

1. はじめに

医用画像や細胞画像のセグメンテーションは、生物学的に関連する形態学的情報を画素レベルで識別することができるため、正確な診断や治療計画のサポートに用いられる。一方で、細胞画像には多くの細胞が含まれているため、個々の細胞を正確に分離してセグメンテーションをすることは難しい。そこで、指定した領域ごとにセグメンテーションできる Segment Anything Model (SAM) [1] が注目されている。SAM は、モデルに与えた指示 (プロンプト) に基づいて前景と背景にセグメンテーションを行う基盤モデルである。例えば、プロンプトにバウンディングボックス (bbox) を使用すると、bbox で囲んだ領域をセグメンテーションすることができる。SAM は、様々な形状に対応することもできるため、細胞画像のセグメンテーションへの応用も可能である。一方、プロンプトを人が定義したり、物体検出結果を用いて与えると、サイズずれが生じることがある。SAM はプロンプトに基づいてセグメンテーションを行うため、プロンプトにサイズずれが生じた場合、セグメンテーションの精度が低下する。

そこで本研究では、SAM にサイズずれを許容する新たなトークンを追加する。さらに、繰り返し推論による最適化を導入し、SAM によるセグメンテーション精度の向上を図る。

2. 関連研究

セグメンテーションの基盤モデルである SAM は、Image encoder, Prompt encoder, Mask decoder から構成される。SAM には、プロンプトチューニング用の学習パラメータ (Learned embedding) が存在しており、Prompt encoder の Positional encoding を行った埋め込みベクトルに加算する。SAM は学習の際に、入力プロンプトに対して、最大 20pixel のランダムなサイズずれを加えているため、0~20pixel 程度のサイズずれであれば許容できる。

Mazurowski らは、SAM のプロンプトとセグメンテーション能力に関する傾向を明らかにした [4]。主な傾向は以下の通りである。1 つ目は、SAM のセグメンテーションの精度はデータセットごとに差がある。2 つ目は、プロンプトが正確である場合は適切なセグメンテーションが得られる。3 つ目は、bbox をプロンプトとした場合、point をプロンプトにした場合よりも高精度なセグメンテーション結果が得られる。4 つ目は、複数の point をプロンプトに利用した場合、セグメンテーションの精度がわずかに向上する。これらの傾向から、SAM は与えるプロンプトが重要であることが分かる。

3. 提案手法

本研究では、サイズずれが生じた bbox に対して適切なセグメンテーションを出力するために、Prompt encoder に新たなトークンを導入したプロンプトチューニングと、繰り返し推論を提案する。

3.1 プロンプトチューニング

プロンプトチューニングは、入力データに学習可能なパラメータを追加し、学習する手法である [5]。プロンプトチューニングでは、プロンプトチューニング用のトークンをモデルに追加するのが一般的である。そこで、提案手法ではサイズずれを考慮するための学習パラメータを新たにトークンとして SAM の Prompt encoder に追加する。追加したトークンを用いたプロンプトチューニングを図 1 に示す。また、このトークンを全結合層により 128 次元から 256 次元に変換することで、トークンはより複雑な表現が可能になる。これにより、サイズずれに対する頑健性に獲得が可能になる。学習時は、60% の確率で 0%、20% の確率で $\pm 10\%$ 、 $\pm 20\%$ のサイズずれした bbox をプロンプトとして入力する。学習時にサイズずれした bbox を学習することで、推論時にサイズずれした bbox への頑健性が向

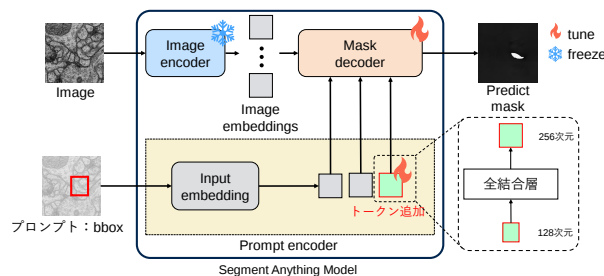


図 1: トークンの追加とプロンプトチューニング

上する。

3.2 繰り返し推論

推論時にサイズずれした不正確な bbox のプロンプトを用いた場合、1 度の推論で良いセグメンテーション結果を出力することは難しい。そこで推論時に、繰り返し推論を行う。繰り返し推論を行うことで、サイズずれした bbox を修正することができる。また、Prompt encoder は繰り返し推論の回数分の入力をする必要があるが、パラメータ数の多い Image encoder への入力は一度の入力でよいため、繰り返し推論の時間が大きく増えない。繰り返し推論の流れを図 2 に示す。繰り返し推論は以下のステップで行う。

- step1** 入力画像とプロンプトを用いて推論を行い、セグメンテーションマスク M_{t-1} を求める。プロンプトにはランダムなサイズずれを含む bbox を用いる。
- step2** 推論した M_{t-1} から外接矩形を bbox として設定する。
- step3** 入力画像と設定した bbox を用いて再度推論を行う。ここで、推論した Predict mask を M_t とする。
- step4** 推論した M_t と 1 回目推論した M_{t-1} の変化率を求める。ここで、 $H_b W_b$ は bbox の面積、 HW は画像サイズ、 τ はしきい値である。

$$\text{変化率} = \frac{1}{H_b W_b} \sum |M_t - M_{t-1}| \quad (1)$$

- step5** 変化率に対して、しきい値より上の場合、step2 に戻る。しきい値以下の場合、Predict mask を最終的なセグメンテーション結果として出力する。

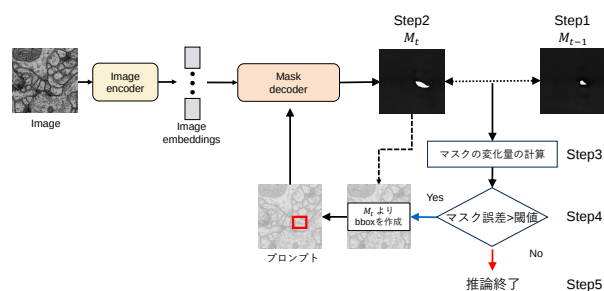


図 2: 繰り返し推論

4. 評価実験

本実験では、電子顕微鏡細胞画像を用いた細胞のセグメンテーションのデータセットである電子顕微鏡データセット、ISBI [2]、Electron Microscopy Dataset [3] で評価実験を行う。これらのデータセットは、bbox が用意されていないため、各細胞の Ground truth のセグメンテーションをもとに作成した外接矩形をプロンプトとして使用する。ベースラインは、SAM をファインチューニングしたモデルとする。SAM のファインチューニングでは、Mask decoder のみを学習する。プロンプトチューニングは、Prompt encoder の

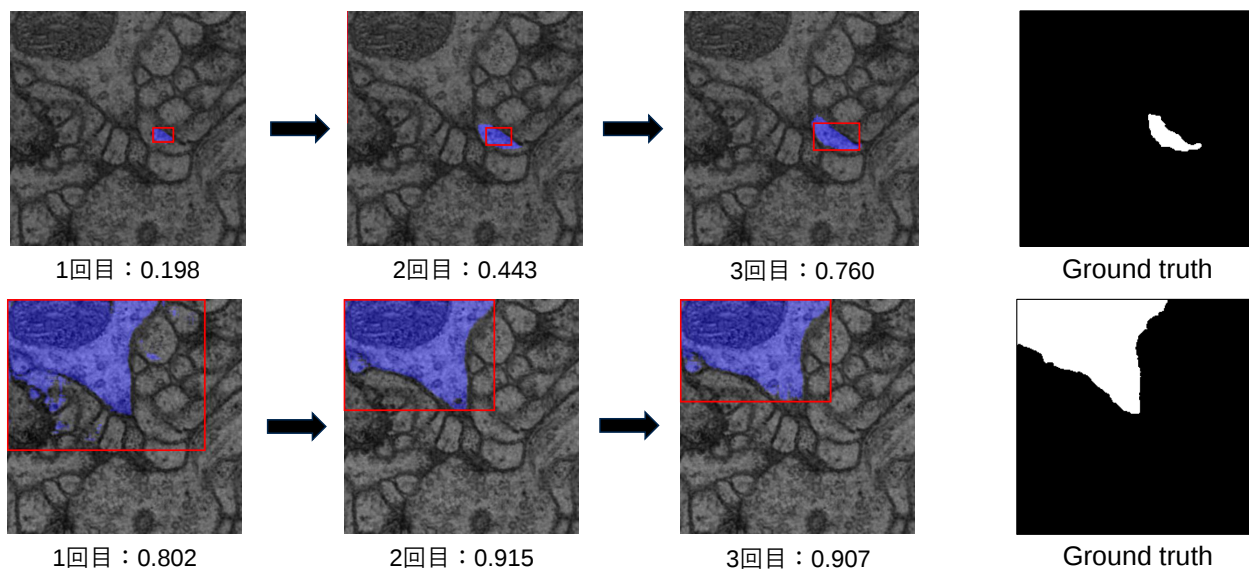


図 3：ISBI データセットの繰り返し推論による Predict mask と mIoU

表 1：電子顕微鏡データセットの実験結果

手法	P	K	bbox のサイズずれ				
			0%	10%	20%	30%	40%
SAM			0.827	0.822	0.788	0.710	0.595
	✓		0.827	0.822	0.787	0.709	0.594
提案手法	✓		0.828	0.823	0.788	0.714	0.600
	✓	✓	0.828	0.823	0.787	0.734	0.653
繰り返し推論回数の平均			1	1	1.02	2.42	3.33

表 2：ISBI データセットの実験結果

手法	P	K	bbox のサイズずれ				
			0%	10%	20%	30%	40%
SAM			0.822	0.810	0.774	0.699	0.604
	✓		0.825	0.810	0.777	0.702	0.606
提案手法	✓		0.830	0.820	0.785	0.714	0.622
	✓	✓	0.830	0.817	0.784	0.727	0.650
繰り返し推論回数の平均			1	1.01	1.05	2.13	3.16

Learned embedding を学習に追加したモデルである。提案手法のプロンプトチューニングは、Learned embedding のパラメータを固定し、代わりに追加したトークンと Mask decoder を学習する。また、通常のプロンプトチューニングは、既存のモデルのパラメータを凍結させる。しかし、SAM は Mask decoder の学習を行うことで精度が向上することが分かっており、Mask decoder は軽量であるため同時に学習を行う。サイズずれに対する評価では、0%から40%のサイズずれした bbox をプロンプトとして入力した場合の Predict mask と Ground truth 間の mIoU を比較する。繰り返し推論ではしきい値を 0.05 とする。

4.1 実験結果

電子顕微鏡データセットの実験結果を表 1, ISBI データセットでの実験結果を表 2, Electron Microscopy Dataset での実験結果を表 3 に示す。ここで、各表の P はプロンプトチューニング、 K は繰り返し推論を示す。提案手法のプロンプトチューニングのみを導入した場合、全てのデータセットにおいてサイズずれを 30% から 40% 加えた場合に SAM を超える認識性能を達成した。どのデータセットにおいてもサイズずれを 30% 以上加えた場合で SAM を超える認識性能を達成していることから、プロンプトチューニング時に学習していないサイズずれ度合いに対して頑健であると言える。繰り返し推論では、10% から 20% で精度が若干低下した。しかし、学習を行っていない範囲で大きく精度が向上した。プロンプトチューニングに加えて繰り返し推論を導入することで、さらに認識性能が改善することから、繰り返し推論も有効であると言える。

4.2 繰り返し推論過程の可視化

繰り返し推論における 1 回目から 3 回目の Predict mask (青色領域)、Ground truth を図 3 に示す。推論を繰り返すことで Predict mask が Ground truth へ近づいていくことから、繰り返し推論が有効であると言える。細胞よりも小

表 3：Electron Microscopy Dataset の実験結果

手法	P	K	bbox のサイズずれ				
			0%	10%	20%	30%	40%
SAM			0.885	0.875	0.844	0.772	0.673
	✓		0.885	0.874	0.844	0.772	0.673
提案手法	✓		0.884	0.874	0.845	0.779	0.683
	✓	✓	0.882	0.867	0.855	0.824	0.762
繰り返し推論回数の平均			1.03	1.37	1.54	2.72	3.11

さい bbox の繰り返し推論では、bbox を適切な大きさに修正できたため精度向上した。オブジェクトよりも大きい bbox の繰り返し推論では、2 回目の繰り返し推論時に高いセグメンテーション結果が得られた。また、サイズずれが 30% の時の繰り返し推論回数の平均は 2.4 回、40% の場合は 3.2 回となり、サイズずれが大きいほど繰り返し推論回数が増加する結果となった。

5. おわりに

本研究では SAM におけるサイズずれのある不正確なプロンプトによる性能低下を改善するプロンプトチューニングと繰り返し推論を提案した。トークンのプロンプトチューニングはサイズずれした bbox に対して精度低下を低減した。また、繰り返し推論をすることにより、さらに精度低下を低減した。繰り返し推論を可視化した結果、bbox が最適化されていることを確認した。今後は、マスクから作成した bbox に対する損失設計によるプロンプトの最適化を行う。

参考文献

- [1] A. Kirillov, *et al.*, “Segment anything”, ICCV, 2023.
- [2] A. Cardona, *et al.*, “An integrated micro-and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy”, PLOS Biology, vol.8, no.10, pp.1-17, 2010.
- [3] A. Lucchi, *et al.*, “Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features”, IEEE Trans Med Imaging, vol.31, no.2, pp.474-486, 2012.
- [4] M. Mazurowski, *et al.*, “Segment Anything Model for Medical Image Analysis: an Experimental Study”, arXiv preprint arXiv:2304.10517, 2023.
- [5] B. Lester, *et al.*, “The Power of Scale for Parameter-Efficient Prompt Tuning”, ICLR, 2021.

研究業績

- [1] 船井祥吾, 平川翼, 山下隆義, 藤吉弘亘, “SAM のプロンプトチューニングと繰り返し推論による細胞画像セグメンテーションの高精度化”, 動的画像処理実利用化ワークショップ, 2024.