 2016:29.
100. Ding W, Li L, Zhuang X, Huang L. Cross-Modality Multi-Atlas Segmentation via Deep Registration and Label Fusion. *IEEE J Biomed Health Inform.* 2022;26(7):3104–3115.
101. Yang H, Sun J, Li H, Wang L, Xu Z. *Deep fusion net for multi-atlas segmentation: application to cardiac MR images. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer; 2016:521–528.
102. Yang H, Sun J, Li H, Wang L, Xu Z. Neural multi-atlas label fusion: Application to cardiac MR images. *Med Image Anal.* 2018;49:60–75.
103. Bhalodia R, Elhabian SY, Kavan L, Whitaker RT. *DeepSSM: a deep learning framework for statistical shape modeling from raw images. Shape in Medical Imaging: International Workshop, ShapeMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*. Springer; 2018:244–257.
104. Yang X, Song Z, King I, Xu Z. A survey on deep semi-supervised learning. *arXiv preprint arXiv:*2103.00550. Published online 2021.
105. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:*1802.07934. Published online 2018.
106. Fang K, Li WJ. *DMNet: difference minimization network for semi-supervised segmentation in medical images. International conference on medical image computing and computer-assisted intervention*. Cham: Springer International Publishing; 2020:532–541.
107. Huang R, Zheng Y, Hu Z, Zhang S, Li H. *Multi-organ segmentation via co-training weight-averaged models from few-organ datasets. Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part IV 23*. Springer; 2017:146–155.
108. Soberanis-Mukul RD, Navab N, Albarqouni S. Uncertainty-based graph convolutional networks for organ segmentation refinement. In: *Medical Imaging with Deep Learning. PMLR.* 2020:755–769.
109. Lee DH. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: *Workshop on Challenges in Representation Learning, ICML*. Vol 3. Atlanta; 2013:896.
110. Lee HH, Tang Y, Tang O, et al. Semi-supervised multi-organ segmentation through quality assurance supervision. In: *Proceedings of SPIE–the International Society for Optical Engineering*. Vol 11313. ; 2020:113131I.
111. Raju A, Cheng CT, Huo Y, et al. *Co-heterogeneous and adaptive segmentation from multi-source and multi-phase CT imaging data: a study on pathological liver and lesion segmentation. European Conference on Computer Vision*. Springer; 2020:448–465.
112. Huang H, Chen Q, Lin L, et al. MTL-ABS^3 net: atlas-based semi-supervised organ segmentation network with multi-task learning for medical images. *IEEE J Biomed Health Inf.* 2022;26(8):3988–3998.
113. Xia Y, Yang D, Yu Z, et al. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Med Image Anal.* 2020;65:101766.
114. Zhou Y, Wang Y, Tang P, et al. *Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE; 2019:121–140.
115. Fei-Fei L, Fergus R, Perona P. One-shot learning of object categories. *IEEE Trans Pattern Anal Mach Intell.* 2006;28(4):594–611.
116. Roy AG, Siddiqui S, Polsterl S, Navab N, Wachinger C. "Squeeze & excite" guided few-shot segmentation of volumetric images. *Med Image Anal.* 2020;59:101587.
117. Toro OJ, Müller H, Krenn M, et al. Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. *IEEE Trans Med Imaging.* 2016;35(11):2459–2475.
118. Sun L, Li C, Ding X, et al. Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Comput Biol Med.* 2022;140:105067.
119. Cui H, Wei D, Ma K, Gu S, Zheng Y. A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE Trans Med Imaging.* 2020;40(10):2656–2671.
120. Makarevich A, Farshad A, Belagiannis V, Navab N. MetaMedSeg: Volumetric Meta-learning for Few-Shot Organ Segmentation. *arXiv preprint arXiv:*2109.09734. Published online 2021.
121. Zhang G, Yang Z, Huo B, Chai S, Jiang S. Multiorgan segmentation from partially labeled datasets with conditional nnU-Net. *Comput Biol Med.* 2021;136:104658.
122. Petit O, Thome N, Soler L. Iterative confidence relabeling with deep ConvNets for organ segmentation with partial labels. *Comput Med Imag Grap.* 2021;91:101938.
123. Zhou Y, Li Z, Bai S, et al. *Prior-aware neural network for partially-supervised multi-organ segmentation. Proceedings of the IEEE/CVF international conference on computer vision*. 2019; 2019:10672–10681.
124. Jiang X, Zhang D, Li X, Liu K, Cheng KT, Yang X. Labeled-to-unlabeled distribution alignment for partially-supervised multi-organ medical image segmentation. *Med Image Anal.* 2025;99:103333.
125. Bitarafan A, Nikdan M, Baghshah MS. 3D image segmentation with sparse annotation by self-training and internal registration. *IEEE J Biomed Health Inf.* 2020;25(7):2665–2672.
126. Natarajan N, Dhillon IS, Ravikumar PK, Tewari A. Learning with noisy labels. *Adv Neural Inf Process Syst.* 2013:26.
127. Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med Image Anal.* 2020;65:101759.
128. Min S, Chen X, Zha ZJ, Wu F, Zhang Y. A two-stream mutual attention network for semi-supervised biomedical segmentation with noisy labels. *Proc AAAI Conf Artif Intell.* 2019;33(1):4578–4585 Springer.
129. Wang S, Li C, Wang R, et al. Annotation-efficient deep learning for automatic medical image segmentation. *Nat Commun.* 2021;12(1):1–13.
130. Wang L, Guo D, Wang G, Zhang S. Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning. *IEEE Trans Med Imaging.* 2020;40(10):2795–2807.
131. Shi J, Guo C, Wu J. A hybrid robust-learning architecture for medical image segmentation with noisy labels. *Future Internet.* 2022;14(2):41.
132. Huang L, Li M, Gou S, Zhang X, Jiang K. Automated segmentation method for low field 3D stomach MRI using transferred learning image enhancement network. *Biomed Res Int.* 2021;1:6679603.
133. Zhu JY, Park T, Isola P, Efros AA. *Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision*. 2017; 2017:2223–2232.
134. Dai X, Lei Y, Wynne J, et al. Synthetic CT-aided multiorgan segmentation for CBCT-guided adaptive pancreatic radiotherapy. *Med Phys.* 2021;48(11):7063–7073.
135. Yuan W, Wei J, Wang J, Ma Q, Tasdizen T. Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. *Med Image Anal.* 2020;64:101731.
136. Chen C, Dou Q, hen CH, Qin J, Heng PA. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans Med Imaging.* 2020;39(7):2494–2505.
137. Yang Y, Wang H, Ji C, Niu Y. Artificial intelligence-driven diagnostic systems for early detection of diabetic retinopathy: Integrating retinal imaging and clinical data. *SHIFAA.* 2023;2023:83–90.
138. Ali MA, Alqaraghuli A. A Survey on the Significance of Artificial intelligence (AI) in Network cybersecurity. *Babylonian Journal of Networking.* 2023;2023:21–29.
139. Heimann T, Ginneken B, Van,Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging.* 2009;28(8):1251–1265.
140. Bilic P, Christ PF, Vorontsov E. The liver tumor segmentation benchmark (lits). *arXiv preprint arXiv:*1901.04056. Published online 2019.
141. Soler L, Hostettler A, Agnus V, et al. 3D image reconstruction for comparison of algorithm database: a patient specific anatomical and medical image database. *IRCAD Strasbg Fr Tech Rep.* 2010;1(1).
142. Antonelli M, Reinke A, Bakas S, et al. The medical segmentation decathlon. *arXiv preprint arXiv:*2106.05735. Published online 2021.
143. Heller N, Isensee F, Maier-Hein KH, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med Image Anal.* 2021;67:101821.
144. Landman B, Xu Z, Igelsias JE, Styner M, Langerak TR, Klein A. *2015 MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge.* 2015; 2015.
145. Kavur AE, Gezer NS, Barış M, et al. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal.* 2021;69:101950.
146. Ma J, Zhang Y, Gu S, et al. Abdomenct-1K: is abdominal organ segmentation a solved problem? *IEEE Trans Pattern Anal Mach Intell.* 2021;44(10):6695–6714.