

Review

Not peer-reviewed version

---

# Deep Learning for Medical Image Segmentation without Full Supervision: An Overview

---

[Xinyue Zhang](#), [Jianfeng Wang](#)<sup>\*</sup>, [Jinqiao Wei](#), [Xinyu Yuan](#), [Ming Wu](#)<sup>\*</sup>

Posted Date: 8 April 2025

doi: 10.20944/preprints202504.0460.v1

Keywords: Medical image segmentation; Semi-supervised Learning; Weakly Supervised Learning; Unsupervised Learning; Survey





Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Deep Learning for Medical Image Segmentation without Full Supervision: An Overview

Xinyue Zhang <sup>1</sup>, Jianfeng Wang <sup>2,\*</sup> , Jinqiao Wei <sup>2</sup>, Xinyu Yuan <sup>2</sup> and Ming Wu <sup>3,4,\*</sup> 

<sup>1</sup> College of Artificial Intelligence, Taiyuan University of Technology

<sup>2</sup> School of Software, Taiyuan University of Technology

<sup>3</sup> Department of Computer Science, KU Leuven

<sup>4</sup> Department of Mechanical Engineering, KU Leuven

\* Correspondence: wangjianfeng@tyut.edu.cn; ming.wu1@kuleuven.be

**Abstract:** Medical image segmentation, a critical task in medical image analysis, aims to precisely delineate regions of interest (ROIs) such as organs, lesions, and cells, and is crucial for applications including computer-aided diagnosis, surgical planning, radiation therapy, and pathological analysis. While fully supervised deep learning methods have demonstrated remarkable performance in this domain, their reliance on large-scale, pixel-level annotated datasets—a significant label scarcity challenge—severely hinders their widespread deployment in clinical settings. Addressing this limitation, this review focuses on non-fully supervised learning paradigms, systematically investigating the application of semi-supervised, weakly supervised, and unsupervised learning techniques for medical image segmentation. We delve into the theoretical foundations, core advantages, typical application scenarios, and representative algorithmic implementations associated with each paradigm. Furthermore, this paper compiles and critically reviews commonly utilized benchmark datasets within the field. Finally, we discuss future research directions and challenges, offering insights for advancing the field and reducing dependence on extensive annotations.

**Keywords:** medical image segmentation; semi-supervised learning; weakly supervised learning; unsupervised learning; survey

## 1. Introduction

Medical image segmentation, a critical component of medical image analysis, involves the application of specific algorithms and techniques to accurately partition target organs, tissues, or lesions from the background within medical images such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), subsequently assigning semantic labels to each segmented region. Early investigations predominantly employed traditional methods that leveraged low-level image features and prior knowledge, including pixel intensity thresholding, region growing, edge detection, and active contour models. While effective for simple images, these methods struggled with real-world clinical complexities like ambiguous boundaries, noise, and anatomical variations.

In recent years, deep learning has significantly advanced medical image segmentation. Convolutional Neural Networks (CNNs) [1], leveraging capabilities such as local feature extraction via convolutional operations, have demonstrated the capacity to automatically learn highly discriminative feature representations from large-scale medical imaging datasets, thereby significantly enhancing segmentation performance. The introduction of Fully Convolutional Networks (FCNs) [2] established a foundational paradigm for semantic segmentation by replacing fully connected layers in traditional CNNs with convolutional layers, enabling end-to-end pixel-level prediction. Building upon this, the U-Net architecture [3] further propelled the domain by employing an encoder-decoder structure with skip connections to effectively fuse multi-scale features. Moreover, the integration of Transformers [4], utilizing their potent global context modeling capabilities to compensate for the limitations of CNNs in capturing long-range dependencies, has emerged as a prominent research trajectory. More recently,

the Segment Anything Model (SAM) [5] has garnered considerable attention within the broader image segmentation community, with its strong zero-shot transfer capabilities presenting novel opportunities for medical image segmentation under the fully supervised paradigm.

However, despite the demonstrated potency of these fully supervised deep learning methods, their profound dependence on extensive, pixel-wise, accurately annotated datasets significantly constrains their clinical applicability. The annotation of medical images constitutes a high-cost, time-intensive, and error-prone undertaking, further compounded by inter-observer variability, data scarcity, and patient privacy concerns. To surmount these challenges, researchers have proposed non-fully supervised medical image segmentation approaches. Such methods aim to train models utilizing limited, incomplete, or coarse-grained annotation information, thereby alleviating the demand for pixel-level ground truth, enhancing annotation efficiency, and facilitating the deployment of medical image segmentation in practical clinical settings. Currently, weakly supervised learning, semi-supervised learning, and unsupervised learning represent the principal research directions within this domain, each endeavoring to learn effective segmentation models from distinct types of incomplete annotation data.

Early approaches to non-fully supervised medical image segmentation predominantly relied upon traditional image processing techniques augmented with limited annotations [6,7]. With the advent of deep learning, various non-fully supervised learning methodologies have progressively become mainstream. Among these, weakly supervised learning methods initially gained prominence, leveraging more readily obtainable annotation forms such as image-level labels [8], bounding boxes [9], scribbles [10], and point annotations [11]. These methods often incorporate strategies like Class Activation Maps (CAM)[8], iterative mining, and adversarial learning [12] to maximize segmentation accuracy while minimizing the annotation burden. To further harness the wealth of unlabeled data, semi-supervised learning methods were introduced into the medical image segmentation domain, significantly enhancing model performance by training on a combination of a small number of labeled samples and a large volume of unlabeled samples. In scenarios characterized by extremely scarce annotations, unsupervised learning methods, particularly unsupervised anomaly segmentation and unsupervised domain adaptation, have also undergone development. Furthermore, techniques such as transfer learning [13], self-supervised learning [14], multi-modal fusion, and the utilization of prior knowledge [15] are frequently integrated with these non-fully supervised approaches to improve segmentation performance and robustness further.

This review aims to provide a comprehensive overview and discussion of the recent advancements in deep learning-based medical image segmentation under the non-fully supervised paradigm(as illustrated in Figure 1). Specifically, Section 1 introduces the background of the medical image segmentation field. Section 2 compiles commonly utilized datasets for medical image segmentation. Sections 3 through 5 delve into non-fully supervised medical image segmentation methods, systematically presenting semi-supervised (Section 3), weakly supervised (Section 4), and unsupervised (Section 5) approaches, categorized based on the type and quantity of annotation information employed during model training. Section 6 discusses the non-fully supervised medical image segmentation methodologies reviewed herein, addressing their clinical applications and outlining future research directions. Finally, Section 7 concludes the paper.

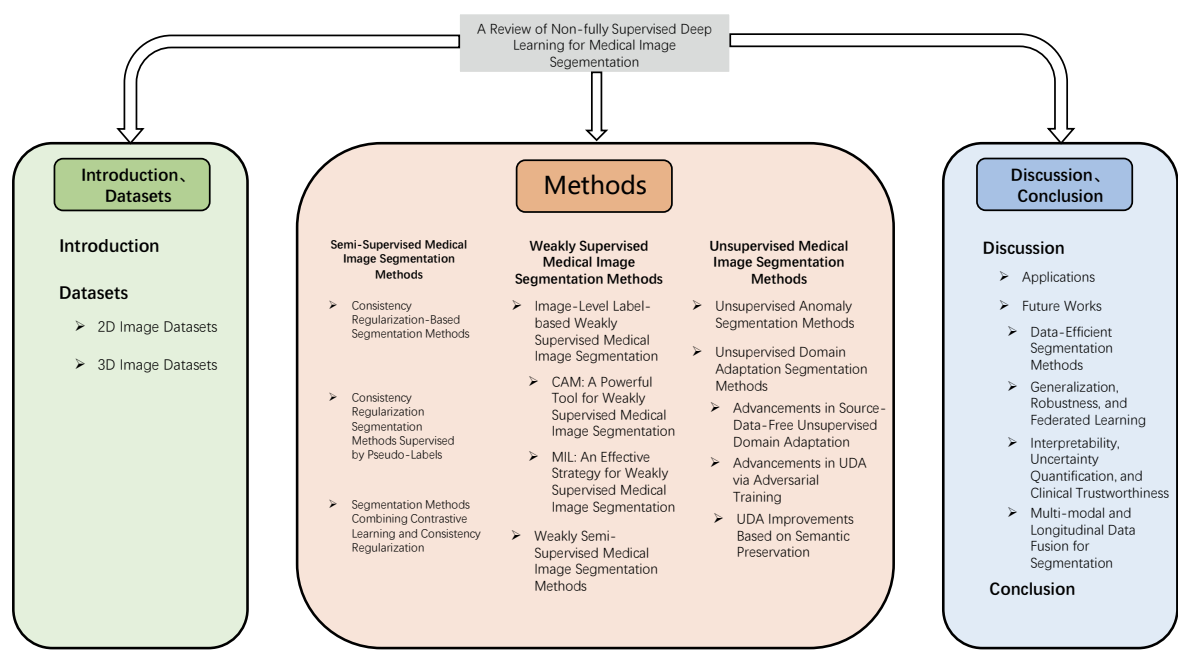


Figure 1. Structure and Coverage of This Review.

2. Datasets

This section provides a systematic overview of benchmark datasets widely utilized in the field of medical image segmentation. These datasets not only furnish the fundamental data support for the training and performance evaluation of segmentation algorithms but also serve as critical resources driving technological advancements and methodological innovation within this domain. Based on data dimensionality, they are principally categorized into two major types: 2D (pixel-based) images and 3D (voxel-based) volumes, with their specific characteristics and representative examples detailed in Sections 2.1 and 2.2, respectively. For ease of reference, Table 1 and Table 2 (continued) summarize the core datasets discussed herein, outlining their imaging modalities, primary anatomical regions covered (e.g., colon, breast, skin), and typical application scenarios.

2.1. 2D Image Datasets

**ACDC Dataset[16]** : The Automated Cardiac Diagnosis Challenge (ACDC) dataset, released as part of the MICCAI 2017 challenge and provided by Pierre-Marc Jodoin, Alain Lalande, and Olivier Bernard, comprises multi-slice 2D cardiac cine-Magnetic Resonance Imaging (cine-MRI) samples from 100 patients. It is primarily utilized for the segmentation of the left ventricle (LV), right ventricle (RV), and myocardium (MYO) at the End-Diastole (ED) and End-Systole (ES) phases. For semi-supervised learning applications, it is commonly partitioned into a training set (70 scans), a validation set (10 scans), and a test set (20 scans). The ACDC dataset is crucial for developing accurate cardiac image segmentation methods to assess cardiac function, providing key information for the diagnosis and treatment of cardiac diseases.

**Colorectal Adenocarcinoma Gland (CRAG) Dataset [17]:** The CRAG dataset focuses on the task of gland segmentation in colorectal adenocarcinoma histopathology images, aiming to foster the development of relevant medical image segmentation algorithms. It consists of 213 H&E stained colorectal adenocarcinoma tissue section images from diverse centers and equipment, with 173 designated for training and 40 for testing. The dataset provides complete instance-level gland annotations, including precise segmentation masks. CRAG represents a significant benchmark resource for the medical image processing field, particularly for research on colorectal cancer gland segmentation.

**IU Chest X-ray Dataset [18]:** Released through a collaboration between Indiana University and the Open-i laboratory, this dataset is a commonly used benchmark in medical image analysis, especially for tasks related to lung disease detection and classification. It includes 3,700 high-resolution chest

X-ray images with annotation information covering 14 common pulmonary conditions and normal status labels. Annotations were performed by radiologists, providing binary labels for disease presence, with some samples containing lesion location information. The standard split comprises a training set (2,590 images), a validation set (370 images), and a test set (740 images).

Table 1. Medical Imaging Datasets Summary

Dataset	Modality	Anatomical Area	Application Scenarios
ACDC[16]	MRI	Heart (left and right ventricles)	Cardiac function analysis, ventricular segmentation
Colorectal adenocarcinoma glands [17]	Pathology Sections (H&E Staining)	Colorectal tissue	Segmentation of the glandular structure
IU Chest X-ray [18]	X-ray (chest x-ray)	Chest (cardiopulmonary area)	Classification of lung diseases
MIMIC-CXR [19]	X-ray (chest x-ray) + clinical report	Chest	Automatic diagnosis of multiple diseases
COV-CTR [20]	CT (chest)	Lung	COVID-19 severity rating
MS-CXR-T [21]	X-ray (chest x-ray)	Chest	Multilingual report generation
NIH-AAPM-Mayo Clinical LDCT [22]	Low-dose CT (chest)	Lung	Lung nodule detection
LoDoPaB [23]	Low-dose CT (Simulation)	Body	CT reconstruction algorithm development
LDCT [24]	Low-dose CT	Chest/abdomen	Radiation dose reduction studies
LA [25]	MRI	Heart (left atrium)	Surgical planning for atrial fibrillation
Pancreas-CT [26]	CT (abdomen)	Pancreas	Pancreatic tumor segmentation
BraTS [27]	Multiparametric MRI	Brain (glioma)	Brain tumor segmentation
ATLAS [28]	MRI (T1)	Brain (stroke lesions)	Stroke analysis
ISLES [29–31]	MRI (multiple sequences)	Brain	Ischemic stroke segmentation

Table 2. Medical Imaging Datasets Summary(continued)

Dataset	Modality	Anatomical Area	Application Scenarios
AISD [32]	Ultrasonic	Abdominal organs	Organ boundary segmentation
Cardiac [33]	MRI	Heart	Ventricular division
KiTS19 [34]	CT (abdomen)	Kidney	Segmentation of kidney tumors
UKB [35]	MRI/CT/X-ray	Body	Multi-organ phenotypic analysis
LiTS [36]	CT (abdomen)	Liver	Segmentation of liver tumors
CHAOS [37]	CT/MRI (abdomen)	Multi-organ	Cross-modal organ segmentation



**MIMIC-CXR Dataset [19]:** MIMIC-CXR, a large-scale open-source medical dataset released jointly by MIT and Beth Israel Deaconess Medical Center, concentrates on the intersection of chest X-ray image analysis and natural language processing research. It encompasses 473,057 chest X-ray images, each associated with structured labels (14 common diseases) and free-text radiology reports. The standard partition includes a training set (378,447 images), a validation set (47,305 images), and a test set (47,305 images).

**COV-CTR Dataset [20]:** COV-CTR is a high-quality open-source dataset for COVID-19 lung CT image analysis, led by the Institute of Automation, Chinese Academy of Sciences, and Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, in collaboration with over 20 institutions including Tsinghua University and Fudan University. It contains CT image data from 1,110 confirmed COVID-19 patients, covering both non-contrast and contrast-enhanced CT types. Annotations, marking five typical lesion types, were cross-validated by at least three radiology experts.

**MS-CXR-T Dataset [21]:** MS-CXR-T is a multi-center chest X-ray dataset focusing on Tuberculosis (TB), designed to address the challenges of heterogeneity in TB imaging and the generalizability of diagnostic models. It includes over 5,000 chest X-ray images, with a case distribution of approximately 3,500 TB-positive and 1,500 TB-negative cases. This dataset can be employed to train TB screening models adaptable to multi-center data and to study the radiological presentation differences of TB lesions across various regions.

**NIH-AAPM-Mayo Clinical LDCT Dataset [22]:** Provided through a collaboration among the National Institutes of Health (NIH), the American Association of Physicists in Medicine (AAPM), and the Mayo Clinic, this dataset is primarily intended for lung cancer screening algorithm development and low-dose CT (LDCT) image reconstruction research. It contains LDCT scans from over 1,000 high-risk individuals. Lung nodules  $\geq 3\text{mm}$  were annotated by three radiologists. It can be utilized for developing deep learning-based nodule detection systems and exploring high-resolution reconstruction algorithms under low-dose conditions.

**LoDoPaB Dataset [23]:** The Low-Dose Parallel Beam (LoDoPaB) dataset is an internationally recognized benchmark for low-dose CT projection data, specifically designed for research on low-dose CT image reconstruction algorithms. While primarily serving reconstruction studies, it may also provide reference value for segmentation research. The dataset contains low-dose CT projection data from over 35,000 clinical patients, covering multiple anatomical regions including the chest, abdomen, and head. It can be used to train deep learning models for direct image reconstruction from projection data and to investigate personalized radiation dose allocation strategies.

**LDCT Datasets [24]:** These low-dose CT medical imaging datasets focus on optimizing image quality while reducing radiation dose, with their core value lying in balancing diagnostic efficacy and patient safety. Particularly noteworthy is their unique application value in the field of medical image segmentation: by providing standardized imaging data under low-dose conditions, they not only support traditional image reconstruction research (including determining minimum effective dose, developing reconstruction algorithms for low signal-to-noise ratio images, and quantum noise suppression), but also provide training and validation benchmarks for precise segmentation of key anatomical structures (such as pulmonary nodules and liver lesions) in low-radiation environments. Typical datasets generally include scan data from key anatomical regions like the chest, abdomen, and heart, enabling them to serve both image quality optimization research and support the development and evaluation of downstream intelligent analysis tasks.

## 2.2. 3D Image Datasets

**LA Dataset [25]:** The Left Atrium (LA) benchmark dataset consists of Late Gadolinium Enhanced MRI (LGE-MRI) scans from patients with atrial fibrillation, specifically curated for the precise segmentation of the left atrium and its associated scar tissue. Data originated from the University of Utah, Beth Israel Deaconess Medical Center, and King's College London, acquired using diverse scanning equipment and resolutions to ensure model generalizability across varied clinical settings. The dataset includes 100 3D LGE-MRI scans, typically used for training, with image resolutions standardized to

$0.625 \times 0.625 \times 0.625 \text{ mm}^3$ . The LA dataset is a widely adopted benchmark within the semi-supervised medical image segmentation domain.

**Pancreas-CT Dataset [26]:** Provided by the National Institutes of Health (NIH), this dataset facilitates research on pancreas segmentation in CT scans. It contains 80 (originally 82, with 2 removed due to duplication) abdominal contrast-enhanced 3D CT scans from 53 male and 27 female subjects (age range 18-76 years, mean age  $46.8 \pm 16.7$  years). The CT scans have a resolution of  $512 \times 512$  pixels, with variations in pixel size and slice thickness; the slice thickness ranges between 1.5 and 2.5 mm. Data were acquired using Philips and Siemens MDCT scanners (120 kVp tube voltage). Slice-by-slice segmentation of the pancreas was performed manually by a medical student and subsequently verified or revised by senior radiologists. This dataset is extensively utilized for research and algorithm development in pancreas segmentation tasks.

**BraTS Dataset [27]:** The Multimodal Brain Tumor Segmentation (BraTS) dataset is a public resource specifically designed for brain tumor segmentation using multi-modal medical images. Data originate from multiple hospitals and include four MRI modalities: T1-weighted (T1), Gadolinium-enhanced T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid-Attenuated Inversion Recovery (T2-FLAIR). All scans underwent rigorous review by neuroradiologists and are accompanied by expert annotations. For semi-supervised learning scenarios, it is commonly partitioned into a training set (250 scans), a validation set (25 scans), and a test set (60 scans). The dataset aims to foster the development and evaluation of automated brain tumor segmentation algorithms.

**ATLAS Dataset [28]:** The Anatomical Tracings of Lesions After Stroke (ATLAS) dataset aggregates MRI brain scans from multiple centers globally to evaluate automated stroke lesion segmentation methods, primarily serving stroke rehabilitation research. It provides a total of 1271 images, with 955 publicly available (comprising 655 training images and 300 test images with hidden annotations), and an additional 316 forming an independent generalization test set. Most images were acquired using 3T MRI scanners at  $1 \text{ mm}^3$  resolution, with a few cases scanned at 1.5 T with  $1\text{-}2 \text{ mm}^3$  resolution. ATLAS offers a valuable data resource for advancing automated segmentation of subacute and chronic stroke lesions.

**ISLES Dataset [29–31]:** The Ischemic Stroke Lesion Segmentation (ISLES) dataset focuses on segmenting ischemic stroke lesions, aiming to automatically delineate acute to subacute ischemic stroke lesions using multi-modal imaging. ISLES can be used for training and validating relevant segmentation algorithms. Taking ISLES22 as an example, this dataset compiles 400 multi-center, multi-device MRI cases, partitioned into a public training set of 250 cases and a test set of 150 cases (with non-public annotations) for online evaluation. This dataset assists researchers in developing more accurate and reliable algorithms to improve diagnosis and treatment planning for stroke patients.

**AISD Dataset [32]:** The Acute Ischemic Stroke Dataset (AISD) is a comprehensive resource integrating clinical information, imaging data, and follow-up data from patients with acute ischemic stroke, designed to provide high-quality data for related scientific research. It contains 397 non-contrast CT scans acquired within 24 hours of stroke onset, with 345 used for model training and validation, and 52 reserved for testing. AISD aims to promote the research, development, and clinical translation of CT-based acute stroke segmentation techniques.

**Cardiac (M&Ms) Dataset [33]:** The dataset—the Multi-Center, Multi-Vendor & Multi-Disease Cardiac Segmentation (M&Ms) dataset—focuses on segmenting the left atrium (and other specified cardiac structures) from single-modality MR images, with its application context clearly defined in the original literature. All imaging data were normalized to the  $[0,1]$  range, comprising 30 clinical cases officially divided into a training set of 20 cases and a test set of 10 cases. Segmentation results for the test set can be submitted for evaluation through the official platform.

**KiTS19 Dataset [34]:** The Kidney Tumor Segmentation Challenge 2019 (KiTS19) dataset focuses on the segmentation of kidneys and kidney tumors, aiming to drive the progress of related medical image segmentation algorithms. It comprises CT images and corresponding semantic segmentation labels for 300 cases from various centers and devices, with 210 allocated for training and 90 for testing.

The dataset provides fine-grained segmentation labels for both the kidney and tumor, includes clinical attributes for some cases, exhibits high diversity and challenge, and represents an important research resource in the kidney tumor segmentation field.

**UKB Dataset [35]:** The UK Biobank (UKB) is a large-scale, multidimensional biomedical database and research platform designed to investigate the complex relationships between genetic factors, lifestyle, and health status, thereby advancing understanding of various chronic diseases (including cancer, heart disease, diabetes, mental disorders, etc.). This database contains deep phenotyping and health data from approximately 500,000 volunteers, encompassing genetic information, biological samples, imaging data (e.g., brain and cardiac MRI), and detailed health records, providing a crucial foundation for large-scale, interdisciplinary health research. Furthermore, its rich collection of semi-structured diagnostic reports paired with corresponding imaging data offers unique resources for developing weakly supervised segmentation methods. These capabilities are concretely manifested in: constructing universal medical image feature extractors, enabling quantitative analysis of anatomical structure-disease risk correlations, and supporting knowledge distillation training based on clinical reports.

**LiTS Dataset [36]:** The Liver Tumor Segmentation Challenge (LiTS) dataset concentrates on the segmentation of the liver and liver tumors in CT images, aiming to advance relevant segmentation algorithms and promote research into automated diagnostic systems. It collects CT scan data from 7 different medical centers, containing 131 training datasets and 70 test datasets. LiTS finds wide application value in medical image segmentation, computer-aided diagnosis, development of medical image analysis tools, and academic research and education, supporting researchers in developing and validating new segmentation algorithms.

**CHAOS Dataset [37]:** The Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) dataset focuses on abdominal organ segmentation, providing paired multi-modal (CT and MR) data with annotations to foster the development of abdominal organ segmentation algorithms. CHAOS includes 40 paired CT and MR images, divided into 20 training sets and 20 test sets. The training set provides annotations for four abdominal organs (liver, left kidney, right kidney, spleen), where CT images are annotated only for the liver, while MR images have annotations for all four organs. Test set annotations are not publicly available. This dataset supports the development and validation of multi-modal learning and segmentation algorithms.

### 3. Semi-Supervised Medical Image Segmentation Methods

Given the abundance of unlabeled data in clinical practice, semi-supervised learning (SSL) methods have emerged as a prominent and highly promising research direction. SSL effectively integrates labeled data with large amounts of unlabeled data for model training, thereby reducing the dependency on large-scale annotated datasets.

A substantial body of research has emerged within the domain of semi-supervised segmentation, with the majority of methods concentrating on two predominant paradigms: pseudo-labeling [38–40] and consistency regularization[41]. Pseudo-labeling approaches leverage the model's predictions on unlabeled data, generating pseudo-labels by applying thresholding or other selection mechanisms. These pseudo-labeled samples are subsequently combined with manually annotated data for further model training and refinement. Conversely, the fundamental principle of consistency regularization posits that for an unlabeled sample, the model's predictions should remain consistent across its different perturbed versions. The objective is to minimize the discrepancy between the model's predictions on different perturbed versions of unlabeled data. Establishing high-quality consistency targets during training is crucial for achieving optimal performance.

In recent years, semi-supervised medical image segmentation models have primarily adopted consistency regularization strategies. Among these, the Mean Teacher (MT) model [42] represents a classic method for implementing consistency regularization. It effectively utilizes unlabeled data by applying weak and strong augmentations and enforcing consistency between the model's predictions



on these different augmented versions. As illustrated in Figure 2, the Mean Teacher (MT) model employs a dual-model architecture comprising a student model (represented by orange weights) and a teacher model (represented by blue weights), typically sharing an identical network structure (e.g., U-Net or V-Net). During the processing of a labeled training sample, both models perform inference on the input, incorporating stochastic perturbations (denoted as  $\eta$  and  $\eta'$ , respectively). The student model's weights ( $\theta$ ) are optimized via backpropagation using a composite loss function. This function includes: a classification loss (classification cost), quantifying the discrepancy between the student's prediction (orange probability distribution) and the ground truth label, and a consistency loss (consistency cost), measuring the divergence between the predictions of the student and teacher models (blue probability distribution). Following the gradient descent update of the student weights, the teacher model's weights ( $\theta'$ ) are updated not through gradient computation but as an exponential moving average (EMA) of the student weights ( $\theta \rightarrow \theta'$ ), as depicted. This EMA update mechanism imparts stability to the teacher model, enabling it to progressively aggregate knowledge acquired by the student throughout training. For unlabeled data, the optimization relies solely on the consistency loss, thereby encouraging the student model's outputs to align with the more stable and reliable predictions generated by the teacher model. Consequently, the teacher model implicitly furnishes "pseudo-label" supervision, guiding the student model to learn latent structural information and data distribution characteristics from abundant unlabeled data, ultimately enhancing the model's generalization capability and segmentation accuracy.

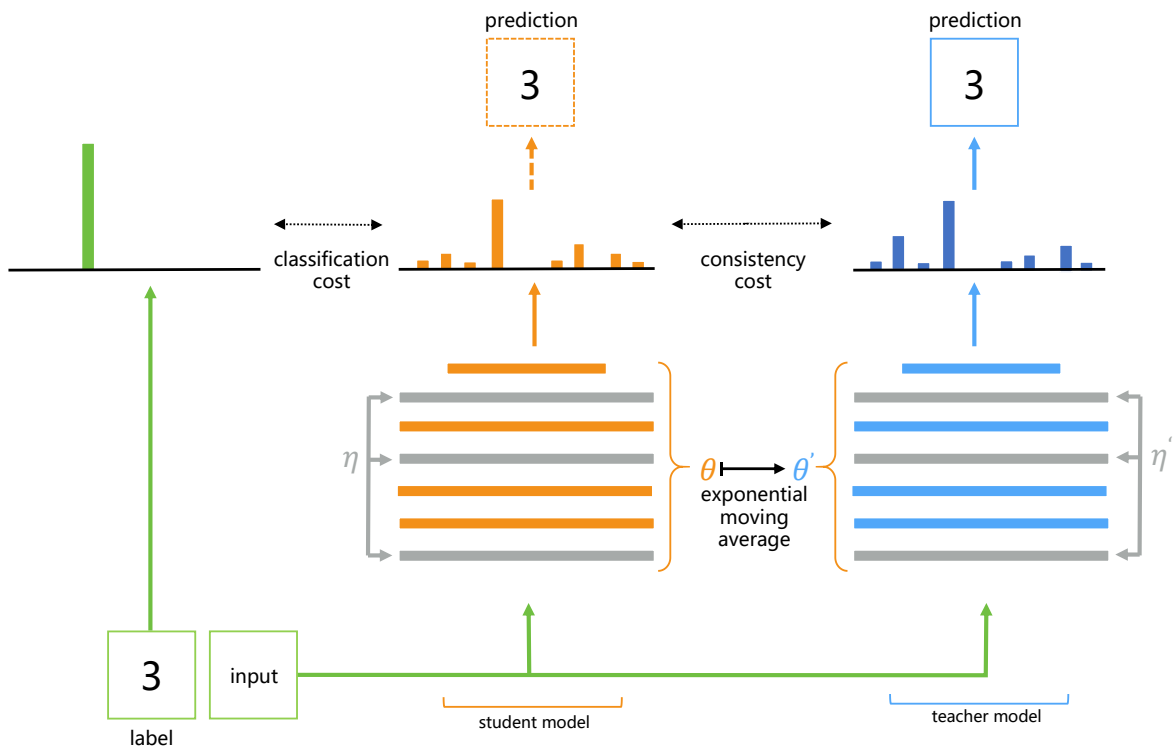


Figure 2. The framework diagram of Mean Teacher [42].

Recently, consistency regularization methods supervised by pseudo-labels have achieved significant success in semi-supervised segmentation [43,44]. Concurrently, approaches combining contrastive learning strategies with consistency regularization methods are continually emerging [45,46]. Based on these evolving trends, this paper categorizes and reviews semi-supervised medical image segmentation methods into the following three classes: consistency regularization methods, consistency regularization methods incorporating pseudo-labeling, and methods combining contrastive learning with consistency regularization.

### 3.1. Consistency Regularization-Based Segmentation Methods

This subsection focuses on semi-supervised segmentation methods that directly apply the principle of consistency regularization without relying on explicit pseudo-labels, aiming to enhance model performance by effectively leveraging unlabeled data.

The Ambiguity-Consensus Mean-Teacher (AC-MT) [47] is an enhancement of the fundamental MT model. Inheriting the student-teacher architecture and EMA weight update mechanism from MT, AC-MT's core innovation lies in the introduction of an ambiguity identification module. This module assesses the prediction "ambiguity" (i.e., uncertainty) for each pixel in the unlabeled data using strategies such as calculating entropy, model uncertainty, or employing prototype/class conditioning for noisy label identification. Unlike MT, which computes consistency loss across all pixels, AC-MT calculates and imposes consistency loss only on pixels identified as having high ambiguity. This forces the student model to achieve consensus with the teacher model, specifically in these challenging yet informative regions. This selective consistency learning strategy enables AC-MT to extract critical information from unlabeled data more precisely, thereby further improving segmentation performance.

During training, the AC-MT model utilizes labeled data for supervised learning of the student model (calculating standard segmentation loss). Unlabeled data first passes through the ambiguity identification module to filter high-ambiguity pixels. These data are then fed into both the student and teacher models, with consistency loss calculated and used to update the student model only on these ambiguous pixels. Finally, the teacher model is updated via the EMA of the student model's weights. This process iterates, achieving more efficient semi-supervised learning by selectively focusing on ambiguous regions within the unlabeled data. Compared to the baseline Mean-Teacher model and other state-of-the-art semi-supervised learning methods, AC-MT demonstrates more effective utilization of unlabeled data, particularly with limited labeled data (e.g., 10% or 20%). It achieves significant improvements in key segmentation accuracy metrics such as the Dice Similarity Coefficient (DSC) and Jaccard index, and maintains robust performance even in scenarios with extremely scarce labeled data (e.g., 2.5%).

AAU-Net [48] similarly represents an enhancement to the standard MT model, yet its distinguishing characteristic lies in the utilization of anatomical prior knowledge to address the challenge of unreliable predictions on unlabeled data within the MT framework. Whereas AC-MT concentrates on the "ambiguity" associated with predictions, AAU-Net places greater emphasis on quantifying the deviation of predictions from expected anatomical structures.

AAU-Net introduces a pre-trained Denoising Autoencoder (DAE) to capture anatomical prior knowledge. This DAE can map any predicted segmentation mask  $P_t$  to a more anatomically plausible segmentation  $P'_t$ , denoted as:

$$P'_t = \text{DAE}(P_t) \quad (1)$$

By utilizing the DAE module, AAU-Net estimates uncertainty based on the discrepancy between the teacher model's prediction  $P_t$  and its "anatomically corrected" version  $P'_t$ , rather than directly using the difference between the raw predictions of the student and teacher models. This uncertainty is then incorporated into the calculation of the consistency loss.

Building upon this representation prior, AAU-Net further proposes an anatomy-aware uncertainty estimation mechanism as its core component. This mechanism fully exploits the anatomical prior knowledge provided by the DAE to assess the reliability of the teacher model's predictions  $P_t$ . By calculating the pixel-wise difference between  $P_t$  and  $P'_t$ , an uncertainty map  $U$  is constructed:

$$U = \|P_t - P'_t\|^2 \quad (2)$$

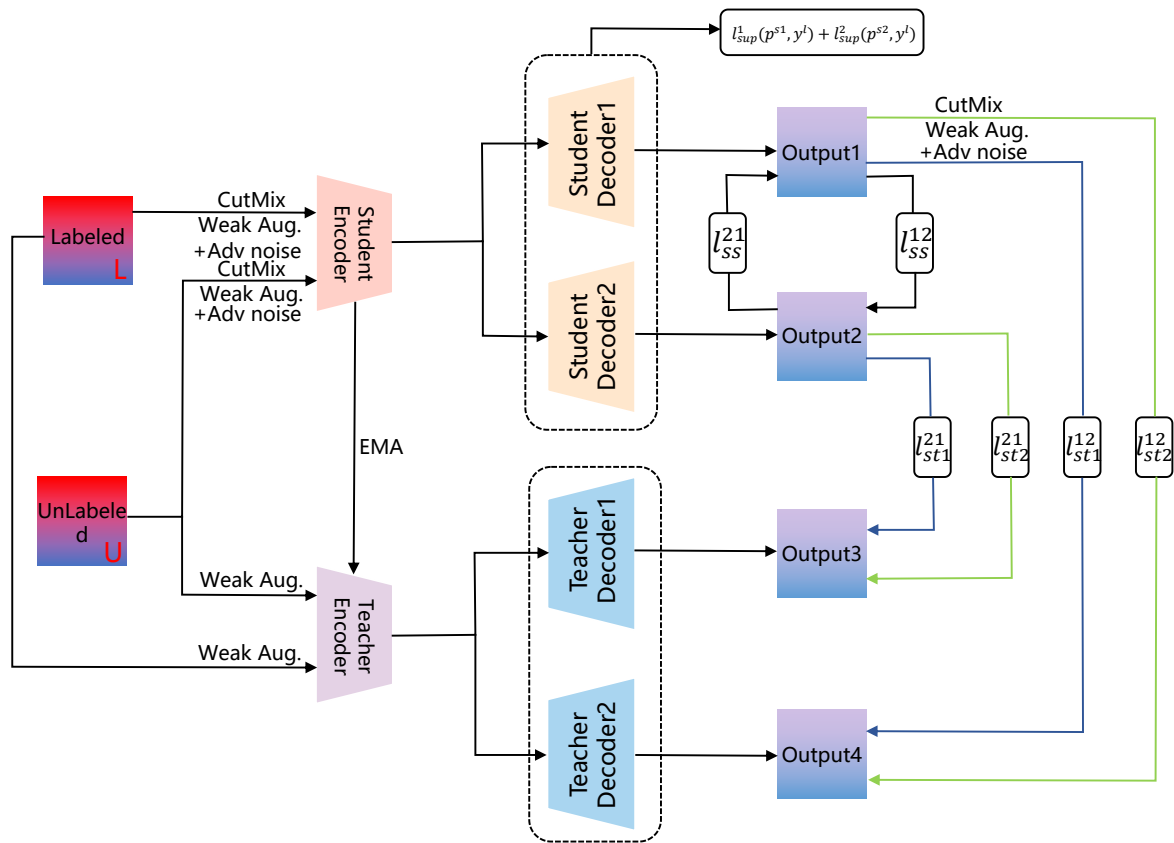
The map  $U$  reflects the inconsistency between the model's predictions and the anatomical priors: a larger difference indicates a greater deviation of the prediction at that pixel from the anatomical structure, thus implying higher uncertainty. This uncertainty estimation method effectively integrates anatomical knowledge into the semi-supervised learning process, enabling the model to more accu-

rately identify potentially erroneous regions in its predictions and, consequently, utilize unlabeled data more effectively for training.

In the task of abdominal CT multi-organ segmentation, compared to existing state-of-the-art baseline methods such as Uncertainty-Aware Mean Teacher (UAMT) [49] and Uncertainty Rectified Pyramid Consistency (URPC) [50], AAU-Net improved the average DSC by 1.65% and the HD metric by 0.6 mm with a 10% labeling ratio. With a 20% labeling ratio, the average DSC improved by 1.95% and the HD metric by 1 mm. These results demonstrate the method's capability to achieve accurate segmentation with limited labeled data, making it suitable for medical image analysis scenarios involving complex anatomical structures or challenging annotation tasks.

Whereas the former two approaches address the unreliable predictions for unlabeled data by focusing on ambiguity or incorporating anatomical priors, respectively, CMMT-Net [51] instead enhances intrinsic model diversity. This is achieved through the construction of a cross-head mutual mean-teaching architecture, aimed at improving the MT model for more robust utilization of unlabeled data. As illustrated in Figure 3, CMMT-Net is fundamentally characterized by a shared encoder and a dual decoder cross-head design, integrated with a teacher model updated via Exponential Moving Average (EMA). Specifically, both the student network (upper part) and the teacher network (lower part) comprise a shared encoder and two distinct decoders. This dual-decoder configuration introduces feature-level diversity. The weights of the teacher network are an EMA of the corresponding student network weights, ensuring the stability of teacher predictions. During the training procedure, labeled data undergoes strong augmentation (CutMix, Weak Aug. + Adv noise) before being fed into the student network. The outputs  $p^{s1}, p^{s2}$  from its two decoders are used to compute the supervised loss  $l_{sup}$  against the ground truth label  $y^l$ . Unlabeled data is concurrently utilized for cross-head self-training (loss  $l_{ss}$ ) and mutual mean-teaching (loss  $l_{stf}$ ). Specifically, unlabeled data, subjected to weak augmentation (Weak Aug.), is input to the teacher network to generate reliable predictions. Simultaneously, the same unlabeled data, subjected to strong augmentation, is input to the student network. Subsequently, the outputs from each teacher decoder are employed to supervise the outputs of both student decoders (e.g., via losses  $l_{st1}^{12}, l_{st2}^{12}, l_{st1}^{21}, l_{st2}^{21}$ ), enforcing prediction consistency under different perturbations and across different decoding paths, thereby effectively leveraging unlabeled data to enhance segmentation performance.

Furthermore, CMMT-Net incorporates a multi-level perturbation strategy. At the data level, it employs Mutual Virtual Adversarial Training (MVAT) to introduce pixel-level adversarial noise and the Cross-set CutMix technique, which generates novel training samples by blending regions between disparate images. These strong augmentation approaches stand in contrast to the weak augmentation applied to the teacher network's inputs, thereby increasing the diversity and difficulty presented within the training data. At the network level, the teacher-student structure inherently constitutes a form of perturbation, with the EMA-updated teacher network furnishing stable supervisory signals. Enabled by these meticulously designed mechanisms, CMMT-Net effectively leverages both limited labeled data and large volumes of unlabeled data, consequently achieving superior performance on medical image segmentation tasks. Experimental results indicate that the proposed CMMT-Net method yielded substantial performance gains in semi-supervised segmentation on the public LA, Pancreas-CT, and ACDC datasets. Specifically, compared to previous state-of-the-art (SOTA) methods, MC-Net+ [52] and BCP [53], CMMT-Net improved the Dice score by 1.79%, 12.73%, and 1.83% on these respective datasets.



**Figure 3.** An overview of CMMT-Net architecture. [51]

### 3.2. Consistency Regularization Segmentation Methods Supervised by Pseudo-Labels

Complementary to the consistency regularization approaches previously reviewed for semi-supervised medical image segmentation, pseudo-labeling represents another pivotal strategy garnering significant research attention. Hybrid methodologies integrating pseudo-labeling with consistency regularization have emerged as a particularly active research direction. This convergence aims to mitigate the principal limitations inherent to pseudo-labeling, namely the potential unreliability of initially generated labels and the critical challenge of effectively selecting and utilizing high-confidence pseudo-labels during the training process.

To tackle the problem of pseudo-label unreliability, Su et al. (2024) proposed a novel Mutual Learning with Reliable Pseudo-Label (MLRPL) framework [54]. The innovation of this method lies in its construction and co-training of two sub-networks with slightly different architectures. Through meticulously designed reliability assessment strategies, it filters and utilizes high-quality pseudo-labels for model optimization.

Specifically, the framework initially employs two sub-networks (sharing an encoder but with independent decoders) to make independent predictions on the same input image, generating respective preliminary pseudo-labels. Subsequently, a dual reliability assessment mechanism is introduced: first, a "Mutual Comparison" strategy is adopted, comparing the prediction confidences of the two sub-networks pixel-by-pixel and selecting the one with higher confidence as a more reliable pseudo-label candidate; second, an "Intra-class Consistency" metric is proposed, further evaluating pseudo-label reliability by calculating the similarity between a pixel's feature and its predicted class prototype, quantifying this reliability as a weighting coefficient. The loss function design ingeniously integrates both assessment results: the mutual comparison outcome determines whether knowledge transfer occurs between the sub-networks (i.e., using one sub-network's pseudo-label to guide the other only when its prediction is significantly superior), while the intra-class consistency metric serves as a weight to finely adjust the cross-entropy loss, assigning greater influence to highly reliable pseudo-labels. This

dual-guarantee mechanism effectively suppresses noise introduced by unreliable pseudo-labels, significantly enhancing the performance and robustness of semi-supervised medical image segmentation and offering a highly promising solution to address the annotation bottleneck in medical image analysis.

On the Pancreas-CT dataset, using 10% labeled data, MLRPL achieved a Dice coefficient improvement of up to 21.99% compared to the baseline model (V-Net) [55]. Furthermore, MLRPL demonstrated significant advantages over existing semi-supervised methods; for instance, when trained with 10% labeled data on the Pancreas-CT dataset, our method achieved a 2.40% improvement in Dice coefficient compared to URPC [50]. Under certain experimental settings, MLRPL's performance was even comparable to fully supervised models trained with the complete labeled dataset.

Building upon prior work demonstrating the potential of reliability assessment mechanisms for enhancing pseudo-label quality, such as that by Su et al., the Cooperative Rectification Learning Network (CRLN) [56] proposed by Wang et al. (2025) further investigates the generation of more accurate pseudo-labels through prototype learning and explicit pseudo-label rectification, specifically targeting semi-supervised 3D medical image segmentation tasks.

The CRLN method operates on inputs comprising a small set of 3D medical images with voxel-level annotations (labeled data) and a large volume of unlabeled 3D medical images. To enhance model generalization and leverage consistency regularization, unlabeled data undergo two distinct augmentation processes: weak augmentation (e.g., random cropping, flipping) and strong augmentation (e.g., applying Gaussian noise, CutMix in addition to weak augmentations). Within the Mean Teacher (MT) framework, CRLN feeds strongly augmented unlabeled data to the student model for prediction, while weakly augmented unlabeled data are input to the teacher model to generate pseudo-labels. Labeled data are utilized for supervised learning and subsequent prototype learning. Both student and teacher models share an identical backbone network architecture, typically an encoder-decoder structure like VNet or 3D-UNet.

The core innovation of CRLN lies in its proposed prototype learning and pseudo-label rectification mechanism, designed to leverage prior knowledge learned from labeled data to improve pseudo-label quality. This process consists of two main stages: a learning stage and a rectification stage. Specifically, during the learning stage, the model learns multiple prototypes for each class to capture intra-class variations. Through a Dynamic Interaction Module (DIM), these prototypes interact with feature maps extracted from intermediate layers of the student model's decoder, specifically using labeled data features. The DIM employs a Pair-wise Cross Attention mechanism to compute similarities between prototypes and feature maps, subsequently updating the prototype representations. Following this interaction, an aggregation operation incorporating spatial awareness and cross-class reasoning (implemented via shared-parameter convolutional layers) generates a holistic relationship map,  $M(x)$ , which encodes the association degree between each voxel and all class prototypes. The learning of prototypes and the student's DIM component is accomplished implicitly by minimizing the segmentation loss on the labeled data, potentially including a term that optimizes predictions on labeled data using  $M(x)$ .

In the pseudo-label rectification stage, the model leverages the class prototypes learned from labeled data during the learning stage and the EMA-updated teacher DIM to handle unlabeled data. First, the teacher model generates original pseudo-labels  $\bar{y}$  from the weakly augmented unlabeled data. Simultaneously, this unlabeled data (or its weakly augmented version) is input to the teacher DIM to generate the corresponding relationship map  $M(x)$ . Then, voxel-wise refinement is performed on the original pseudo-labels using the rectification formula:

$$\hat{y}_r = \bar{y} + (1 - \mu) \times M(x) \quad (3)$$

where  $\hat{y}_r$  denotes the rectified pseudo-labels, and  $\mu$  is a learnable parameter that adaptively controls the intensity of the rectification guided by  $M(x)$ . The rectified pseudo-labels  $\hat{y}_r$  are regarded as more reliable supervisory signals to supervise the student model's training on strongly augmented unlabeled data.



Experimental results demonstrate that the CRLN method yielded substantial performance improvements on the LA, Pancreas-CT, and BraTS19 datasets. On the Pancreas-CT dataset specifically, compared to the MC-Net+[52] baseline, CRLN improved the Dice score by 11.8% and 4.57% when utilizing 10% and 20% labeled data, respectively. These results highlight the enhanced accuracy and robustness of CRLN for semi-supervised medical image segmentation.

### 3.3. Segmentation Methods Combining Contrastive Learning and Consistency Regularization

To further enhance the utilization efficiency of unlabeled data, a pivotal research direction involves the synergistic integration of Contrastive Learning (CL) and Consistency Regularization (CR). Contrastive learning improves feature discrimination by comparing similar/dissimilar regions, while consistency regularization ensures stable predictions under perturbations. Its integration with consistency regularization enables the imposition of constraints concomitantly within both the feature embedding space and the model prediction space. This dual constraint paradigm aims to improve model generalization and performance.

The CRCFP proposed by Bashir et al. (2024) is built upon the DeepLab-v3 architecture and incorporates multiple techniques to enhance performance.

DeepLab-v3 is a classic semantic segmentation network employing an encoder-decoder structure with atrous convolutions. The encoder, typically a pre-trained ResNet network, extracts image features; the Atrous Spatial Pyramid Pooling (ASPP) module captures multi-scale contextual information using atrous convolutions with different dilation rates; the decoder progressively recovers spatial resolution and fuses multi-level features to generate pixel-level class predictions.

As illustrated in Figure 4, the core of the CRCFP model consists of a shared encoder (h) and decoder (g). For labeled data, the model utilizes a supervised branch (blue path), where the input is processed through the shared backbone network and subsequently fed into the main pixel-wise classifier ( $C_f$ ) to obtain predictions. The cross-entropy loss  $\mathcal{L}_{sup}$  is then computed. For unlabeled data  $x_u$ , the model incorporates two unsupervised pathways:

1. Context-aware Consistency Path (Green path): Two overlapping patches,  $x_{u1}$  and  $x_{u2}$ , cropped from the unlabeled image are passed through the shared backbone network. Their resulting features are mapped through a projection head ( $\Phi$ ) to obtain embeddings  $\phi_{u1}$  and  $\phi_{u2}$ . A contrastive loss,  $\mathcal{L}_{cont}$ , is employed to enforce feature consistency under differing contextual views.
2. Cross-Consistency Training Path (Brown path): Features extracted from the complete unlabeled image  $x_u$  are fed into the main classifier  $C_f$  to yield prediction  $\hat{y}_u$ . Concurrently, these features, subjected to perturbation (P), are input to multiple auxiliary classifiers, producing predictions  $\hat{y}_u^k$ . A cross-consistency loss,  $\mathcal{L}_{cross}$ , enforces consistency between the outputs of the main and auxiliary classifiers.

Furthermore, an entropy minimization loss,  $\mathcal{L}_{ent}$ , is applied to the main classifier's predictions  $\hat{y}_u$  for unlabeled data to enhance prediction confidence. Finally, all constituent loss terms are weighted and combined for end-to-end training.

Experimental results demonstrate that the CRCFP framework exhibits superior performance in semi-supervised semantic segmentation tasks on two public histology datasets, BCSS and MoNuSeg. The advantages of this framework are particularly pronounced in low-data regimes, and its performance using only a fraction of the labeled data approaches that achieved by fully supervised models.

CRCFP, through context-aware consistency and cross-consistency training, effectively leveraged contextual information and feature perturbations from unlabeled data, enhancing model robustness. However, CRCFP primarily focused on consistency at the global feature level, potentially resulting in lower accuracy for crucial edge details significant in medical imaging. Addressing this, Yang et al. (2025) further explored the application of contrastive learning specifically to enhance segmentation accuracy in edge regions, integrating it with feature perturbation consistency within a novel network architecture for semi-supervised medical image segmentation.

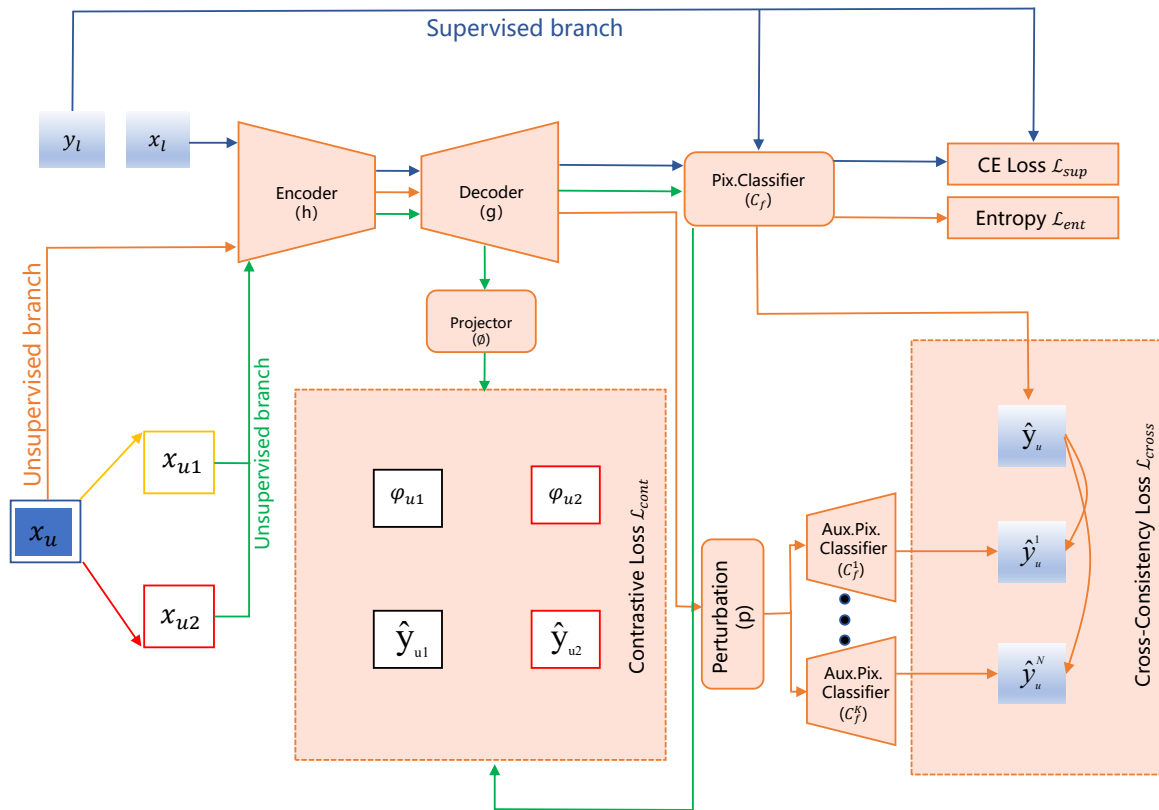


Figure 4. the framework diagram of CRCFP [45].

The methodology [46] proposed by Yang et al. (2025) similarly employs an architecture based on a shared encoder and multiple decoders, diverging from the strategy of Bashir et al., which utilized lightweight auxiliary classifiers; Yang et al. designed multiple complete decoder branches. Its core innovation resides in the introduction of a structured weak-to-strong feature perturbation mechanism. Operating not at the image level, but rather at the feature level, it leverages the statistical information (mean, standard deviation) of feature maps to perform controllable linear transformations, applying perturbations of incrementally increasing intensity across the different decoder branches. This strategy is designed to explore the feature space more systematically and comprehensively, facilitating the learning of representations robust to perturbations.

To effectively leverage this structured perturbation for learning from unlabeled data, the method incorporates a feature perturbation consistency loss, compelling the model to yield consistent predictions for the same input under varying perturbation strengths. Crucially, for generating reliable supervisory targets to compute this consistency loss, the model does not simply average the predictions from the various branches but employs an uncertainty-weighted aggregation strategy. This strategy fuses the prediction results based on the confidence level (derived from uncertainty estimation) associated with each perturbed branch, thereby producing more dependable aggregated pseudo-labels.

Furthermore, specifically addressing the critical and often challenging edge regions in medical image segmentation, Yang et al. designed an Edge-Aware Contrastive Learning (ECL) branch. The novelty of this branch lies in its intelligent sample selection mechanism. It utilizes the prediction results generated by the main segmentation branch along with corresponding uncertainty maps to identify and prioritize the selection of pixels located in edge regions. By constructing positive and negative pairs from these carefully chosen edge pixel features and applying a contrastive loss, the model is explicitly guided to learn more discriminative feature representations pertinent to edge areas, consequently enhancing edge segmentation accuracy.

On the public BraTS2020, LA, and ACDC datasets, the method demonstrably outperformed contemporary baseline models including SFPC [43], PLCT [57], CAML [58], and MC-Net+ [52] in

semi-supervised segmentation tasks, with its advantages being particularly pronounced under low labeled data regimes (e.g., 5%). Moreover, the method demonstrated its capability to significantly enhance segmentation accuracy within challenging edge regions, addressing a prevalent limitation of existing techniques and exhibiting superior precision at object boundaries.

This section reviews semi-supervised medical image segmentation approaches, with a focus on representative methods employing consistency regularization strategies and their integration with pseudo-labeling and contrastive learning. To enable a clear and comprehensive performance comparison of deep learning segmentation models within the semi-supervised paradigm, Tables 3 and 4 respectively summarize the DSC, Jaccard, 95HD, and ASD scores achieved by representative methods on the 2D ACDC2017 dataset and the 3D BraTS2020 dataset, utilizing labeled data proportions of only 5% and 10%.

**Table 3.** Comparison of classical semi-supervised methods on the 2D dataset ACDC2017 [46]

Method	% Labeled Scans	2017 ACDC (2D)			
		DSC (%)	Jaccard (%)	95HD (mm)	ASD (mm)
<i>Using 5% labeled scans</i>					
UAMT [49]	5	51.23± 1.96	41.82± 1.62	17.13± 2.82	7.76± 2.01
SASSNet [59]	5	58.47± 1.74	47.04± 2.02	18.04± 3.63	7.31± 1.53
Tri-U-MT [60]	5	59.15± 2.01	47.37± 1.82	17.37± 2.77	7.34± 1.31
DTC [61]	5	57.09± 1.57	45.61± 1.23	20.63± 2.61	7.05± 1.94
CoraNet [62]	5	59.91± 2.08	48.37± 1.75	15.53± 2.23	5.96± 1.42
SPCL [63]	5	81.82± 1.24	70.62± 1.04	5.96± 1.62	2.21± 0.29
MC-Net+ [52]	5	63.47± 1.75	53.13± 1.41	7.38± 1.68	2.37± 0.32
URPC [50]	5	62.57± 1.18	52.75± 1.36	7.79± 1.85	2.64± 0.36
PLCT [57]	5	78.42± 1.45	67.43± 1.25	6.54± 1.62	2.48± 0.24
DGCL [41]	5	80.57± 1.12	68.74± 0.96	6.04± 1.73	2.17± 0.30
CAML [64]	5	79.04± 0.83	68.45± 0.97	6.28± 1.79	2.24± 0.26
DCNet [40]	5	71.57± 1.58	61.12± 1.19	8.37± 1.92	4.08± 0.84
SFPC [43]	5	80.52± 1.03	68.73± 0.88	6.08± 1.47	2.14± 0.22
<i>Using 10% labeled scans</i>					
UAMT [49]	10	81.86± 1.25	71.07± 1.43	12.92± 1.68	3.49± 0.64
SASSNet [59]	10	84.61± 1.97	74.53± 1.78	6.02± 1.54	1.71± 0.35
Tri-U-MT [60]	10	84.06± 1.69	74.32± 1.77	7.41± 1.63	2.59± 0.51
DTC [61]	10	82.91± 1.65	71.61± 1.81	8.69± 1.84	3.04± 0.59
CoraNet [62]	10	84.56± 1.53	74.41± 1.49	6.11± 1.15	2.35± 0.44
SPCL [63]	10	87.57± 1.15	78.63± 0.89	4.87± 0.79	1.31± 0.27
MC-Net+ [52]	10	86.78± 1.41	77.31± 1.27	6.92± 0.95	2.04± 0.37
URPC [50]	10	85.18± 0.98	74.65± 0.83	5.01± 0.79	1.52± 0.26
PLCT [57]	10	86.83± 1.17	77.04± 0.83	6.62± 0.86	2.27± 0.42
DGCL [41]	10	87.74± 1.06	78.82± 1.22	4.74± 0.73	1.56± 0.24
CAML [64]	10	87.67± 0.83	78.70± 0.91	4.97± 0.62	1.35± 0.17
DCNet [40]	10	87.81± 0.88	78.96± 0.94	4.84± 0.81	1.23± 0.21
SFPC [43]	10	87.76± 0.92	78.94± 0.83	4.90± 0.74	1.28± 0.23

**Table 4.** Comparison of classic semi-supervised methods on the 3D dataset BraTS2020 [46]

Method	% Labeled	BraTS2020 (3D)			
		Scans	DSC (%)	Jaccard (%)	95HD (mm)
Using 5% labeled scans					
UAMT [49]	5	49.46± 2.51	38.46± 1.86	19.57± 3.28	6.54± 0.86
SASSNet [59]	5	51.82± 1.74	43.93± 1.42	23.47± 2.83	7.47± 1.09
Tri-U-MT [60]	5	53.95± 1.97	44.33± 2.18	19.68± 3.06	7.29± 0.84
DTC [61]	5	56.72± 2.04	45.78± 1.67	17.38± 4.31	6.28± 1.22
CoraNet [62]	5	57.97± 1.83	46.40± 1.64	19.52± 2.80	5.83± 0.85
SPCL [63]	5	78.73± 1.54	67.90± 1.29	16.26± 1.68	4.47± 1.08
MC-Net+ [52]	5	58.91± 1.47	47.24± 1.36	20.82± 3.35	7.14± 1.12
URPC [50]	5	60.48± 2.01	50.69± 1.99	18.21± 3.27	7.12± 0.95
PLCT [57]	5	65.74± 2.17	55.40± 1.85	16.61± 3.04	6.85± 1.39
DGCL [41]	5	80.21± 0.75	68.86± 0.63	14.91± 1.53	4.63± 1.16
CAML [64]	5	77.86± 0.96	66.42± 1.37	15.21± 1.74	5.10± 1.12
DCNet [40]	5	78.52± 1.21	67.81± 1.07	17.37± 1.48	4.32± 0.96
SFPC [43]	5	80.76± 0.74	69.18± 0.83	14.87± 1.92	4.02± 0.75
Using 10% labeled scans					
UAMT [49]	10	81.04± 1.46	68.88± 1.57	17.27± 3.35	6.25± 1.63
SASSNet [59]	10	82.36± 2.08	71.03± 2.35	14.80± 3.72	4.11± 1.54
Tri-U-MT [60]	10	82.83± 1.35	71.52± 1.21	15.19± 2.86	3.57± 1.30
DTC [61]	10	81.98± 2.41	70.41± 2.73	16.27± 3.62	3.62± 1.71
CoraNet [62]	10	81.38± 1.68	70.01± 1.83	13.94± 2.72	3.95± 1.26
SPCL [63]	10	84.65± 1.16	73.91± 1.19	12.24± 1.47	3.28± 0.42
MC-Net+ [52]	10	83.93± 1.73	72.34± 1.69	13.52± 2.74	3.37± 1.13
URPC [50]	10	84.23± 1.41	72.37± 1.26	11.52± 1.79	3.26± 1.14
PLCT [57]	10	83.66± 1.82	71.99± 1.67	13.68± 1.29	3.59± 1.02
DGCL [41]	10	84.02± 1.24	72.16± 1.07	12.98± 1.28	3.02± 0.96
CAML [64]	10	84.34± 1.03	73.84± 0.92	12.02± 1.84	3.31± 0.58
DCNet [40]	10	83.39± 0.97	71.94± 0.88	11.93± 1.24	3.50± 0.33
SFPC [43]	10	85.01± 0.89	74.67± 1.14	10.73± 1.36	3.03± 0.31

4. Weakly Supervised Medical Image Segmentation Methods

To mitigate the reliance of fully supervised learning on large-scale, high-quality pixel-level annotations, research in Weakly Supervised Learning (WSL) has garnered significant attention. WSL aims to replace precise pixel-level segmentation masks with easily obtainable coarse-grained annotations, such as image-level labels, bounding boxes, or scribbles. This approach significantly enhances the efficiency of medical image annotation, reduces labeling costs, and offers more feasible solutions for clinical applications.

The field of weakly supervised medical image segmentation has witnessed rapid development in recent years, with the emergence of diverse methods utilizing various types of weak labels. Early research primarily focused on leveraging image-level labels, generating Class Activation Maps (CAMs) [8] to localize target regions. However, CAMs typically highlight only the most discriminative parts of the target, often leading to incomplete segmentation results. To address this limitation, subsequent studies have explored various strategies, including incorporating saliency information [65], iterative region mining [12], and employing adversarial learning [66]. Furthermore, some methods utilize stronger forms of weak labels, such as bounding boxes [67], scribbles [68], and point annotations, to provide more precise localization information.

Recently, approaches combining multiple forms of weak labels, leveraging prior knowledge (e.g., target size, shape), and introducing self-supervised learning [69] have gained increasing attention. These methods continuously improve the performance of weakly supervised medical image segmentation, advancing towards greater precision and robustness.

To clearly delineate the developmental trajectory of weakly supervised medical image segmentation and provide an in-depth exploration of the advantages and disadvantages of different approaches, this paper will subsequently classify and summarize existing methods based on the type of weak annotation information utilized and the learning paradigm employed. Specifically, we will first focus on weakly supervised methods that rely solely on image-level labels (Section 4.1). These methods entail the lowest annotation cost but typically yield relatively lower segmentation accuracy. Subsequently, we will delve into weakly semi-supervised methods (Section 4.2) capable of leveraging a small amount of sparse annotations (e.g., scribbles, point annotations) combined with large volumes of unlabeled data. Such approaches strike a more effective balance between annotation cost and segmentation accuracy.

4.1. Image-Level Label-Based Weakly Supervised Medical Image Segmentation

In the domain of weakly supervised medical image segmentation relying solely on image-level labels, various effective methods have emerged. Among these, Class Activation Mapping (CAM) [8] and Multiple Instance Learning (MIL) [70] represent two of the most representative and widely adopted techniques. CAM provides weak supervision signals for segmentation by visualizing the internal activations of convolutional neural networks, thereby revealing regions within the image pertinent to specific classes. Conversely, MIL treats an image as a ‘bag’ and pixels as ‘instances,’ inferring pixel-level segmentation results by learning from bag-level labels. A detailed exposition of recent research based on these two categories of methods follows.

4.1.1. CAM: A Powerful Tool for Weakly Supervised Medical Image Segmentation

The network architecture for Class Activation Mapping (CAM) is illustrated in Figure 5. An input image is first processed through a series of convolutional layers to extract features. The key modification resides in the final part of the network: following the last convolutional layer (which outputs a set of feature maps), the conventional fully connected layers are removed, and a Global Average Pooling (GAP) layer is applied directly. The GAP layer computes the spatial average over each feature map from the last convolutional layer, compressing it into a feature vector whose dimension equals the number of channels (n) in that layer. This feature vector, output by GAP, is then fed directly into the final fully connected output layer (e.g., a softmax layer for classification). For any given class (e.g., ‘Australian terrier’), the final score is obtained as a weighted sum of the elements in the GAP output vector, where the weights ( $W_1, W_2, \dots, W_n$ ) represent the connection strength between the average activation of each feature map and the output node for that class. This architecture allows the output layer weights to be projected back onto the feature maps of the last convolutional layer, thereby generating Class Activation Maps (CAM), which intuitively localize the image regions most contributory to the classification of a specific class.

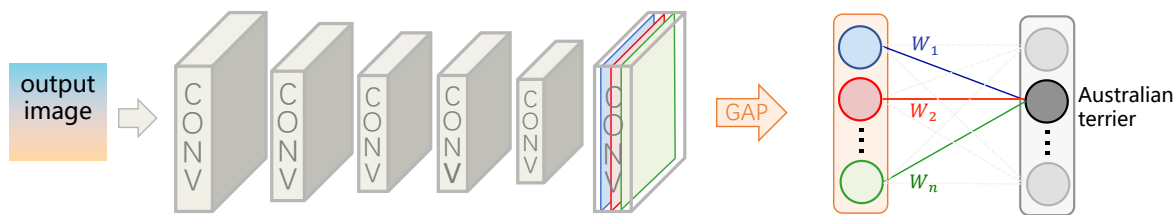


Figure 5. The Architecture Diagram of CAM Network [8]

Leveraging the principle of CAM, which utilizes image-level supervision to generate localization maps, Chikontwe et al. (2022) further developed a weakly supervised learning framework specifically addressing the challenges of segmenting whole-slide histopathology images (WSIs) in digital pathol-



ogy [71]. To manage the substantial memory requirements of WSIs while retaining global contextual information, the method initially employs a neural compression technique. An encoder, trained on image patches via unsupervised contrastive learning, is subsequently utilized to compress the entire WSI into a fixed-size feature map of significantly reduced dimensionality that preserves critical spatial information, thus circumventing the limitations inherent in conventional patch-based processing.

Building upon this compressed representation, the core of the framework entails a single-stage weakly supervised segmentation process, augmented by an innovative self-supervised Class Activation Map (CAM) refinement mechanism. A segmentation network initially generates a preliminary CAM based on the compressed features. Rather than directly utilizing this initial CAM, the method refines it via two key modules: first, the “ $\tau$ -MaskOut” technique identifies and masks input features corresponding to low-confidence regions in the initial CAM, serving as a form of spatial regularization; second, a Pixel Correlation Module (PCM) employs self-attention to compute the correlation between the masked features and the initial CAM, thereby promoting activation expansion to generate a more comprehensive refined CAM.

To facilitate end-to-end training and effectively leverage weak labels alongside self-supervised signals, the framework employs a specifically designed composite loss function. This incorporates a standard classification loss, associating the global average pooled outputs of both the initial and refined CAMs with image-level labels; an equivariant regularization loss, enforcing consistency between the initial and refined CAMs; and a conditional entropy minimization loss, aimed at mitigating prediction uncertainty. Minimization of this composite objective, which integrates classification, consistency, and uncertainty constraints, enables the learning of high-quality segmentation masks conditioned solely on image-level labels.

Experimental results demonstrate that the proposed method, utilizing only image-level labels, achieves segmentation accuracy remarkably close to that of a fully supervised UNet model trained on the same compressed data. Specifically, the reported Dice Similarity Coefficient (DSC) gaps were approximately 1.6% on the Set-I dataset and 8.5% on the Set-II dataset. Furthermore, the majority of these performance improvements were verified as statistically significant ( $p < 0.05$ ), robustly validating the efficacy of the proposed self-supervised CAM refinement framework.

However, CAM-based approaches typically encounter a limitation: the generated activation maps often highlight only the most discriminative regions of the target object, potentially overlooking other less salient but equally relevant portions. To address this issue, G. Patel et al. (2022) introduced a novel multi-modal learning strategy [72] that leverages both intra-modal and cross-modal equivariant constraints to enhance CAMs. This approach is based on the observation that while different modalities emphasize distinct tissue characteristics, they should yield consistent segmentations over the same underlying anatomical structures. Building upon this insight, G. Patel et al. devised a composite loss function incorporating terms for intra-modal equivariance, cross-modal equivariance, and KL divergence, integrated with the standard image-level classification objective. This formulation aims to produce CAMs that are both more complete and more accurate.

Central to the training process (detailed in Algorithm 1) is the self-supervised refinement of Class Activation Maps (CAMs) by leveraging multi-modal data and spatial transformations. Specifically, the training involves  $K$  neural networks, one for each modality, parameterized by  $\Theta_{m_k}$ . In each training iteration, the algorithm processes a minibatch of data. Initially, the same random spatial transformation  $\pi$  is applied to the images of all modalities within the minibatch. Subsequently, forward propagation is performed on both the original and transformed images to obtain their respective CAMs (denoted as  $M$  and  $M_\pi$ ) and softmax probability outputs ( $P$  and  $P_\pi$ ). The crucial step involves computing a composite loss  $L(\Theta_{m_k})$  for each modality  $k$ , which is a weighted sum formed by the standard classification loss  $L_c$  utilizing the image-level label  $y$ , the within-modality equivariance loss  $L_{ER}$  enforcing consistency between  $M$  and  $M_\pi$  under the transformation  $\pi$ , the cross-modal knowledge distillation loss  $L_{KD}$  employing KL divergence to encourage alignment between  $P$  and  $P_\pi$  across different modalities, and the central cross-modal equivariance loss  $L_{CMER}$  enforcing consistency between  $M$  and  $M_\pi$  of different

modalities when subjected to the *same* transformation  $\pi$ . Finally, gradients are computed based on this composite loss, and the parameters  $\Theta_{m_k}$  for each network are updated accordingly. Experiments on the BraTS brain tumor segmentation and prostate segmentation datasets demonstrate that the proposed method significantly outperforms standard CAM, GradCAM++ [73], as well as state-of-the-art weakly supervised segmentation methods such as SEAM [69].

---

**Algorithm 1** Training algorithm.

---

**Require:** Training dataset  $\mathcal{D}$

```

1:  $K = \text{Number of image modalities}$ 
2:  $\Pi = \text{Set of transformations}$ 
3:  $T = \text{Total number of Epochs}$ 
4: for  $k$  in  $[1, K]$  do
5:   Initialize  $\Theta_{m_k}$ 
6: end for
7: for  $t$  in  $[1, T]$  do
8:   for every minibatch  $B$  in  $\mathcal{D}$  do
9:      $\pi \leftarrow \pi \sim \Pi$ 
10:     $\mathcal{M} \leftarrow \{\{\mathbf{M}_i^{m_k}\}_{k=1}^K\}_{i \in B}$ 
11:     $\mathcal{P} \leftarrow \{\{\mathbf{P}_i^{m_k}\}_{k=1}^K\}_{i \in B}$ 
12:     $\mathcal{D}_\pi \leftarrow \{\{\pi(\mathbf{X}_i^{m_k})\}_{k=1}^K, \mathbf{y}_i\}_{i \in B}$ 
13:     $\mathcal{M}_\pi \leftarrow \{\{\mathbf{M}_i^{m_k}\}_{k=1}^K\}_{i \in B}$ 
14:     $\mathcal{P}_\pi \leftarrow \{\{\mathbf{P}_i^{m_k}\}_{k=1}^K\}_{i \in B}$ 
15:    for  $k$  in  $[1, K]$  do
16:      Compute  $\mathcal{L}_C(\Theta_{m_k}), \mathcal{L}_{ER}(\Theta_{m_k}), \mathcal{L}_{KD}(\Theta_{m_k}, m_{l \neq k}), \mathcal{L}_{CMER}(\Theta_{m_k}, m_{l \neq k})$ .
17:      Compute  $\mathcal{L}(\Theta_{m_k})$ 
18:       $\text{loss} \leftarrow \frac{1}{|B|} \sum_{i \in B} \mathcal{L}(\Theta_{m_k})$ 
19:      Compute gradients of loss w.r.t  $\Theta_{m_k}$ 
20:      Update  $\Theta_{m_k}$  using the optimizer
21:    end for
22:  end for
23: end for

```

---

While prior methods often address single-class segmentation scenarios, medical images frequently contain multiple lesions with diverse morphologies. To overcome this challenge, Yang et al. (2024) introduced an Anomaly-Guided Mechanism (AGM) [74] for multi-class lesion segmentation in Optical Coherence Tomography (OCT) images.

AGM initially employs a GANomaly network, trained on normal images, to generate a pseudo-healthy counterpart for each input OCT image. An anomaly-discriminative representation, highlighting abnormal regions, is then produced by computing the difference between the original image and its pseudo-healthy counterpart. AGM utilizes a dual-branch architecture: a backbone branch processes concatenated information from the original and pseudo-healthy images, while an Anomaly Self-Attention Module (ASAM) branch processes the anomaly-discriminative representation. The ASAM branch leverages self-attention to capture global contextual information and pixel dependencies within abnormal patterns, particularly focusing on small lesions. Feature maps from the two branches are fused (e.g., via element-wise multiplication) and subsequently processed through Global Max Pooling (GMP) and Fully Connected (FC) layers for multi-label classification and initial Class Activation Map (specifically GradCAM) generation.

A key component is an iterative refinement learning stage: CAMs and classification predictions from the preceding iteration are used to create a weighted Region of Interest (ROI) mask. This mask enhances the input to the backbone branch in the subsequent training iteration, thereby guiding the model to focus more precisely on potential lesion areas. Finally, the refined CAMs undergo post-processing (including retina mask extraction, thresholding, and class selection) to yield high-quality pseudo pixel-level labels, which are then used to train a standard segmentation network.

By integrating anomaly detection and self-attention within the Weakly Supervised Semantic Segmentation (WSSS) framework and incorporating iterative refinement, AGM aims to enhance localization accuracy, especially when dealing with small, low-contrast, and co-existing lesions common in medical images. The method achieved state-of-the-art (SOTA) performance across multiple datasets. For instance, on the public RESC and Duke SD-OCT datasets, as well as a private retina OCT dataset employed in the study, AGM demonstrated significantly superior performance in terms of pseudo-label quality (measured by mean Intersection over Union, mIoU) and final segmentation results compared to baseline methods such as SEAM [75], ReCAM [76].

#### 4.1.2. MIL: An Effective Strategy for Weakly Supervised Medical Image Segmentation

Multiple Instance Learning (MIL) offers an effective weakly supervised strategy for medical image segmentation using only image-level labels. Within the MIL framework, an image is treated as a "bag," and each pixel (or region) within the image is considered an "instance." The objective of the model is to predict instance-level labels (i.e., whether each pixel/region belongs to a lesion) based on the bag-level label (e.g., whether the image contains a lesion).

To overcome a common limitation of traditional MIL methods – overlooking long-range dependencies among pixels in histopathology image segmentation – Li et al. (2023) proposed SA-MIL [77], a weakly supervised segmentation method based on self-attention. Instead of directly feeding pixel features into a classifier for prediction, as is common in conventional MIL approaches, SA-MIL integrates self-attention modules at multiple stages of feature extraction. As illustrated in Figure 6, SA-MIL employs the first three convolutional stages of VGGNet as its backbone network (represented by the sequence of purple 3x3 Conv modules). A Self-Attention Module (SAM, orange module) is inserted between the convolutional and pooling layers within each stage. The core component of the SAM is the Criss-Cross Attention (CCA) module, which aggregates contextual information for each pixel along its horizontal and vertical directions through two recurrent operations, thereby establishing long-range dependencies among pixels. This mechanism effectively enhances the feature representation capability, enabling the model to better distinguish between foreground (cancerous regions) and background.

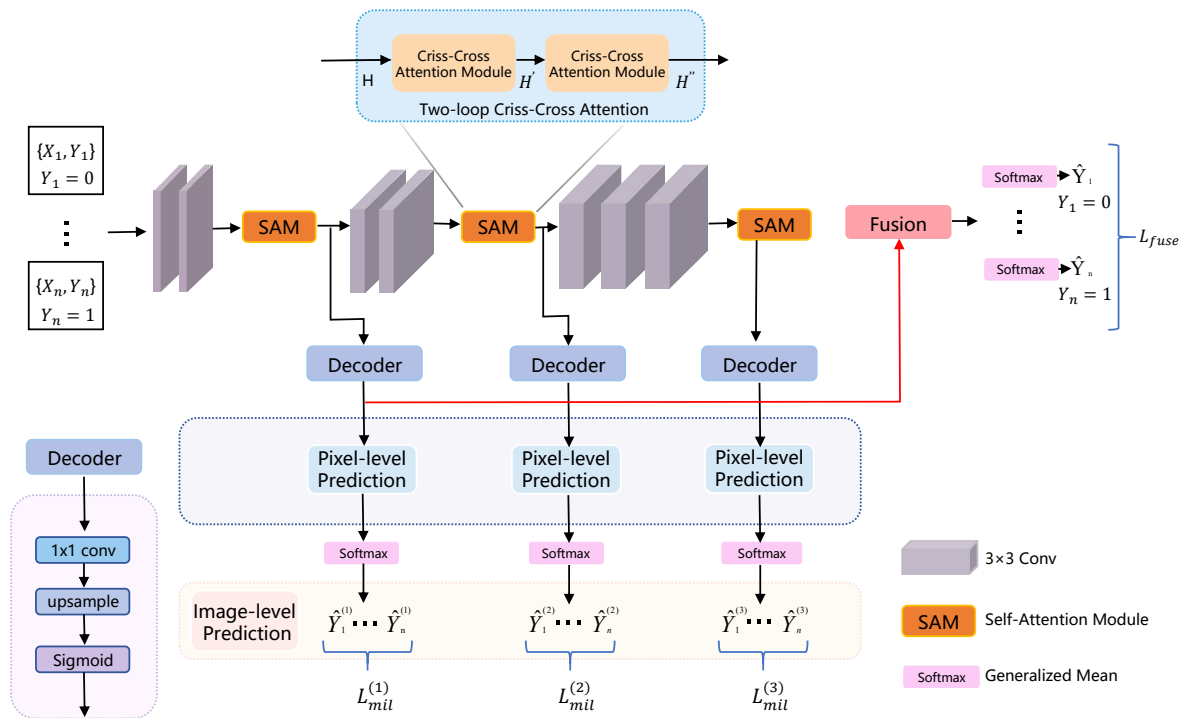


Figure 6. The framework of SA-MIL [77]

Furthermore, SA-MIL adopts a deep supervision strategy to fully leverage the limited image-level annotation information. Following each SAM module, a decoder branch (light blue module) is connected. This decoder generates a pixel-level prediction map corresponding to that stage. These pixel-level prediction maps are subsequently aggregated via Softmax activation and Generalized Mean (GeM) pooling operations to produce an image-level prediction for that stage. The corresponding Multiple Instance Learning loss (MIL Loss) is then computed using the image-level ground truth label  $Y$ . Finally, the output features from the last SAM module, along with the pixel-level prediction maps generated by all intermediate decoders, are jointly fed into a Fusion module (red module). A final fusion loss,  $L_{fuse}$ , is utilized to supervise this ultimate segmentation result.

SA-MIL was extensively validated through experiments on two histopathology image datasets (colon cancer tissue and cervical cancer cells) and compared against various weakly supervised and fully supervised methods. The experimental results demonstrate that SA-MIL, utilizing only image-level labels, significantly outperforms weakly supervised methods such as PSPS [78] and Implicit PointRend [79] across multiple metrics, including F1 score, Hausdorff distance, mIoU, and mAP. Furthermore, its performance approaches, and in some cases matches, that of fully supervised methods like U-Net.

Seeböck et al.(2024) introduced a novel strategy, termed Anomaly Guided Segmentation (ANGUS) [80], which utilizes the output of a pre-trained anomaly detection model as supplementary semantic context to enhance lesion segmentation in retinal Optical Coherence Tomography (OCT) images. This approach leverages weak spatial information derived from anomaly detection, differing from methods relying solely on specific target lesion annotations.

Specifically, the method is implemented in three stages: First, an anomaly detection model (e.g., WeakAnD), pre-trained on a dataset of healthy OCT images, is applied to the segmentation training set to generate pixel-wise weak anomaly maps. These maps encode regions deviating from normal patterns. Second, the manually annotated ground truth masks for the target lesions are merged with these weak anomaly maps to construct an expanded annotation scheme. This scheme adds an 'anomaly' class to the original lesion classes, representing areas identified as abnormal by the detector but not explicitly labeled as a target lesion, potentially encompassing other pathologies or variations. Finally, a segmentation network (e.g., U-Net) is trained using this expanded annotation system, typically employing a weighted cross-entropy loss function. This training strategy compels the network not only to learn features of the annotated lesions but also to discriminate between normal tissue, known lesions, and unannotated anomalies, thereby improving robustness to complex lesions, inter-class variability, and real-world data variations without requiring additional manual annotation effort.

Experimental results demonstrated that across two in-house and two public datasets, targeting various lesion types (including IRC, SRF, PED, HRF, SHRM, etc.), the ANGUS method consistently, and often statistically significantly, outperformed standard U-Net baselines trained solely on target lesion annotations. This was evidenced by improvements in metrics such as Dice coefficient, precision, and recall, enabling accurate segmentation of lesions in retinal OCT images.

#### 4.2. Weakly Semi-Supervised Medical Image Segmentation Methods

Weakly semi-supervised learning aims to enhance medical image segmentation performance by leveraging a small amount of weak annotations (e.g., scribbles, points) and a large volume of unlabeled data, effectively mitigating the annotation bottleneck. This section focuses on representative methods, SOUSA and Point SEGTR, whose core idea involves integrating supervision from sparse annotations with consistency constraints derived from unlabeled data, albeit with differing implementation emphases.

Proposed by Gao et al. (2022), Segmentation Only Uses Sparse Annotations (SOUSA) [58] is a framework for medical image segmentation that integrates weakly supervised and semi-supervised consistency learning (WSCL). SOUSA leverages sparse semantic information from scribble annotations alongside consistency priors inherent in unlabeled data. The framework is based on the Mean Teacher (MT) architecture, comprising a student network and a teacher network. The student network features

dual output heads: a primary segmentation head predicting pixel-level segmentation masks, and an auxiliary regression head tasked with predicting Geodesic distance maps, which are pre-computed offline based on the input image and its corresponding scribble annotations.

For images with scribble annotations, the student network is supervised using two loss components: the segmentation head employs a Partial Cross-Entropy (PCE) loss, calculated only on scribble pixels, while the regression head utilizes a regression loss comparing its predicted distance map with the pre-computed Geodesic distance map. The incorporation of the Geodesic distance map aims to more fully exploit the sparse scribble information by providing spatial context that guides the model's focus toward the target regions.

For unlabeled images, a consistency regularization strategy is applied. The same unlabeled image, subjected to different perturbations (e.g., random noise, linear transformations), is fed into the student and teacher networks, respectively. Consistency between their segmentation predictions is enforced using two loss functions: the standard Mean Squared Error (MSE) loss and a novel Multi-angle Projection Reconstruction (MPR) loss. The MPR loss first randomly rotates the segmentation output maps from the student and teacher networks by the same angle, then projects the rotated maps onto the horizontal and vertical axes, and finally computes the consistency between these projection vectors. Compared to MSE, this projection mechanism is more sensitive to prediction discrepancies at boundaries and in small, discrete regions, addressing a limitation of MSE in penalizing such errors effectively.

Throughout the training process, the total loss function combines the supervised loss (for labeled data) and the consistency loss (for unlabeled data, comprising MSE and MPR components). The weight of the unlabeled data loss is modulated by a function that changes over the course of training (e.g., a Gaussian ramp-up). The student network's parameters are optimized via backpropagation, while the teacher network's parameters are updated using an Exponential Moving Average (EMA) of the student network's parameters (momentum update).

Experimental results on the ACDC cardiac dataset and an in-house colon tumor dataset validated the efficacy of SOUSA. Compared to weakly supervised methods using only scribble annotations (e.g., Scribble2Label, MAAG), SOUSA significantly improved segmentation accuracy (e.g., achieving approximately 5% higher Dice score than PCE+CRF on the ACDC dataset with 10% labeled data). Furthermore, when compared to standard semi-supervised methods adapted for the WSCL setting (e.g., ICT, Uncertainty-aware MT), SOUSA not only yielded higher Dice scores (e.g., 3.55% higher than ICT) but also demonstrated significantly lower Hausdorff Distance (HD) and Average Symmetric Surface Distance (ASSD) metrics (e.g., reductions of 26.20mm and 6.88mm compared to ICT, respectively), indicating more accurate boundary delineation and fewer false positive regions in its segmentation outputs.

While SOUSA demonstrates the potential of weakly semi-supervised learning in medical image segmentation, Point SEGTR [81], proposed by Shi et al. (2023), adopts an alternative strategy by integrating a small amount of pixel-level annotations with a large volume of more readily available point-level annotations. Point SEGTR employs a teacher-student learning framework, wherein the teacher model is based on the Point DETR architecture, incorporating a Point Encoder for encoding point annotations, an Image Encoder (CNN+Transformer) for extracting image features, a Transformer Decoder for fusing information via attention mechanisms, and a Segmentation Head for outputting the segmentation results. During training, Point SEGTR first utilizes pixel-level annotated data to initialize the teacher model, enhancing its robustness to variations in point annotation location by introducing a Multi-Point Consistency (MPC) loss, which enforces consistent segmentation predictions for different points sampled within the same target object. Subsequently, the teacher model is fine-tuned using a large amount of point-annotated data, concurrently applying a Symmetric Consistency (SC) loss. This SC loss encourages consistent predictions for the same input subjected to transformations or perturbations, thereby improving generalization and better leveraging the weak annotations. Finally, the optimized teacher model is utilized to generate high-quality pseudo-segmentation labels for all



point-annotated data. These pseudo-labels, combined with the original pixel-level annotations, are used to jointly train a student network (e.g., Mask R-CNN), which serves as the final model for inference. Through this weakly semi-supervised process incorporating MPC and SC regularization, Point SEGTR achieves competitive segmentation performance while significantly reducing the dependency on pixel-level annotations.

Experiments conducted on three endoscopic datasets (CVC, ETIS, and NASOP) demonstrated that the Point SEGTR framework, when augmented with MPC and SC regularization, enabled the teacher model to achieve performance comparable to or even exceeding that of baseline models trained with 100% pixel-level annotations, even when using only a limited fraction (e.g., 50%) of such annotations. Furthermore, student models trained using pseudo-labels generated by this enhanced teacher model exhibited significantly improved segmentation accuracy compared to baselines. These results validate the effectiveness of the proposed regularization strategies in reducing annotation requirements and enhancing weakly semi-supervised segmentation performance.

SOUA and Point SEGTR represent two distinct weakly semi-supervised strategies. SOUSA emphasizes consistency between scribble annotations and unlabeled data, whereas Point SEGTR focuses on fusing pixel-level and point-level annotations. Both approaches underscore the efficacy of combining sparse annotations with consistency constraints. Future research could explore the integration of additional types of weak annotations with consistency learning, as well as investigate more effective model architectures and training strategies to further advance the performance of weakly semi-supervised medical image segmentation.

This section systematically reviews weakly supervised medical image segmentation methods, focusing on key advancements that leverage image-level labels (e.g., CAM- and MIL-based approaches) as well as weak semi-supervised learning strategies combining sparse annotations (such as scribbles or points) with consistency regularization (e.g., SOUSA, Point SEGTR). To facilitate quantitative performance evaluation of existing weakly supervised semantic segmentation (WSSS) methods, Table 5 systematically presents the DSC and mIoU scores achieved by representative approaches on the RESC and Duke datasets across different lesion regions.

**Table 5.** Performance Comparison of weakly supervised medical image segmentation methods [74]

Dataset Lesions Metrics	RESC						Duke			
	BG		SRF		PED		BG		Fluid	
	DSC	mIoU	DSC	mIoU	DSC	mIoU	DSC	mIoU	DSC	mIoU
IRNet[82]	98.88%	97.78%	49.18%	33.75%	22.98%	14.66%	99.02%	98.10%	17.79%	20.45%
SEAM[83]	98.69%	97.43%	46.44%	34.13%	28.09%	10.71%	98.48%	97.03%	25.48%	17.87%
ReCAM[75]	98.81%	97.66%	31.19%	14.23%	31.99%	19.11%	98.16%	96.41%	18.91%	11.67%
WSMIS[84]	96.90%	95.64%	45.91%	24.64%	10.34%	2.96%	98.16%	96.41%	0.42%	0.42%
MSCAM[85]	98.59%	97.25%	18.52%	10.14%	17.03%	11.97%	98.98%	98.00%	29.93%	17.98%
TransWS [86]	99.07%	98.18%	52.44%	34.88%	30.28%	17.22%	99.06%	98.15%	37.58%	27.01%
DFP [87]	98.83%	97.72%	20.39%	6.40%	31.39%	15.64%	99.10%	98.24%	27.53%	15.14%
AGM [74]	99.15%	98.34%	57.84%	43.94%	34.03%	22.33%	99.13%	98.29%	40.17%	30.06%

5. Unsupervised Medical Image Segmentation Methods

Unsupervised segmentation methods have garnered significant attention owing to their independence from pixel-level annotations, particularly in scenarios where labeled data is scarce or prohibitively expensive to acquire. These methods aim to automatically identify and delineate regions of interest (ROIs), such as lesions or specific organs, within images algorithmically, without reliance on manually annotated training data. Initially, researchers primarily explored traditional image processing techniques, including clustering, region growing, and thresholding, to achieve unsupervised segmentation. These approaches typically operate based on low-level image features (e.g., pixel intensity, texture, edges) and do not require training data. However, such methods are often constrained by hand-crafted features and predefined rules, limiting their ability to effectively handle the complex

anatomical structures and pathological variations inherent in medical imaging. In recent years, the advent of deep learning has spurred significant advancements in unsupervised segmentation methods based on Autoencoders (AEs) and their variants, which exhibit powerful feature representation and learning capabilities [3,6,7]. These techniques often operate by learning the latent distribution of normal images, thereby enabling the identification and segmentation of anomalous regions that deviate from this learned normality model, offering valuable support for clinical diagnosis. Building upon this developmental trajectory, contemporary unsupervised medical image segmentation approaches are predominantly categorized into two main classes: Unsupervised Anomaly Segmentation (UAS) and Unsupervised Domain Adaptation (UDA) for segmentation. These two classes address distinct challenges and have demonstrated considerable success in practical applications.

### 5.1. Unsupervised Anomaly Segmentation Methods

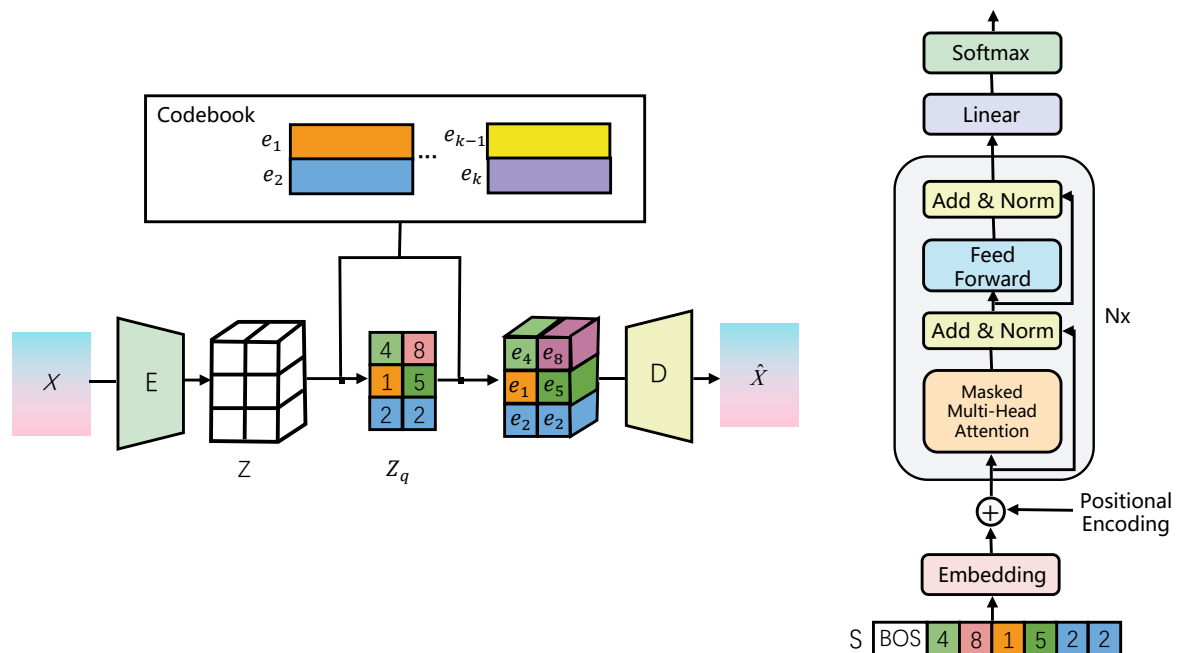
Unsupervised anomaly detection and segmentation in medical image analysis aim to identify and delineate pathological manifestations that deviate from normal anatomical structures. The heterogeneity of pathologies makes it difficult to capture all possible variations with labeled examples, while acquiring large-scale annotated datasets is generally challenging and costly. These factors limit the feasibility of fully supervised approaches and thus motivate the use of unsupervised methods, which obviate the need for explicit anomaly labels and have consequently garnered significant attention. Nevertheless, effectively modeling the complex distribution of normal anatomy to accurately differentiate diverse and potentially subtle deviations (stemming from the aforementioned heterogeneity) remains a considerable challenge for these unsupervised techniques. Recent advancements in deep learning have spurred progress in this area, with methods based on Autoencoders (AEs) and their variants emerging as a prominent research focus due to their potent feature representation capabilities.

Silva-Rodríguez et al. (2022) introduced a novel framework [88] based on constrained optimization applied to attention mechanisms derived from a Variational Autoencoder (VAE). Instead of relying solely on reconstruction error, this approach leverages attention maps extracted from the VAE encoder's intermediate layers to identify anomalies. While initially investigating Gradient-Weighted Class Activation Mapping (Grad-CAM), the authors found non-gradient-weighted Activation Maps (AMs) preferable. A key innovation is the formulation of a constraint loss designed to ensure comprehensive attention coverage over the entire context in normal images. Crucially, this is implemented not as a pixel-wise equality constraint forcing maximum activation everywhere, but as a global *inequality* constraint on the overall activation level of the attention map. This formulation grants the model greater flexibility in learning the distribution of normal patterns.

To address the optimization challenge posed by the inequality constraint, the work explores the extended log-barrier method. This technique incorporates the constraint into the objective function via a smooth, differentiable barrier term. Furthermore, the study proposes and ultimately favors an alternative regularization strategy: maximizing the Shannon entropy of the (softmax-normalized) attention map. This encourages a diffuse attention distribution across normal images and concurrently reduces the number of hyperparameters. At the inference stage, the trained VAE generates the (softmax-normalized) activation map, which serves directly as the anomaly saliency map. Thresholding this saliency map produces the final anomaly segmentation mask. Experimental evaluation highlighted the superior performance of the AMCons strategy (employing Shannon entropy maximization), which demonstrably enhances the separation between the activation distributions corresponding to normal and anomalous patterns, reducing their histogram overlap to a remarkable 10.6%.

Contrasting with the approach of Silva-Rodríguez et al. centered on attention mechanisms and global constraints, Pinaya et al. (2022) investigated an alternative strategy leveraging the synergy between Vector Quantized Variational Autoencoders (VQ-VAEs) and Transformers for unsupervised 3D neuroimaging anomaly detection and segmentation [89] (illustrated schematically in their Figure 7 A and B). The initial stage employs a VQ-VAE architecture (Figure 7 A) to acquire a discrete latent representation  $Z_q$  of the input brain image  $X$ . An encoder network  $E$  maps the input to a continuous latent space  $Z$ , followed by vector quantization against a learnable codebook  $e_k$  to yield the

discrete index grid  $Z_q$ . A decoder network  $D$  reconstructs the image  $\hat{X}$  from the quantized representation. Trained exclusively on healthy brain images, this phase aims to derive a high-fidelity, compact discrete encoding.



**Figure 7.** Framework for Brain Data Processing Based on VQ-VAE and Transformer[89]

Subsequently, the learned discrete latent representation  $Z_q$  is serialized into a one-dimensional sequence  $S$  (Figure 7 B). This sequence forms the input for the second stage, which employs an autoregressive Transformer model. This Transformer learns the sequential dependencies inherent in latent sequences derived from normal brain data by modeling the conditional probability distribution  $p(s_i | s_{<i})$ . Training involves maximizing the log-likelihood across sequences from the healthy training set, enabling the Transformer to internalize the statistical patterns characteristic of normal brain structure in the discrete latent space.

At the inference stage, the trained Transformer identifies indices  $s_i$  in the latent sequence  $S$  exhibiting probabilities  $p(s_i | s_{<i})$  below a predefined threshold, forming a resampling mask  $m$ . This mask guides the generation of a "healed" latent sequence  $\hat{s}$  (by resampling anomalous indices) and its corresponding decoded image  $\hat{x}_{healed}$ . Crucially, the upscaled and smoothed resampling mask  $m$  is then used to filter the pixel/voxel-wise residual map  $|x - \hat{x}_{healed}|$ , leveraging the Transformer's probabilistic anomaly judgment to suppress reconstruction artifacts. Thresholding this filtered residual map yields the final anomaly segmentation.

On synthetic data as well as four real-world 2D/3D brain lesion datasets (UKB, MSLUB, BRATS, WMH), the proposed framework integrating VQ-VAE, Transformer, the residual mask filtering technique, and a multi-view ensemble strategy demonstrated superior performance in unsupervised anomaly segmentation. It consistently and significantly outperformed established state-of-the-art unsupervised methods, including Autoencoder variants (AE) [90], Variational Autoencoders (VAE)[91], and f-AnoGAN [92], achieving the highest reported DICE and AUPRC scores across the evaluated benchmarks.

## 5.2. Unsupervised Domain Adaptation Segmentation Methods

Unsupervised Domain Adaptation (UDA) represents a critical research paradigm for addressing the challenge of domain shift in medical image segmentation. This phenomenon—the significant performance degradation of deep learning models when applied to data distributions differing from their training source (e.g., arising from multi-center, multi-modal, or multi-protocol variations)—constitutes

a primary impediment to clinical translation. Given the prohibitive cost and practical difficulties (including privacy constraints) associated with acquiring fine-grained annotations for the target domain, the core objective of UDA is to effectively adapt models trained on a labeled source domain to an unlabeled target domain, thereby achieving robust generalization performance. This necessitates UDA methods capable of effectively handling complex distribution discrepancies, learning feature representations that are both robust to domain shifts and discriminative for the segmentation task, while mitigating the risk of negative transfer. Successful UDA is therefore critical for enhancing the generalization, robustness, and deployment feasibility of segmentation models in real-world clinical settings. Addressing the core challenges of UDA, numerous methodological advancements have been proposed in recent years. This subsection provides a systematic review of these developments, categorized along three primary lines of methodological progress.

### 5.2.1. Advancements in Source-Data-Free Unsupervised Domain Adaptation

Conventional UDA methods typically necessitate concurrent access to both source and target domain data. However, this assumption often proves impractical in real-world scenarios due to constraints such as data privacy regulations and restrictive sharing protocols. To address the challenge of limited source domain data accessibility during adaptation, Stan and Rostami (2024) and Liu et al. (2023) have independently proposed distinct UDA strategies that operate without requiring access to the source domain data.

Stan and Rostami (2024) introduced a strategy centered on latent feature distribution alignment [93]. The core principle involves leveraging information about the source domain's latent feature distribution during the adaptation phase, coupled with Optimal Transport (OT) theory, to align the target domain features with the source feature distribution.

Specifically, the methodology commences by training a semantic segmentation network on labeled source domain data. Subsequently, this network is employed to extract latent features from the source data, and a Gaussian Mixture Model (GMM) is learned to approximate the distribution of these source latent features. This GMM encapsulates prior knowledge from the source domain and serves as its surrogate during the subsequent adaptation stage.

In the adaptation phase, where the model lacks access to source data and relies solely on unlabeled target domain data, a distribution alignment strategy grounded in OT theory is proposed. This involves employing the Sliced Wasserstein Distance (SWD) as the metric to quantify the discrepancy between the target domain feature distribution and the learned GMM distribution. SWD approximates the Wasserstein distance by projecting high-dimensional distributions onto multiple one-dimensional directions and averaging the corresponding one-dimensional Wasserstein distances. This approach facilitates end-to-end optimization via backpropagation, circumventing the computational challenges associated with directly calculating high-dimensional Wasserstein distances. By minimizing the SWD loss, the method effectively pulls the target domain's latent feature distribution towards the pre-learned GMM distribution, thereby achieving domain alignment. To further enhance performance on the target domain, a regularization term is introduced for fine-tuning the classifier.

The efficacy of this method was validated on two public cardiac image datasets and one abdominal image dataset. Experimental results demonstrate that, even without access to source domain data, the proposed approach achieves performance comparable to, or even surpassing, existing UDA methods for medical image segmentation that utilize source data, underscoring its potential for effective domain adaptation while preserving data privacy.

Alternatively, the OSUDA framework proposed by Liu et al. (2023) presents a "plug-and-play" UDA strategy [94]. This approach similarly obviates the need for source domain data access, relying solely on a segmentation model pre-trained on the source domain. The central concept of OSUDA is to leverage the statistical information contained within Batch Normalization (BN) layers, which are widely utilized in pre-trained models. BN layers encapsulate two categories of statistics: low-order statistics (mean and variance) and high-order statistics (scaling and shifting factors). Research indicates that low-order statistics exhibit domain specificity, whereas high-order statistics tend to be

domain-invariant. Consequently, the OSUDA framework adapts to the target domain by progressively adjusting the low-order statistics while maintaining the high-order statistics.

Specifically, OSUDA employs a progressive adaptation strategy for low-order statistics based on Exponential Momentum Decay (EMD). During adaptation, the target domain's BN statistics gradually converge towards the statistics of the current batch with exponentially decaying momentum, facilitating a smooth transition. For the high-order statistics, OSUDA introduces a High-order BN Statistics Consistency Loss (LHBS) to explicitly constrain them from changing during adaptation.

To further bolster target domain performance, OSUDA incorporates an adaptive channel weighting mechanism. The transferability of each channel is assessed by calculating the difference in low-order statistics between domains and considering the scaling factors from the high-order statistics. Channels deemed more transferable are assigned higher weights within the LHBS loss calculation, thereby encouraging the model to prioritize these informative channels.

Furthermore, OSUDA integrates unsupervised Self-Entropy (SE) minimization and a novel Memory-Consistent Self-Training (MCSF) strategy based on a queue. SE minimization promotes high-confidence predictions on the target domain. MCSF utilizes a dynamic queue to store historical prediction information, enabling the selection of reliable pseudo-labels for self-training by enforcing consistency between current and past predictions, which enhances both performance and stability on the target domain.

OSUDA was validated across multiple medical image segmentation tasks, including cross-modality and cross-subtype brain tumor segmentation, as well as cardiac MR-to-CT segmentation. Experimental results indicate that OSUDA outperforms existing Source-Relaxed UDA methods and achieves performance comparable to UDA methods that utilize source data. This highlights OSUDA's capability to effectively transfer knowledge while preserving source domain data privacy, demonstrating significant practical applicability.

### 5.2.2. Advancements in UDA via Adversarial Training

Distinct from the source-data-free methodologies detailed in the preceding section, another significant category of UDA enhancements focuses on learning domain-invariant features through adversarial training. The primary objective of these approaches is to mitigate the discrepancies between domains, thereby enabling the model to attain robust performance on the target domain.

The ODADA framework proposed by Sun et al. (2022) represents a characteristic adversarial learning-based UDA methodology [95]. However, diverging from conventional adversarial training approaches, ODADA separates features into domain-invariant and domain-specific components, using adversarial training to enhance generalizability. ODADA decomposes features into a Domain-Invariant Representation (DIR) and a Domain-Specific Representation (DSR), employing a novel orthogonality loss function to encourage independence between the DIR and DSR components. As illustrated in Figure 8, the ODADA architecture utilizes a shared Feature Extractor to process input images from both source and target domains, generating an initial mixed feature representation, denoted as  $F_{both}$ . The core of this framework lies in the explicit decomposition of  $F_{both}$ : a dedicated, learnable module termed the DIR Extractor is responsible for extracting the domain-invariant features (DIR),  $F_{di}$ , from  $F_{both}$ . Concurrently, the domain-specific features (DSR),  $F_{ds}$ , are computed via a non-parametric difference layer (Diff Layer), where

$$F_{ds} = F_{both} - F_{di} \quad (4)$$

This design ensures that the decomposition process yielding  $F_{ds}$  is lossless and does not require learning independent parameters. The extracted  $F_{di}$  is utilized in two downstream tasks: firstly, it is fed into a Segmentor to predict segmentation results, supervised solely on the source domain via a Segmentation Loss; secondly, it is passed through a Gradient Reversal Layer (GRL) before being input to a first Domain Classifier, where an Adversarial Loss compels the alignment of  $F_{di}$ , enforcing its domain invariance. Conversely, the computed  $F_{ds}$  is directly input to a second Domain Classifier, which is also trained using an adversarial loss. Critically, however, no GRL is employed



here; the objective is to maximize this classifier's capability to distinguish domains based on  $F_{ds}$ . This mechanism inversely incentivizes the DIR Extractor to isolate as much domain-specific information as possible into  $F_{ds}$ , thereby purifying  $F_{di}$ . Finally, an Orthogonal Loss is imposed between  $F_{di}$  and  $F_{ds}$  to explicitly promote the independence of these two representation components. Through this combination of orthogonal decomposition, targeted adversarial training, and independence constraints, the entire architecture aims to learn purer and more effective domain-invariant features suitable for cross-domain segmentation.

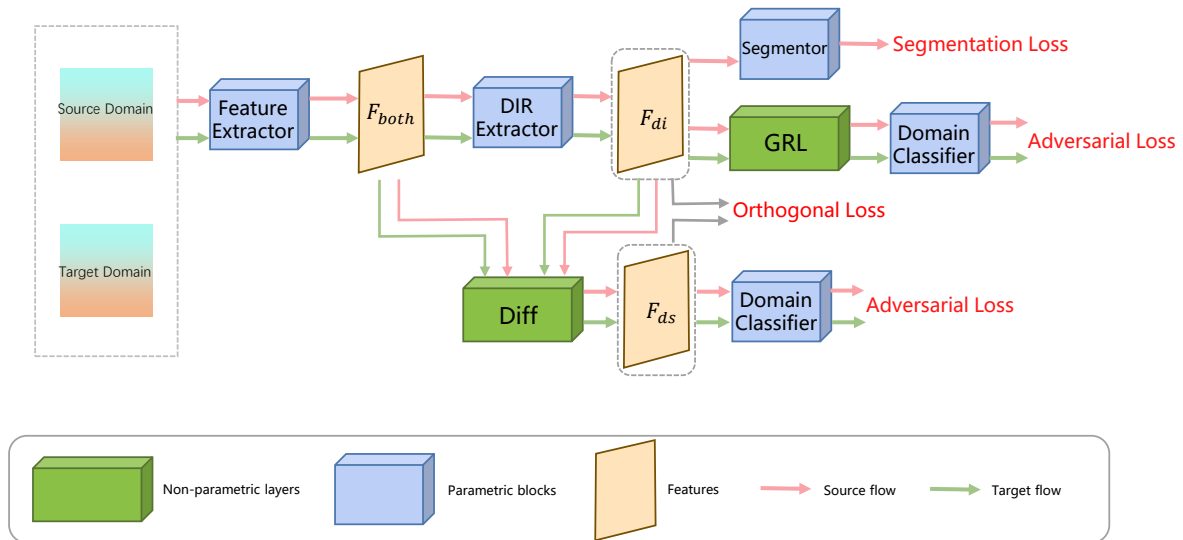


Figure 8. An overview of ODADA architecture [95]

The ODADA model demonstrated superior experimental outcomes across several challenging medical image segmentation tasks. On three public datasets—cross-center prostate MRI segmentation, cross-center COVID-19 CT lesion segmentation, and cross-modality cardiac (MRI/CT) segmentation—ODADA significantly outperformed conventional adversarial learning-based UDA methods (e.g., DANN, ADDA) as well as image translation-based approaches (e.g., CycleGAN). For instance, on the prostate dataset, ODADA achieved a 7.73% improvement in Dice score compared to DANN. Furthermore, experiments validated ODADA's plug-and-play capability; integrating it as an adversarial module into other state-of-the-art UDA frameworks led to further performance enhancements, highlighting the method's effectiveness and versatility.

The SMEDL method [96] proposed by Cai et al. (2025) builds upon adversarial training by further incorporating concepts of style mixup and dual domain invariance learning. The central tenet of SMEDL is to enhance model generalization by implicitly generating mixed domains with diverse styles and to improve robustness through dual domain invariance learning. SMEDL employs a Disentangled Style Mixup (DSM) strategy, utilizing a shared feature extractor alongside independent content and style extractors to decompose images into content and style features. Subsequently, convex combinations of style features from different domains are performed to generate mixed-style features, thereby implicitly creating multiple style-mixed domains.

Moreover, SMEDL introduces a dual domain invariance learning mechanism, comprising: Intra-domain Contrastive Learning (ICL), which performs contrastive learning within the source and target domains separately, encouraging the model to learn invariance between features with the same semantic content but perturbed by different styles; and Inter-domain Adversarial Learning (IAL), which conducts adversarial learning between two style-mixed domains. This involves training a discriminator to distinguish between the mixed domains while simultaneously training the feature extractor to generate domain-invariant features, thus capturing invariance between features possessing the same mixed style but different semantic content. This approach enables SMEDL to leverage both intra-domain and inter-domain variations for learning robust domain-invariant representations.

SMEDL achieves domain adaptation through style mixup and dual domain invariance learning without requiring image translation or diversity regularization, offering a more concise and efficient solution.

SMEDL was comprehensively validated on two public cardiac datasets and one brain dataset. Experimental results indicated that SMEDL achieved significant performance improvements compared to state-of-the-art UDA methods for medical image segmentation, demonstrating its effectiveness in addressing cross-modality medical image segmentation tasks.

### 5.2.3. UDA Improvements Based on Semantic Preservation

Diverging from source-free and adversarial training-based strategies, the Dual Domain Distribution Disruption with Semantics Preservation (DDSP) method [97] proposed by Zheng et al. (2024) adopts a unique perspective for unsupervised domain adaptation. The core strategy of DDSP distinguishes itself from traditional Generative Adversarial Network (GAN) approaches that aim for precise synthesis of target domain images. Instead, DDSP introduces a distribution disruption module to actively and broadly alter the image distributions of both source and target domains, while employing strong constraints based on semantic information. This design compels the model to move beyond reliance on specific domain distribution details and focus on intrinsic, domain-invariant anatomical structural information, thereby achieving domain-agnostic capabilities.

In its implementation, DDSP utilizes a dual domain distribution disruption strategy, simultaneously perturbing source and target domain images via non-learning transformation functions such as Shuffle Remap. A key aspect is the asymmetric application of perturbation magnitude: greater perturbation is applied to the source domain, which possesses label information, leveraging strong semantic supervision (segmentation loss) to force the model to learn distribution robustness. Conversely, lesser perturbation is applied to the unlabeled target domain, coupled with a semantic consistency loss (Lsec) — constraining prediction consistency before and after perturbation — thereby guiding the model to adapt to target domain characteristics while avoiding the introduction of excessive noise. Furthermore, DDSP innovatively introduces a Cross-Channel Similarity Feature Alignment (IFA) mechanism. This mechanism leverages the prior knowledge that cross-domain anatomical structures exhibit consistency in semantic information and relative volume ratios. By aligning the channel-wise similarity matrices of source and target domain feature maps, IFA encourages target domain features across different channels to reflect a structural emphasis consistent with the source domain, consequently significantly improving the accuracy of the shared classifier when processing target domain features.

In bidirectional cross-modality segmentation experiments on the MMWHS17 cardiac dataset, DDSP demonstrated exceptional performance. Compared to the preceding state-of-the-art (SOTA) method, ODADA [95], DDSP achieved an average Dice score improvement ranging from 7.9% to 9.6%, significantly narrowing the gap to the fully supervised baseline. On the PRO12 prostate cross-center segmentation task, DDSP also outperformed existing UDA methods, yielding Dice scores of 76.6% and 83.0%. These quantitative results, spanning multiple tasks and datasets, consistently validate the effectiveness of the DDSP framework. They indicate that DDSP achieves SOTA performance in overcoming GAN limitations and enhancing cross-domain medical image segmentation, while demonstrating the potential to approach fully supervised performance levels.

This section reviews unsupervised domain adaptation (UDA) methods in the field of medical image segmentation, focusing on three primary strategies: source-free approaches, adversarial training-based methods, and techniques based on distribution disruption and semantic preservation. To facilitate a quantitative comparison of the performance across various representative UDA techniques, Table 6 presents the Dice coefficient and Average Symmetric Surface Distance (ASSD) metrics achieved by these methods on the MM-WHS challenge dataset.

**Table 6.** The quantitative comparison results of unsupervised domain adaptation methods for medical image segmentation on the MM-WHS challenge dataset.[96]

DataSets Methods	Cardiac MRI → Cardiac CT		Cardiac CT → Cardiac MRI	
	AA		AA	
	Dice(%)	ASSD(mm)	Dice(%)	ASSD(mm)
Supervised training (Upper bound)	92.0 ± 7.2	1.5 ± 0.8	80.12 ± 4.0	4.2 ± 1.9
Without adaptation (Lower bound)	0.1 ± 0.1	51.0 ± 9.1	18.1 ± 13.7	32.9 ± 4.7
One-shot Finetune	46.2 ± 9.2	10.7 ± 2.1	39.9 ± 11.2	8.2 ± 1.5
Five-shot Finetune	73.1 ± 3.4	8.6 ± 1.7	39.5 ± 10.3	8.5 ± 1.2
PnP-AdaNet [98]	74.0 ± 21.1	24.9 ± 6.7	43.7 ± 6.2	3.1 ± 2.2
AdvEnt [99]	84.2 ± 3.0	9.1 ± 4.1	53.0 ± 5.9	6.9 ± 1.7
SIFA [100]	81.3 ± 5.7	7.9 ± 2.7	65.3 ± 10.9	7.3 ± 5.0
VarDA [101]	81.9 ± 9.1	8.1 ± 5.0	54.6 ± 9.3	15.5 ± 4.5
BMCAN [102]	83.0 ± 6.8	5.8 ± 4.1	72.2 ± 4.3	3.7 ± 2.6
DAAM [75]	87.0 ± 2.1	5.4 ± 3.0	76.0 ± 7.3	6.8 ± 3.2
ADR [95]	87.9 ± 3.6	5.9 ± 4.4	69.7 ± 4.2	5.1 ± 2.1
MPSCL [102]	86.8 ± 2.6	7.7 ± 3.9	64.6 ± 4.7	4.5 ± 2.3
SMEDL [96]	88.3 ± 3.5	4.3 ± 2.3	80.12 ± 4.0	4.2 ± 1.9

6. Discussion

Recent years have witnessed significant advancements in deep learning-based medical image segmentation approaches operating under limited supervision. These methods have effectively mitigated the reliance on the large-scale, pixel-precise annotations typically demanded by fully supervised techniques and have found widespread application across various medical image analysis tasks. Nevertheless, critical challenges remain, including constrained segmentation accuracy, a frequent limitation to single-class object segmentation, and the inadequate modeling of long-range dependencies between pixels. In light of these issues, this section will first provide an overview of representative application scenarios (Section 6.1), followed by an in-depth exploration of future research directions and development trends within this domain (Section 6.2).

6.1. Applications

Deep learning (DL) has emerged as the dominant technology in medical image segmentation, and the rise of paradigms employing limited supervision provides a crucial foundation for the widespread deployment and application of these techniques. This section initially presents a comparative analysis of representative methods within semi-supervised learning (SSL), weakly supervised learning (WSL), and unsupervised learning (UL), focusing on three dimensions: core characteristics, performance advantages, and applicable scenarios (as shown in Tables 7 and 8). Building upon this comparison, the typical application contexts for these methodologies are outlined below.

Table 7. Medical Image Segmentation Methods Summary

Method	Authors (Year)	Key Feature	Application Domain(s)	Strengths
AC-MT [47]	Xu et al. (2023)	Ambiguity recognition module selectively calculates consistency loss	Medical image segmentation	High-ambiguity pixels screening with entropy and selective consistency learning improves segmentation index
AAU-Net [48]	Adiga V. et al. (2024)	Uncertainty estimation of anatomical prior (DAE)	Abdominal CT multi-organ segmentation	Denoising autoencoder optimizes prediction anatomy rationality and improves DSC/HD
CMMT-Net [51]	Li et al. (2024)	Cross-head mutual-aid mean teaching and multi-level perturbations	Medical image segmentation on LA, Pancreas-CT, ACDC	Multi-head decoder enhances prediction diversity and improves Dice
MLRPL [54]	Su et al. (2024)	Collaborative learning framework with dual reliability evaluation	Medical image segmentation (e.g., Pancreas-CT)	Dual decoders with mutual comparison strategy achieves near fully-supervised performance
CRLN [56]	Wang et al. (2025)	Prototype learning and dynamic interaction correction pseudo-labeling	3D medical image segmentation (LA, Pancreas-CT, BraTS19)	Multi-prototype learning captures intra-class diversity to enhance generalization
CRCFP [45]	Bashir et al. (2024)	Exponential Momentum Context-aware contrast and cross-consistency training	Histopathology image segmentation (BCSS, MoNuSeg)	Dual-path unsupervised learning with lightweight classifier achieves near fully-supervised performance
AGM [74]	Yang et al. (2024)	Iterative refinement learning stage	Handling small size, low contrast, and multiple co-existing lesions in medical images	Enhances lesion localization accuracy
SA-MIL [77]	Li et al. (2023)	Criss-Cross Attention (CCA)	Better differentiation between foreground (e.g., cancerous regions) and background	Enhances feature representation capability

Table 8. Medical Image Segmentation Methods Summary (continued)

Method	Authors (Year)	Key Feature	Application Domain(s)	Strengths
SOUSA [58]	Gao et al. (2022)	Multi-angle projection reconstruction loss	More accurate segmentation boundaries, fewer false positive regions	Significantly improves segmentation accuracy
Point SEGTR [81]	Shi et al. (2023)	Fuses limited pixel-level annotations with abundant point-level annotations	Endoscopic image analysis	Significantly reduces dependency on pixel-level annotations
VAE [88]	Silva-Rodríguez et al. (2022)	Attention mechanism (Grad-CAM) + Extended log-barrier method	Unsupervised Anomaly Detection and Segmentation (UAS); Lesion detection & localization	Effectively separates activation distributions of normal and abnormal patterns
OSUDA [94]	Liu et al. (2023)	Exponential Momentum Decay (EMD); Consistency loss on Higher-order BN Statistics (LHBS)	Source-Free Unsupervised Domain Adaptation (SFUDA); Privacy-preserving knowledge transfer	Improves performance and stability in the target domain
ODADA [95]	Sun et al. (2022)	Domain-Invariant Representation (DIR) and Domain-Specific Representation (DSR) decomposition	Scenarios with significant domain shift; Unsupervised Domain Adaptation (UDA)	Learns purer and more effective domain-invariant features
SMEDL [96]	Cai et al. (2025)	Disentangled Style Mixup (DSM) strategy	Cross-modal medical image segmentation tasks	Leverages both intra-domain and inter-domain variations to learn robust representations
DDSP [97]	Zheng et al. (2024)	Dual Domain Distribution Disruption strategy; Inter-channel Feature Alignment (IFA) mechanism	Scenarios with complex domain shift; Unsupervised Domain Adaptation (UDA) tasks	Significantly improves shared classifier accuracy for target domains

**Auxiliary Diagnosis.** In auxiliary diagnosis, approaches under limited supervision significantly mitigate the dependency on large volumes of pixel-level ground truth annotations. WSL enables the training of segmentation models using more readily available weak labels, such as image-level tags, bounding boxes, or point annotations. For instance, employing Class Activation Maps (CAMs) or Multiple Instance Learning (MIL) frameworks allows for the localization and preliminary segmentation of lesion areas based solely on image-level labels indicating the presence or absence of pathology, thereby providing valuable indicative information for radiologists, particularly suitable for large-scale screening or the detection of atypical lesions [103,104]. SSL, conversely, markedly improves segmentation performance by integrating a small quantum of precisely annotated data with substantial



amounts of unlabeled data. Strategies including consistency regularization or pseudo-labeling have demonstrated efficacy in tasks such as lung nodule, skin lesion, and retinal vessel segmentation, achieving competitive accuracy at a considerably lower annotation cost compared to fully supervised methods, thus facilitating earlier and more accurate identification of disease indicators [49].

**Surgical Planning.** Surgical planning necessitates precise segmentation of patient-specific anatomical structures (e.g., tumors, organs, vessels). Methods utilizing limited supervision can significantly accelerate this process and enhance adaptability to diverse data sources. SSL allows the leveraging of existing, albeit limited, high-quality segmentation data (potentially from disparate sources or prior cases) in conjunction with the current patient's unlabeled pre-operative images to rapidly generate personalized 3D anatomical models, which is vital for accommodating individual anatomical variations [105]. WSL, particularly when combined with interactive methodologies (e.g., clinician-provided sparse clicks or bounding boxes), can yield segmentation results adequate for planning requirements within minutes, considerably faster than exhaustive manual delineation, while ensuring the accuracy of critical structures [106]. Furthermore, Unsupervised Domain Adaptation (UDA) technology is crucial for successfully deploying models trained on standard datasets or specific imaging devices to distinct surgical cases (potentially involving different scanning parameters or equipment); it facilitates model adjustment in the absence of target case labels to curtail performance degradation attributable to domain shift, thereby ensuring planning reliability [107].

**Treatment Response Assessment & Longitudinal Monitoring.** Tracking disease progression or evaluating therapeutic efficacy requires consistent and reproducible segmentation across serial imaging time points. Manual processing of extensive longitudinal datasets is practically infeasible. SSL proves particularly effective in this context; by exploiting annotation information from select time points (e.g., baseline) and incorporating temporal consistency constraints (e.g., assuming minimal or smooth structural changes over short intervals), it enables accurate segmentation of images from other unlabeled time points, facilitating reliable quantitative tracking of metrics such as lesion volume and morphology [108]. WSL can also leverage coarse information regarding changes (e.g., clinician assessments of "increase/stable/decrease") or global measurements to guide the segmentation model. Unsupervised change detection methods permit direct comparison of images from different time points, highlighting regions exhibiting significant structural or intensity alterations without necessitating prior definition or segmentation of specific structures, thus aiding in the rapid identification of abnormal changes or the assessment of treatment-induced tissue alterations [109].

**Image Data Standardization & Quality Control.** The inherent heterogeneity of medical image data—stemming from variations in imaging devices, centers, and protocols—constitutes a primary obstacle to the generalization capability and widespread clinical adoption of DL models. UDA represents a key technology to address this challenge. By aligning feature distributions or image styles between the source domain (labeled) and the target domain (unlabeled), UDA can significantly boost model performance on unseen target domain data [110,111]. This is essential for developing tools amenable to reliable utilization in multi-center studies or stable operation across diverse clinical environments, ensuring the consistency and comparability of analysis results. For example, UDA methods predicated on adversarial learning or feature moment matching have been extensively applied in cross-device segmentation tasks involving brain, cardiac, and abdominal organs, effectively enhancing model robustness and the degree of standardization [112].

## 6.2. Future Works

This subsection systematically analyzes and envisages several key future research directions and challenges in the field of medical image segmentation, integrating current technological advancements with clinical requirements.

### 6.2.1. Data-Efficient Segmentation Methods

The acquisition of large-scale, high-quality pixel-level annotations for medical images remains a critical bottleneck constraining the advancement of deep learning models. Consequently, investigating

methods to achieve precise segmentation under conditions of limited or even absent annotations—that is, developing data-efficient learning paradigms—constitutes a significant future research direction. This encompasses, but is not limited to: exploring more effective semi-supervised learning (SSL) strategies to fully leverage abundant unlabeled data; advancing weakly supervised learning (WSL) research to utilize readily obtainable weak information such as image-level labels, bounding boxes, point annotations, or scribbles for pixel-level prediction; developing self-supervised learning approaches to mine supervisory signals from the data itself for pre-training or direct application in segmentation tasks; and investigating few-shot and even zero-shot segmentation techniques to enable models to rapidly adapt to novel segmentation tasks where annotations are scarce. Such research endeavors promise to substantially reduce data annotation costs and accelerate the application of models across diverse clinical scenarios.

#### 6.2.2. Generalization, Robustness, and Federated Learning

Medical imaging data inherently suffer from the *domain shift* problem, where models trained on one dataset or center may experience a sharp decline in performance when applied to another dataset or center due to discrepancies in imaging devices, protocols, or patient populations. Enhancing model generalization capability on unseen data and robustness against various interferences (e.g., noise, artifacts) represents a core challenge for achieving widespread clinical deployment. Future research must prioritize unsupervised domain adaptation (UDA), domain generalization (DG) techniques, and strategies capable of handling multi-center, heterogeneous data. Concurrently, with increasingly stringent data privacy and security regulations, federated learning (FL), as a distributed, privacy-preserving training framework, exhibits immense potential within medical image segmentation. Key research questions will involve how to effectively conduct model training and aggregation under the federated setting, and how to address data heterogeneity (*Non-IID data*).

#### 6.2.3. Interpretability, Uncertainty Quantification, and Clinical Trustworthiness

The "black-box" nature of deep learning models restricts their application in high-reliability medical decision-making contexts. Improving the interpretability or explainability (XAI) of segmentation models, enabling clinicians to comprehend the rationale behind specific segmentation decisions made by the model, is crucial for establishing trust. This necessitates the development of techniques capable of generating visual explanations (e.g., saliency maps, Class Activation Maps (CAMs)) or providing rule-based or concept-based interpretations. Concurrently, quantifying the uncertainty (UQ) associated with model predictions is equally critical; it informs users about regions where the model lacks confidence in its segmentation results, thereby prompting manual review or mitigating potential errors. Research into reliable uncertainty estimation methods and their effective integration into clinical workflows represents a vital direction for enhancing model safety and practical utility.

#### 6.2.4. Multi-Modal and Longitudinal Data Fusion for Segmentation

In clinical practice, physicians often integrate information from multiple imaging modalities (e.g., CT, MRI, PET) and refer to patient historical images (longitudinal data) for diagnosis and assessment. Current segmentation models predominantly focus on single-modality, single-timepoint analysis. Future research should concentrate on developing effective multi-modal fusion strategies to fully exploit the complementary information provided by different modalities, thereby achieving more precise and comprehensive segmentation. Simultaneously, for longitudinal data, it is imperative to devise segmentation models capable of capturing spatio-temporal dynamics, such as modeling lesion evolution or organ changes using Recurrent Neural Networks (RNNs), Transformers, or Graph Neural Networks (GNNs). This holds significant importance for accurate treatment response assessment and disease progression monitoring. Effectively fusing multi-source, heterogeneous spatio-temporal information remains a core challenge in this direction.

## 7. Conclusion

This review investigates deep learning-based techniques for medical image segmentation, emphasizing non-fully supervised learning paradigms designed to overcome the reliance of fully supervised methods on large-scale, pixel-level annotations. By systematically examining the core principles, representative algorithms, and application contexts of semi-supervised learning (utilizing strategies like consistency regularization and pseudo-labeling with limited labeled and abundant unlabeled data), weakly supervised learning (employing coarse-grained annotations such as image-level labels, bounding boxes, or scribbles), and unsupervised learning (including anomaly segmentation and domain adaptation to handle data heterogeneity and privacy concerns), this work underscores the substantial value of these methods in significantly reducing annotation costs and promoting the clinical translation of advanced segmentation technologies. Although significant progress has been achieved, challenges persist for these non-fully supervised approaches, particularly concerning closing the performance gap with fully supervised methods, improving model robustness and generalization, enhancing interpretability, and effectively fusing multi-modal/longitudinal data. Continuous innovation in non-fully supervised learning remains crucial for accelerating the application of AI in medical image analysis and improving diagnosis and treatment.

**Acknowledgments:** We are grateful to the reviewers and colleagues for their insightful comments and suggestions, which greatly improved the manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Author Contributions:** Conceptualization, X.Z. and J.F.W.; Methodology, X.Z. and M.W.; Validation, X.Z., J.Q.W. and X.Y.; Investigation, X.Y.; Writing – Original Draft Preparation, X.Z.; Writing – Review & Editing, X.Z. and J.Q.W.; Visualization, X.Y.; Supervision, J.W. and M.W.; Project Administration, M.W.

## References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **2012**, *60*, 84–90. <https://doi.org/10.1145/3065386>.
2. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation, 2015.
3. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2015**, *9351*, 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
4. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Łukasz Kaiser.; Polosukhin, I. Attention Is All You Need. *Advances in Neural Information Processing Systems* **2017**, 2017-December, 5999–6009.
5. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything, 2023.
6. Baur, C.; Wiestler, B.; Albarqouni, S.; Navab, N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2019**, *11383 LNCS*, 161–169. [https://doi.org/10.1007/978-3-030-11723-8\\_16/TABLES/1](https://doi.org/10.1007/978-3-030-11723-8_16/TABLES/1).
7. Yarkony, J.; Wang, S.; Inc, B. Accelerating Message Passing for MAP with Benders Decomposition **2018**.
8. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization, 2016.
9. Dai, J.; He, K.; Sun, J. BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation, 2015.
10. Ballan, L.; Castaldo, F.; Alahi, A.; Palmieri, F.; Savarese, S. Knowledge transfer for scene-specific motion prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2016**, *9905 LNCS*, 697–713. [https://doi.org/10.1007/978-3-319-46448-0\\_42/FIGURES/9](https://doi.org/10.1007/978-3-319-46448-0_42/FIGURES/9).
11. Tanaka, K. Minimal networks for sensor counting problem using discrete Euler calculus. *Japan Journal of Industrial and Applied Mathematics* **2017**, *34*, 229–242. <https://doi.org/10.1007/S13160-017-0243-2/METRICS>.

12. Wei, Y.; Feng, J.; Liang, X.; Cheng, M.M.; Zhao, Y.; Yan, S. Object Region Mining With Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach, 2017.
13. Hu, F.; Wang, Y.; Ma, B.; Wang, Y. Emergency supplies research on crossing points of transport network based on genetic algorithm. *Proceedings - 2015 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2015* **2016**, pp. 370–375. <https://doi.org/10.1109/ICITBS.2015.98>.
14. Gannon, S.; Kulosman, H. The condition for a cyclic code over  $Z_4$  of odd length to have a complementary dual **2019**.
15. Abraham, N.; Khan, N.M. A novel focal tversky loss function with improved attention u-net for lesion segmentation. *Proceedings - International Symposium on Biomedical Imaging* **2019**, 2019-April, 683–687. <https://doi.org/10.1109/ISBI.2019.8759329>.
16. Bernard, O.; Lalande, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.A.; Cetin, I.; Lekadir, K.; Camara, O.; Ballester, M.A.G.; et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Transactions on Medical Imaging* **2018**, 37, 2514–2525. <https://doi.org/10.1109/TMI.2018.2837502>.
17. Graham, S.; Chen, H.; Gamper, J.; Dou, Q.; Heng, P.A.; Snead, D.; Tsang, Y.W.; Rajpoot, N. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical image analysis* **2019**, 52, 199–211. <https://doi.org/10.1016/J.MEDIA.2018.12.001>.
18. Demner-Fushman, D.; Kohli, M.D.; Rosenman, M.B.; Shooshan, S.E.; Rodriguez, L.; Antani, S.; Thoma, G.R.; McDonald, C.J. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association : JAMIA* **2016**, 23, 304–310. <https://doi.org/10.1093/JAMIA/OCV080>.
19. Johnson, A.E.; Pollard, T.J.; Berkowitz, S.J.; Greenbaum, N.R.; Lungren, M.P.; ying Deng, C.; Mark, R.G.; Horng, S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data* **2019**, 6. <https://doi.org/10.1038/S41597-019-0322-0>.
20. Li, Z.; Li, D.; Xu, C.; Wang, W.; Hong, Q.; Li, Q.; Tian, J. TFCNs: A CNN-Transformer Hybrid Network for Medical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, 13532 LNCS, 781–792. [https://doi.org/10.1007/978-3-031-15937-4\\_65](https://doi.org/10.1007/978-3-031-15937-4_65).
21. Bannur, S.; Hyland, S.; Liu, Q.; Pérez-García, F.; Ilse, M.; Castro, D.C.; Boecking, B.; Sharma, H.; Bouzid, K.; Thieme, A.; et al. Learning to Exploit Temporal Structure for Biomedical Vision-Language Processing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2023**, 2023-June, 15016–15027. <https://doi.org/10.1109/CVPR52729.2023.01442>.
22. McCollough, C.H.; Bartley, A.C.; Carter, R.E.; Chen, B.; Drees, T.A.; Edwards, P.; Holmes, D.R.; Huang, A.E.; Khan, F.; Leng, S.; et al. Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge. *Medical physics* **2017**, 44, e339–e352. <https://doi.org/10.1002/MP.12345>.
23. Leuschner, J.; Schmidt, M.; Baguer, D.O.; Maass, P. LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction. *Scientific Data* **2021**, 8, 1–12. <https://doi.org/10.1038/s41597-021-00893-z>.
24. Moen, T.R.; Chen, B.; Holmes, D.R.; Duan, X.; Yu, Z.; Yu, L.; Leng, S.; Fletcher, J.G.; McCollough, C.H. Low-dose CT image and projection dataset. *Medical physics* **2021**, 48, 902–911. <https://doi.org/10.1002/MP.14594>.
25. Xiong, Z.; Xia, Q.; Hu, Z.; Huang, N.; Bian, C.; Zheng, Y.; Vesal, S.; Ravikumar, N.; Maier, A.; Yang, X.; et al. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical Image Analysis* **2021**, 67, 101832. <https://doi.org/10.1016/J.MEDIA.2020.101832>.
26. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging* **2013**, 26, 1045–1057. <https://doi.org/10.1007/S10278-013-9622-7/METRICS>.
27. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* **2015**, 34, 1993–2024. <https://doi.org/10.1109/TMI.2014.2377694>.
28. Liew, S.Q.; Ngoh, G.C.; Yusoff, R.; Teoh, W.H. Acid and Deep Eutectic Solvent (DES) extraction of pectin from pomelo (*Citrus grandis* (L.) Osbeck) peels. *Biocatalysis and Agricultural Biotechnology* **2018**, 13, 1–11. <https://doi.org/10.1016/J.BCAB.2017.11.001>.



29. Petzsche, M.R.H.; de la Rosa, E.; Hanning, U.; Wiest, R.; Valenzuela, W.; Reyes, M.; Meyer, M.; Liew, S.L.; Kofler, F.; Ezhov, I.; et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific data* **2022**, *9*. <https://doi.org/10.1038/S41597-022-01875-5>.
30. Maier, O.; Menze, B.H.; von der Gablentz, J.; Häni, L.; Heinrich, M.P.; Liebrand, M.; Winzeck, S.; Basit, A.; Bentley, P.; Chen, L.; et al. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis* **2017**, *35*, 250–269. <https://doi.org/10.1016/J.MEDIA.2016.07.009>.
31. Hakim, A.; Christensen, S.; Winzeck, S.; Lansberg, M.G.; Parsons, M.W.; Lucas, C.; Robben, D.; Wiest, R.; Reyes, M.; Zaharchuk, G. Predicting Infarct Core From Computed Tomography Perfusion in Acute Ischemia With Machine Learning: Lessons From the ISLES Challenge. *Stroke* **2021**, *52*, 2328–2337. [https://doi.org/10.1161/STROKEAHA.120.030696/SUPPL\\_FILE/SUPPLEMENTAL](https://doi.org/10.1161/STROKEAHA.120.030696/SUPPL_FILE/SUPPLEMENTAL).
32. Liang, K.; Han, K.; Li, X.; Cheng, X.; Li, Y.; Wang, Y.; Yu, Y. Symmetry-Enhanced Attention Network for Acute Ischemic Infarct Segmentation with Non-contrast CT Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2021**, 12907 LNCS, 432–441. [https://doi.org/10.1007/978-3-030-87234-2\\_41](https://doi.org/10.1007/978-3-030-87234-2_41).
33. Campello, V.M.; Gkontra, P.; Izquierdo, C.; Martin-Isla, C.; Sojoudi, A.; Full, P.M.; Maier-Hein, K.; Zhang, Y.; He, Z.; Ma, J.; et al. Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The MMs Challenge. *IEEE Transactions on Medical Imaging* **2021**, *40*, 3543–3554. <https://doi.org/10.1109/TMI.2021.3090082>.
34. Heller, N.; Isensee, F.; Maier-Hein, K.H.; Hou, X.; Xie, C.; Li, F.; Nan, Y.; Mu, G.; Lin, Z.; Han, M.; et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *ElsevierN Heller, F Isensee, KH Maier-Hein, X Hou, C Xie, F Li, Y Nan, G Mu, Z Lin, M Han, G YaoMedical image analysis, 2021•Elsevier* **2021**, *67*, 101821. <https://doi.org/10.1016/j.media.2020.101821>.
35. Littlejohns, T.J.; Holliday, J.; Gibson, L.M.; Garratt, S.; Oesingmann, N.; Alfaro-Almagro, F.; Bell, J.D.; Boultonwood, C.; Collins, R.; Conroy, M.C.; et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *nature.com/TJ Littlejohns, J Holliday, LM Gibson, S Garratt, N Oesingmann, F Alfaro-Almagro, JD BellNature communications, 2020•nature.com* **2020**, *11*. <https://doi.org/10.1038/s41467-020-15948-9>.
36. Bilic, P.; Christ, P.; Li, H.B.; Vorontsov, E.; Ben-Cohen, A.; Kaissis, G.; Szeskin, A.; Jacobs, C.; Mamani, G.E.H.; Chartrand, G.; et al. The Liver Tumor Segmentation Benchmark (LiTS). *Medical Image Analysis* **2023**, *84*, 102680. <https://doi.org/10.1016/J.MEDIA.2022.102680>.
37. Kavur, A.E.; Gezer, N.S.; Barış, M.; Aslan, S.; Conze, P.H.; Groza, V.; Pham, D.D.; Chatterjee, S.; Ernst, P.; Özkan, S.; et al. CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. *Medical image analysis* **2021**, *69*. <https://doi.org/10.1016/J.MEDIA.2020.101950>.
38. Zhang, P.; Zhang, B.; Chen, D.; Yuan, L.; Wen, F. Cross-domain correspondence learning for exemplar-based image translation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2020**, pp. 5142–5152. <https://doi.org/10.1109/CVPR42600.2020.00519>.
39. Sohn, K.; Berthelot, D.; Li, C.L.; Zhang, Z.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Zhang, H.; Raffel, C. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *Advances in Neural Information Processing Systems* **2020**, 2020-December.
40. Chen, H.; Tao, R.; Fan, Y.; Wang, Y.; Wang, J.; Schiele, B.; Xie, X.; Raj, B.; Savvides, M. SoftMatch: Addressing the Quantity-Quality Trade-off in Semi-supervised Learning. *11th International Conference on Learning Representations, ICLR 2023* **2023**.
41. Wang, X.; Tang, F.; Chen, H.; Cheung, C.Y.; Heng, P.A. Deep semi-supervised multiple instance learning with self-correction for DME classification from OCT images. *Medical Image Analysis* **2023**, *83*, 102673. <https://doi.org/10.1016/J.MEDIA.2022.102673>.
42. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems* **2017**, 2017-December, 1196–1205.
43. Yang, L.; Qi, L.; Feng, L.; Zhang, W.; Shi, Y. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation, 2023.
44. Lyu, F.; Ye, M.; Carlsen, J.F.; Erleben, K.; Darkner, S.; Yuen, P.C. Pseudo-Label Guided Image Synthesis for Semi-Supervised COVID-19 Pneumonia Infection Segmentation. *IEEE Transactions on Medical Imaging* **2023**, *42*, 797–809. <https://doi.org/10.1109/TMI.2022.3217501>.



45. Bashir, R.M.S.; Qaiser, T.; Raza, S.E.; Rajpoot, N.M. Consistency regularisation in varying contexts and feature perturbations for semi-supervised semantic segmentation of histology images. *Medical Image Analysis* **2024**, *91*, 102997. <https://doi.org/10.1016/J.MEDIA.2023.102997>.
46. Yang, Y.; Sun, G.; Zhang, T.; Wang, R.; Su, J. Semi-supervised medical image segmentation via weak-to-strong perturbation consistency and edge-aware contrastive representation. *Medical Image Analysis* **2025**, *101*, 103450. <https://doi.org/10.1016/J.MEDIA.2024.103450>.
47. Xu, X.; Chen, Y.; Wu, J.; Lu, J.; Ye, Y.; Huang, Y.; Dou, X.; Li, K.; Wang, G.; Zhang, S.; et al. A novel one-to-multiple unsupervised domain adaptation framework for abdominal organ segmentation. *Medical Image Analysis* **2023**, *88*, 102873. <https://doi.org/10.1016/J.MEDIA.2023.102873>.
48. V., S.A.; Dolz, J.; Lombaert, H. Anatomically-aware uncertainty for semi-supervised image segmentation. *Medical Image Analysis* **2024**, *91*, 103011. <https://doi.org/10.1016/J.MEDIA.2023.103011>.
49. Yu, L.; Wang, S.; Li, X.; Fu, C.W.; Heng, P.A. Uncertainty-Aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2019**, 11765 LNCS, 605–613. [https://doi.org/10.1007/978-3-030-32245-8\\_67/TABLES/2](https://doi.org/10.1007/978-3-030-32245-8_67/TABLES/2).
50. Luo, X.; Wang, G.; Liao, W.; Chen, J.; Song, T.; Chen, Y.; Zhang, S.; Metaxas, D.N.; Zhang, S. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis* **2022**, *80*, 102517. <https://doi.org/10.1016/J.MEDIA.2022.102517>.
51. Li, W.; Bian, R.; Zhao, W.; Xu, W.; Yang, H. Diversity matters: Cross-head mutual mean-teaching for semi-supervised medical image segmentation. *Medical Image Analysis* **2024**, *97*, 103302. <https://doi.org/10.1016/J.MEDIA.2024.103302>.
52. Wu, Y.; Ge, Z.; Zhang, D.; Xu, M.; Zhang, L.; Xia, Y.; Cai, J. Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis* **2022**, *81*, 102530. <https://doi.org/10.1016/J.MEDIA.2022.102530>.
53. Bai, Y.; Chen, D.; Li, Q.; Shen, W.; Wang, Y. Bidirectional Copy-Paste for Semi-Supervised Medical Image Segmentation, 2023.
54. Su, J.; Luo, Z.; Lian, S.; Lin, D.; Li, S. Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. *Medical Image Analysis* **2024**, *94*, 103111. <https://doi.org/10.1016/J.MEDIA.2024.103111>.
55. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016* **2016**, pp. 565–571. <https://doi.org/10.1109/3DV.2016.79>.
56. Wang, Y.; Song, K.; Liu, Y.; Ma, S.; Yan, Y.; Carneiro, G. Leveraging labelled data knowledge: A cooperative rectification learning network for semi-supervised 3D medical image segmentation. *Medical Image Analysis* **2025**, *101*, 103461. <https://doi.org/10.1016/J.MEDIA.2025.103461>.
57. Chaitanya, K.; Erdil, E.; Karani, N.; Konukoglu, E. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Medical Image Analysis* **2023**, *87*, 102792. <https://doi.org/10.1016/J.MEDIA.2023.102792>.
58. Gao, F.; Hu, M.; Zhong, M.E.; Feng, S.; Tian, X.; Meng, X.; Yi, D.; Li, N.; Jia, T.; M.; Huang, Z.; Lv, M.; Song, T.; et al. Segmentation only uses sparse annotations: Unified weakly and semi-supervised learning in medical images. *Medical Image Analysis* **2022**, *80*, 102515. <https://doi.org/10.1016/J.MEDIA.2022.102515>.
59. Li, S.; Zhang, C.; He, X. Shape-Aware Semi-supervised 3D Semantic Segmentation for Medical Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2020**, 12261 LNCS, 552–561. [https://doi.org/10.1007/978-3-030-59710-8\\_54](https://doi.org/10.1007/978-3-030-59710-8_54).
60. Wang, R.; Chen, S.; Ji, C.; Fan, J.; Li, Y. Boundary-aware context neural network for medical image segmentation. *Medical Image Analysis* **2022**, *78*, 102395. <https://doi.org/10.1016/J.MEDIA.2022.102395>.
61. Luo, X.; Chen, J.; Song, T.; Wang, G. Semi-supervised Medical Image Segmentation through Dual-task Consistency. *Proceedings of the AAAI Conference on Artificial Intelligence* **2021**, *35*, 8801–8809. <https://doi.org/10.1609/AAAI.V35I10.17066>.
62. Jie Shi, G.; Wei Gao, D. Transverse ultimate capacity of U-type stiffened panels for hatch covers used in ship cargo holds. *Ships and Offshore Structures* **2021**, *16*, 608–619. <https://doi.org/10.1080/17445302.2020.1765691>.
63. Peng, J.; Wang, P.; Desrosiers, C.; Pedersoli, M. Self-Paced Contrastive Learning for Semi-supervised Medical Image Segmentation with Meta-labels. *Advances in Neural Information Processing Systems* **2021**, *20*, 16686–16699.

64. Gao, S.; Zhang, Z.; Ma, J.; Li, Z.; Zhang, S. Correlation-Aware Mutual Learning for Semi-supervised Medical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2023**, 14220 LNCS, 98–108. [https://doi.org/10.1007/978-3-031-43907-0\\_10](https://doi.org/10.1007/978-3-031-43907-0_10).
65. Oh, S.J.; Benenson, R.; Khoreva, A.; Akata, Z.; Fritz, M.; Schiele, B. Exploiting saliency for object segmentation from image level labels. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017**, 2017-January, 5038–5047. <https://doi.org/10.1109/CVPR.2017.535>.
66. Durieux, G.; Irles, A.; Miralles, V.; Peñuelas, A.; Perelló, M.; Pöschl, R.; Vos, M. The electro-weak couplings of the top and bottom quarks — Global fit and future prospects. *Journal of High Energy Physics* **2019**, 2019, 1–44. [https://doi.org/10.1007/JHEP12\(2019\)098](https://doi.org/10.1007/JHEP12(2019)098).
67. Kervadec, H.; Dolz, J.D.; Montral, D.; Wang, S.; Granger, E.G.; Montral, G.; Ben, I.; Ayed, A.; Montral, A. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. *Proceedings of Machine Learning Research* **2020**, 121, 365–380.
68. Lin, D.; Dai, J.; Jia, J.; He, K.; Sun, J. ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016**, 2016-December, 3159–3167. <https://doi.org/10.1109/CVPR.2016.344>.
69. Wang, Y.; Zhang, J.; Kan, M.; Shan, S.; Chen, X. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation, 2020.
70. Dietterich, T.G.; Lathrop, R.H.; Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* **1997**, 89, 31–71. [https://doi.org/10.1016/S0004-3702\(96\)00034-3](https://doi.org/10.1016/S0004-3702(96)00034-3).
71. Chikontwe, P.; Sung, H.J.; Jeong, J.; Kim, M.; Go, H.; Nam, S.J.; Park, S.H. Weakly supervised segmentation on neural compressed histopathology with self-equivariant regularization. *Medical Image Analysis* **2022**, 80, 102482. <https://doi.org/10.1016/J.MEDIA.2022.102482>.
72. Patel, G.; Dolz, J. Weakly supervised segmentation with cross-modality equivariant constraints. *Medical Image Analysis* **2022**, 77, 102374. <https://doi.org/10.1016/J.MEDIA.2022.102374>.
73. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018* **2018**, 2018-January, 839–847. <https://doi.org/10.1109/WACV.2018.00097>.
74. Yang, J.; Mehta, N.; Demirci, G.; Hu, X.; Ramakrishnan, M.S.; Naguib, M.; Chen, C.; Tsai, C.L. Anomaly-guided weakly supervised lesion segmentation on retinal OCT images. *Medical Image Analysis* **2024**, 94, 103139. <https://doi.org/10.1016/J.MEDIA.2024.103139>.
75. Chen, Z.; Wang, T.; Wu, X.; Hua, X.S.; Zhang, H.; Sun, Q. Class Re-Activation Maps for Weakly-Supervised Semantic Segmentation, 2022.
76. Zhang, W.; Zhu, L.; Hallinan, J.; Zhang, S.; Makmur, A.; Cai, Q.; Ooi, B.C. BoostMIS: Boosting Medical Image Semi-Supervised Learning With Adaptive Pseudo Labeling and Informative Active Annotation, 2022.
77. Li, K.; Qian, Z.; Han, Y.; Chang, E.I.; Wei, B.; Lai, M.; Liao, J.; Fan, Y.; Xu, Y. Weakly supervised histopathology image segmentation with self-attention. *Medical Image Analysis* **2023**, 86, 102791. <https://doi.org/10.1016/J.MEDIA.2023.102791>.
78. Yao, T.; Pan, Y.; Li, Y.; Ngo, C.W.; Mei, T. Wave-ViT: Unifying Wavelet and Transformers for Visual Representation Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, 13685 LNCS, 328–345. [https://doi.org/10.1007/978-3-031-19806-9\\_19](https://doi.org/10.1007/978-3-031-19806-9_19).
79. Cheng, B.; Parkhi, O.; Kirillov, A. Pointly-Supervised Instance Segmentation, 2022.
80. Seeböck, P.; Orlando, J.I.; Michl, M.; Mai, J.; Schmidt-Erfurth, U.; Bogunović, H. Anomaly guided segmentation: Introducing semantic context for lesion segmentation in retinal OCT using weak context supervision from anomaly detection. *Medical Image Analysis* **2024**, 93, 103104. <https://doi.org/10.1016/J.MEDIA.2024.103104>.
81. Shi, Y.; Wang, H.; Ji, H.; Liu, H.; Li, Y.; He, N.; Wei, D.; Huang, Y.; Dai, Q.; Wu, J.; et al. A deep weakly semi-supervised framework for endoscopic lesion segmentation. *Medical Image Analysis* **2023**, 90, 102973. <https://doi.org/10.1016/J.MEDIA.2023.102973>.
82. Ahn, J.; Shin, S.Y.; Shim, J.; Kim, Y.H.; Han, S.J.; Choi, E.K.; Oh, S.; Shin, J.Y.; Choe, J.C.; Park, J.S.; et al. Association between epicardial adipose tissue and embolic stroke after catheter ablation of atrial fibrillation. *Journal of Cardiovascular Electrophysiology* **2019**, 30, 2209–2216. <https://doi.org/10.1111/JCE.14154>.

83. Wang, Y.; Zhang, J.; Kan, M.; Shan, S.; Chen, X. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2020**, pp. 12272–12281. <https://doi.org/10.1109/CVPR42600.2020.01229>.
84. Viniavskyi, O.; Dobko, M.; Dobosevych, O. Weakly-Supervised Segmentation for Disease Localization in Chest X-Ray Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2020**, 12299 LNAI, 249–259. [https://doi.org/10.1007/978-3-030-59137-3\\_23](https://doi.org/10.1007/978-3-030-59137-3_23).
85. Ma, X.; Ji, Z.; Niu, S.; Leng, T.; Rubin, D.L.; Chen, Q. MS-CAM: Multi-Scale Class Activation Maps for Weakly-Supervised Segmentation of Geographic Atrophy Lesions in SD-OCT Images. *IEEE journal of biomedical and health informatics* **2020**, 24, 3443–3455. <https://doi.org/10.1109/JBHI.2020.2999588>.
86. Zhang, S.; Zhang, J.; Xia, Y. TransWS: Transformer-Based Weakly Supervised Histology Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, 13583 LNCS, 367–376. [https://doi.org/10.1007/978-3-031-21014-3\\_38](https://doi.org/10.1007/978-3-031-21014-3_38).
87. Wang, T.; Niu, S.; Dong, J.; Chen, Y. Weakly Supervised Retinal Detachment Segmentation Using Deep Feature Propagation Learning in SD-OCT Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2020**, 12069 LNCS, 146–154. [https://doi.org/10.1007/978-3-030-63419-3\\_15](https://doi.org/10.1007/978-3-030-63419-3_15).
88. Silva-Rodríguez, J.; Naranjo, V.; Dolz, J. Constrained unsupervised anomaly segmentation. *Medical Image Analysis* **2022**, 80, 102526. <https://doi.org/10.1016/J.MEDIA.2022.102526>.
89. Pinaya, W.H.; Tudosiu, P.D.; Gray, R.; Rees, G.; Nachev, P.; Ourselin, S.; Cardoso, M.J. Unsupervised brain imaging 3D anomaly detection and segmentation with transformers. *Medical Image Analysis* **2022**, 79, 102475. <https://doi.org/10.1016/J.MEDIA.2022.102475>.
90. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, 313, 504–507. [https://doi.org/10.1126/SCIENCE.1127647/SUPPL\\_FILE/HINTON.SOM.PDF](https://doi.org/10.1126/SCIENCE.1127647/SUPPL_FILE/HINTON.SOM.PDF).
91. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* **2013**. <https://doi.org/10.61603/ceas.v2i1.33>.
92. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Langs, G.; Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis* **2019**, 54, 30–44. <https://doi.org/10.1016/J.MEDIA.2019.01.010>.
93. Stan, S.; Rostami, M. Unsupervised model adaptation for source-free segmentation of medical images. *Medical Image Analysis* **2024**, 95, 103179. <https://doi.org/10.1016/J.MEDIA.2024.103179>.
94. Liu, X.; Xing, F.; Fakhri, G.E.; Woo, J. Memory consistent unsupervised off-the-shelf model adaptation for source-relaxed medical image segmentation. *Medical Image Analysis* **2023**, 83, 102641. <https://doi.org/10.1016/J.MEDIA.2022.102641>.
95. Sun, Y.; Dai, D.; Xu, S. Rethinking adversarial domain adaptation: Orthogonal decomposition for unsupervised domain adaptation in medical image segmentation. *Medical Image Analysis* **2022**, 82, 102623. <https://doi.org/10.1016/J.MEDIA.2022.102623>.
96. Cai, Z.; Xin, J.; You, C.; Shi, P.; Dong, S.; Dvornek, N.C.; Zheng, N.; Duncan, J.S. Style mixup enhanced disentanglement learning for unsupervised domain adaptation in medical image segmentation. *Medical Image Analysis* **2025**, 101, 103440. <https://doi.org/10.1016/J.MEDIA.2024.103440>.
97. Zheng, B.; Zhang, R.; Diao, S.; Zhu, J.; Yuan, Y.; Cai, J.; Shao, L.; Li, S.; Qin, W. Dual domain distribution disruption with semantics preservation: Unsupervised domain adaptation for medical image segmentation. *Medical Image Analysis* **2024**, 97, 103275. <https://doi.org/10.1016/J.MEDIA.2024.103275>.
98. Dou, Q.; Ouyang, C.; Chen, C.; Chen, H.; Glocker, B.; Zhuang, X.; Heng, P.A. PnP-AdaNet: Plug-and-Play Adversarial Domain Adaptation Network with a Benchmark at Cross-modality Cardiac Segmentation **2018**.
99. Vu, T.H.; Jain, H.; Bucher, M.; Cord, M.; Perez, P. ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation, 2019.
100. Chen, C.; Dou, Q.; Chen, H.; Qin, J.; Heng, P.A. Unsupervised Bidirectional Cross-Modality Adaptation via Deeply Synergistic Image and Feature Alignment for Medical Image Segmentation. *IEEE Transactions on Medical Imaging* **2020**, 39, 2494–2505. <https://doi.org/10.1109/TMI.2020.2972701>.
101. Wu, F.; Zhuang, X. CF Distance: A New Domain Discrepancy Metric and Application to Explicit Domain Adaptation for Cross-Modality Cardiac Image Segmentation. *IEEE Transactions on Medical Imaging* **2020**, 39, 4274–4285. <https://doi.org/10.1109/TMI.2020.3016144>.

102. Liu, Z.; Zhu, Z.; Zheng, S.; Liu, Y.; Zhou, J.; Zhao, Y. Margin Preserving Self-Paced Contrastive Learning Towards Domain Adaptation for Medical Image Segmentation. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26*, 638–647. <https://doi.org/10.1109/JBHI.2022.3140853>.
103. Chen, J.; Huang, W.; Zhang, J.; Debattista, K.; Han, J. Addressing inconsistent labeling with cross image matching for scribble-based medical image segmentation. *IEEE Transactions on Image Processing* **2025**. <https://doi.org/10.1109/TIP.2025.3530787>.
104. Gao, W.; Wan, F.; Pan, X.; Peng, Z.; Tian, Q.; Han, Z.; Zhou, B.; Ye, Q. TS-CAM: Token Semantic Coupled Attention Map for Weakly Supervised Object Localization, 2021.
105. Mahapatra, D. Generative Adversarial Networks And Domain Adaptation For Training Data Independent Image Registration **2019**.
106. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; et al. Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging* **2018**, *37*, 1562–1573. <https://doi.org/10.1109/TMI.2018.2791721>.
107. Chen, C.; Dou, Q.; Chen, H.; Qin, J.; Heng, P.A. Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**, *33*, 865–872. <https://doi.org/10.1609/AAAI.V33I01.3301865>.
108. Lei, T.; Zhang, D.; Du, X.; Wang, X.; Wan, Y.; Nandi, A.K. Semi-Supervised Medical Image Segmentation Using Adversarial Consistency Learning and Dynamic Convolution Network. *IEEE Transactions on Medical Imaging* **2023**, *42*, 1265–1277. <https://doi.org/10.1109/TMI.2022.3225687>.
109. Kalinicheva, E.; Ienco, D.; Sublime, J.; Trocan, M. Unsupervised Change Detection Analysis in Satellite Image Time Series Using Deep Learning Combined with Graph-Based Approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2020**, *13*, 1450–1466. <https://doi.org/10.1109/JSTARS.2020.2982631>.
110. Kamnitsas, K.; Baumgartner, C.; Ledig, C.; Newcombe, V.; Simpson, J.; Kane, A.; Menon, D.; Nori, A.; Criminisi, A.; Rueckert, D.; et al. Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2017**, *10265 LNCS*, 597–609. [https://doi.org/10.1007/978-3-319-59050-9\\_47](https://doi.org/10.1007/978-3-319-59050-9_47).
111. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.Y.; Isola, P.; Saenko, K.; Efros, A.; Darrell, T. CyCADA: Cycle-Consistent Adversarial Domain Adaptation, 2018.
112. Zhang, Y.; Miao, S.; Mansi, T.; Liao, R. Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *11071 LNCS*, 599–607. [https://doi.org/10.1007/978-3-030-00934-2\\_67/TABLES/1](https://doi.org/10.1007/978-3-030-00934-2_67/TABLES/1).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.