Conditional Diffusion Prompting for Ambiguous Medical Image Segmentation*

Hongkai Zhao[1], Jun Gao[2], Qingbo Kang[3], and Qicheng Lao[1*]

[1] Beijing University of Posts and Telecommunications, Beijing, China
zhk2336258035@bupt.edu.cn, qicheng.lao@gmail.com
[2] Stork Healthcare, China
[3] West China Hospital, Sichuan University, Chengdu, China

Abstract. Ambiguous medical image segmentation often admits multiple clinically plausible delineations due to substantial inter-expert variability, and such inherent uncertainty cannot be faithfully represented by a single deterministic mask. Promptable foundation models such as SAM offer strong representations, yet under ambiguous boundaries, their predictions can be highly sensitive to prompt perturbations, and existing stochastic prompting strategies are often coarse and weakly coupled to image-side ambiguity, yielding diverse samples that may deviate from image-consistent boundaries. We propose Conditional Diffusion Prompting (CDP), which performs diffusion prompting in dense prompt embedding space prior to SAM-style decoding and conditions prompt-space denoising on stochastic image embeddings to couple prompt ambiguity with image ambiguity. Experiments on multiple ambiguous medical image segmentation benchmarks demonstrate improved distribution matching and sample quality, achieving lower GED and higher HM-IOU and $D_{\max}$. CDP provides a structured sampling mechanism for promptable decoders, enabling diverse yet image-consistent segmentations and improving practical reliability in ambiguous clinical settings.

Keywords: Ambiguous segmentation · Promptable segmentation · Diffusion models · Medical imaging

1 Introduction

Ambiguous medical image segmentation (AMIS) is inherently a *one-to-many* problem: a single image can admit multiple clinically plausible delineations due to weak/missing boundaries, low contrast, partial volume, and inter-annotator variability [15,2]. Such non-uniqueness is a routine characteristic of real-world clinical data. Consequently, a segmentation model must go beyond point estimation. Rather than producing a single deterministic mask, a reliable AMIS framework should learn the conditional distribution of plausible segmentations and generate predictions that are both accurate and sufficiently diverse [20,25].

* Code: https://github.com/shannan-zhk/Conditional-Diffusion-Prompting

To model this distributional nature, prior work formulates AMIS as a conditional generative process. Probabilistic segmentation models employ latent-variable decoders and structured uncertainty modeling to enable stochastic sampling [15,2,20]. Diffusion-based approaches provide a principled sampling framework via iterative denoising [9] and have recently been explored for ambiguous segmentation [21,23]. More recently, stochastic extensions of promptable foundation models such as Segment Anything (SAM), e.g., $\mathcal{A}$-SAM [16], have been introduced to support distributional prediction under ambiguity. Despite these advances, two limitations remain salient for promptable decoders in AMIS: (i) prompt-side uncertainty is often introduced in a coarse and weakly structured manner, limiting its ability to capture complex boundary variability; (ii) prompt uncertainty is often decoupled from image ambiguity, so increasing diversity can produce delineations that deviate from image-consistent boundaries.

Motivated by these limitations and inspired by diffusion-based latent modeling, we reformulate prompt generation itself as probabilistic distribution modeling in dense prompt embedding space, while maintaining a promptable decoding architecture for AMIS. Specifically, we propose *Diffusion Prompting (DP)*, which performs iterative denoising over dense prompt embeddings to model their underlying distribution and generate stochastic prompt latents. To further align prompt uncertainty with image ambiguity, we extend DP to *Conditional Diffusion Prompting (CDP)*. CDP conditions prompt-space diffusion on stochastic image embeddings through latent-variable generative modeling, explicitly coupling prompt variability with image-level uncertainty and producing diverse yet image-consistent segmentation candidates—a key advance over prior promptable AMIS methods. Our contributions can be summarized as follows:

- We introduce Diffusion Prompting (DP) to model stochastic dense prompt embeddings via diffusion denoising for promptable ambiguous segmentation.
- We propose Conditional Diffusion Prompting (CDP) by conditioning prompt-space diffusion on stochastic image latents, coupling prompt uncertainty with image ambiguity for improved distributional alignment.
- We conduct extensive experiments on multiple ambiguous medical segmentation benchmarks, demonstrating consistent improvements in distribution matching and sample quality, outperforming state-of-the-art baselines.

2 Related Work

Ambiguous segmentation, diffusion, and promptable decoders. Ambiguous segmentation is often framed as learning a conditional mask distribution from which multiple plausible delineations can be sampled, using latent-variable decoders and structured uncertainty modeling [15,2,20,11]. Representative variants include conditional generative modeling for structured outputs [22], hierarchical latent-variable decoders [14], autoregressive PixelSeg [25], and strong U-Net backbones [10,3]. Diffusion models have also been adopted for diverse medical mask generation via iterative refinement under uncertain boundaries [9,21,23]. In parallel, promptable foundation models such as SAM [13] have been adapted

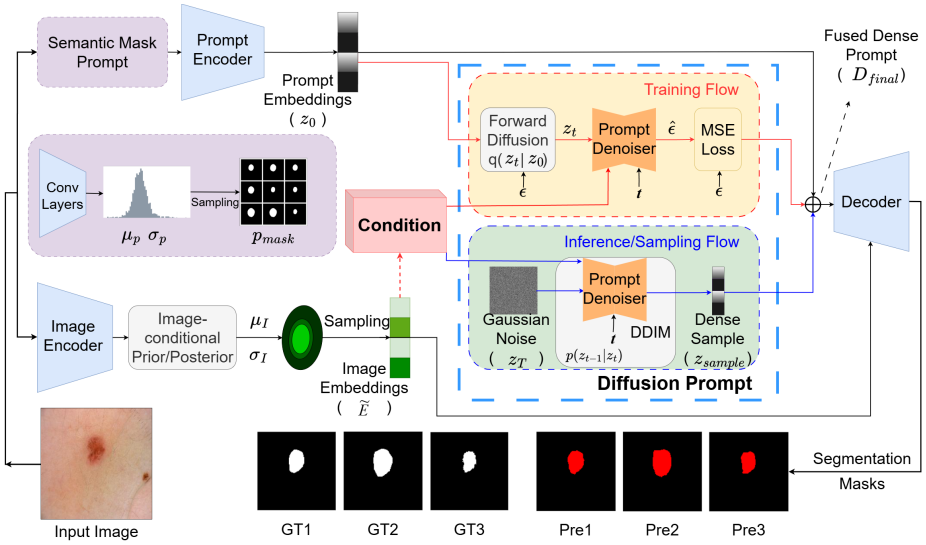

Fig. 1. Overview of our diffusion prompting framework for ambiguous segmentation. A semantic mask prompt is sampled and encoded by a SAM prompt encoder into sparse prompt tokens and a dense prompt embedding (prompt latent z_0). DP performs diffusion denoising in dense prompt space to sample a stochastic prompt latent, which is aggregated into the final dense prompt for decoding. The blue dashed box delineates the DP module; integrating the adjacent conditioning branch yields the full CDP framework. CDP further conditions prompt-space diffusion on stochastic image embeddings $\tilde{E}$, coupling image ambiguity with prompt uncertainty. A SAM-style mask decoder then generates multiple segmentation samples from the image embeddings and the diffusion-refined dense prompt.

to medical imaging [5,24,4,7,19], and A-SAM introduces stochastic embedding sampling to support distributional outputs [16].

Gap and our approach. Existing probabilistic and diffusion approaches for AMIS primarily model uncertainty in mask or pixel space and do not leverage prompt-guided decoding. In contrast, promotable SAM variants either rely on manual prompts or inject unstructured embedding noise, without explicitly coupling prompt-side uncertainty with image-side ambiguity. We therefore perform diffusion in dense prompt embedding space (DP) and condition prompt denoising on stochastic image embeddings (CDP), enabling diverse, image-consistent sampling without manual prompting.

3 Method

3.1 Problem Formulation and Method Overview

Ambiguous medical image segmentation is inherently one-to-many. For a given image x, multiple clinically plausible annotations $\{y_1, \ldots, y_n\}$ may coexist due to

weak boundaries, low contrast, and inter-annotator variability. A natural view is to regard the annotations as samples from an unknown conditional distribution $p^*\{y \mid x\}$ and to learn a model $p_\theta\{y \mid x\}$ that supports sampling:

$$y_i^* \sim p^*\{y \mid x\}, \quad \hat{y}^{(j)} \sim p_\theta\{y \mid x\}. \tag{1}$$

This formulation emphasizes distributional prediction. Rather than collapsing to a single mask, the model is expected to generate a set of plausible candidates whose distribution matches the multi-annotator reference distribution.

As illustrated in Fig. 1, our framework retains a promptable decoding architecture while reformulating prompt generation as probabilistic distribution modeling in the latent space. We begin by proposing *Diffusion Prompting (DP)*, which models the distribution of dense prompt embeddings through iterative denoising in prompt space, thereby generating stochastic prompt latents. To better align prompt uncertainty with image ambiguity, we further develop *Conditional Diffusion Prompting (CDP)*, which augments prompt-space diffusion with stochastic image embeddings as conditioning signals, explicitly linking prompt variability to image-level uncertainty and yielding diverse yet image-consistent segmentation candidates.

3.2 Diffusion Prompting for Stochastic Prompt Latent Modeling

We first obtain the initial prompt embeddings without manual interaction by sampling a semantic mask prompt from a frozen probabilistic mask generator [17] and encoding it with a SAM prompt encoder [13]:

$$\hat{m} \sim p_{\text{mask}}\{m \mid x\}, \quad \{S, D\} = \text{Enc}_{\text{prompt}}\{\hat{m}\}. \tag{2}$$

Here, $p_{\text{mask}}\{m \mid x\}$ is instantiated as a low-rank Gaussian in logit-mask space. Given an input image x, the generator predicts μ, d, and V, and samples $z \sim \mathcal{N}\{\mu, \text{diag}\{d\} + VV^\top\}$, which is mapped to a probabilistic mask by $\sigma\{z\}$ and resized to form $\hat{m}$. The prompt encoder $\text{Enc}_{\text{prompt}}\{\cdot\}$ then maps $\hat{m}$ to sparse tokens S and a dense prompt embedding D, where D serves as the *prompt latent* that provides spatial guidance to the SAM-style mask decoder.

Diffusion Prompting To model prompt uncertainty, we define an implicit conditional distribution over dense prompt embeddings centered at the per-image anchor D. Specifically, we treat D as the clean latent z_0 and adopt a standard diffusion formulation in prompt space. The forward noising process is defined as:

$$z_t = \sqrt{\bar{\alpha}_t}\, z_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \quad z_0 \equiv D, \quad \epsilon \sim \mathcal{N}(0, I), \tag{3}$$

where $\{\bar{\alpha}_t\}$ denotes the cumulative noise schedule.

A lightweight denoiser ϵ_θ is trained to predict the injected noise without image conditioning: $\hat{\epsilon} = \epsilon_\theta(z_t, t)$, and the model is optimized using the standard noise-prediction objective:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{D,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2\right], \tag{4}$$

where the expectation is taken over prompt samples, diffusion steps, and injected noise. Minimizing $\mathcal{L}_{\text{diff}}$ learns a reverse denoising process in dense prompt space.

At inference time, we sample a stochastic prompt latent $\tilde{D} \sim p_\theta(\tilde{D} \mid D)$ by initializing the reverse process from a heavily noised latent $z_T = \sqrt{\bar{\alpha}_T} D + \sqrt{1-\bar{\alpha}_T}\epsilon$ at step T, followed by deterministic DDIM sampling: $\tilde{D} = \text{DDIM}_\theta(z_T)$.

To stabilize sampling while preserving stochastic diversity, we construct the final prompt embedding via a conservative aggregation: $D_{\text{final}} = (D + \tilde{D})/2$. The SAM-style mask decoder then produces a segmentation prediction conditioned on image embeddings, sparse tokens S, and D_{final}.

3.3 Conditional Diffusion Prompting with Stochastic Image Latent

Stochastic Image Latent To represent image-side ambiguity, we adopt a VAE-style stochastic latent variable [12]. Following $\mathcal{A}$-SAM [16], we parameterize an image-conditional *prior* as a diagonal Gaussian: $q_\phi(z_I \mid x) = \mathcal{N}(\mu_\phi(x), \text{diag}(\sigma_\phi^2(x)))$, and introduce a *posterior* available only during training to incorporate the ground-truth mask y and stabilize learning under ambiguity: $q_\psi(z_I \mid x, y) = \mathcal{N}(\mu_\psi(x, y), \text{diag}(\sigma_\psi^2(x, y)))$. To ensure that the prior can serve as a reliable approximation at inference time, we regularize it toward the posterior using a KL divergence term:

$$\mathcal{L}_{\text{reg}} = \text{KL}\{q_\psi\{z_I \mid x, y\} \| q_\phi\{z_I \mid x\}\}. \tag{5}$$

The image latent z_I can be sampled via reparameterization, e.g.,

$$z_I = \mu_\phi(x) + \sigma_\phi(x) \odot \eta, \quad \eta \sim \mathcal{N}(0, I). \tag{6}$$

Image-Conditioned Prompt Diffusion CDP conditions prompt-space diffusion on the stochastic image latent, enabling image-aware modeling of prompt uncertainty. Given x, the image encoder produces embeddings E, we sample z_I, and a fusion module injects it to obtain perturbed image embeddings:

$$E = \text{Enc}_{\text{img}}\{x\}, \quad z_I \sim q_\phi\{z_I \mid x\}, \quad \tilde{E} = \text{Fcomb}_o\{E, z_I\}. \tag{7}$$

Here, z_I represents image-side ambiguity and $\tilde{E}$ serves as an image-consistent stochastic conditioning signal for prompt denoising. Notably, z_I does not perturb the forward noising of the prompt latent z_t; it only modulates the denoiser through conditioning. The diffusion denoiser introduced in Sec. 3.2 is therefore extended as:

$$\hat{\epsilon} = \epsilon_\theta\{z_t, \tilde{E}, t\}, \tag{8}$$

incorporating $\tilde{E}$ as an additional conditioning input.

By conditioning prompt-space diffusion on $\tilde{E}$, CDP explicitly couples image ambiguity with prompt uncertainty, leading to improved distributional alignment and more image-consistent stochastic segmentation predictions.

Table 1. Quantitative comparison of segmentation methods across four datasets using GED, HM-IoU, and $D_{\max}$ metrics. The best results are highlighted in **bold**.

LIDC				**ISIC**			
Method	GED ↓	HM-IOU ↑	$D_{\max}$ ↑	Method	GED ↓	HM-IOU ↑	$D_{\max}$ ↑
Prob UNet [15]	0.343	0.708	0.865	Prob UNet [15]	0.488	0.554	0.893
SSN [20]	0.327	0.736	0.774	SSN [20]	0.532	0.733	0.782
CAR [11]	0.456	0.612	0.878	CAR [11]	0.442	0.759	0.887
Mose [8]	0.326	0.723	0.934	Mose [8]	0.346	0.621	0.908
A-SAM [16]	0.358	0.768	0.961	A-SAM [16]	0.300	0.756	0.894
CDP	**0.323**	**0.839**	**0.972**	**CDP**	**0.265**	**0.791**	**0.913**
Prostate				**Pancreatic-lesion**			
Method	GED ↓	HM-IOU ↑	$D_{\max}$ ↑	Method	GED ↓	HM-IOU ↑	$D_{\max}$ ↑
Prob UNet [15]	0.3667	0.633	0.780	Prob UNet [15]	0.190	0.777	0.802
SSN [20]	0.452	0.523	0.723	SSN [20]	0.206	0.646	0.717
CAR [11]	0.420	0.583	0.786	CAR [11]	0.224	0.606	0.730
Mose [8]	0.343	0.650	0.825	Mose [8]	0.180	0.862	0.908
A-SAM [16]	0.390	0.586	0.826	A-SAM [16]	0.183	0.878	0.905
CDP	**0.341**	**0.668**	**0.885**	**CDP**	**0.170**	**0.880**	**0.907**

3.4 Training Objective

The training objective combines supervised segmentation loss with augmented terms for diffusion prompting, controlled by a single hyperparameter λ:

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda \left(\mathcal{L}_{\text{diff}} + \mathcal{L}_{\text{reg}} \right). \tag{9}$$

Here, $\mathcal{L}_{\text{seg}}$ denotes the supervised loss between predicted logits and y, instantiated as the sum of cross entropy, Dice, and sigmoid focal losses, i.e., $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{Focal}}$. The regularization term $\mathcal{L}_{\text{reg}}$ matches the image-conditional prior to the posterior over z_I via KL divergence, while $\mathcal{L}_{\text{diff}}$ corresponds to the diffusion noise-prediction loss introduced in Eq. (4).

4 Experiment

4.1 Experimental Setup

Datasets We evaluate CDP on four public datasets widely used for ambiguous segmentation, each representing distinct imaging modalities and clinical objectives: LIDC [1], ISIC [6], Prostate [18], and Pancreatic-lesion [18]. These datasets present challenges such as ambiguous boundaries and inter-annotator variability.
Evaluation Metrics We report three evaluation metrics: Generalized Energy Distance (GED), Hausdorff Metric Intersection over Union (HM-IoU), and Maximum Distance ($D_{\max}$). GED measures distributional discrepancy using a $1-$IoU

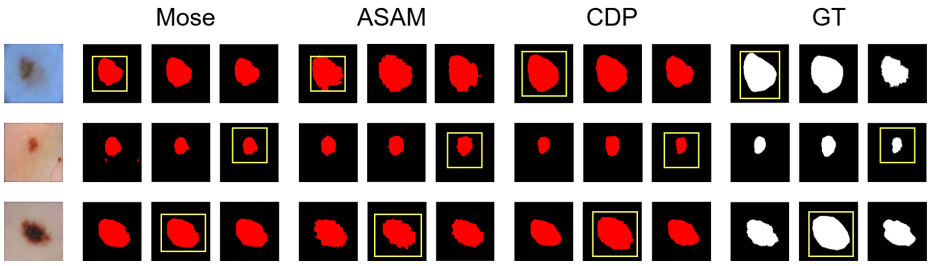

Fig. 2. Visual comparison of segmentation diversity on the ISIC dataset.

distance, where lower values indicate better alignment. HM-IoU utilizes the Hungarian algorithm to match predicted samples with annotations, with higher values reflecting better performance. $D_{\max}$ reports the Dice coefficient of the best-matched annotation, where higher values indicate greater segmentation accuracy.
Implementation Details We use Adam with learning rate 1×10^{-4}. For the prompt diffusion, we use DDIM sampling with 100 reverse steps. For each test image, we generate 16 segmentation samples via stochastic inference to evaluate both sample fidelity and distribution congruence. Following [16], we also apply ℓ_2 regularization to the prior, posterior, and fusion modules to stabilize training. CDP incurs higher computation than A-SAM (49.54 vs. 22.66 GFLOPs/sample) under the 100-step setting, but this overhead is tunable: reducing to 10 DDIM steps lowers the runtime from 3.464 s to 0.721 s with 16 samples, while peak memory remains around 9.42 GB, offering a favorable cost-quality trade-off.

4.2 Experimental Results

Quantitative Comparison Table 1 reports quantitative results on LIDC, ISIC, Prostate, and Pancreatic-lesion under GED, HM-IOU, and $D_{\max}$. We compare CDP with established ambiguous segmentation baselines, including latent-variable models Probabilistic U-Net [15], PHiSeg [2], and SSN [20], as well as CAR [11] and Mose [8], and the promptable baseline A-SAM [16]. All baseline results are obtained by running their official released code under the same evaluation protocol. CDP achieves the lowest GED on all datasets, which indicates improved distribution matching to the multi-annotator reference. CDP also yields strong HM-IOU and $D_{\max}$, reflecting competitive sample fidelity and boundary quality alongside improved distributional alignment.

Qualitative Comparison Fig. 2 qualitatively compares Mose [8], A-SAM [16], and our proposed CDP on the ISIC dataset. Mose shows limited variation across samples, while A-SAM often yields less sharp boundaries. CDP produces diverse candidates that better align with the reference masks.

Ablation Study Table 2 presents an ablation study on LIDC to explicitly quantify the contribution of each component, namely diffusion prompting in

Table 2. Ablation results on LIDC. **Base** denotes a SAM-style mask decoder with a fixed semantic mask prompt generator and stochastic image embeddings. DP denotes diffusion prompting in prompt space without image conditioning. CDP denotes conditional diffusion prompting with image-conditioned guidance.

Base	DP	CDP	GED ↓	HM-IOU ↑	$D_{\max}$ ↑
✓			0.3400 ± 0.0036	0.7823 ± 0.0012	0.9643 ± 0.0006
✓	✓		0.3297 ± 0.0048	0.8363 ± 0.0058	0.9720 ± 0.0012
✓	✓	✓	**0.3233 ± 0.0064**	**0.8493 ± 0.0045**	**0.9747 ± 0.0015**

Table 3. Effect of sampling count on the LIDC and ISIC datasets. All experiments use the full CDP pipeline.

Dataset	Metric	16	32	64
LIDC	GED ↓	0.323	0.318	0.312
	HM-IOU ↑	0.849	0.857	0.862
	$D_{\max}$ ↑	0.974	0.976	0.978
ISIC	GED ↓	0.265	0.261	0.258
	HM-IOU ↑	0.791	0.811	0.823
	$D_{\max}$ ↑	0.913	0.916	0.918

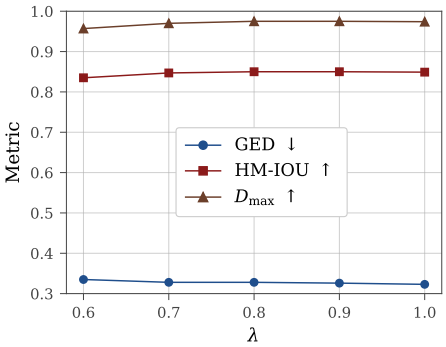

Fig. 3. Hyperparameter sensitivity.

the dense prompt space (denoted as 'DP') and image-conditioned guidance (denoted as 'CDP'). Results are reported as mean±standard deviation across three runs. 'Base' is without diffusion prompting in prompt space; instead, it derives prompt embeddings from semantic mask prompts and combines them with stochastic image embeddings for SAM-style decoding. Introducing DP yields consistent improvements over 'Base' across GED, HM-IoU, and $D_{\max}$. This confirms that modeling prompt uncertainty through diffusion is itself effective, even without image-conditioned guidance. Building upon DP, CDP further conditions the prompt denoising process on stochastic image latents, which brings further performance improvements across all metrics, achieving the best overall results. These improvements demonstrate that image-conditioned diffusion prompting provides complementary benefits beyond standalone diffusion modeling.

Sensitivity Analyses Table 3 studies the effect of sampling count. We observe consistent performance across the tested sampling counts, with modest gains as the number of samples increases. Fig. 3 reports sensitivity to the auxiliary loss weight λ and shows that performance remains stable over the evaluated range, indicating that CDP is not overly sensitive to this hyperparameter.

5 Conclusion

Promptable foundation models offer strong transferable representations, yet in medical imaging their predictions can vary substantially under ambiguous boundaries and small prompt perturbations, motivating distributional prediction. We propose Conditional Diffusion Prompting (CDP), which performs diffusion prompting in dense prompt embedding space and conditions prompt denoising on stochastic image embeddings for image-consistent sampling. Experiments on four public ambiguous segmentation benchmarks show improved distribution matching and sample quality, reflected by lower GED and higher HM-IOU and $D_{\max}$, and ablations confirm complementary gains from diffusion prompting and image-conditioned guidance. Future work will improve the efficiency of prompt-space diffusion sampling and explore broader ambiguity-aware clinical tasks.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics **38**(2), 915–931 (2011)
2. Baumgartner, C.F., Tezcan, K.C., Chaitanya, K., Hötker, A.M., Muehlematter, U.J., Schawkat, K., Becker, A.S., Donati, O., Konukoglu, E.: PHiSeg: Capturing uncertainty in medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. LNCS, vol. 11765, pp. 119–127. Springer, Cham (2019)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision (ECCV). pp. 205–218 (2022)
4. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: SAM-Adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3367–3375 (2023)
5. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., Sun, H., He, J., Zhang, S., Zhu, M., Qiao, Y.: SAM-Med2D. arXiv preprint arXiv:2308.16184 (2023)
6. Codella, N.C.F., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1902.03368 (2019)
7. Gao, Y., Xia, W., Hu, D., Gao, X.: DeSAM: Decoupling segment anything model for generalizable medical image segmentation. arXiv preprint arXiv:2306.00499 (2023)
8. Gao, Z., Chen, Y., Zhang, C., He, X.: Modeling multimodal aleatoric uncertainty in segmentation with mixture of stochastic expert. arXiv preprint arXiv:2212.07328 (2022)

9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (2020)
10. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: UNet 3+: A full-scale connected UNet for medical image segmentation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1055–1059 (2020)
11. Kassapis, E., Dikov, G., Gupta, D.K., Nugteren, C.: Calibrated adversarial refinement for stochastic semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7057–7067 (2021)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (ICLR) (2014)
13. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4015–4026 (2023)
14. Kohl, S.A.A., Romera-Paredes, B., Maier-Hein, K.H., Rezende, D.J., Eslami, S.M.A., Kohli, P., Zisserman, A., Ronneberger, O.: A hierarchical probabilistic U-Net for modeling multi-scale ambiguities. arXiv preprint arXiv:1905.13077 (2019)
15. Kohl, S.A.A., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S.M.A., Jimenez Rezende, D., Ronneberger, O.: A probabilistic U-Net for segmentation of ambiguous images. In: Advances in Neural Information Processing Systems. pp. 6965–6975 (2018)
16. Li, C., Huang, Y., Li, W., Liu, H., Liu, X., Xu, Q., Chen, Z., Huang, Y., Yuan, Y.: Flaws can be applause: Unleashing potential of segmenting ambiguous objects in SAM. In: Advances in Neural Information Processing Systems (NeurIPS) (2024)
17. Li, F., Li, X., Su, X., Qiu, X., Dong, S., Wang, W., Wang, K., Luo, G., Li, S.: Ambiguity-aware truncated flow matching for ambiguous medical image segmentation. arXiv preprint arXiv:2511.06857 (2025)
18. Li, H., Navarro, F., Ezhov, I., Bayat, A., Das, D., Kofler, F., et al.: QUBIQ: Uncertainty quantification for biomedical image segmentation challenge. arXiv preprint arXiv:2405.18435 (2024)
19. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
20. Monteiro, M., Le Folgoc, L., Coelho de Castro, D., Pawlowski, N., Marques, B., Kamnitsas, K., van der Wilk, M., Glocker, B.: Stochastic segmentation networks: Modelling spatially correlated aleatoric uncertainty. In: Advances in Neural Information Processing Systems. vol. 33 (2020)
21. Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11536–11546 (2023)
22. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 28 (2015)
23. Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y.: MedSegDiff: Medical image segmentation with diffusion probabilistic model. In: Medical Imaging with Deep Learning. Proceedings of Machine Learning Research, vol. 227, pp. 1623–1639 (2024)
24. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical SAM adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)

25. Zhang, W., Zhang, X., Huang, S., Lu, Y., Wang, K.: PixelSeg: Pixel-by-pixel stochastic semantic segmentation for ambiguous medical images pp. 4742–4750 (2022)