U-Net: バイオメディカル画像セグメンテーションのための畳み込みネットワーク

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

概要。ディープネットワークの学習を成功させるには、数千の注釈付き学習サンプルが必要であることは、大きな同意を得ている。本論文では、利用可能な注釈付きサンプルをより効率的に使用するために、データ補強の強力な利用に依存するネットワークと学習戦略を提示する。このアーキテクチャは、コンテキストをキャプチャするための収縮パスと、正確なローカライゼーションを可能にする対称的な拡張パスから構成される。このようなネットワークは、非常に少ない画像からエンドツーエンドで学習でき、電子顕微鏡スタックにおけるニューロン構造のセグメンテーションのためのISBIチャレンジにおいて、先行する最良の方法(スライディングウィンドウ畳み込みネットワーク)を凌駕することを示す。透過光顕微鏡画像(位相差とDIC)で学習させた同じネットワークを用いて、これらのカテゴリーにおけるISBI細胞追跡チャレンジ2015で大きな差をつけて優勝した。さらに、このネットワークは高速である。512x512の画像のセグメンテーションは、最近のGPUで1秒未満しかかからない。完全な実装(Caffeベース)と学習済みネットワークはhttp://lmb.informatik.uni-freiburg.de/people/ronneber/u-netで公開されている。

1 Introduction

この2年間で、深層畳み込みネットワークは、例えば[7,3]のような多くの視覚認識タスクにおいて、最先端技術を凌駕している。畳み込みネットワークはすでに長い間存在していたが[8]、利用可能な学習セットのサイズと考慮されたネットワークのサイズのため、その成功は限定的であった。Krizhevskyら[7]によるブレークスルーは、100万枚の学習画像を持つImageNetデータセットに対して、8層、数百万のパラメータを持つ大規模ネットワークの教師あり学習によるものである。それ以来、さらに大規模で深いネットワークが学習されている[12]。

畳み込みネットワークの典型的な使用法は、画像への出力が単一のクラスラベルである分類タスクである。しかし、多くの視覚タスク、特に生物医学的画像処理では、所望の出力はローカライゼーションを含むべきである、すなわち、クラスラベルは各ピクセルに割り当てられるべきである。さらに、何千枚もの学習画像は、通常、生物医学的なタスクでは手の届かないものである。したがって、Ciresanら[1]は、入力としてそのピクセルの周りの局所領域(パッチ)を提供することによって、各ピクセルのクラスラベルを予測するために、スライディングウィンドウセットアップでネットワークを訓練した。

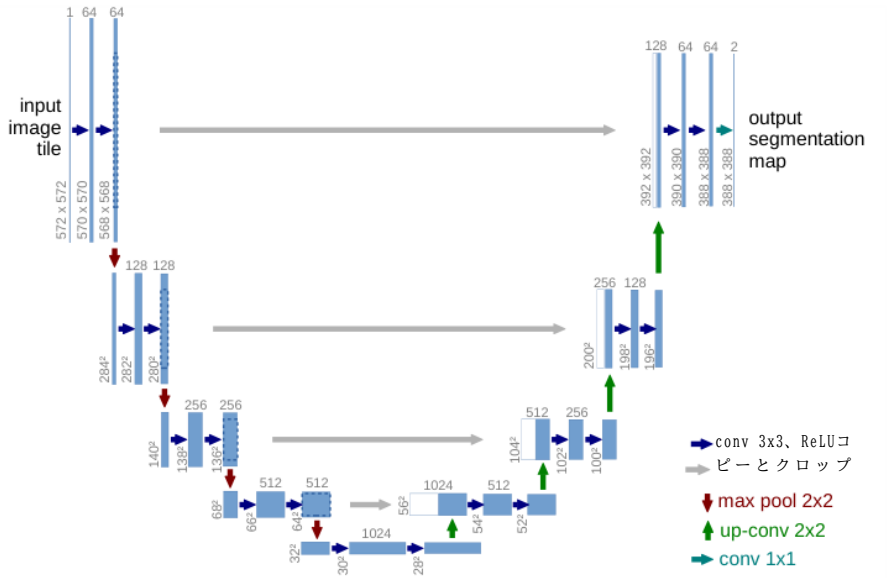

図1. U-netアーキテクチャ(最低解像度の32x32ピクセルの例)。各青枠はマルチチャンネル特徴マップに対応する。チャネルの数はボックスの上に示されている。x-y-sizeはボックスの左下端に記載されている。白いボックスはコピーされた特徴マップを表す。矢印は異なる操作を示す。

まず、このネットワークはローカライズできる。第二に、パッチ単位での学習データは、学習画像数よりもはるかに大きい。その結果、ISBI 2012のEMセグメンテーションチャレンジで大差をつけて優勝した。

明らかに、Ciresanら[1]の戦略には2つの欠点がある。まず、パッチごとにネットワークを別々に実行しなければならず、パッチの重なりによる冗長性が多いため、かなり時間がかかる。第二に、ローカライゼーションの精度とコンテキストの使用との間にトレードオフがある。より大きなパッチは、ローカライゼーションの精度を低下させるマックスプーリング層をより多く必要とするが、小さなパッチは、ネットワークがわずかなコンテキストしか見ることを可能にする。より最近のアプローチ[11,4]では、複数の層からの特徴を考慮した分類器出力が提案されている。良好なローカライゼーションとコンテキストの利用が同時に可能である。

本論文では、よりエレガントなアーキテクチャ、いわゆる「完全畳み込みネットワーク」[9]をベースに構築する。我々はこのアーキテクチャを修正・拡張し、非常に少ない学習画像で動作し、より正確なセグメンテーションが得られるようにした(図1参照)。9]の主なアイデアは、通常の収縮ネットワークを、プーリング演算子をアップサンプリング演算子に置き換えた連続的な層で補うことである。したがって、これらの層は出力の解像度を向上させる。

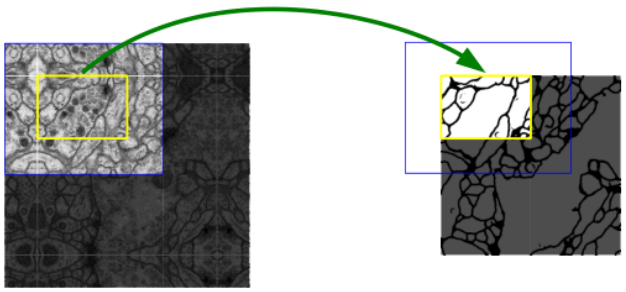

図2. 任意の大きな画像のシームレスなセグメンテーションのためのオーバーラップタイル戦略(ここではEMスタック中の神経細胞構造のセグメンテーション)。黄色の領域におけるセグメンテーションの予測には、青色の領域内の画像データが入力として必要である。欠落した入力データはミラーリングによって外挿される

ローカライズするために、収縮パスからの高解像度特徴は、アップサンプリングされた出力と組み合わされる。連続する畳み込み層は、この情報に基づいて、より正確な出力を組み立てることを学習することができる。

我々のアーキテクチャにおける重要な修正点の1つは、アップサンプリング部分では、ネットワークがより高解像度のレイヤーにコンテキスト情報を伝播することを可能にする、多数の特徴チャンネルも持っていることである。その結果、拡大経路は縮小経路と多かれ少なかれ対称的であり、U字型のアーキテクチャが得られる。ネットワークは完全連結層を持たず、各畳み込みの有効な部分のみを使用する。すなわち、セグメンテーションマップは、入力画像で完全なコンテキストが利用可能なピクセルのみを含む。この戦略により、オーバーラップタイル戦略による任意の大きさの画像のシームレスなセグメンテーションが可能になります(図2参照)。画像の境界領域の画素を予測するために、入力画像をミラーリングすることで、欠落したコンテキストを外挿する。このタイリング戦略は、大きな画像にネットワークを適用するために重要である。そうでなければ、解像度はGPUメモリによって制限されるからである。
我々のタスクでは、利用可能な学習データが非常に少ないため、利用可能な学習画像に弾性変形を適用することで、過剰なデータ補強を行う。これにより、ネットワークは、注釈付き画像コーパスにこのような変換を見ることなく、このような変形に対する不変性を学習することができる。以前は組織で最も一般的な変形であったが、現実的な変形を効率的にシミュレートすることができるため、これはバイオメディカルセグメンテーションにおいて特に重要である。Dosovitskiyら[2]では、教師なし特徴学習の範囲において、不変性を学習するためのデータ補強の価値が示されている。

多くの細胞セグメンテーションタスクにおけるもう一つの課題は、同じクラスの接触物体の分離である。この目的のために、我々は重み付き損失の使用を提案する。ここで、接触する細胞間の背景ラベルの分離は、損失関数において大きな重みを得る。

得られたネットワークは、様々な生物医学的セグメンテーション問題に適用可能である。本論文では、EMスタックにおける神経細胞構造のセグメンテーションに関する結果を示す(IS

BI 2012で開始された進行中のコンペティション)。さらに、ISBI細胞追跡チャレンジ2015の光学顕微鏡画像における細胞セグメンテーションの結果を示す。ここでは、最も困難な2D透過光データセット2つで大きなマージンを獲得した。

2 ネットワークアーキテクチャ

ネットワーク・アーキテクチャを図1に示す。これは、収縮経路(左側)と拡大経路(右側)から構成される。契約経路は畳み込みネットワークの典型的なアーキテクチャに従う。2つの3x3畳み込み(パディングされていない畳み込み)を繰り返し適用し、それぞれに整流線形ユニット(ReLU)とダウンサンプリングのためのストライド2の2x2最大プーリング演算を繰り返す。各ダウンサンプリングステップで、特徴チャンネル数を2倍にする。拡大パスの各ステップは、特徴マップのアップサンプリングと、それに続く特徴チャンネル数を半分にする2x2畳み込み(「アップコンボリューション」)、縮小パスから対応する切り出された特徴マップとの連結、2つの3x3畳み込み(それぞれReLUが続く)で構成される。すべての畳み込みで境界画素が失われるため、切り出しが必要である。最終層では、1x1畳み込みが、各64成分特徴ベクトルを所望のクラス数にマッピングするために使用される。ネットワークは合計で23の畳み込み層を持つ。
出力セグメンテーションマップ(図2参照)のシームレスなタイリングを可能にするために、すべての2x2マックスプーリング演算が、偶数x、yサイズのレイヤーに適用されるような入力タイルサイズを選択することが重要である。

3 Training

入力画像とそれに対応するセグメンテーションマップは、Caffe [6]の確率的勾配降下法の実装でネットワークを学習するために使用される。パディングされていない畳み込みにより、出力画像は入力よりも一定の境界幅だけ小さくなる。オーバーヘッドを最小化し、GPUメモリを最大限に活用するために、大きなバッチサイズよりも大きな入力タイルを優先し、その結果、バッチを単一画像に縮小する。したがって、我々は、以前に見たトレーニングサンプルの多くが、現在の最適化ステップにおける更新を決定するような、高い運動量(0.99)を使用する。
エネルギー関数は、クロスエントロピー損失関数と組み合わせた、最終的な特徴マップに対するピクセル単位のソフトマックスによって計算される。ここで、a_k (x) は画素位置 x∈Ω における特徴チャネル k の活性化を表し、Ω ∼$\frac{2}{Z}$ である。K はクラス数、p_k (x) は近似された最大関数である。すなわち、活性化a_k (x) が最大となるkに対してp_k (x) ≒ 1、それ以外のkに対してp_k (x) ≒ 0となる。クロスエントロピーは、各位置でp_{`(x)} (x)の1からの偏差にペナルティを与える。

$$E = \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x})) \tag{1}$$

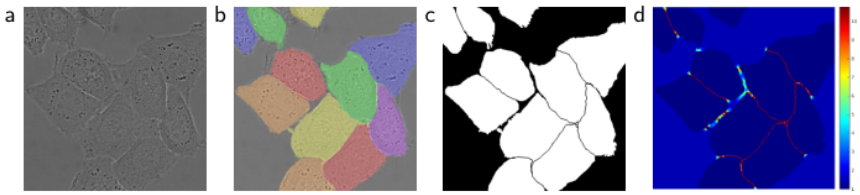

図3. DIC(微分干渉コントラスト)顕微鏡で記録したガラス上のHeLa細胞。(a) 生画像。(b) グランドトゥルースセグメンテーションとのオーバーレイ。色の違いは、HeLa細胞の異なるインスタンスを示す。(c) 生成されたセグメンテーションマスク(白:前景、黒:背景)。(d) 境界画素を学習させるために、画素単位の損失重みを持つマップ。

ここで ` : Ω → {1,.. K}は各画素の真のラベルであり、w : Ω → Rは学習においていくつかの画素をより重要視するために導入したウェイトマップである。

学習データセットの特定のクラスからのピクセルの異なる頻度を補正し、ネットワークが接触セル間に導入する小さな分離境界を学習するように、各グランドトゥルースセグメンテーションのウェイトマップを事前に計算する(図3cとdを参照)。

分離境界はモルフォロジー演算を用いて計算される。そして、ウェイトマップは次のように計算される。

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right) \tag{2}$$

ここで、w_c : Ω → R はクラス頻度のバランスをとるためのウェイトマップ、d_1 : Ω → R は最も近いセルの境界までの距離、d_2 : Ω → R は2番目に近いセルの境界までの距離を表す。実験では、w_0 = 10、σ ∼ 5ピクセルとした。

多くの畳み込み層とネットワークを通る異なる経路を持つディープネットワークでは、重みの良い初期化が非常に重要である。そうでなければ、ネットワークの一部が過剰な活性化を与えるかもしれないが、他の部分は決して寄与しない。理想的には、ネットワークの各特徴マップがほぼ単位分散を持つように、初期重みを適合させるべきである。我々のアーキテクチャ(畳み込み層とReLU層を交互に配置)を持つネットワークでは、標準偏差^p 2/Nのガウス分布から初期重みを引くことで実現できる。例:3x3畳み込みと前の層の64の特徴チャンネルの場合 N = 9 - 64 = 576.

3.1 データの拡張

データ増強は、利用可能な学習サンプルが少ない場合に、望ましい不変性と頑健性の特性をネットワークに教えるために不可欠である。

微視的な画像の場合、主にシフトや回転の不変性、変形やグレー値の変動に対するロバスト性が必要である。特に、学習サンプルのランダムな弾性変形は、非常に少ない注釈付き画像でセグメンテーションネットワークを学習するためのキーコンセプトであると思われる。粗い3×3グリッド上のランダムな変位ベクトルを用いて滑らかな変形を生成する。変位は標準偏差10ピクセルのガウス分布からサンプリングされる。画素ごとの変位はバイキュービック補間を用いて計算される。契約パスの末尾にあるドロップアウト層は、さらに暗黙的なデータ補強を行う。

4 Experiments

u-netを3つの異なるセグメンテーションタスクに適用することを実証する。最初の課題は、電子顕微鏡記録における神経細胞構造のセグメンテーションである。データセットと我々のセグメンテーションの例を図2に示す。全結果を補足資料として提供する。このデータセットは、ISBI 2012 で開始された EM セグメンテーションチャレンジ [14] によって提供され、まだ新しい貢献が期待できる。学習データは、ショウジョウバエの初齢幼虫腹神経索(VNC)の連続切片透過電子顕微鏡による30枚の画像(512x512ピクセル)のセットである。各画像には、細胞(白)と膜(黒)に対応する完全に注釈されたグランドトゥルースのセグメンテーションマップが付属している。テストセットは公開されているが、セグメンテーションマップは秘密にされている。予測された膜の確率マップをオーガナイザーに送ることで、評価を得ることができる。評価は、10段階の異なるレベルでマップを閾値処理し、「ワーピングエラー」、「ランドエラー」、「ピクセルエラー」を計算することによって行われる[14]。

u-net(入力データの7つの回転バージョンの平均)は、それ以上の前処理や後処理をすることなく、0.0003529のワーピング誤差(新しいベストスコア、表1参照)と0.0382のランド誤差を達成した。
これは、Ciresanら[1]によるスライディングウィンドウ畳み込みネットワークの結果よりも有意に優れており、その最良の提出物はワーピング誤差0.000420、ランド誤差0.0504であった。

表1. EMセグメンテーションチャレンジ[14](2015年、行進6位)のランキング、ワーピングエラーでソート。

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

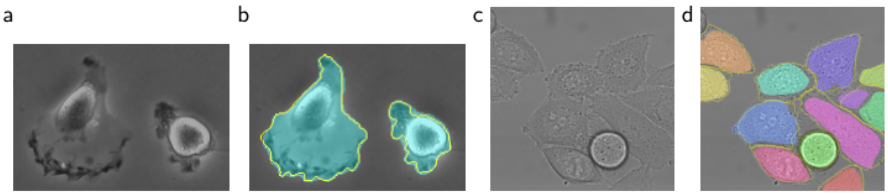

図4. ISBI細胞追跡チャレンジの結果。(a) "PhC-U373 "データセットの入力画像の一部。(b)セグメンテーション結果(シアンマスク)と手動グランドトゥルース(黄色枠) (c) "DIC-HeLa "データセットの入力画像。(d) セグメンテーション結果(ランダムな色のマスク)と手動によるグランドトゥルース(黄色の境界線)。

表2. ISBI細胞追跡チャレンジ2015でのセグメンテーション結果(IOU)。

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

ランド誤差の観点からは、このデータセットで唯一性能の良いアルゴリズムは、Ciresanら[1]の確率マップに適用された、データセットに特化した後処理法[1]を使用している。

また、u-netを光学顕微鏡画像の細胞セグメンテーションタスクに適用した。このセグメンテーションタスクは、ISBI細胞追跡チャレンジ2014と2015の一部である[10,13]。最初のデータセット "PhC-U373"[2]には、位相差顕微鏡で記録されたポリアクリルイミド基質上の膠芽腫-星細胞腫U373細胞が含まれている(図4a,bおよび補足資料参照)。35枚の部分的に注釈された学習画像を含む。ここでは、平均92%のIOU("intersection over union")を達成し、2番目に優れたアルゴリズムである83%よりも大幅に優れている(表2参照)。第二のデータセット "DIC-HeLa"[3]は、微分干渉コントラスト(DIC)顕微鏡で記録した平板ガラス上のHeLa細胞である(図3、図4c,d、補足資料参照)。部分的にアノテーションされた20枚の学習画像を含む。ここでは平均77.5%のIOUを達成し、2番目に優れたアルゴリズムである46%を大幅に上回る。

5 Conclusion

u-netアーキテクチャは、非常に異なる生物医学的セグメンテーションアプリケーションで非常に優れた性能を達成した。弾性変形によるデータ補強のおかげである。

[1] このアルゴリズムの著者は、この結果を達成するために78種類の解決策を提出した。

[2] Sanjay Kumar博士から提供されたデータセット。カリフォルニア大学バークレー校バイオエンジニアリング学部。バークレーカリフォルニア州(米国)

[3] Data set provided by Dr. Gert van Cappellen Erasmus Medical Center. Rotterdam. The Netherlands

の場合、アノテーション画像はほとんど必要なく、NVidia Titan GPU (6GB)での学習時間はわずか10時間と非常に合理的である。Caffe[6]ベースの実装と学習済みネットワーク[4]を提供する。u-netアーキテクチャは、より多くのタスクに容易に適用できることを確信している。

謝辞(謝辞

本研究は、ドイツ連邦政府および州政府の Excellence Initiative(EXC 294)および BMBF(Fkz 0316185B)の支援を受けた。

References

1. Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS. pp. 2852–2860 (2012)
2. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
4. Hariharan, B., Arbelez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization (2014), arXiv:1411.5752 [cs.CV]
5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification (2015), arXiv:1502.01852 [cs.CV]
6. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding (2014), arXiv:1408.5093 [cs.CV]
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012)
8. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989)
9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038 [cs.CV]
10. Maska, M., (...), de Solorzano, C.O.: A benchmark for comparison of cell tracking algorithms. Bioinformatics 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. In: Computer Vision (ICCV), 2013 IEEE International Conference on. pp. 2168–2175 (2013)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014), arXiv:1409.1556 [cs.CV]
13. WWW: Web page of the cell tracking challenge, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
14. WWW: Web page of the em segmentation challenge, http://brainiac2.mit.edu/isbi_challenge/

[4] U-net implementation, trained networks and supplementary material available at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net