This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

SemiGDA: Generative Dual-distribution Alignment for Semi-Supervised Medical Image Segmentation

Kaiwen Huang[1] Yi Zhou[2] Yizhe Zhang[1] Jingxiong Li[1] Tao Zhou[1*]
[1]Nanjing University of Science and Technology, Nanjing, China
[2]Southeast University, Nanjing, China
kwhuang@njust.edu.cn, taozhou.ai@gmail.com

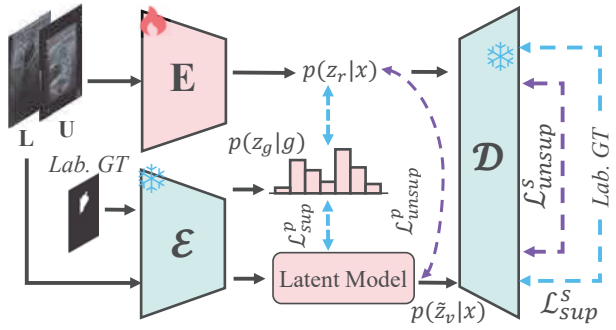
Abstract

Semi-supervised learning addresses label scarcity and high annotation costs in medical image segmentation by exploiting the latent information in unlabeled data to enhance model performance. Traditional discriminative segmentation relies on segmentation masks, neglecting feature-level distribution constraints. This limits robust semantic representation learning and adaptive modeling of unlabeled data in scenarios with few labels. To address these limitations, we propose SemiGDA, a novel Generative Dual-distribution Alignment framework for semi-supervised medical image segmentation. Our SemiGDA overcomes the reliance of discriminative methods on large labeled datasets by aligning feature and semantic distributions to boost semantic learning and scene adaptability. Specifically, we propose a Dual-distribution Alignment Module (DAM), which employs two structurally distinct encoders to model image and mask feature distributions. It enforces their alignment in the latent space via distributional constraints, establishing structured feature consistency. Moreover, we design a Consistency-Driven Skip Adapter (CDSA) strategy, which introduces dual skip adapters (Image and Mask) to fuse multi-scale features via skip connections. Using a consistency loss, CDSA enhances cross-branch semantic alignment and reinforces fine-grained semantic consistency. Experimental results on diverse medical datasets show that our method outperforms other state-of-the-art semi-supervised segmentation methods. ***Code is released at: https://github.com/taozh2017/SemiGDA***.

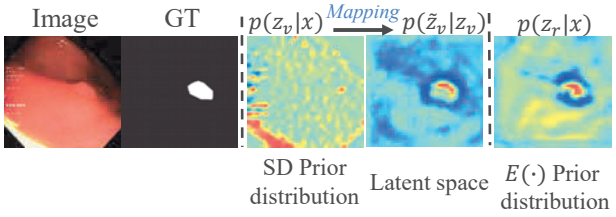
1. Introduction

Medical image segmentation plays a crucial role in disease diagnosis, treatment planning, and surgical guidance [23, 31]. However, its fully-supervised paradigms [5, 28, 35] require large, expertly-annotated datasets, which are costly and difficult to acquire. Semi-supervised Medical Image Segmentation (SMIS) [8, 17, 45] has thus emerged as a promising alternative, leveraging both limited labeled data and abundant unlabeled data to reduce annotation dependency while achieving robust performance.

Existing SMIS methods are primarily based on pseudo-labeling and consistency learning. Pseudo-labeling methods [43, 46] generate labels for unlabeled data and iteratively refine the model, but their performance can be hampered by noisy initial predictions, which affects training stability. In contrast, consistency learning methods [18, 42]



(a) Overview of the proposed framework



(b) Visualization of feature evolution

Figure 1. Illustration of the proposed SemiGDA. (a) Overview of the prior distribution-based semi-supervised architecture. By utilizing feature-level mask constraints and generative segmentation, the model offers high robustness to limited hard labels. (b) Feature evolution: The Latent Mapping Model aligns the Stable Diffusion (SD) prior to a valid manifold, preventing feature collapse and ensuring stable training.

*Corresponding author.

improve model generalization by enforcing output invariance under different perturbations applied to inputs or features. The representative Mean Teacher framework [37] enforces consistency between teacher and student models. Moreover, some methods [14, 39, 41] utilize dual-stream networks to capture different data features, employing mutual learning by enforcing cognitive consistency across the same input. However, these approaches are primarily based on the discriminative paradigm, focusing on per-pixel classification, which restricts their robustness, particularly in semi-supervised tasks with limited labeled data. Furthermore, when addressing complex segmentation tasks, they often fail to effectively capture image structures and integrate contextual semantic information.

To address these limitations, generative models offer a promising alternative. Representative generative models, including Generative Adversarial Networks (GANs) [15] and Variational Autoencoders (VAEs) [24], have been widely adopted for data augmentation. In semi-supervised tasks, these generative methods are widely applied to use adversarial mechanisms to make the prediction distribution of unlabeled images as close as possible to that of labeled images [6, 33]. These methods typically involve a discriminator to differentiate between the predictions of labeled and unlabeled data, thus optimizing the generative model's performance [16]. However, such adversarial training often faces convergence challenges [23].

To this end, we propose a novel Generative Dual-distribution Alignment framework for semi-supervised medical image segmentation (SemiGDA). Its core innovation is to map images and masks into a latent space and align their prior distributions, leveraging these aligned variables to synthesize high-quality segmentation masks with limited labeled data. To enable generative segmentation in semi-supervised scenarios, we propose a Dual-distribution Alignment Module (DAM), which involves a dual-encoder structure to enforce consistency on the same input, as illustrated in Fig. 1(a). Image features are then translated into mask features via a learnable encoder and a latent space mapping module, effectively aligning the prior distribution to a valid manifold, as shown in Fig. 1(b). Furthermore, to overcome the generative decoder's limitations in capturing global context and semantic detail, we present a Consistency-Driven Skip Adapter (CDSA) strategy integrating multi-scale adapters during decoding to enhance robustness and semantic alignment. Finally, an Annotation Conversion and Reversion (ACR) strategy is designed to resolve dimensionality mismatches with ground truth, ensuring compatibility while preserving semantic integrity.

To summarize, our main contributions are four-fold:

- To the best of our knowledge, the proposed SemiGDA is the first work to directly address the overfitting and poor generalization of discriminative models under limited annotations through a generative paradigm with dual-distribution alignment.
- We present the dual-distribution alignment module to model the distribution transformation between images and masks, facilitating effective alignment of image and mask feature distributions in a latent space.
- A consistency-driven skip adapter is proposed to enhance the decoder's capacity for capturing semantic details and improving robustness, enabling the generative model to be better adapted to the segmentation task.
- Experiments results on four distinct medical benchmarks demonstrate that the proposed method outperforms existing state-of-the-art SMIS approaches.

2. Related Works

2.1. Semi-supervised Medical Image Segmentation

Semi-supervised medical image segmentation leverages a small amount of labeled data and a large volume of unlabeled data to train models, enhancing segmentation performance by utilizing the potential of unlabeled data [23]. Traditional approaches to semi-supervised medical image segmentation predominantly rely on pseudo-labeling [22, 26, 49] and consistency learning [3, 36, 39, 42]. Pseudo-labeling uses an initially trained model to predict labels for unlabeled data, generating pseudo-labels that are iteratively refined to improve model accuracy. zeng *et al.* [46] proposed a Semi-Supervised Tri-Branch Network (SS-TBN) that combines segmentation and classification tasks with pseudo labeling and inter-slice consistency to improve COVID-19 diagnosis with limited annotated data. Consistency learning, widely used in semi-supervised learning, is exemplified by the Mean Teacher (MT) [37] model, where the teacher model's parameters are updated via Exponential Moving Average (EMA), and the student model is optimized by minimizing the output discrepancy between the teacher and student networks. However, the student model is often influenced by the cumulative errors of the teacher model, leading to the development of various derivatives of the teacher-student framework [19, 45, 51].

2.2. Generative Models for Image Processing

In medical image processing, generative models have demonstrated great potential in generating synthetic images, improving resolution, and reconstructing damaged areas [10]. In scenarios with limited medical annotations, many methods [44, 48] expand the dataset for downstream tasks by generating new data using generative models. Specifically, VAE [24] can achieve efficient image generation and reconstruction through variational inference, but faces limitations in detail and diversity. Generative Adversarial Networks (GANs) [15] learn the data's latent distribution and generate realistic high-dimensional sam-

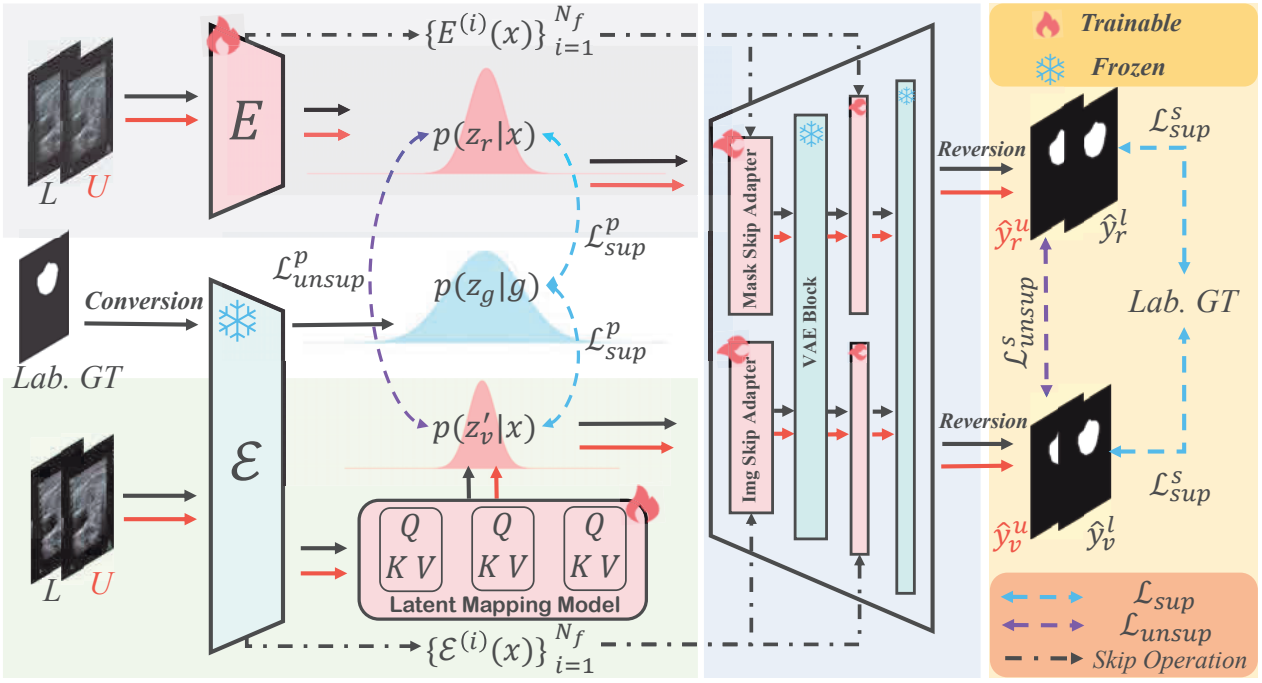


Figure 2. Overview of the proposed framework. The trainable components include a trainable encoder (*i.e.* ResNet as backbone), a mapping network, a lightweight skip adapter module, and a VAE network with frozen parameters. In the first stage, the encoder and mapping network align the input image with the label's prior distribution. In the second stage, the transformed features are passed to the VAE decoder. The skip adapter module is added to improve model performance within the segmentation framework. Additionally, label transformations and inverse transformations are applied to ensure the ground truth conforms to VAE input requirements.

ples through adversarial training, but suffer from issues like mode collapse and vanishing gradients during training. Meanwhile, the generative approach based on diffusion models trained on large-scale datasets, such as Stable Diffusion [34], alleviates the issue of data scarcity [48] and has yielded promising results. However, the iterative training process is complex and time-consuming. Currently, the direct application of generative models to segmentation research is quite limited. GMS [20] utilizes a pre-trained vision model to extract latent representations and learn a mapping function to generate segmentation masks, improving generalization and reducing overfitting. GSS [7] frames semantic segmentation as a mask generation problem by modeling the posterior distribution of segmentation masks and using a conditioning network to optimize the divergence between mask and image priors. GMMSeg [27] combines generative and discriminative models by using Gaussian Mixture Models (GMMs) for class-conditional densities and end-to-end discriminative training.

3. Proposed Method

Setting and Overview. In the context of semi-supervised learning, the training dataset consists of labeled data $\mathcal{D}^L = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$ with N_l samples and unlabeled data $\mathcal{D}^U = \{x_i^u\}_{i=N_l+1}^{N_l+N_u}$ with N_u samples, where $N_l \ll N_u$. Here, $x_i \in \mathbb{R}^{H\times W\times 3}$ denotes an input image, and $y_i \in \{0,1\}^{K\times H\times W}$ represents the corresponding ground-truth annotation with K classes.

Fig. 2 illustrates the overall pipeline of the proposed method. First, both labeled and unlabeled data are input into a frozen VAE encoder. The outputs of the VAE encoder are then passed into the latent mapping model to align their distribution with the label distribution. Meanwhile, another encoding branch inputs the image into a learnable encoder, where consistency learning is applied to the unlabeled data using two distinct mapping strategies. The obtained features are then passed to the decoder, which includes two adapters (*i.e.*, Mask Adapter and Image Adapter). Finally, the features are processed by the reverse transformation to generate the predicted segmentation maps.

3.1. Dual-distribution Alignment Module

Traditional discriminative segmentation methods rely on large labeled datasets and optimize models through maximum likelihood estimation. However, these methods are generally ineffective at capturing global image structures and contextual semantics, particularly when applied to complex or annotation-limited medical scenes. In contrast, gen-

erative segmentation enhances performance by aligning latent spaces, where the model learns the posterior distribution of latent variables, thereby capturing underlying structural information and improving the understanding of intricate image details and fine-grained semantics.

Unlike prior decoder-multiplexing paradigms, we utilize two distinct encoders to model the feature distribution transformation from images to masks. Specifically, the input image $x \in \mathbb{R}^{H\times W\times 3}$ is first processed through a pre-trained VAE [34] encoder $\mathcal{E}(\cdot)$, to capture the prior distribution $p(z_v|x)$ of the image, which can be expressed as follows:

$$p(z_v|x) = \mathcal{E}(x) := \mathcal{N}(z_v; \mu_{z_v}(x;\theta_v), \sigma_{z_v}(x;\theta_v)), \quad (1)$$

where $\mu_{z_v}(x;\theta_v)$ and $\sigma_{z_v}(x;\theta_v)$ are the mean and standard deviation of the Gaussian distribution, respectively, parameterized by the VAE encoder with parameters θ_v. We then utilize a latent mapping model $\mathcal{M}(\cdot)$ that leverages the self-attention mechanism to map image features to a low-dimensional latent space, effectively capturing the global dependencies between the images and masks, expressed by

$$p(\tilde{z}_v|z_v) = \mathcal{M}(z_v) := \mathcal{N}(\tilde{z}_v; \mu_{\tilde{z}_v}(z_v;\theta_m), \sigma_{\tilde{z}_v}(z_v;\theta_m)), \quad (2)$$

where $\mu_{\tilde{z}_v}$ and $\sigma_{\tilde{z}_v}$ are the mean and standard deviation of the Gaussian distribution, parameterized by the learnable parameters θ_m of the latent mapping model and dependent on z_v. To leverage unlabeled data and explore inherent data patterns through consistency learning, we employ two encoders with significant structural differences to provide diverse feature perspectives. Therefore, the same input image $x \in \mathbb{R}^{H\times W\times 3}$ is also fed into a trainable encoder $E(\cdot)$ (as shown in Fig. 2), which focuses on extracting fine-grained structural and discriminative features. The prior distribution $p(z_r|x)$ is expressed as follows:

$$p(z_r|x) = E(x) := \mathcal{N}(z_r; \mu_{z_r}(x;\theta_r), \sigma_{z_r}(x;\theta_r)), \quad (3)$$

where $\mu_{z_r}(x;\theta_r)$ and $\sigma_{z_r}(x;\theta_r)$ are the mean and standard deviation of the Gaussian distribution, respectively, parameterized by the learnable parameters θ_r of the trainable encoder $E(\cdot)$. To constrain the prior distribution of segmentation masks with image feature distributions, we employ the VAE encoder $\mathcal{E}$ for mask processing, and for a ground-truth mask $g \in \mathbb{R}^{H\times W}$, only available for labeled data, its prior distribution $p(z_g|g)$ is modeled as:

$$p(z_g|g) = \mathcal{E}(g) := \mathcal{N}(z_g; \mu_{z_g}(g;\theta_v), \sigma_{z_g}(g;\theta_v)), \quad (4)$$

where μ_{z_g} and σ_{z_g} are the mean and standard deviation, parameterized by the VAE encoder's parameters θ_v.

To ensure consistency between the two-branch feature distributions and the mask prior in the latent space, as well as enhance branch alignment, we design a prior distribution alignment loss. Feature-level supervision, which is more informative than mask-level supervision, employs MSE to quantify the discrepancies among the distributions $p(\tilde{z}_v|z_v)$, $p(z_r|x)$, and the mask prior distribution $p(z_g|g)$. For labeled data, the loss is formulated by

$$\mathcal{L}^p_{sup} = \|\tilde{z}^l_v - z_g\|^2_2 + \|z^l_r - z_g\|^2_2, \quad (5)$$

where $\tilde{z}^l_v$ and z^l_r denote the two image branches' feature distributions under labeled data, respectively. Moreover, for unlabeled data, consistency learning is employed to constrain the distribution loss, which is expressed as follows:

$$\mathcal{L}^p_{unsup} = \|\tilde{z}^u_v - z^u_r\|^2_2, \quad (6)$$

where $\tilde{z}^u_v$ and $\tilde{z}^u_v$ denote the feature distributions of the two image branches for unlabeled data, respectively.

3.2. Consistency-Driven Skip Adapter

In the segmentation task, while prior distribution alignment achieves image-mask feature-level consistency, gaps persist in multi-scale feature utilization and task-specific structural adaptation. Conventional VAE-based pipelines, despite effective distribution matching, often neglect fine-grained multi-scale details critical for segmentation. Moreover, image-centric and mask-centric feature distributions inherently differ in semantic granularity and domain specificity. To this end, we propose the Consistency-Driven Skip Adapter, which fuses multi-scale features via adapter-based skip connections to enhance cross-branch consistency.

Specifically, the module consists of two parallel components: an Image Skip Adapter and a Mask Skip Adapter, tailored for image distribution and mask distributions, respectively. As shown in Fig. 3, each adapter takes as input the hierarchical multi-scale features from its corresponding encoder at the skip connection points. We define the resulting two feature banks as follows:

$$S_v = \left\{\mathcal{E}^{(i)}(x)\right\}_{i=1}^{N_f}, \quad S_r = \left\{E^{(i)}(x)\right\}_{i=1}^{N_f}, \quad (7)$$

where N_f denotes the number of multi-scale levels, S_v represents the hierarchical multi-scale feature bank from the image-distribution encoder, and S_r corresponds to that from the mask-distribution encoder. Note that the adapters are named based on the intrinsic differences between their feature distributions. Specifically, the Image Skip Adapter processes outputs from $\mathcal{E}^{(i)}(x)$, which retains image-distribution characteristics as VAE-compressed features. Conversely, the Mask Skip Adapter processes the outputs from $E^{(i)}(x)$, which inherits mask-distribution attributes through direct constraint by the label feature distribution.

To enhance the intrinsic consistency of cross-branch features, we employ a consistency loss to enforce distribution alignment between the two adapters' outputs, ensuring their feature representations converge to a unified semantic space. For this constraint, we adopt the Dice loss, formally defined as follows:

$$\mathcal{L}_{dice}(\hat{y}, y) = 1 - \frac{2 \cdot \sum(\hat{y} \odot y)}{|\hat{y}|_1 + |y|_1}, \tag{8}$$

where $\hat{y}$ and y respectively denote the prediction and ground-truth. For labeled data, each adapter's output is supervised by the ground-truth, which can be expressed by

$$\mathcal{L}^s_{sup} = \mathcal{L}_{dice}(\hat{y}^l_v, y) + \mathcal{L}_{dice}(\hat{y}^l_r, y), \tag{9}$$

where $\hat{y}^l_v$ and $\hat{y}^l_r$ represent the final predictions from the Image Skip Adapter and Mask Skip Adapter, respectively. For unlabeled data, the consistency constraint is imposed directly between the outputs of the two adapters, and the cross-branch consistency loss is formulated as follows:

$$\mathcal{L}^s_{unsup} = \mathcal{L}_{dice}(\hat{y}^u_v, \hat{y}^u_r) + \mathcal{L}_{dice}(\hat{y}^u_r, \hat{y}^u_v). \tag{10}$$

where $\hat{y}^u_v$ and $\hat{y}^u_r$ denote the predictions for the unlabeled data, respectively. This paradigm, tailored to segmentation's demand for multi-scale and spatially coherent features, preserves fine-grained details and coarse-grained semantics at low computational cost and reinforces dual-branch consistency on unlabeled samples.

3.3. Annotation Conversion and Reversion

We present an Annotation Conversion and Reversion (ACR) strategy to preprocess ground-truth (GT) annotations for compatibility with the VAE input in the segmentation paradigm. Specifically, for the input GT, we first normalize the original pixel values to the range [0, 1] by dividing them by the total number of classes K, then map the result to [-1, 1] to meet the annotation input requirements of the VAE. The complete processing can be expressed as:

$$G' = 2 \cdot \frac{G}{K} - 1, \tag{11}$$

where G denotes the raw segmentation mask (with pixel values ranging from 0 to $K-1$ for K-class tasks), and G' represents the processed GT fed into the VAE. This two-step transformation (normalization and range mapping) ensures compatibility with the input distribution characteristics of the VAE. For the reversion step, the adopted end-to-end inverse calculation directly mirrors the forward conversion process, enabling full reversibility between the two operations.It not only guarantees strict consistency between the mask processing and reversion workflows but also preserves the semantic integrity of the segmentation mask throughout the pipeline, supporting reliable supervised training.

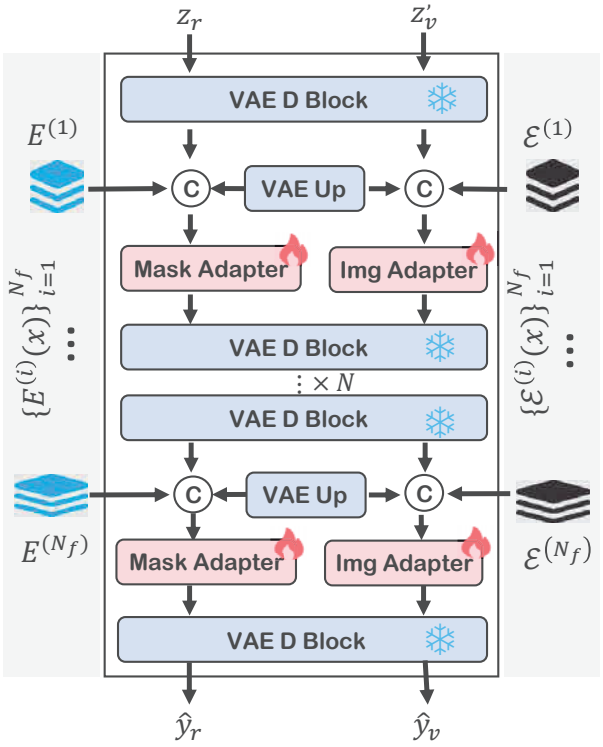


Figure 3. Structure of the skip connection adapter. "VAE D Block" and "VAE Up" represent the frozen VAE decoding block and upsampling block (denoted by blue blocks), where "Img Adapter" and "Mask Adapter" employ lightweight convolutional layers to efficiently adapt multi-scale skip connections.

3.4. Overall Loss Function

The overall loss function of the framework is composed of supervised and unsupervised components Specifically, the supervised loss uses labels as supervision signals, including latent space prior distribution loss (described in Sec. 3.1) to constrain the distribution of latent variables and segmentation loss (described in Sec. 3.2) to measure the discrepancy with the ground truth. The overall expression is:

$$\mathcal{L}_{sup} = \mathcal{L}^p_{sup} + \mathcal{L}^s_{sup}. \tag{12}$$

For unlabeled data, model optimization relies on unsupervised consistency constraints. This is accomplished by enforcing output consistency and distributional consistency loss, which can be formulated as:

$$\mathcal{L}_{unsup} = \mathcal{L}^p_{unsup} + \mathcal{L}^s_{unsup}. \tag{13}$$

Finally, the total loss function of the framework is the sum of the supervised and unsupervised components, which is formulated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_u \mathcal{L}_{unsup}, \tag{14}$$

Table 1. Quantitative results on the colonoscopy and ISIC-2018 datasets.

Methods		10% labeled			30% labeled				10% labeled			30% labeled		
		Dice (%)	IoU (%)	95HD	Dice (%)	IoU (%)	95HD		Dice (%)	IoU (%)	95HD	Dice (%)	IoU (%)	95HD
UA-MT [45] [MICCAI'19]	CVC-300	40.15	32.46	4.97	80.95	70.77	3.19	CVC-ClinicDB	74.51	67.14	4.05	78.89	72.61	3.68
DTC [29] [AAAI'21]		42.81	33.35	5.42	77.44	67.43	3.32		67.82	57.96	4.61	74.28	66.21	3.97
MC-Net [40] [MICCAI'21]		69.89	61.58	3.93	81.36	72.11	3.33		75.21	67.87	4.10	83.53	76.78	3.56
URPC [30] [MIA'22]		58.83	50.81	4.30	77.05	66.75	3.65		76.42	69.33	3.78	80.70	74.58	3.64
MCF [39] [CVPR'23]		67.58	57.85	3.91	79.51	69.03	3.27		70.83	62.41	4.06	80.41	74.50	3.63
CauSSL [32] [ICCV'23]		64.93	55.06	4.33	71.92	62.02	3.71		76.23	68.44	3.99	75.18	68.08	3.89
CDMA [52] [MICCAI'23]		59.44	48.93	4.34	66.38	52.84	4.77		74.18	65.31	4.09	80.81	72.79	3.81
BS-Net [16] [TMI'24]		65.46	56.31	4.09	80.53	70.99	3.33		74.83	67.44	4.03	80.32	74.15	3.62
PMT [13] [ECCV'24]		59.94	51.16	4.02	80.58	72.08	3.15		72.17	64.62	3.76	74.35	67.17	3.55
VCLIPSeg [25] [MICCAI'24]		69.14	61.29	3.61	78.66	70.80	3.14		76.48	68.85	3.70	81.61	75.37	**3.36**
UnCo [47] [TMI'25]		77.56	69.64	3.97	82.02	73.84	3.12		78.01	71.61	3.39	79.29	72.45	3.58
SKCDF [50] [CVPR'25]		57.28	48.24	4.03	75.68	64.09	3.62		70.70	63.25	3.67	78.31	71.11	3.62
CSCPA [11] [CVPR'25]		76.97	65.99	3.57	81.54	70.97	3.08		78.75	70.64	**3.38**	79.66	70.75	4.34
Ours ——		**84.34**	**76.28**	**3.19**	**86.75**	**79.26**	**3.07**		**79.37**	**72.10**	3.88	**83.88**	**77.22**	3.56
UA-MT [45] [MICCAI'19]	Kvasir	73.15	64.12	5.06	83.71	75.95	4.51	ISIC-2018	80.15	71.48	5.37	83.86	75.54	4.81
DTC [29] [AAAI'21]		66.79	56.00	5.63	77.73	68.78	4.73		80.02	71.74	5.19	83.53	75.21	4.89
MC-Net [40] [MICCAI'21]		78.03	69.21	4.89	85.51	78.07	4.46		80.98	72.03	5.31	83.21	74.42	4.98
URPC [30] [MIA'22]		79.14	70.53	4.65	84.47	76.63	4.34		79.38	70.36	5.37	83.18	74.85	4.91
MCF [39] [CVPR'23]		74.01	64.65	4.99	83.25	75.42	4.46		80.76	71.88	5.09	83.50	75.13	4.86
CauSSL [32] [ICCV'23]		76.48	67.61	4.93	81.03	73.20	4.67		79.73	71.19	5.34	84.53	75.88	4.93
CDMA [52] [MICCAI'23]		74.98	65.06	5.07	82.43	73.47	4.63		76.70	66.58	5.84	83.27	74.88	4.78
BS-Net [16] [TMI'24]		77.18	68.22	4.99	82.50	74.98	4.45		81.70	73.68	5.05	84.22	75.83	4.85
PMT [13] [ECCV'24]		75.16	66.18	4.49	81.05	73.16	4.35		82.37	74.18	4.80	84.99	76.73	4.59
VCLIPSeg [25] [MICCAI'24]		77.22	67.75	4.96	83.79	76.01	4.26		79.37	70.32	5.39	83.60	74.91	4.99
UnCo [47] [TMI'25]		81.19	73.79	4.60	85.04	77.99	4.24		85.59	77.56	4.77	87.06	79.59	4.62
SKCDF [50] [CVPR'25]		74.14	64.43	5.11	81.72	73.57	4.56		84.86	77.45	4.74	86.06	77.92	4.59
CSCPA [11] [CVPR'25]		81.60	72.14	4.54	87.31	78.56	4.08		85.75	78.25	4.81	86.47	78.48	4.59
Ours ——		**83.03**	**72.91**	**4.39**	**89.01**	**80.69**	**3.99**		**86.28**	**78.47**	**4.70**	**87.62**	**80.51**	**4.53**

where λ_u is governed by a Gaussian warm-up schedule defined as $\lambda_u(t) = \beta \cdot e^{-5(1-t/t_{max})^2}$, where β is set to 0.1 and t_{max} represents the total number of training iterations.

4. Experiments

4.1. Datasets

We conduct experiments on four medical image segmentation tasks: • **Colonoscopy**: We use CVC-ClinicDB [4], Kvasir [21], and CVC-300 [38]. Following [12], 1,450 images from ClinicDB and Kvasir form the training set (10% for validation). The remaining images and the entirety of CVC-300 are used for testing. • **ISIC-2018**: This dataset [9] contains an extensive collection of dermoscopy images, with $2,075$ images for training, 519 images for validation, and $1,000$ images for testing. • **BCSS** [2]: We crop 151 pathological images into 3,888 patches (1024×1024), treating only tumor regions as foreground. The data is split into 70% training, 10% validation, and 20% testing. • **BUSI** [1]: This dataset [1] aggregates 780 images from 600 female participants with an average size of 500×500 pixels. The samples include 133 normal, 210 malignant, and 437 benign cases. Our study utilizes the 647 abnormal samples, which are split into training and test subsets at 80% and 20%, respectively.

4.2. Implementation Details

All experiments are conducted on two NVIDIA 4090 GPUs using PyTorch 2.4.1 and CUDA 11.2 settings. We use the Stable Diffusion (SD) VAE weights [34] as pretrained encoder $\mathcal{E}(\cdot)$ and decoder $\mathcal{D}(\cdot)$, leveraging their rich latent representation for strong zero-shot generalization. To better adapt the generative model to downstream segmentation tasks, we divide the training into two stages. We first pretrain the mapping network $\mathcal{M}(\cdot)$ and the encoder $E(\cdot)$, and then we fine-tune the entire model in the second stage. This strategy enhances training stability. The batch size is set to 4, including 2 labeled and 2 unlabeled samples. The input images are resized to 224×224. Moreover, the pretraining phase consists of 200 epochs, followed by full training for 350 epochs. During the inference phase, we use the average of the two predictions as the final result.

4.3. Comparison with State-of-the-art Methods

We compare the proposed model with 11 SOTA semi-supervised medical image segmentation methods. To eval-

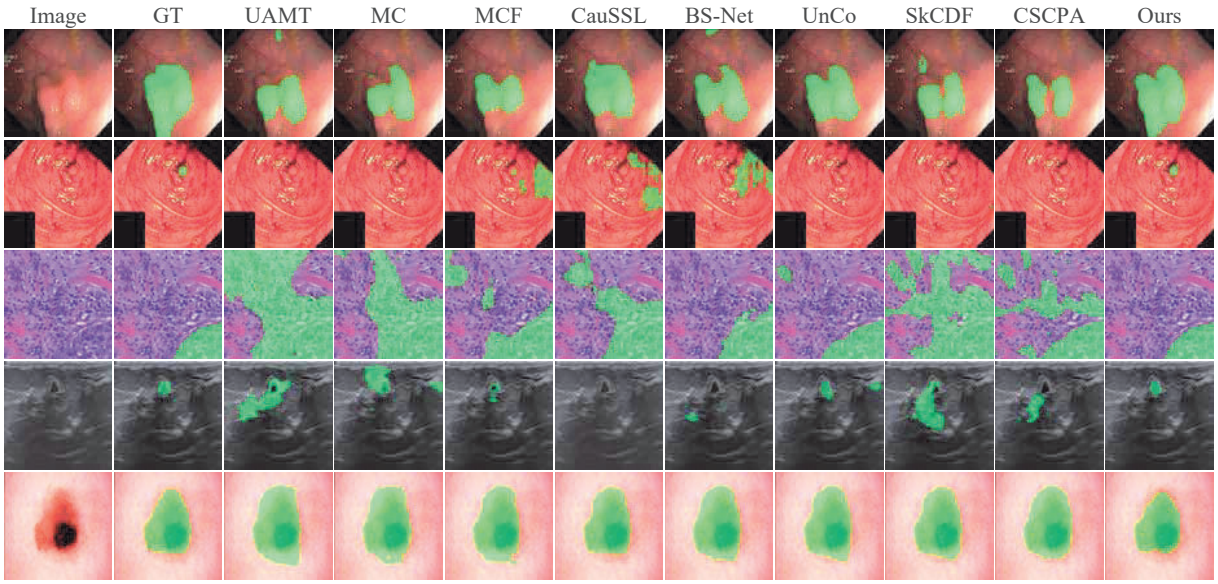


Figure 4. Visual comparisons of our model and other state-of-the-art semi-supervised medical segmentation methods.

Table 2. Quantitative results on the BCSS and BUSI datasets.

Methods	BCSS (10%)			BCSS (30%)		
	Dice (%)	IoU (%)	95HD	Dice (%)	IoU (%)	95HD
UA-MT [45] [MICCAI'19]	63.05	49.65	9.28	67.00	54.28	8.81
DTC [29] [AAAI'21]	63.60	50.19	9.36	69.44	57.07	8.32
MC-Net [40] [MICCAI'21]	63.67	50.12	9.21	69.55	57.33	8.36
URPC [30] [MIA'22]	63.93	50.15	9.23	68.92	56.11	8.63
MCF [39] [CVPR'23]	65.52	52.14	8.98	70.49	57.91	8.30
CDMA [52] [MICCAI'23]	65.58	52.25	8.92	67.28	54.57	8.65
CauSSL [32] [ICCV'23]	62.88	49.35	9.45	71.12	58.69	8.32
BS-Net [16] [TMI'24]	64.47	52.37	8.28	71.30	58.95	8.05
PMT [13] [ECCV'24]	68.97	56.37	8.06	71.45	59.75	7.48
VCLIPSeg [25] [MICCAI'24]	68.14	55.80	7.84	71.77	60.07	7.46
UnCo [47] [TMI'25]	68.66	56.62	7.69	71.56	60.17	7.38
SKCDF [50] [CVPR'25]	70.31	57.65	7.67	73.84	62.07	7.26
CSCPA [11] [CVPR'25]	71.95	59.48	7.57	73.40	61.89	7.13
Ours ——	**74.05**	**62.68**	**7.05**	**75.65**	**64.59**	**6.90**
	BUSI (10%)			BUSI (30%)		
UA-MT [45] [MICCAI'19]	57.85	49.07	4.96	67.82	58.66	4.67
DTC [29] [AAAI'21]	52.73	42.69	5.26	63.67	53.57	4.76
MC-Net [40] [MICCAI'21]	61.72	52.77	4.42	69.61	60.13	4.92
URPC [30] [MIA'22]	57.29	48.64	4.54	64.65	53.10	5.78
MCF [39] [CVPR'23]	64.13	54.57	4.95	73.24	64.68	4.57
CDMA [52] [MICCAI'23]	62.32	51.26	5.20	73.78	64.07	4.70
CauSSL [32] [ICCV'23]	58.12	47.87	5.07	65.94	57.13	4.68
BS-Net [16] [TMI'24]	63.01	52.48	5.24	72.28	63.60	4.56
PMT [13] [ECCV'24]	60.49	52.63	3.91	69.48	61.93	4.63
VCLIPSeg [25] [MICCAI'24]	59.01	49.68	4.90	72.09	63.45	4.52
UnCo [47] [TMI'25]	62.24	54.42	4.27	70.24	62.83	4.66
SKCDF [50] [CVPR'25]	60.99	50.68	5.16	72.91	64.04	4.74
CSCPA [11] [CVPR'25]	65.16	55.17	4.87	73.53	65.05	4.61
Ours ——	**75.57**	**65.72**	**4.70**	**78.74**	**69.49**	**4.52**

uate these methods, we employ three commonly used metrics, namely the Dice coefficient (Dice), Intersection over Union (IoU), and 95% Hausdorff Distance (95HD).

Results on Colonoscopy and ISIC Datasets: Tab. 1 presents quantitative results on colonoscopy datasets (CVC-300, CVC-ClinicDB) and ISIC-2018 under 10% and 30% labeled data. It can be observed that our model achieves promising performance than other comparison methods. For example, on Kvasir with 10% labels, our method enhances Dice by 1.84% and 8.89% compared to UnCo and SKCDF, respectively. On ISIC-2018 with 10% labels, compared to CSCPA, our method increases Dice from 85.75% to 86.28% and decreases 95HD from 4.81 to 4.70. As shown in Fig. 4, our model segments target regions with more precise boundaries and improved completeness compared to methods like UA-MT and MC.

Results on BCSS and BUSI: Tab. 2 presents quantitative results on the BCSS (pathology) and BUSI (ultrasound) datasets. Our model achieves leading performance across various labeling ratios on two distinct domains. On the BCSS dataset with 10% labeled data, our model achieves 74.05% Dice and 62.68% IoU, surpassing the second-best method (CSCPA) in terms of Dice and IoU by 2.10% and 3.20%, respectively. This superiority is consistent on the BUSI dataset, where our model shows significant gains in both segmentation accuracy and boundary precision. As shown in Fig. 4, our model achieves more precise segmentation in complex regions compared to other methods.

4.4. Ablation Study

Effects of Key Components. Given the specific nature of our generative framework, traditional semi-supervised architectures like UAMT [45] or MC [40] are not directly applicable. We therefore establish a baseline model by removing the feature-layer constraints and skip adapter, retaining only the final mask supervision. We then sequentially integrate our two core components: the DAM and the CDSA. As shown in Tab. 3, each component delivers signif-

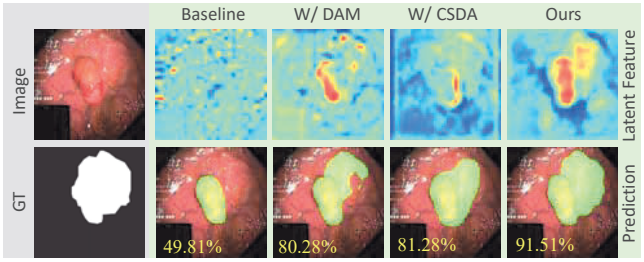


Figure 5. Visual maps of latent features and segmentation results. Yellow numbers denote the sample-level Dice scores.

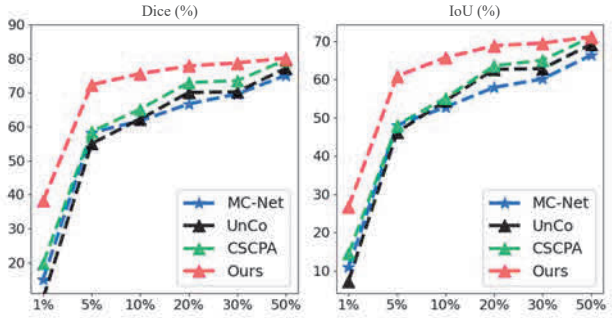


Figure 6. Performance comparison with different labeled ratios on the BUSI dataset.

icant performance gains. For example, on the BUSI dataset with 10% labeled data, the Dice coefficient improves from 70.48% for the baseline to 73.07% with DAM, and further to 75.25% with the addition of the CDSA. Moreover, we provide a step-wise visualization in Fig. 5. It shows that while the features of *baseline (Col. 2)* are scattered and noisy, the proposed *DAM (Col. 3)* effectively rectifies the distribution to focus on the lesion, and *CSDA (Col. 4)* further sharpens the boundaries.

Effects of Two Skip Adapters in CDSA. To investigate the effectiveness of skip adapters in applying generative models to segmentation, we conduct an ablation study in which the baseline retains all settings except for the removal of the two skip adapters. As shown in Tab. 4, on BUSI with 10% labeled data, compared to the baseline, the Dice coefficient improves from 73.07% to 75.57%, and the 95HD decreases from 5.30 to 4.70. This confirms that skip adapters are crucial for enhancing segmentation accuracy and improving the utilization of sparse labeled data.

Effects of Loss Functions. To evaluate the loss function, our baseline retains all structural components (*e.g.* skip adapters) but excludes the unnecessary loss. As shown in Tab. 5, incorporating the unsupervised loss significantly boosts the Dice score (*e.g.* from 79.34% to 83.03% on Kvasir with 10% labels). This confirms that our generative approach enhances pseudo-label quality, demonstrating strong robustness in leveraging unlabeled data.

Performance under Different Labeled Ratios. We further evaluate the performance under different labeled ratios.

Table 3. Ablation study on each key component in our method.

Base.	GDA		ClinicDB (10%)		Kvasir (10%)		BUSI (10%)	
	DAM	CDSA	Dice (%)	95HD	Dice (%)	95HD	Dice (%)	95HD
✓			74.23	4.47	77.01	5.03	70.48	5.44
✓	✓		75.92	4.21	80.02	4.70	73.07	5.30
✓		✓	76.83	4.07	81.84	4.54	75.25	4.77
✓	✓	✓	**79.37**	**3.88**	**83.03**	**4.39**	**75.57**	**4.70**

Table 4. Ablation study on the two skip adapters in CDSA.

Settings		ClinicDB (10%)		Kvasir (10%)		BUSI (10%)	
Image	Mask	Dice (%)	95HD	Dice (%)	95HD	Dice (%)	95HD
		75.92	4.21	80.02	4.70	73.07	5.30
✓		76.98	4.25	81.95	4.69	74.92	4.75
	✓	77.14	4.11	81.79	4.52	73.91	4.88
✓	✓	**79.37**	**3.88**	**83.03**	**4.39**	**75.57**	**4.70**

Table 5. Ablation study on loss functions (10% labeled data).

Datasets	Loss Functions				Metrics	
	$\mathcal{L}^s_{sup}$	$\mathcal{L}^p_{sup}$	$\mathcal{L}^s_{unsup}$	$\mathcal{L}^p_{unsup}$	Dice (%)	HD95
ClinicDB	✓	✓			78.61	4.97
	✓	✓	✓		78.77	4.90
	✓	✓		✓	78.80	4.78
	✓	✓	✓	✓	**79.37**	**3.88**
Kvasir	✓	✓			79.34	4.63
	✓	✓	✓		82.64	4.52
	✓	✓		✓	81.10	4.57
	✓	✓	✓	✓	**83.03**	**4.39**
BUSI	✓	✓			75.08	4.89
	✓	✓	✓		75.36	4.75
	✓	✓		✓	75.32	4.88
	✓	✓	✓	✓	**75.57**	**4.70**

As shown in Fig. 6, our model consistently outperforms both representative (*e.g.*, MC-Net) and state-of-the-art (*e.g.*, UnCo and CSCPA) approaches. Notably, our model excels with extremely limited labeled data (*i.e.*, 1%), underscoring the superior robustness of generative segmentation over discriminative approaches in low-label scenarios.

5. Conclusion

In this paper, we propose SemiGDA, a novel generative segmentation framework to address overfitting and poor pseudo-label quality in settings with limited labeled data. The framework consists of two key components: DAM and CDSA. DAM uses two learnable components to map the prior distributions from images to masks, improving model robustness through feature distribution-level supervision. CDSA introduces skip adapters into the decoder of a pre-trained generative model, enhancing its adaptation to domain-specific segmentation tasks. Additionally, we resolve the compatibility issue between ground truth and the generative model's input-output structure with a training-free conversion and reversion strategy. Experimental results on four segmentation tasks show that SemiGDA outperforms existing SOTA semi-supervised methods.

6. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos. 62576153, 62476054).

References

[1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020. 6

[2] Mohamed Amgad, Habiba Elfandy, Hagar Hussein, Lamees A Atteya, Mai AT Elsebaie, Lamia S Abo Elnasr, Rokia A Sakr, Hazem SE Salem, Ahmed F Ismail, Anas M Saad, et al. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*, 35(18): 3461–3467, 2019. 6

[3] Hritam Basak, Rajarshi Bhattacharya, Rukhshanda Hussain, and Agniv Chatterjee. An exceedingly simple consistency regularization method for semi-supervised medical image segmentation. In *International Symposium on Biomedical Imaging*, pages 1–4. IEEE, 2022. 2

[4] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015. 6

[5] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*, pages 205–218. Springer, 2022. 1

[6] Gaoxiang Chen, Jintao Ru, Yilin Zhou, Islem Rekik, Zhifang Pan, Xiaoming Liu, Yezhi Lin, Beichen Lu, and Jialin Shi. Mtans: multi-scale mean teacher combined adversarial network with shape-aware embedding for semi-supervised brain lesion segmentation. *NeuroImage*, 244:118568, 2021. 2

[7] Jiaqi Chen, Jiachen Lu, Xiatian Zhu, and Li Zhang. Generative semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7111–7120, 2023. 3

[8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1

[9] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 6

[10] Sanuwani Dayarathna, Kh Tohidul Islam, Sergio Uribe, Guang Yang, Munawar Hayat, and Zhaolin Chen. Deep learning based synthesis of mri, ct and pet: Review and analysis. *Medical image analysis*, 92:103046, 2024. 2

[11] Zhenhui Ding, Guilian Chen, Qin Zhang, Huisi Wu, and Jing Qin. Csc-pa: Cross-image semantic correlation via prototype attentions for single-network semi-supervised breast tumor segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 15632–15641, 2025. 6, 7

[12] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 263–273. Springer, 2020. 6

[13] Ning Gao, Sanping Zhou, Le Wang, and Nanning Zheng. PMT: Progressive mean teacher via exploring temporal consistency for semi-supervised medical image segmentation. *European Conference on Computer Vision*, 2024. 6, 7

[14] Shengbo Gao, Ziji Zhang, Jiechao Ma, Zihao Li, and Shu Zhang. Correlation-aware mutual learning for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 98–108. Springer, 2023. 2

[15] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural Information Processing Systems*, 27, 2014. 2

[16] Along He, Tao Li, Juncheng Yan, Kai Wang, and Huazhu Fu. Bilateral supervision network for semi-supervised medical image segmentation. *IEEE Transactions on Medical Imaging*, 43(5):1715–1726, 2024. 2, 6, 7

[17] Kaiwen Huang, Tao Zhou, Huazhu Fu, Yizhe Zhang, Yi Zhou, Chen Gong, and Dong Liang. Learnable prompting sam-induced knowledge distillation for semi-supervised medical image segmentation. *IEEE Transactions on Medical Imaging*, 44(5):2295–2306, 2025. 1

[18] Kaiwen Huang, Tao Zhou, Huazhu Fu, Yizhe Zhang, Yi Zhou, and Xiao-Jun Wu. Uncertainty-aware cross-training for semi-supervised medical image segmentation. *IEEE Transactions on Image Processing*, 34:5543–5556, 2025. 1

[19] Kaiwen Huang, Yi Zhou, Huazhu Fu, Yizhe Zhang, Chen Gong, and Tao Zhou. Text-driven multiplanar visual interaction for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 604–613. Springer, 2025. 2

[20] Jiayu Huo, Xi Ouyang, Sébastien Ourselin, and Rachel Sparks. Generative medical segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3851–3859, 2025. 3

[21] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *MMM*, pages 451–462. Springer, 2020. 6

[22] Siyao Jiang, Huisi Wu, Junyang Chen, Qin Zhang, and Jing Qin. PH-Net: Semi-supervised breast lesion segmentation via patch-wise hardness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11418–11427, 2024. 2

[23] Rushi Jiao, Yichi Zhang, Le Ding, Bingsen Xue, Jicong Zhang, Rong Cai, and Cheng Jin. Learning with limited annotations: a survey on deep semi-supervised learning for

medical image segmentation. *Computers in Biology and Medicine*, page 107840, 2023. 1, 2

[24] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[25] Lei Li, Sheng Lian, Zhiming Luo, Beizhan Wang, and Shaozi Li. Vclipseg: Voxel-wise clip-enhanced model for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 692–701. Springer, 2024. 6, 7

[26] Ning Li, Lianjin Xiong, Wei Qiu, Yudong Pan, Yiqian Luo, and Yangsong Zhang. Segment anything model for semi-supervised medical image segmentation via selecting reliable pseudo-labels. In *International Conference on Neural Information Processing*, pages 138–149. Springer, 2023. 2

[27] Chen Liang, Wenguan Wang, Jiaxu Miao, and Yi Yang. Gmmseg: Gaussian mixture based generative semantic segmentation models. *Advances in Neural Information Processing Systems*, 35:31360–31375, 2022. 3

[28] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. 1

[29] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8801–8809, 2021. 6, 7

[30] Xiangde Luo, Guotai Wang, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Dimitris N Metaxas, and Shaoting Zhang. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis*, 80:102517, 2022. 6, 7

[31] Jiaxin Mei, Tao Zhou, Kaiwen Huang, Yizhe Zhang, Yi Zhou, Ye Wu, and Huazhu Fu. A survey on deep learning for polyp segmentation: Techniques, challenges and future trends. *Visual Intelligence*, 3(1):1, 2025. 1

[32] Juzheng Miao, Cheng Chen, Furui Liu, Hao Wei, and Pheng-Ann Heng. Caussl: Causality-inspired semi-supervised learning for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21426–21437, 2023. 6, 7

[33] Himashi Peiris, Zhaolin Chen, Gary Egan, and Mehrtash Harandi. Duo-segnet: adversarial dual-views for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 428–438. Springer, 2021. 2

[34] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 3, 4, 6

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 1

[36] Yucheng Shu, Hengbo Li, Bin Xiao, Xiuli Bi, and Weisheng Li. Cross-mix monitoring for medical image segmentation with limited supervision. *IEEE Transactions on Multimedia*, 25:1700–1712, 2022. 2

[37] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2

[38] David Vázquez, Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Antonio M López, Adriana Romero, Michal Drozdzal, and Aaron Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017(1):4037190, 2017. 6

[39] Yongchao Wang, Bin Xiao, Xiuli Bi, Weisheng Li, and Xinbo Gao. Mcf: Mutual correction framework for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15651–15660, 2023. 2, 6, 7

[40] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang. Semi-supervised left atrium segmentation with mutual consistency training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 297–306. Springer, 2021. 6, 7

[41] Chenchu Xu, Yuan Yang, Zhiqiang Xia, Boyan Wang, Dong Zhang, Yanping Zhang, and Shu Zhao. Dual uncertainty-guided mixing consistency for semi-supervised 3d medical image segmentation. *IEEE Transactions on Big Data*, 9(4): 1156–1170, 2023. 2

[42] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023. 1, 2

[43] Huifeng Yao, Xiaowei Hu, and Xiaomeng Li. Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3099–3107, 2022. 1

[44] Jiarong Ye, Haomiao Ni, Peng Jin, Sharon X Huang, and Yuan Xue. Synthetic augmentation with large-scale unconditional pre-training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 754–764. Springer, 2023. 2

[45] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 605–613. Springer, 2019. 1, 2, 6, 7

[46] Ling-Li Zeng, Kai Gao, Dewen Hu, Zhichao Feng, Chenping Hou, Pengfei Rong, and Wei Wang. SS-TBN: A semi-supervised tri-branch network for covid-19 screening and lesion segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10427–10442, 2023. 1, 2

[47] Xiang Zeng, Shengwu Xiong, Jinming Xu, Guangxing Du, and Yi Rong. Uncertainty co-estimator for improving semi-supervised medical image segmentation. *IEEE Transactions on Medical Imaging*, 44(9):3870–3881, 2025. 6, 7

[48] Li Zhang, Basu Jindal, Ahmed Alaa, Robert Weinreb, David Wilson, Eran Segal, James Zou, and Pengtao Xie. Genera-

tive ai enables medical image segmentation in ultra low-data regimes. *Nature Communications*, 16(1):6486, 2025. 2, 3

[49] Zhenxi Zhang, Chunna Tian, Harrison X Bai, Zhicheng Jiao, and Xilan Tian. Discriminative error prediction network for semi-supervised colon gland segmentation. *Medical Image Analysis*, 79:102458, 2022. 2

[50] Zheng Zhang, Guanchun Yin, et al. A semantic knowledge complementarity based decoupling framework for semi-supervised class-imbalanced medical image segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25940–25949, 2025. 6, 7

[51] Zhen Zhao, Zicheng Wang, Longyue Wang, Dian Yu, Yixuan Yuan, and Luping Zhou. Alternate diverse teaching for semi-supervised medical image segmentation. In *European Conference on Computer Vision*, pages 227–243. Springer, 2024. 2

[52] Lanfeng Zhong, Xin Liao, Shaoting Zhang, and Guotai Wang. Semi-supervised pathological image segmentation via cross distillation of multiple attentions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 570–579. Springer, 2023. 6, 7