ScribblePrompt: Fast and Flexible Interactive Segmentation for Any Biomedical Image

Hallee E. Wong[1,2], Marianne Rakic[1,2], John Guttag[1], and Adrian V. Dalca[1,2,3]

[1] MIT CSAIL, Cambridge, MA, USA
[2] Martinos Center, Massachusetts General Hospital, Charlestown, MA, USA
[3] Harvard Medical School, Boston, MA, USA
{hallee,mrakic,guttag,adalca}@mit.edu

Abstract. Biomedical image segmentation is a crucial part of both scientific research and clinical care. With enough labelled data, deep learning models can be trained to accurately automate specific biomedical image segmentation tasks. However, manually segmenting images to create training data is highly labor intensive and requires domain expertise. We present *ScribblePrompt*, a flexible neural network based interactive segmentation tool for biomedical imaging that enables human annotators to segment previously unseen structures using scribbles, clicks, and bounding boxes. Through rigorous quantitative experiments, we demonstrate that given comparable amounts of interaction, ScribblePrompt produces more accurate segmentations than previous methods on datasets unseen during training. In a user study with domain experts, ScribblePrompt reduced annotation time by 28% while improving Dice by 15% compared to the next best method. ScribblePrompt's success rests on a set of careful design decisions. These include a training strategy that incorporates both a highly diverse set of images and tasks, novel algorithms for simulated user interactions and labels, and a network that enables fast inference. We showcase ScribblePrompt in an interactive demo, provide code, and release a dataset of scribble annotations at https://scribbleprompt.csail.mit.edu

Keywords: Interactive Segmentation · Biomedical Imaging · Scribbles

1 Introduction

Biomedical image segmentation is an essential step in a wide range of biomedical research and clinical care pipelines. Deep learning has become the predominant method to automate existing segmentation tasks [43, 78, 92, 106]. Biomedical researchers and clinicians often encounter novel segmentation tasks involving either new regions of interest [119] or new image modalities [77,97]. Unfortunately, supervised training of accurate models for new domains requires diverse images with careful annotations by skilled experts.

Most widely used interactive segmentation systems for biomedical imaging provide minimal, intensity-based, algorithmic assistance [118]. Despite a growing literature, learning-based systems are less widely used in practice, perhaps

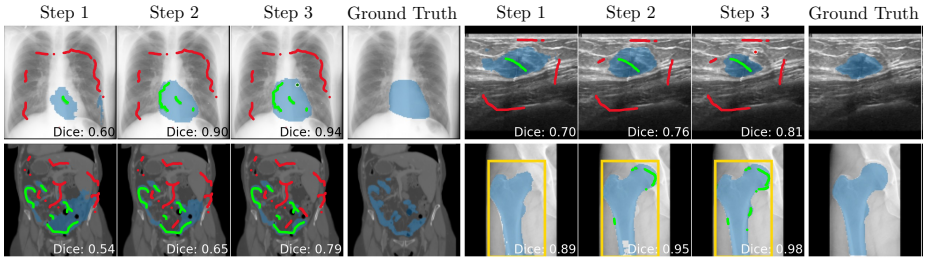


Fig. 1: ScribblePrompt enables rapid iterative interactive segmentation of *unseen* tasks using bounding boxes, clicks, and scribbles. We show predictions from ScribblePrompt with iterative interaction steps on examples from datasets unseen during training. At each step, we visualize positive scribble and click inputs in green, negative scribble and click inputs in red, bounding box inputs in yellow, and the predicted segmentation in blue. Scribble thickness is enlarged for visual clarity. See Supplementary Material for more examples.

because they are generally focused on specific tasks or modalities, which limits their applicability [68, 83, 95, 109, 110, 128]. Recent vision foundation models target broad applicability, but require fine-tuning to the medical domain [47, 112, 113, 129]. Most interactive segmentation models require specific interactions, like carefully placed clicks, making them easy to develop but difficult to use in practice. The result of these shortcomings is that existing learning-based models are not widely used in practice.

In this work, we present *ScribblePrompt*, a general model for interactive biomedical image segmentation that

1. enables users to rapidly and accurately accomplish any biomedical image segmentation task, outperforming state-of-the-art models, particularly for *unseen* labels and image types;
2. is flexible to different annotation styles, including bounding boxes, clicks, *and* scribbles,
3. is computationally efficient, enabling fast inference, even on a single CPU.

To achieve this, we focused on design decisions aimed at ease of use and realistic interactions. To train our model, we started by gathering a large corpus of biomedical imaging datasets. We then designed task augmentation through synthesis strategies to encourage generalization to unseen tasks. Finally, we designed *new* interaction simulation strategies for training. The model itself uses an architecture optimized for efficient inference.

To evaluate ScribblePrompt, we compared its performance to that of existing interactive segmentation methods on datasets unseen during training. These experiments involved manually-collected scribbles, simulated interactions (Fig. 1), and a user study in which experienced annotators were asked to segment images. In the user study, ScribblePrompt reduced annotation time by 28% while increasing Dice by 15% compared to the next best method, the Segment Anything Model (SAM) [47]. Given similar amounts of interaction, ScribblePrompt

achieved consistently better Dice scores than other methods. We release an interactive tool, code, model weights, and our dataset of manually-collected scribbles at https://scribbleprompt.csail.mit.edu.

2 Related Works

Interactive Biomedical Image Segmentation. Early research into interactive segmentation of biomedical images focused on traditional intensity-based methods [13, 21, 33, 107, 118]. Recent deep-learning based techniques [68, 83, 109, 110] use human interaction to improve the accuracy of segmentation tasks where labelled data exists, but fully-automatic segmentation methods fail to produce accurate segmentations. This approach assumes that the model will be used to perform the same segmentation task(s) it was trained on, leading to domain-specific solutions [6, 62, 80, 89, 110, 111]. A few methods generalize to new, but similar, classes [95] and modalities [68, 109] to those seen in training. In contrast, ScribblePrompt is designed to help annotators segment a wide range of biomedical images and (potentially unseen) tasks at inference time.

Foundation Models. Recent vision foundation models employ prompting to enable generalization to new tasks. These models, trained on large collections of (natural) image data, segment potentially-unseen structures specified by spatial prompts [47, 129], text [47, 87, 129], or examples [15, 22, 112, 113]. Some of these models can be used for interactive segmentation with spatial prompts.

Fine-tuning models initially developed using natural images is often unhelpful for *specific* tasks in the biomedical imaging domain, compared to training from scratch [88]. The Segment Anything Model (SAM) [47], a foundation model trained on natural images, performs well at biomedical image segmentation tasks with clear boundaries (*e.g.* organs in abdominal CT), but performs poorly on tasks and modalities involving more subtle delineations (*e.g.* deep structures in Brain MRI) [36, 41, 76, 100]. Several recent or concurrent papers fine-tune SAM for specific biomedical tasks or modalities [40, 46, 59, 81, 108, 120]. A few fine-tune SAM to segment medical images from multiple-modalities with bounding boxes [69], clicks [115] or both [19]. We show in our experiments that these models don't perform well on many *unseen* tasks and are limiting in their required interactions.

Several natural image methods simulate *iterative* user interactions and condition the model on previous predictions to train a single network to make the initial prediction and perform refinement [14, 27, 47, 61, 102, 103]. This approach is appropriate for natural images where there is ample data and typically one modality. However, such methods have not previously been developed for the biomedical imaging domain, which is more fragmented because of specialized regions of interest and different modalities. We use iterative ideas in ScribblePrompt, but we facilitate iterative interactive segmentation of unseen biomedical imaging tasks.

User Interaction. Users can prompt interactive segmentation systems in many different ways, such as bounding boxes [9, 47, 89, 94, 109, 117], scribbles

[18, 110, 111], or clicks [9, 61, 68, 103, 110]. Few interactive segmentation models incorporate multiple types of inputs. Some works have explored training with more specialized types of clicks [122], such as extreme points [72, 93], interior margin points [68], and center clicks [103, 116], to maximize the information per interaction. Specialized interactions often require more user time, and models trained only with such interactions are less robust to deviations from the interaction protocol [9, 103, 117]. In contrast, we focus on a user-friendly model that performs well under a variety of interaction scenarios.

Scribbles are an intuitive form of interaction, however few interactive segmentation works employ them, because acquiring *realistic* scribbles for training and evaluation is challenging. Previous deep-learning models trained for interactive scribble-based segmentation either relied on collecting large datasets of manually drawn scribbles [53, 57, 128] or simplistic approaches such as sampling random points [5, 7, 110] or Bezier curves [2]. One work has explored simulating more complex scribbles for interactive segmentation of natural images [18]. We build on these concepts to create a new scribble simulation engine for training and evaluation.

Scribble-supervised learning methods use scribble annotations as *supervision* to train automatic segmentation models for predicting segmentation given only an input image [32, 56, 57, 66, 121]. However, these methods require the manual scribble-annotation of many training images and retraining for each new task. In contrast, ScribblePrompt is trained on a large corpus of datasets once and then can be used to perform new segmentation tasks at inference time without retraining, using scribbles as *input*.

3 ScribblePrompt Approach

We present an interactive segmentation method that can generalize to new biomedical imaging modalities and regions of interest, while staying focused on practical usability. We describe the problem formulation, and present the important aspects of the ScribblePrompt framework: (i) simulation of realistic interactions during training, (ii) augmentation with synthetic labels during training to encourage generalization, and (iii) an efficient architecture for fast inference. We show an overview of training in Fig. 2 and 4.

3.1 Problem Formulation

Let t be a segmentation task consisting of image and segmentation pairs, $\{(x^t, y^t)_j\}_{j=1}^N$. At step i, given an image x^t, a set of user interactions u_i, and previous prediction $\hat{y}_{i-1}^t$, we learn function $f_\theta(x^t, u_i, \hat{y}_{i-1}^t)$ with parameters θ that produces a segmentation $\hat{y}_i$. The set of interactions u_i, which may include positive or negative scribbles, positive or negative clicks, and bounding boxes, is provided by a user who has access to the image x^t and previous prediction $\hat{y}_{i-1}^t$.

We minimize the difference between the true segmentation y^t and each of the k iterative predictions $\hat{y}_1, \ldots, \hat{y}_k$,

$$\mathcal{L}(\theta; \mathcal{T}) = \mathbb{E}_{t \in \mathcal{T}} \left[\mathbb{E}_{(x^t, y^t) \in t} \left[\sum_{i=1}^{k} \mathcal{L}_{Seg} \left(y^t, f_\theta(x^t, u_i, \hat{y}^t_{i-1}) \right) \right] \right], \tag{1}$$

where $\mathcal{L}_{Seg}$ is a supervised segmentation loss.

During training, we sample a task t from training task collection $\mathcal{T}$, from which we sample an image x^t, and segmentation map y^t. We simulate a possible set of interactions u_i based on y^t, and predict $\hat{y}^t_i = f_\theta(x^t, u_i, \hat{y}^t_{i-1})$. We simulate the next set of interactions u_{i+1} based on the error region $\varepsilon^t_i = y^t - \hat{y}^t_i$, and repeat for k iterations. In the following sections, we describe the core aspects of the framework: strategies for simulating u_i, collecting $\mathcal{T}$, sampling (x^t, y^t), building f_θ, and optimizing $\mathcal{L}(\theta; \mathcal{T})$.

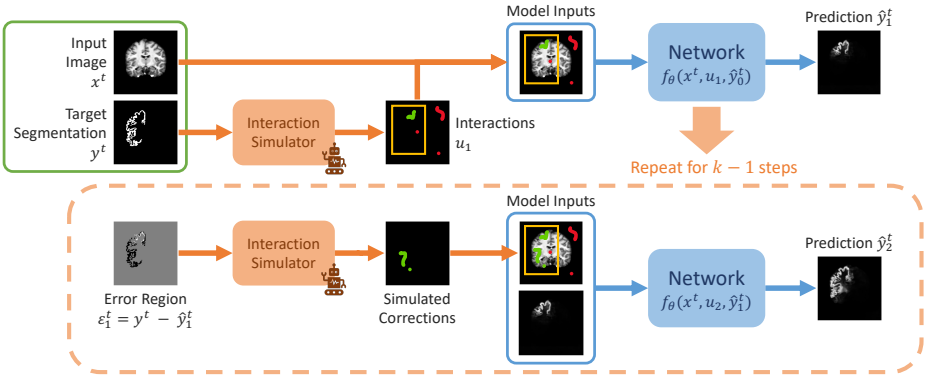


Fig. 2: Training. We simulate k consecutive steps of interactive segmentation. Given an image segmentation pair (x^t, y^t), we first simulate a set of initial interactions u_1, which may contain bounding boxes, clicks, and/or scribbles. We predict segmentation $\hat{y}^t_1 := f_\theta(x^t, u_1, \hat{y}^t_0)$ where the initial prediction $\hat{y}^t_0$ is set to zeros. In the second step, we simulate corrections using the error region ε^t_1 between the previous prediction $\hat{y}^t_1$ and ground truth y^t, and add them to the set of initial interactions u_1 to get u_2. We predict segmentation $\hat{y}^t_2 := f_\theta(x^t, u_2, \hat{y}^t_1)$ and repeat to produce a series of predictions, $\hat{y}^t_1, \ldots, \hat{y}^t_k$. We learn θ to minimize $\sum_{i=1}^{k} \mathcal{L}_{seg}(y^t, \hat{y}^t_i)$, the sum of losses between the target segmentation y^t and iterative predictions $\hat{y}^t_1, \ldots, \hat{y}^t_k$.

3.2 Prompt Simulation

To enable a practical, easy-to-use model, we encourage robustness to different types of user interactions u_i. We introduce algorithms for simulating scribbles, clicks, and bounding box inputs during training. Each scribble and click strategy (Fig. 3) can be applied to a ground truth segmentation label y^t to simulate positive interactions, or to the background $1 - y^t$ to simulate negative interactions.

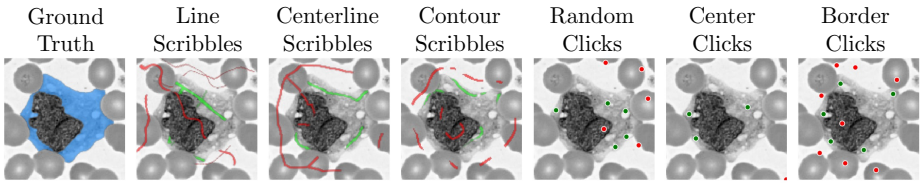


Fig. 3: Simulated scribbles and clicks. Positive interactions (green) are simulated on the segmentation label y^t (blue), while negative interactions (red) are simulated on the background $1 - y^t$. Scribble thickness is enlarged for visual clarity.

We simulate positive and negative correction scribbles or clicks by applying the same strategies to the false negative region $\varepsilon_i^t > 0$ and false positive region $\varepsilon_i^t < 0$ for error region ε_i^t.

Scribbles. Given binary mask $y \in \{0,1\}^{h \times w}$, we simulate a scribble mask $s \in [0,1]^{h \times w}$ by first generating clean scribbles, and then corrupting them to account for user behavior and variability. We illustrate this process in Supplementary Material Sec. B.1. To generate the clean scribbles, we start with one of these strategies:

(i) **Line Scribbles**: We draw random lines by connecting two end points sampled from $\{(u,v)|y_{uv} = 1\}$.
(ii) **Centerline Scribbles**: We simulate scribbles in the center of label y using a thinning algorithm [123] that reduces the label to a 1-pixel wide skeleton.
(iii) **Contour Scribbles**: We simulate a rough contour of the desired segmentation within the boundaries of the mask. We first blur the mask to reduce the size of the label such that $\tilde{y} = \min(y, y \circ G_k)$, where G_k is a Gaussian blur kernel. Then we apply a threshold $\tilde{y} < h$ sampled in some intensity range $h \sim U[\tilde{y}_{min}, \tilde{y}_{max}]$ and extract a contour inside the boundary of the mask.

To limit the size and complexity of the centerline and contour scribbles, we apply a random mask that breaks the scribbles into smaller parts. We generate the random mask by sampling a smooth noise image p, where each pixel is sampled independently from $\mathcal{N}(\mu_p, \sigma_p^2)$, applying Gaussian blur, and then thresholding it at μ_p. We warp the resulting scribble mask s using a random deformation field ϕ to vary the scribble shape and thickness. We ensure the resulting scribble is consistent with mask y, by multiplying the warped scribble mask $s \circ \phi$ by y.

Clicks. Given a binary mask $y \in \{0,1\}^{h \times w}$, we simulate $n \sim U[n_{min}, n_{max}]$ clicks at a time using one of three strategies, illustrated in Fig. 3:

(i) **Random clicks**: We randomly sample clicks from all pixels in the given region $\{(u,v)|y_{uv} = 1\}$.
(ii) **Center clicks**: We sample clicks from the set of points at the center of each disconnected component of the label. First, we create a multi-label mask $m \in \{1, \ldots, C\}^{h \times w}$ identifying the C components of label y. We identify the center of each components using the euclidean distance transform [103,116].

(iii) **Interior border region clicks**: We sample clicks from a border region inside the boundary of the mask. We first blur the mask y to reduce the size of the label such that $\tilde{y} = \min(y, y \circ G_k)$ where G_k is a Gaussian blur kernel. We then sample click coordinates from $\{(u, v) | \tilde{y}_{uv} \in [a, b]\}$, where $a, b \sim U[\tilde{y}_{min}, \tilde{y}_{max})$ are thresholds sampled in some intensity range.

Since users are inclined to spread out their clicks, we impose a minimum separation of a few pixels between random clicks and border region clicks.

Bounding Boxes. We compute the minimum bounding box that encloses the label y^t, and enlarge each dimension by $r \sim U[0, 20]$ pixels to account for human variability.

Iterative Training. During the first step ($i = 1$), we sample the combination of interactions and the number of initial positive and negative interactions $n_{pos}, n_{neg} \sim U[n_{min}, n_{max}]$. The initial interactions u_1 are simulated using the ground truth label y^t. In subsequent steps, correction scribbles or clicks are sampled from the error region ε^t_{i-1} between the last prediction $\hat{y}^t_{i-1}$ and the ground truth y^t. Since a user can make multiple corrections in each step, we sample $n_{cor} \sim U[n_{min}, n_{max}]$ corrections (scribbles or clicks) per step.

3.3 Data

We build on large dataset gathering efforts like MegaMedical [15,90] to compile a collection of 77 open-access biomedical imaging datasets for training and evaluation, covering over 54k scans, 16 image types, and 711 labels. The collection includes a diverse array of biomedical domains, such as eyes [39,55,70,85,105], thorax [86,96,98,99,101], spine [64,98,114,126], cells [16,17,28,29,63,127], skin [20], abdominal [11,12,37,44,45,51,52,54,60,67,71,86,91,101,104], neck [48,79,82,84], brain [3, 8, 31, 38, 49, 50, 73–75, 101], bones [34, 98, 114], teeth [1, 65] and lesions [4, 101, 124, 125].

We define a 2D segmentation task as a combination of dataset, axis (for 3D modalities), and label. For datasets with multiple segmentation labels, we consider each label separately as a binary segmentation task and for 3D modalities we use the slice with maximum label area and the middle slice from each volume. We provide more details on the data in Supplementary Material Sec. C.

Task Diversity. During training we sample image, segmentation pairs hierarchically – by dataset and modality, axis, and then label – to balance training on datasets of different sizes. To increase the diversity of segmentation tasks, we apply data augmentation [15] to both the input image and sampled segmentation prior to simulating the user interactions.

Synthetic Labels. To limit overfitting to specific segmentation tasks, we introduce a mechanism to generate synthetic labels (Fig. 4). During training, for a given sample (x_0, y_0), with probability p_{synth}, we replace y with a synthetic label y_{synth}. Given an image x_0, we generate a synthetic label y_{synth} by applying a superpixel algorithm [25] with scale parameter $\lambda \sim U[1, \lambda_{max}]$ to partition the image into a multi-label mask of k superpixels $z \in \{1, \ldots, k\}^{n \times n}$, and then randomly select a superpixel $y_{synth} = \mathbb{1}(z = c)$. We conduct experiments varying p_{synth} and show examples in Supplementary Material Sec. B.3.

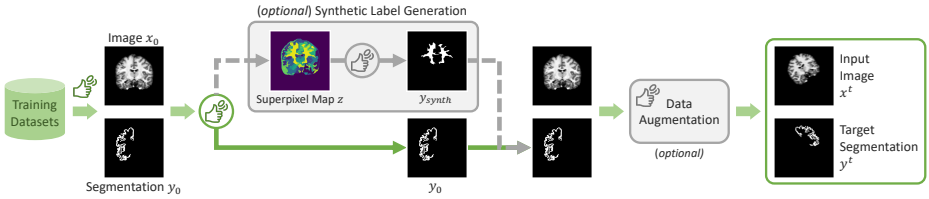


Fig. 4: Task sampling and augmentation. During training, we sample an image and segmentation pair (x_0, y_0). With probability p_{synth}, we replace y_0 with a synthetic label y_{synth}. We generate y_{synth} by applying a superpixel algorithm to the image x_0 to generate a map z of potential synthetic labels (superpixels), and then sampling one label. Finally, we apply random data augmentations to get (x^t, y^t).

3.4 Network

Motivated by producing a practical tool, we primarily demonstrate ScribblePrompt using an efficient fully-convolutional architecture similar to a UNet [92]. We also demonstrate ScribblePrompt using a vision transformer architecture [47].

ScribblePrompt-UNet: We use an 8 layer CNN following a decoder-encoder structure similar to the popular UNet architecture [92] without Batch Norm. Each convolutional layer has 192 features and uses PReLu activation [35].

ScribblePrompt-SAM: We take the smallest SAM model (ViT-b) [47] and fine-tune the decoder. SAM takes bounding box and click inputs as lists of (x, y) coordinates, and the low-resolution logits of the previous prediction $\hat{y}_{i-1}$. To adapt SAM for scribbles, we consider each scribbled pixel as a click.

Loss. We minimize eq. (1) where $\mathcal{L}_{Seg}$ is a linear combination of soft Dice Loss [24] and Focal Loss [58]. In preliminary experiments (Supplementary Material Sec. B.2), training with Soft Dice Loss alone or a linear combination of Soft Dice Loss and Binary-Cross Entropy Loss resulted in slightly lower Dice scores.

4 Experimental Setup

We compare ScribblePrompt-UNet and ScribblePrompt-SAM to previous methods through experiments with manual scribbles, simulated interactions, and a user study with experienced annotators. Lastly, we report on inference runtime and ablation experiments.

Data. We use 65 (out of 77) datasets during training. We partition another nine datasets, ACDC [10], BUID [4], BTCV Cervix [52], DRIVE [105], HipXRay [34], PanDental [1], SCD [86], SpineWeb [126], and WBC [127], each into validation (used in model selection, but not training) and test (used only for final evaluation). Three additional datasets, TotalSegmentator [114], SCR [30], and COBRE [3, 23, 26], were used only for final evaluation.

We report final results on 12 sets of data (the test splits of the 9+3 evaluation datasets). The evaluation datasets were not seen during training, and cover 608 tasks and 8 modalities, including unseen image types and unseen labels. We

selected these 12 evaluation datasets to cover a variety of modalities (MRI, CT, ultrasound, fundus photography, microscopy) and anatomical regions of interest (brain, teeth, bones, abdominal organs, muscles, heart, thorax, cells), including both healthy anatomy and lesions.

Training. During training, we simulated five steps of interactive segmentation per example and set the maximum number of interactions per step to three. We set the minimum number of initial negative prompts to zero, and the minimum initial positive and correction prompts to one. During the first step, we sample the combination of prompt types (clicks, scribbles, bounding boxes) and then sample the number of positive and negative interactions for each type. In each subsequent steps, we simulate either scribble corrections or click corrections with equal probability. We selected these values based on what we believe to be reasonable interactions for a user to perform.

Baselines. We compare to existing generalist methods for interactive segmentation, with a focus on methods developed for biomedical imaging.

SAM [47]: We evaluate the smallest (ViT-b) and largest (ViT-h) versions of the Segment Anything Model (SAM) trained on natural images. SAM takes bounding boxes, clicks, and the logits of the previous prediction as input.

SAM-Med2D [19]: SAM-Med2D is a SAM ViT-b model with additional adapter layers in the image encoder. SAM-Med2D was fine-tuned, using bounding boxes and iterative clicks as input, on a collection of biomedical imaging datasets containing 4.6M images and 19.7M segmentations. Following [19], we evaluate SAM-Med2D both with and without adapter layers.

MedSAM [69]: MedSAM is a SAM ViT-B model fine-tuned with bounding box prompts on a collection of biomedical imaging datasets containing 1.5M image segmentation pairs. We evaluate MedSAM with bounding boxes only, because we found it to perform poorly when given point or mask prompts.

MIDeepSeg [68]: MIDeepSeg is an interactive segmentation framework designed to generalize to unseen tasks on medical images. MIDeepSeg takes interior margin points (positive clicks) as initial inputs, crops the image based on those points, and uses a CNN to make an initial prediction. Given additional clicks, the prediction can be refined using conditional random fields. We evaluate the pre-trained model, which was developed on placenta T2 MRI.

5 Evaluation

We evaluate all methods using both manual and simulated interactions, with a focus on scribbles. Simulated interactions enable us to test on a large volume of images and tasks. However, simulations do not always match user behavior. We address this by (1) collecting a diverse dataset of manual scribbles for evaluation, and (2) conducting an interactive user study.

5.1 Manual Scribbles

Setup. We evaluate on two datasets of manual scribbles. First, we collected a dataset (MedScribble) of manual scribbles from three annotators for seven

segmentation tasks from unseen datasets, with a total of 31 examples. The annotators were shown five training examples with the ground truth segmentation per task and instructed to draw positive and negative scribbles to indicate the region of interest on new images. We provide more details on MedScribble and release the data in Supplementary Material Sec. E.1. Second, we used 380 slices from the ACDC dataset [10], which has scribble annotations for three labels and background.

In this evaluation, we make a prediction given a set of positive and negative scribbles, simulating a non-iterative scenario where the annotator draws several scribbles before running inference. We evaluate accuracy using Dice score [24] and 95th percentile Hausdorff Distance [42]. Since MIDeepSeg's initial prediction network only takes positive inputs, we report results on both MIDeepseg's initial prediction from the positive scribbles, and after applying MIDeepSeg's refinement procedure using the negative scribbles as corrections. For SAM and variants fine-tuning SAM on clicks, we convert each scribble to a series of points. For MedSAM, we fit a bounding box to the positive scribbles.

Results. ScribblePrompt-UNet and ScribblePrompt-SAM produce the most accurate segmentations in a single step of manual scribbles on both our manual scribbles dataset and the ACDC scribbles dataset (Tab. 1). We show example predictions in Fig. 5 and additional comparisons to scribble-supervised learning in Supplementary Material Sec. E.3.

SAM and SAM-Med2D do not generalize well to scribble inputs (which they were not trained for). MedSAM has better predictions than other SAM baselines using the SAM architecture, however it is not able to make use of the negative scribbles and thus often misses segmentations with holes in them (Fig. 5). The initial predictions from MIDeepSeg's network are poor, but improve after applying the refinement procedure.

Table 1: Manual scribbles. Mean Dice and HD95 with 95% CI of segmentations predicted from manually-collected scribbles. Best and second best are highlighted.

	MedScribble (n=93)		ACDC (n=1,140)	
Model	↑ Dice Score	↓ HD95	↑ Dice Score	↓ HD95
SAM (ViT-b)	0.40 ± 0.05	32.58 ± 3.80	0.20 ± 0.01	108.39 ± 0.85
SAM (ViT-h)	0.56 ± 0.05	14.61 ± 3.18	0.42 ± 0.02	59.78 ± 2.02
SAM-Med2D w.o. adapter	0.52 ± 0.05	30.34 ± 3.29	0.16 ± 0.01	107.25 ± 1.12
SAM-Med2D w/ adapter	0.30 ± 0.06	19.04 ± 3.63	0.17 ± 0.02	41.01 ± 3.04
MIDeepSeg	0.68 ± 0.04	6.94 ± 1.06	0.58 ± 0.01	5.78 ± 0.26
MIDeepSeg w/ refinement	0.81 ± 0.03	3.10 ± 0.67	0.73 ± 0.01	2.71 ± 0.18
MedSAM (box)	0.70 ± 0.04	7.54 ± 1.35	0.70 ± 0.02	3.53 ± 0.28
ScribblePrompt-UNet	0.84 ± 0.02	2.92 ± 0.88	0.84 ± 0.01	1.80 ± 0.11
ScribblePrompt-SAM	0.87 ± 0.02	2.61 ± 0.84	0.77 ± 0.01	4.17 ± 0.32

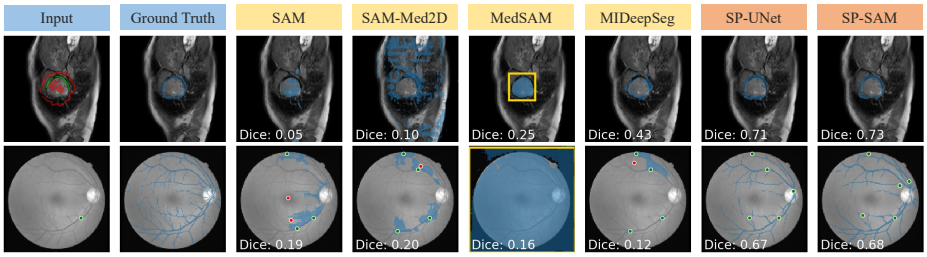


Fig. 5: Example predictions. SP = ScribblePrompt. Top: predictions after one step of manual scribbles. Bottom: predictions after five steps of simulated interactions (one center click followed by one center correction click per step).

5.2 Simulated Interactions

Setup. We use our interaction simulator to evaluate the performance of ScribblePrompt and baselines with iterative scribbles across the 12 evaluation sets.

For scribble-focused prompting procedures, we use:

- **Line Scribbles**: Three positive line scribbles and three negative line scribbles to start, followed by one correction line scribble at each step.
- **Centerline Scribbles**: One positive and one negative centerline scribble to start, followed by one correction centerline scribble at each step.
- **Contour Scribbles**: One positive and one negative contour scribble to start, followed by one correction contour scribble at each step.

We limit each scribble to cover a maximum of w pixels, where w is the maximum dimension of the image. For centerline and contour scribbles, a scribble might contain multiple disconnected components.

For click-focused prompting procedures, we use:

- **Center Clicks**: One positive click in the center of the largest component to start, followed by one (positive or negative) correction click per step in the center of the largest component of the error region.
- **Random Clicks**: One random positive click to start, followed by one (positive or negative) correction click per step, randomly sampled from the error region.
- **Random Warm Start**: Three random positive clicks and three random negative clicks to start, followed by one (positive or negative) correction click per step in the center of the largest component of the error region.

For each example and interaction procedure, we simulate a series of iterative interactions with five random seeds. Since MIDeepSeg cannot make accurate predictions with only a few clicks, we exclude it from simulations that start with a single click. For MedSAM, we show results for a bounding box prompt.

Results. Fig. 6 shows that both versions of ScribblePrompt outperform baseline methods for all simulated interaction procedures at all numbers of interactions.

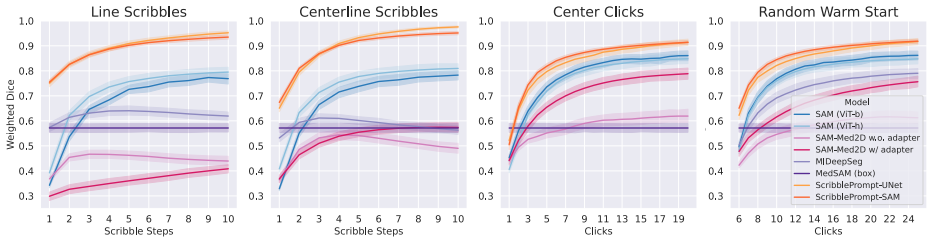


Fig. 6: Simulated clicks and scribbles. We simulate interactions following three scribble protocols and three click protocols. We show more results and example predictions, with similar trends, in Supplementary Material Sec. F.2.

We show examples with simulated interactions in Fig. 1 and Fig. 5 (bottom). We show results with similar trends using bounding boxes and by dataset with fully-supervised baselines in Supplementary Material Sec. F.

Discussion. Comparing the simulated scribble and click results (Fig. 6) highlights the efficiency of scribble-based interactions. Both ScribblePrompt models reach an average Dice above 0.8 in two scribble steps. It takes five steps of carefully-placed center clicks or eight random clicks to reach the same average Dice score with ScribblePrompt-SAM.

Although SAM-Med2D was trained on a large biomedical imaging collection, it does not perform as well as SAM at refining its predictions on held-out evaluation sets. MedSAM only uses bounding box prompts and is not able to refine predictions, limiting its performance and usability.

5.3 User Study

To assess the practical utility of ScribblePrompt, we conducted a user study with experienced annotators. We compare ScribblePrompt-UNet to SAM (ViT-b). We selected ScribblePrompt-UNet for the user study because it had similar performance and lower latency compared to ScribblePrompt-SAM. We selected SAM (ViT-b) because it had the highest Dice score in our experiments with clicks and lower latency than the next closest baseline, SAM (ViT-h).

Setup. Study participants ($n = 16$) were neuroimaging researchers at an academic hospital. They used each model to segment a series of nine test images from different evaluation datasets. For each segmentation task, participants were shown the target segmentation and were asked to prompt the model until the predicted segmentation closely matched the target, or they could no longer improve the prediction. We provided participants with the target segmentation to disentangle the cognitive process of identifying the region of interest from prompting the model to achieve the desired segmentation. We provide additional details and visualizations in Supplementary Material Sec. G.

Results. Participants produced more accurate segmentations (0.84 vs. 0.73 Dice; $p = 0.001$ using a paired t-test) using ScribblePrompt (Tab. 2). Partici-

Table 2: User study. Mean ± std. for Dice score, HD95 and total time per segmentation. Median (and max) number of times the participants refreshed the prediction.

	Dice Score (↑)	HD95 (↓)	Sec./task (↓)	Iter./task (↓)
SAM (ViT-b)	0.73 ± 0.30	16.1 ± 37.1	131.7 ± 168.5	6 (47)
ScribblePrompt-UNet	0.84 ± 0.13	4.5 ± 9.1	94.2 ± 89.0	3 (19)

pants spent ≈ 1.5 minutes per segmentation on average using ScribblePrompt, compared to over 2 minutes per segmentation with SAM ($p = 0.02$ using a paired t-test). While using ScribblePrompt, participants updated the prediction fewer times before being satisfied with the segmentation.

Upon completion, 15 out of 16 participants reported they preferred using ScribblePrompt to SAM and one participant had no preference. All participants reported it was easier to achieve the target segmentation using ScribblePrompt compared to SAM. 93.8% of participants reported ScribblePrompt was better than SAM at refining its predictions in response to scribbles. 87.5% of participants preferred using ScribblePrompt over SAM for clicks.

Discussion. The most common factors that influenced participant preference for ScribblePrompt were 1) being able get accurate predictions from multiple types of prompts, including scribbles and 2) responsiveness to their corrections, which enabled fine-grained control over the next prediction. For some tasks, such as retinal vein segmentation, SAM was unable to make accurate predictions even with many corrections. As a result, there was more variability in Dice score and time per segmentation, for SAM than ScribblePrompt.

5.4 Inference Runtime

Setup. We evaluate computational efficiency by measuring inference time on a single CPU for one prediction with a scribble input covering 128 pixels. Performance on a CPU reflects practical utility. Because of a variety of barriers such as protected health information, users may not be able to send their data to a server for computation, and must rely on local, most often CPU-only, computing.

Results. On a single CPU, ScribblePrompt-UNet requires 0.27 ± 0.04 sec per prediction, enabling the model to be used even in low-resource environments. The next most accurate model (Fig. 6), SAM ViT-h, requires over 2 minutes per prediction on a CPU (130.79 ± 7.96 sec). SAM ViT-b takes 13.59 ± 0.77 seconds per prediction. We report results for all baselines and on GPU hardware with similar trends in Supplementary Material Sec. H.

6 Ablations

We conduct two ablations of important design decisions: (1) synthetic label inputs used during training, and (2) types of prompts simulated during training. We report results on the validation splits of nine datasets unseen during training.



(a) Probability of synthetic labels. **(b) Interactions during training.**

Fig. 7: Ablations. We report Dice after five steps of simulated interactions following three inference-time interaction procedures. Error bars show 95% CI. **(a)** shows mean change in Dice relative to ScribblePrompt-UNet trained without any synthetic labels.

6.1 Synthetic Labels

Fig. 7a shows the effect of varying the probability of sampling a synthetic label during training for ScribblePrompt-UNet. Training with both real and synthetic labels improves generalization to new datasets, compared to training with only real labels. Using $p_{synth} = 0.5$ results in the highest Dice. Training with *only* synthetic labels results in worse Dice scores. We show similar results for ScribblePrompt-SAM and other interactions in Supplementary Material Sec. I.1.

6.2 Prompt Types

Setup. We evaluate ScribblePrompt-UNet models trained with different combinations of prompts, compared to the complete ScribblePrompt-UNet:

- **ScribblePrompt-UNet (scribbles)** trained on boxes and scribbles.
- **ScribblePrompt-UNet (clicks)** trained on boxes and clicks.
- **ScribblePrompt-UNet (random clicks)** trained on boxes and random clicks.

Results. Fig. 7b shows ScribblePrompt-UNet trained with scribbles, clicks and bounding boxes predicts segmentations more accurately than do ablated versions of ScribblePrompt-UNet. We show results for other interaction procedures with similar trends in Supplementary Material Sec. I.2.

7 Conclusion

We present ScribblePrompt, a practical framework for interactive segmentation that enables users to segment diverse medical images with scribbles, clicks, and bounding boxes. We introduce methods for simulating realistic user interactions and generating synthetic labels. These methods enable us to train models that generalize to unseen segmentation tasks and datasets. ScribblePrompt is more accurate than existing baselines, and ScribblePrompt-UNet is computationally efficient, even on a CPU. Our user study shows that nearly all users prefer ScribblePrompt and achieve segmentations with 15% higher Dice with less effort than the next most accurate baseline. ScribblePrompt promises to significantly reduce the burden of manual segmentation in biomedical imaging.

Acknowledgements

This work was supported in part by funding from Quanta Computer Inc., the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard and the Wistron Corporation. Research reported in this paper was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R01EB033773. Much of the computation resources required for this research was performed on computational hardware generously provided by the Massachusetts Life Sciences Center.

References

1. Abdi, A.H., Kasaei, S., Mehdizadeh, M.: Automatic segmentation of mandible in panoramic x-ray. Journal of Medical Imaging **2**(4), 044003 (2015)
2. Agustsson, E., Uijlings, J.R., Ferrari, V.: Interactive Full Image Segmentation by Considering All Regions Jointly. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11614–11623. IEEE, Long Beach, CA, USA (Jun 2019). https://doi.org/10.1109/CVPR.2019.01189, https://ieeexplore.ieee.org/document/8953526/
3. Aine, C.J., Bockholt, H.J., Bustillo, J.R., Cañive, J.M., Caprihan, A., Gasparovic, C., Hanlon, F.M., Houck, J.M., Jung, R.E., Lauriello, J., Liu, J., Mayer, A.R., Perrone-Bizzozero, N.I., Posse, S., Stephen, J.M., Turner, J.A., Clark, V.P., Calhoun, V.D.: Multimodal Neuroimaging in Schizophrenia: Description and Dissemination. Neuroinformatics **15**(4), 343–364 (Oct 2017). https://doi.org/10.1007/s12021-017-9338-9, http://link.springer.com/10.1007/s12021-017-9338-9
4. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in Brief **28**, 104863 (2020). https://doi.org/https://doi.org/10.1016/j.dib.2019.104863, https://www.sciencedirect.com/science/article/pii/S2352340919312181
5. Asad, M., Fidon, L., Vercauteren, T.: ECONet: Efficient Convolutional Online Likelihood Network for Scribble-based Interactive Segmentation. In: International Conference on Medical Imaging with Deep Learning. pp. 35–47. PMLR (2022), http://arxiv.org/abs/2201.04584, arXiv:2201.04584 [cs, eess]
6. Atzeni, A., Peter, L., Robinson, E., Blackburn, E., Althonayan, J., Alexander, D.C., Iglesias, J.E.: Deep active learning for suggestive segmentation of biomedical image stacks via optimisation of Dice scores and traced boundary length. Medical Image Analysis **81** (2022)
7. Bai, J., Wu, X.: Error-Tolerant Scribbles Based Interactive Image Segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 392–399. IEEE, Columbus, OH, USA (Jun 2014). https://doi.org/10.1109/CVPR.2014.57, https://ieeexplore.ieee.org/document/6909451
8. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
9. Benenson, R., Popov, S., Ferrari, V.: Large-scale interactive object segmentation with human annotators. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11700–11709 (2019)

10. 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 **37**(11), 2514–2525 (2018)
11. Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056 (2019)
12. Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K.: Nci-isbi 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive **370**(6), 5 (2015)
13. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 1, pp. 105–112. IEEE Comput. Soc, Vancouver, BC, Canada (2001). https://doi.org/10.1109/ICCV.2001.937505, http://ieeexplore.ieee.org/document/937505/
14. Bredell, G., Tanner, C., Konukoglu, E.: Iterative interaction training for segmentation editing networks. In: Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9. pp. 363–370. Springer (2018)
15. Butoi*, V.I., Ortiz*, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Universeg: Universal medical image segmentation. In: ICCV (2023)
16. Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., Rohban, M., Singh, S., Carpenter, A.E.: Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nature Methods **16**(12), 1247–1253 (Dec 2019). https://doi.org/10.1038/s41592-019-0612-7, https://doi.org/10.1038/s41592-019-0612-7
17. Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., Hartenstein, V.: An integrated micro-and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. PLoS biology **8**(10), e1000502 (2010)
18. Chen, X., Cheung, Y.S.J., Lim, S.N., Zhao, H.: ScribbleSeg: Scribble-based Interactive Image Segmentation (Mar 2023), http://arxiv.org/abs/2303.11320, arXiv:2303.11320 [cs]
19. 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 (Aug 2023), http://arxiv.org/abs/2308.16184, arXiv:2308.16184 [cs]
20. Codella, N.C.F., Gutman, D.A., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N.K., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). CoRR **abs/1710.05006** (2017), http://arxiv.org/abs/1710.05006
21. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic Image Segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) Computer Vision – ECCV 2008, vol. 5302, pp. 99–112. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_9, http://link.springer.com/10.1007/978-3-540-88682-2_9, series Title: Lecture Notes in Computer Science

22. Czolbe, S., Dalca, A.V.: Neuralizer: General neuroimage analysis without retraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6217–6230 (2023)
23. Dalca, A.V., Guttag, J., Sabuncu, M.R.: Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9290–9299 (2018)
24. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
25. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International journal of computer vision **59**, 167–181 (2004)
26. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
27. Forte, M., Price, B., Cohen, S., Xu, N., Pitié, F.: Getting to 99% Accuracy in Interactive Segmentation (Mar 2020), http://arxiv.org/abs/2003.07932, arXiv:2003.07932 [cs]
28. Gamper, J., Koohbanani, N., Benes, K., Graham, S., Jahanifar, M., Khurram, S., Azam, A., Hewitt, K., Rajpoot, N.: Pannuke dataset extension, insights and baselines. arxiv. 2020 doi: 10.48550. ARXIV (2003)
29. Gerhard, S., Funke, J., Martel, J., Cardona, A., Fetter, R.: Segmented anisotropic ssTEM dataset of neural tissue (11 2013). https://doi.org/10.6084/m9.figshare.856713.v1, https://figshare.com/articles/dataset/Segmented_anisotropic_ssTEM_dataset_of_neural_tissue/856713
30. van Ginneken, B., Stegmann, M.B., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Medical Image Analysis **10**(1), 19–40 (2006). https://doi.org/https://doi.org/10.1016/j.media.2005.02.002, https://www.sciencedirect.com/science/article/pii/S1361841505000368
31. Gollub, R.L., Shoemaker, J.M., King, M.D., White, T., Ehrlich, S., Sponheim, S.R., Clark, V.P., Turner, J.A., Mueller, B.A., Magnotta, V., et al.: The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. Neuroinformatics **11**, 367–388 (2013)
32. Gotkowski, K., Lüth, C., Jäger, P.F., Ziegler, S., Krämer, L., Denner, S., Xiao, S., Disch, N., Maier-Hein, K.H., Isensee, F.: Embarrassingly simple scribble supervision for 3d medical segmentation. arXiv preprint arXiv:2403.12834 (2024)
33. Grady, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11), 1768–1783 (Nov 2006). https://doi.org/10.1109/TPAMI.2006.233, http://ieeexplore.ieee.org/document/1704833/
34. Gut, D.: X-ray images of the hip joints **1** (Jul 2021). https://doi.org/10.17632/zm6bxzhmfz.1, https://data.mendeley.com/datasets/zm6bxzhmfz/1, publisher: Mendeley Data
35. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)
36. He, S., Bao, R., Li, J., Stout, J., Bjornerud, A., Grant, P.E., Ou, Y.: Computer-vision benchmark segment-anything model (sam) in medical images: Accuracy in 12 datasets (2023)
37. 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. Medical Image Analysis p. 101821 (2020)

38. Hernandez Petzsche, M.R., 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 **9**(1), 762 (2022)
39. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical imaging **19**(3), 203–210 (2000)
40. Hu, X., Xu, X., Shi, Y.: How to efficiently adapt large segmentation model(sam) to medical images (2023)
41. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., Chi, H., Hu, X., Fan, D.P., Dong, F., Ni, D.: Segment anything model for medical images? (2023)
42. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Transactions on pattern analysis and machine intelligence **15**(9), 850–863 (1993)
43. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z, http://www.nature.com/articles/s41592-020-01008-z
44. Ji, Y., Bai, H., Yang, J., Ge, C., Zhu, Y., Zhang, R., Li, Z., Zhang, L., Ma, W., Wan, X., et al.: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. arXiv preprint arXiv:2206.08023 (2022)
45. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021). https://doi.org/https://doi.org/10.1016/j.media.2020.101950, https://www.sciencedirect.com/science/article/pii/S1361841520303145
46. Kim, S., Oh, H.J., Min, S., Jeong, W.K.: Evaluation and improvement of segment anything model for interactive histopathology image segmentation (2023)
47. 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: ICCV (2023)
48. Krönke, M., Eilers, C., Dimova, D., Köhler, M., Buschner, G., Schweiger, L., Konstantinidou, L., Makowski, M., Nagarajah, J., Navab, N., et al.: Tracked 3d ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry. Plos one **17**(7), e0268550 (2022)
49. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE transactions on medical imaging **38**(11), 2556–2568 (2019)
50. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Doria, V., Serag, A., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., et al.: A dynamic 4d probabilistic atlas of the developing brain. NeuroImage **54**(4), 2750–2763 (2011)
51. Lambert, Z., Petitjean, C., Dubray, B., Kuan, S.: Segthor: segmentation of thoracic organs at risk in ct images. In: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2020)

52. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault Workshop Challenge. vol. 5, p. 12 (2015)
53. Lee, H., Jeong, W.K.: Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23. pp. 14–23. Springer (2020)
54. Lemaître, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M., Meriaudeau, F.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: a review. Computers in biology and medicine **60**, 8–31 (2015)
55. Li, M., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q.: Ipn-v2 and octa-500: Methodology and dataset for retinal image segmentation. arXiv preprint arXiv:2012.07261 (2020)
56. Li, Z., Zheng, Y., Luo, X., Shan, D., Hong, Q.: Scribblevc: Scribble-supervised medical image segmentation with vision-class embedding. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3384–3393 (2023)
57. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3159–3167 (2016)
58. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
59. Lin, X., Xiang, Y., Zhang, L., Yang, X., Yan, Z., Yu, L.: Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation (2023)
60. Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical image analysis **18**(2), 359–373 (2014)
61. Liu, Q., Xu, Z., Bertasius, G., Niethammer, M.: Simpleclick: Interactive image segmentation with simple vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22290–22300 (2023)
62. Liu, Q., Xu, Z., Jiao, Y., Niethammer, M.: iSegFormer: Interactive Segmentation via Transformers with Application to 3D Knee MR Images. In: Medical Image Computing and Computer Assisted Intervention (MICCAI) (2022)
63. Ljosa, V., Sokolnicki, K.L., Carpenter, A.E.: Annotated high-throughput microscopy image sets for validation. Nature methods **9**(7), 637–637 (2012)
64. Löffler, M.T., Sekuboyina, A., Jacob, A., Grau, A.L., Scharr, A., El Husseini, M., Kallweit, M., Zimmer, C., Baum, T., Kirschke, J.S.: A vertebral segmentation dataset with fracture grading. Radiology: Artificial Intelligence **2**(4), e190138 (2020)
65. in the Loop, H.: Teeth segmentation dataset, `https://humansintheloop.org/resources/datasets/teeth-segmentation-dataset/`
66. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528–538. Springer (2022)

67. Luo, X., Liao, W., Xiao, J., Song, T., Zhang, X., Li, K., Wang, G., Zhang, S.: Word: Revisiting organs segmentation in the whole abdominal region. arXiv preprint arXiv:2111.02403 (2021)
68. Luo, X., Wang, G., Song, T., Zhang, J., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., Zhang, S.: MIDeepSeg: Minimally Interactive Segmentation of Unseen Objects from Medical Images Using Deep Learning. Medical Image Analysis **72**, 102102 (2021)
69. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 1–9 (2024)
70. Ma, Y., Hao, H., Xie, J., Fu, H., Zhang, J., Yang, J., Wang, Z., Liu, J., Zheng, Y., Zhao, Y.: Rose: a retinal oct-angiography vessel segmentation dataset and new model. IEEE Transactions on Medical Imaging **40**(3), 928–939 (2021). https://doi.org/10.1109/TMI.2020.3042802
71. Macdonald, J.A., Zhu, Z., Konkel, B., Mazurowski, M., Wiggins, W., Bashir, M.: Duke liver dataset (MRI) v2 (Apr 2023). https://doi.org/10.5281/zenodo.7774566, https://doi.org/10.5281/zenodo.7774566
72. Maninis, K.K., Caelles, S., Pont-Tuset, J., Van Gool, L.: Deep Extreme Cut: From Extreme Points to Object Segmentation (2018), arXiv:1711.09081
73. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
74. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al.: The parkinson progression marker initiative (ppmi). Progress in neurobiology **95**(4), 629–635 (2011)
75. Mazurowski, M.A., Clark, K., Czarnek, N.M., Shamsesfandabadi, P., Peters, K.B., Saha, A.: Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data. Journal of neuro-oncology **133**, 27–35 (2017)
76. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: An experimental study. Medical Image Analysis **89**, 102918 (Oct 2023). https://doi.org/10.1016/j.media.2023.102918, https://www.sciencedirect.com/science/article/pii/S1361841523001780
77. Men, J., Huang, Y., Solanki, J., Zeng, X., Alex, A., Jerwick, J., Zhang, Z., Tanzi, R.E., Li, A., Zhou, C.: Optical coherence tomography for brain imaging and developmental biology. IEEE Journal of Selected Topics in Quantum Electronics **22**(4), 120–132 (2015)
78. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. Fourth International Conference on 3D Vision (3DV) pp. 565–571 (2016)
79. Montoya, A., Hasnin, kaggle446, shirzad, Cukierski, W., yffud: Ultrasound nerve segmentation (2016), https://kaggle.com/competitions/ultrasound-nerve-segmentation
80. Ooi, A.Z.H., Embong, Z., Abd Hamid, A.I., Zainon, R., Wang, S.L., Ng, T.F., Hamzah, R.A., Teoh, S.S., Ibrahim, H.: Interactive Blood Vessel Segmentation from Retinal Fundus Image Based on Canny Edge Detector. Sensors **21**(19), 6380 (Sep 2021)

81. Paranjape, J.N., Nair, N.G., Sikder, S., Vedula, S.S., Patel, V.M.: Adaptivesam: Towards efficient tuning of sam for surgical scene segmentation. arXiv preprint arXiv:2308.03726 (2023)
82. Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., Romero, E.: An open access thyroid ultrasound image database. In: Romero, E., Lepore, N. (eds.) 10th international symposium on medical information processing and analysis. vol. 9287, p. 92870W. SPIE / International Society for Optics and Photonics (2015). https://doi.org/10.1117/12.2073532, https://doi.org/10.1117/12.2073532
83. Philbrick, K.A., Weston, A.D., Akkus, Z., Kline, T.L., Korfiatis, P., Sakinis, T., Kostandy, P., Boonrod, A., Zeinoddini, A., Takahashi, N., et al.: Ril-contour: a medical imaging dataset annotation tool for and with deep learning. Journal of digital imaging **32**, 571–581 (2019)
84. Podobnik, G., Strojan, P., Peterlin, P., Ibragimov, B., Vrtovec, T.: HaN-Seg: The head and neck organ-at-risk CT and MR segmentation dataset. Medical Physics **50**(3), 1917–1927 (2023). https://doi.org/https://doi.org/10.1002/mp.16197, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16197, tex.eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.16197
85. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid) (2018). https://doi.org/10.21227/H25W98, https://dx.doi.org/10.21227/H25W98
86. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G.: Evaluation framework for algorithms segmenting short axis cardiac mri. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge **49** (2009)
87. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
88. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems **32** (2019)
89. Rajchl, M., Lee, M.C.H., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M.A., Hajnal, J.V., Kainz, B., Rueckert, D.: DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks. IEEE Transactions on Medical Imaging **36**(2), 674–683 (2017)
90. Rakic, M., Wong, H.E., Ortiz, J.J.G., Cimini, B., Guttag, J.V., Dalca, A.V.: Tyche: Stochastic in-context learning for medical image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2024)
91. Rister, B., Yi, D., Shivakumar, K., Nobashi, T., Rubin, D.L.: CT-ORG, a new dataset for multiple organ segmentation in computed tomography. Scientific Data **7**(1), 381 (Nov 2020). https://doi.org/10.1038/s41597-020-00715-8, https://www.nature.com/articles/s41597-020-00715-8
92. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
93. Roth, H.R., Yang, D., Xu, Z., Wang, X., Xu, D.: Going to Extremes: Weakly Supervised Medical Image Segmentation. Machine Learning and Knowledge Ex-

traction **3**(2), 507–524 (Jun 2021). https://doi.org/10.3390/make3020026, https://www.mdpi.com/2504-4990/3/2/26
94. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" — Interactive Foreground Extraction using Iterated Graph Cuts. ACM Transactions on Graphics **23**, 309–314 (2004)
95. Sakinis, T., Milletari, F., Roth, H., Korfiatis, P., Kostandy, P., Philbrick, K., Akkus, Z., Xu, Z., Xu, D., Erickson, B.J.: Interactive segmentation of medical images through fully convolutional neural networks (2019), arXiv:1903.08205
96. Saporta, A., Gui, X., Agrawal, A., Pareek, A., Truong, S., Nguyen, C., Ngo, V.D., Seekins, J., Blankenberg, F.G., Ng, A., et al.: Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation. MedRxiv (2021)
97. Sati, P., George, I.C., Shea, C.D., Gaitán, M.I., Reich, D.S.: Flair*: a combined mr contrast technique for visualizing white matter lesions and parenchymal veins. Radiology **265**(3), 926–932 (2012)
98. Seibold, C., Reiß, S., Sarfraz, S., Fink, M.A., Mayer, V., Sellner, J., Kim, M.S., Maier-Hein, K.H., Kleesiek, J., Stiefelhagen, R.: Detailed annotations of chest x-rays via ct projection for report understanding. In: Proceedings of the 33th British Machine Vision Conference (BMVC) (2022)
99. Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. Medical image analysis **42**, 1–13 (2017)
100. Shi, P., Qiu, J., Abaxi, S.M.D., Wei, H., Lo, F.P.W., Yuan, W.: Generalist vision foundation models for medical imaging: A case study of segment anything model on zero-shot medical segmentation. Diagnostics **13**(11), 1947 (2023)
101. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
102. Sofiiuk, K., Petrov, I., Barinova, O., Konushin, A.: F-BRS: Rethinking Backpropagating Refinement for Interactive Segmentation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8620–8629. IEEE, Seattle, WA, USA (Jun 2020). https://doi.org/10.1109/CVPR42600.2020.00865, https://ieeexplore.ieee.org/document/9156403/
103. Sofiiuk, K., Petrov, I.A., Konushin, A.: Reviving Iterative Training with Mask Guidance for Interactive Segmentation (Feb 2021), http://arxiv.org/abs/2102.06583, arXiv:2102.06583 [cs]
104. Song, Y., Zheng, J., Lei, L., Ni, Z., Zhao, B., Hu, Y.: CT2US: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. Ultrasonics **122**, 106706 (2022). https://doi.org/https://doi.org/10.1016/j.ultras.2022.106706, https://www.sciencedirect.com/science/article/pii/S0041624X22000191
105. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging **23**(4), 501–509 (2004)
106. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In: 2022 IEEE/CVF Conference on Computer Vision

and Pattern Recognition (CVPR). pp. 20698–20708. IEEE, New Orleans, LA, USA (Jun 2022). https://doi.org/10.1109/CVPR52688.2022.02007, https://ieeexplore.ieee.org/document/9879123/
107. Vezhnevets, V., Konouchine, V.: "GrowCut" - Interactive Multi-Label N-D Image Segmentation By Cellular Automata. proc. of Graphicon **1**(4), 150–156 (2005)
108. Wang, C., Chen, X., Ning, H., Li, S.: Sam-octa: A fine-tuning strategy for applying foundation model to octa image segmentation tasks (2023)
109. Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. IEEE Transactions on Medical Imaging **37**(7), 1562–1573 (2018)
110. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(7), 1559–1572 (2019)
111. Wang, G., Zuluaga, M.A., Pratt, R., Aertsen, M., Doel, T., Klusmann, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S.: Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. Medical Image Analysis **34**, 137–147 (Dec 2016). https://doi.org/10.1016/j.media.2016.04.009
112. Wang, X., Wang, W., Cao, Y., Shen, C., Huang, T.: Images speak in images: A generalist painter for in-context visual learning. arXiv preprint arXiv:2212.02499 (2022)
113. Wang, X., Zhang, X., Cao, Y., Wang, W., Shen, C., Huang, T.: SegGPT: Segmenting Everything In Context (Apr 2023), http://arxiv.org/abs/2304.03284, arXiv:2304.03284 [cs]
114. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5) (2023)
115. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation (Apr 2023), http://arxiv.org/abs/2304.12620, arXiv:2304.12620 [cs]
116. Xu, N., Price, B., Cohen, S., Yang, J., Huang, T.: Deep Interactive Object Selection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 373–381. IEEE, Las Vegas, NV, USA (Jun 2016). https://doi.org/10.1109/CVPR.2016.47, http://ieeexplore.ieee.org/document/7780416/
117. Xu, N., Price, B., Cohen, S., Yang, J., Huang, T.: Deep GrabCut for Object Selection. arXiv (2017), arXiv:1707.00243
118. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)
119. Zhang, J., Ding, X., Hu, D., Jiang, Y.: Semantic segmentation of covid-19 lesions with a multiscale dilated convolutional network. Scientific Reports **12**(1), 1847 (2022)
120. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation (2023)
121. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11656–11665 (2022)

122. Zhang, S., Liew, J.H., Wei, Y., Wei, S., Zhao, Y.: Interactive Object Segmentation With Inside-Outside Guidance. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12231–12241. IEEE (2020)
123. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. Commun. ACM **27**(3), 236–239 (mar 1984). https://doi.org/10.1145/357994.358023, https://doi.org/10.1145/357994.358023
124. Zhang, Y., Xian, M., Cheng, H.D., Shareef, B., Ding, J., Xu, F., Huang, K., Zhang, B., Ning, C., Wang, Y.: Busis: A benchmark for breast ultrasound image segmentation. In: Healthcare. vol. 10, p. 729. MDPI (2022)
125. Zhao, Q., Lyu, S., Bai, W., Cai, L., Liu, B., Wu, M., Sang, X., Yang, M., Chen, L.: A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation. CoRR **abs/2207.06799** (2022)
126. Zheng, G., Chu, C., Belavỳ, D.L., Ibragimov, B., Korez, R., Vrtovec, T., Hutt, H., Everson, R., Meakin, J., Andrade, I.L., et al.: Evaluation and comparison of 3d intervertebral disc localization and segmentation methods for 3d t2 mr data: A grand challenge. Medical image analysis **35**, 327–344 (2017)
127. Zheng, X., Wang, Y., Wang, G., Liu, J.: Fast and robust segmentation of white blood cell images by self-supervised learning. Micron **107**, 55–71 (2018). https://doi.org/https://doi.org/10.1016/j.micron.2018.01.010, https://www.sciencedirect.com/science/article/pii/S0968432817303037
128. Zhou, T., Li, L., Bredell, G., Li, J., Unkelbach, J., Konukoglu, E.: Volumetric memory network for interactive medical image segmentation. Medical Image Analysis **83**, 102599 (Jan 2023). https://doi.org/10.1016/j.media.2022.102599, https://linkinghub.elsevier.com/retrieve/pii/S1361841522002316
129. Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Wang, J., Wang, L., Gao, J., Lee, Y.J.: Segment everything everywhere all at once. Advances in Neural Information Processing Systems **36** (2024)