

敵対的生成モデルを用いた電子顕微鏡画像からの 神経細胞膜セグメンテーション

内橋 堅志 大羽 成征 石井 信
Kenshi Uchihashi Oba Shigeyuki Shin Ishii

京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

In order to reconstruct nano-scale 3D structure of neuronal cells in an animal brain from serial section electron microscopy images, the cell membrane segmentation of 2D slice images is considered as a crucial phase whose performance greatly influences the overall reconstruction accuracy. In this study, we propose a pix-to-pix map trained by a framework of conditional generative adversarial network learning and apply it to the cell membrane segmentation problem. When applied to a benchmark dataset, we achieved a good segmentation accuracy, comparable to the latest methods, with a smaller computational cost.

1. 序論

マイクロコネクティクスと呼ばれる新しい研究分野では、連続切片電子顕微鏡画像に基づいて多数の神経細胞の詳細な三次元形状および局所的な回路構造を再構成することによって、脳における神経細胞の構造、細胞間の局所結合や神経線維の配線構造を網羅的に調べ、それらの機能的意味を理解することを目的としている。このために、Harvard 大学や MIT の研究グループでは段階を踏んだ画像処理パイプラインを採用している [1]。これによれば、第一に、二次元画像に対するパターン認識によって画像上のピクセルを細胞膜上とそうでないものに分類する「細胞膜セグメンテーション」を行い、第二に、二次元画像上で細胞膜で囲まれるブロップと呼ばれる連続領域をまとめ、第三に、ブロップ同士を三次元的に繋ぐことによって神経細胞の三次元構造を決定する。これら三工程のうち、最初に行われる細胞膜セグメンテーションは、その精度が後の工程全てに影響を与えるためその性能向上は重要である。

電子顕微鏡画像に基づく細胞膜セグメンテーションの手法開発は近年盛んに進められており、Deep Contextual Network [2] や U-Net [3] などが人間のエキスパートに迫る精度を達成しているが、煩雑な前処理によるデータ拡張、豊富な計算資源を数日単位で用いた長時間学習と、パラメータチューニングを必要としている。一方で、近年のセグメンテーション精度の向上の中で、多少の精度の向上よりも専門家による誤り訂正のための手間の縮減の方が重要な課題となりつつあり、効率の良い学習法と、それによる簡便な解析手法のニーズが高まっている。

一方で、学習データの分布に関する情報を事前に与えず

連絡先：〒 606-8501 京都市左京区吉田本町 36-1 京都
大学大学院情報学研究科システム科学専攻 <ishii-
seccr@sys.i.kyoto-u.ac.jp>

に分布の形状を学習することができる敵対的生成モデル (generative adversarial network; GAN) [4] が登場し、これを用いて有用な生成モデルの構築が可能になった。この GAN の応用として、画像からエッジ抽出などの処理を施した劣化画像から元の画像を復元するなどが示されている。

本研究ではセグメンテーションの精度と計算コストを同時に改善するために、画像変換モデルの学習を対抗事例学習によって行う条件付き敵対的生成モデル (conditional generative adversarial network; cGAN) [5] と、顕微鏡画像に対する画像変換モデルで実績のある U-Net [3] を組み合わせた手法を提案する。提案手法の特長は、U-Net を用いることで、セグメントの同時分布と多スケールの情報を自然に考慮できている点と、cGAN の枠組みを用いることで少数データからでも効率の良い学習を行うことができる点である。

2. 手法

本研究では、cGAN の一種であって、その生成部 G を U-Net としたモデルにより細胞膜のセグメンテーションを行う。本章では、この提案手法 (図 2) の説明を行う。

2.1 画像変換モデルと U-Net

細胞膜セグメンテーション問題は、電子顕微鏡による二次元 (xy) のグレースケール画像を入力とし、各ピクセルが細胞膜に対応するか否かを確率的に表現したグレースケール画像 (ラベルマップ) を出力する画像変換を、学習データに基づき同定することである。

U-Net の構造は down-convolution 写像によって空間スケールを小さくしてゆく encode 部と、up-convolution 写像によって空間スケールを大きくしてオリジナルに近づけてゆく decode 部からなるが、U-Net では encode 部と decode 部とでスケールが対応する層の間に接続 (バイパス) が追加されている点が特徴である。この工夫により、異なる空間周

波数を持つ特徴量を用いたセグメンテーションを混在させ、互いに協調・競合的に学習させることができるため、比較的少ない教師データの量のもとでも精度の高い学習が可能になる (図 1)。

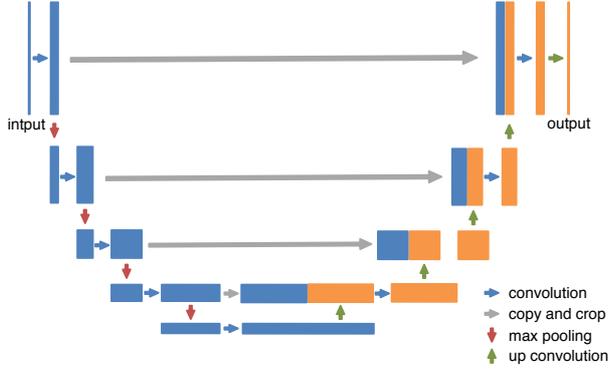


図 1: U-Net のネットワーク構造。encode 部と decode 部の層の数が同じであり、画像内の異なる空間周波数を持つ情報が異なるチャンネルを介して伝送、また、その様式が協調・競合的に学習される。U 字型のアーキテクチャから U-Net と呼ばれる。

2.2 対抗事例学習による条件付き敵対的生成モデル学習の定式化

GAN [4] は、生成モデルの効率的な学習法として近年注目されている枠組みである。生成モデル G と、識別モデル D を用意し、 G は観測データと見分けの付きづらい偽パターン x を生成するように、 D は偽パターンと観測パターンとを高精度で識別するように、互いに対抗させながら学習させることで、精度のよい生成モデル G を得ることができる。

条件付き敵対的生成モデル (cGAN) [5][6] は、条件付き生成モデルの学習に GAN の枠組みを適用するアイデアである。条件付き生成モデル $G(y|x)$ とは、入力画像 x を条件として出力画像 y を確率的に生成する過程のことである。cGAN ではこれを学習で求めるために、入出力ペア (x, y) が G が生成したものか、観測データに含まれるものかを、識別モデル D で高精度で識別するべく GAN 同様に G と D を対抗させながら学習する。

GAN の生成モデルは、入力雑音 z から出力画像 y へのマップ $G: z \rightarrow y$ の学習が目的であった。一方で、cGAN の生成モデルは、観測画像 x と入力雑音 z のペアから出力画像 y への画像変換 $G: \{x, z\} \rightarrow y$ の学習を行う。

cGAN の目的関数は以下のように表される。

$$L_{\text{cGAN}}(G, D) = E_{x, y \sim p_{\text{data}}(x, y)} [\log D(x, y)] + E_{x \sim p_{\text{data}}(x), z \sim p_z(z)} [\log(1 - D(x, G(x, z)))] \quad (1)$$

これを用いて、最適な生成部 G^* は、

$$G^* = \arg \min_G \left[\max_D (L_{\text{cGAN}}(G, D)) \right] \quad (2)$$

と表される。生成部 G が上記の目的関数を最小化しようとする一方で、識別部 D は最大化しようとしているため、生成部 G にとって識別部 D が敵対的な存在となっている。また、先行研究では、cGAN の目的関数に正則化項を加えることの有効性が示されている [7] ことから、本研究では、目的関数に以下のような L1 正則化項を加えることで、過学習の抑制効果に期待する。

$$L_{L1}(G) = E_{x, y \sim p_{\text{data}}(x, y), z \sim p_z(z)} [\|y - G(x, z)\|_1] \quad (3)$$

以上をまとめると、提案手法における生成部 G の最適解 G^* は、

$$G^* = \arg \min_G \left[\max_D (L_{\text{cGAN}}(G, D)) + \lambda L_{L1}(G) \right] \quad (4)$$

で与えられる。

cGAN の生成部 G では、入力雑音 z により出力 y に揺らぎを与えることが通常であった [8]。しかし、最近の研究 [6] において、入力雑音は揺らぎ成分としてさほど効果的ではなく、代わりにいくつかの層でドロップアウトを行ったほうが良い結果が得られている。そこで本研究でも、実際には、入力雑音は用いずにドロップアウトを使用した。

2.3 モデルアーキテクチャ

提案手法の概略は図 2 のようになっている。まず、識別部は畳み込み (convolution) とマックスプーリングを行う層 4 層と全結合層 1 層からなる畳み込みニューラルネットワーク (CNN) である。畳み込みフィルターの大きさは全て 5×5 で、フィルター数は 1 層目から順に 64, 128, 256, 512 とした。活性化関数は ReLU を使用し、ADAM によるミニバッチ確率的勾配法によって最適化を行った。ADAM の 1 つ目のモーメント項の係数は 0.5、2 つ目のモーメント項の係数は 0.999、 ϵ の値は 10^{-8} 、学習率の初期値は 0.0002 とした。次に、生成部は、encode 部、decode 部がそれぞれ 8 層になっている合計 16 層の U-Net である。畳み込みフィルターの大きさは全て 5×5 で、encode 部のフィルター数は 1 層目から順に 64, 128, 256, 512, 512, 512, 512, 512 とし、decode 部のフィルター数はこれらを逆順に並べたものとした。decode 部の最初の 3 層ではドロップアウト率 0.5 とした。正則化項として追加している L1 ノルムの重み減衰パラメータ λ は $\lambda = 100$ とした。また、全ての畳み込み層ではゼロパディングを行うことによって画像の境界領域付近を補完し、入力画像パッチによって次元数に変化しないようにしている。また、バッチ正規化も併用している。

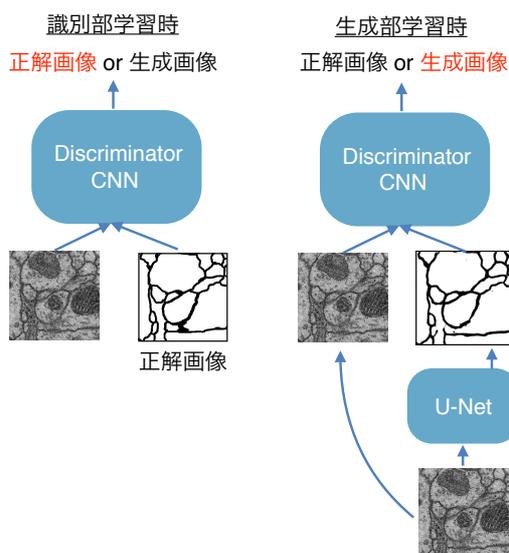


図2: 提案手法は識別部 D と生成部 G を持つ。識別部 D は、入力された2枚の画像が教師データ対（入力画像とエキスパートによるラベルマップ）であるかそうでないかを分類し、生成部 G は D を騙すデータを生成するように学習を行う。この時、通常のGANとは異なり、 G と D とで共通の画像を入力する。

3. 実験

3.1 データセット

評価実験には、ショウジョウバエの第1齢幼虫腹部の連続切片透過電子顕微鏡 (ssTEM) データセットを用いた。このデータセットは、 512×512 ピクセルの二次元 (xy) スライス画像を深さ (z) 方向に連続して30枚ぶんまとめたものであり、同サイズの真のラベル画像セットとして、専門家による細胞膜セグメンテーション結果が付けられている。使用した画像データ全体に対応する（生物学的）領域の大きさは、 $2 \times 2 \times 1.5 \mu\text{m}^3$ であり、xy平面上でのピクセルの大きさは $4 \times 4 \text{nm}$ 、z方向のスライス間ギャップは 50nm である。学習用データはISBI 2012のWebサイト*1上にて公開されている。

3.2 評価尺度

連続切片電子顕微鏡画像からの神経細胞構造再構成において、細胞膜セグメンテーションは、後に専門家がアルゴリズムの誤りを訂正することによって最終的な結果を得る半自動的システムの一工程である。そこで細胞膜セグメンテーション手法の性能評価は、専門家が誤り訂正を行う際に支払うコストの大きさに基づくものが望ましい。

専門家が誤り訂正を行う際、周囲の状況から考えて明らかにノイズと判定できるようなものは簡単に訂正できるが、例えば、本来存在しない細胞を検出しそれを全て細胞膜で

*1 ISBI Challenge: Segmentation of neuronal structures in EM stacks: http://brainiac2.mit.edu/isbi_challenge/

囲ってしまったようなケースは、その訂正は容易ではない。このような観点から、

- 単一の神経細胞が誤って2つ（以上）のセグメントに分割されている
- 2個（以上）の神経細胞が誤って1つのセグメントに結合されている

の二種の誤りを訂正の難しいものと定義した。さらに、これらの誤りに対して罰則を与えるような評価尺度として、本研究では、既存手法の精度評価でもしばしば用いられているランダムFスコア [9] と情報論的Fスコア [10] を用いた。

3.3 実験設定

512×512 のピクセルからなる二次元画像30枚のうち27枚から 256×256 ピクセルからなる画像パッチをxy方向に16ピクセルずつずらしながら取得し、合計7803枚の画像パッチの集合を学習用入力データとした。同様に、ラベル画像を構成し、学習用出力データとした。細胞膜セグメンテーションアルゴリズムは、残る3枚の二次元画像から同様にして取得した画像パッチ867枚を用いて評価した。

実装は全てPythonによって行い、特にモデルの構築はTensorFlowを用いた。また、実験は全てNVIDIA GeForce GTX Titan X GPU 4基の並列計算によって行った。

3.4 結果

ランダムFスコアと情報論的Fスコアを計算し、既存手法の中で代表的なものと比較した。結果を表1に示す。

表1: 既存手法の中で代表的なものとの精度比較結果。

Group name	Rand Score	Information Score
CUMedVision	0.976824580	0.988645822
Image Analysis Lab Freiburg	0.972760748	0.986616590
PyraMiD-LSTM	0.967669864	0.982914537
ours (10 epoch model)	0.934084673	0.977873920
ours (1 epoch model)	0.925437863	0.971674829
simple threshold	0.724521829	0.817598215

表1に示したように、ランダムFスコアにおいては既存手法に及ばないものの、情報論的Fスコアにおいては既存手法にやや劣る程度のスコアとなった。このスコアを算出するために要した時間は以下のとおり。CUMedVisionチームは、4つのモデルのアンサンブルを用いて、各モデルに対しそれぞれ3時間かけて学習を行った。また、Image Analysis Lab Freiburgチームは10時間ほどの学習時間を要したとしている。両者ともに、ガウシアンノイズの印加やマルチプルクロッシング、鏡像反転などに基づくデータ拡張と、同一構造の学習モデルを初期値などの条件を変更して複数個学習しそれらの出力の平均を取るアンサンブル学習などを行っているため、実際の処理の手間・計算時間は、表示されているものよりも大きい可能性がある。一方で、提案手

法は、これらの前処理や後処理を行うことなく、1 epoch の場合で 30 分弱、10 epoch の場合で 2 時間程度の学習時間でこの精度に到達している。

目視によって提案手法の性能を検討する。細胞膜セグメンテーションにおいて、ミトコンドリアやシナプス小胞などの小器官と細胞膜との区別が難しいことが知られている。図 3 上部に細胞が密な領域の画像に対するセグメンテーション結果を示す。元画像中にシナプス小胞が多数存在しディストラクタとなっている箇所において、提案手法によるセグメンテーション結果（ラベルマップ）では、それらが適切に除かれている。図 3 下部のシナプス小胞とミトコンドリアが混在している領域を見ると、シナプス小胞とミトコンドリアのいずれについても誤って細胞膜と認識するケースが少ないことが分かる。

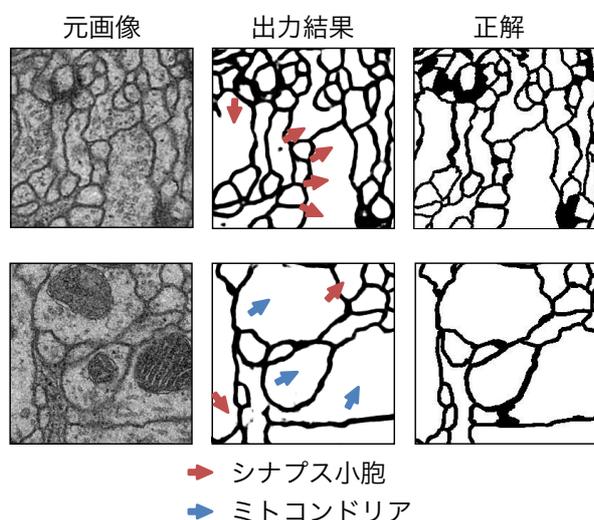


図 3: 提案手法による細胞膜セグメンテーション結果。左から、元の ssTEM 画像、提案手法によるセグメンテーション結果、専門家によるセグメンテーション結果（真のラベルマップ）を示している。赤い矢印と青い矢印は、それぞれシナプス小胞とミトコンドリアが見えている箇所を示している。シナプス小胞とミトコンドリアのいずれについても、誤って細胞膜と認識するケースが少ないことが分かる。

4. 結論

本研究では、連続切片電子顕微鏡画像に対する細胞膜セグメンテーション問題に対して、U-Net 型の深層ニューラルネットワーク構造と敵対的生成モデルに基づく深層学習の組み合わせにより、高効率と高精度を同時に実現する方法を提案した。提案手法を連続切片電子顕微鏡画像に対する細胞膜セグメンテーションのベンチマークデータに適用し、既存手法との性能を比較したところ、特に情報論的 F スコアを用いた評価のもとで最新の手法に匹敵する精度を、最新の手法によるよりも大幅に少ない計算時間の下で達成

できることが確認された。

参考文献

- [1] S. Knowles, M. Roberts, T. R. Jones, N. Kasthuri, E. Miller, J. Lichtman, V. Kaynig, A. Vazquez-Reina, and H. Pfister. Large-scale automatic reconstruction of neuronal process from electron microscopy images. 1303.7186, 2013.
- [2] C. Hao, Q. Xiaojuan, C. Jie-Zhi, and H. Pheng-Ann. Deep contextual networks for neuronal structure segmentation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16*, pages 1167–1173. AAAI Press, 2016.
- [3] T. Brox, O. Ronneberger, P. Fischer. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI), MICCAI'15*, pages 234–241. Springer, 2015.
- [4] Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Ian J Goodfellow, Jean Pouget-Abadie. Generative adversarial nets. pages 2672–2680. *Proceedings of advances in Neural Information Processing Systems (NIPS)*, 2014.
- [5] S. Osindero, M. Mirza. Conditional generative adversarial nets. volume 1411.1784. arXiv, 2014.
- [6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016.
- [7] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [8] Xiaolong Wang and Abhinav Gupta. Generative image modeling using style and structure adversarial networks. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2016.
- [9] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):929–944, 2007.
- [10] Pablo Arbelaez and Michael et al. Maire. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 2011.