

## 深層学習に基づく CT 画像からの複数の解剖学的構造の同時自動認識と抽出

周 向 栄\*<sup>1</sup> 藤田 広志\*<sup>1</sup>

### 要 旨

本稿では、畳み込みニューラルネットワーク (convolutional neural network; ConvNet) に基づく深層学習のアプローチを用いて、CT 画像から広範囲・多種類の解剖学的構造を自動的に認識・抽出する研究について紹介する。具体的には、筆者らの研究グループの事例から、CT 画像における多臓器・組織の自動認識・抽出問題を ConvNet で解決する方法を述べる。そして、中心的に行っている二次元画像に対応する 2D ConvNet とその三次元への拡張版について、それぞれの長所と短所を議論し、最新の研究成果を紹介する。最後に、設計者の直感と経験に依存する従来の設計アプローチと比較して、CT 画像からの多臓器・組織の自動認識・抽出手順の設計に関する深層学習の性能を考察する。紹介する実験の結果から、ConvNet に基づく深層学習が CT 画像における広範囲・多種類の解剖学的構造の自動認識・抽出問題の解決に寄与できることがわかる。

**キーワード**：深層学習，畳み込みニューラルネットワーク，三次元 CT 画像，解剖学的構造の自動認識・抽出

Med Imag Tech 35(4): 187-193, 2017

### 1. はじめに

CT 画像や MR 画像に代表される三次元医用画像は、患者の体内情報を観察する有効な手段であり、医療現場で診断、手術計画、治療の目的で大量に撮影・使用されている。しかしながら、三次元医用画像がこのような初期的な目的で一度使用されたあとには、ほとんど再利用されないのが一般である。よってこれらの膨大な画像データに含まれる貴重な人体情報は“眠っている宝の山”といわれ、計算機による“画像支援診断”システムの開発や、新しい医学的な知恵を生み出す“ビッグデータ分析”の情報源として期待されている。

医用画像に含まれる情報を効率的に活用するためには、計算機による画像情報の定量化が必要である。この定量化処理においては、複数の臓器・組織から構成される人体の解剖学的構造を、画像から自動認識・抽出することが前提で

ある。しかし、医用画像における低コントラスト、画像内に含まれる解剖構造の不完全性、個人差による解剖構造のばらつき、病変による解剖構造の異変などの原因で、医用画像からの臓器・組織の自動認識・抽出処理手法の開発は難しい問題の一つであった。

三次元医用画像から複数の臓器・組織を自動認識・抽出する手法に関する研究はこれまで数十年続き、数多くの論文が掲載されている。しかし、医療現場で三次元医用画像から多臓器・組織を自動的に抽出できるシステムの実用化はみられていない。原因としては、計算機による自動抽出手法の性能（特に汎用性、効率性、および頑丈性）が医用現場での実用レベルにまだ合致していないことが大きな要因と考えられる。一方、最近話題になっている自動車の自動運転システムにおいて、車載カメラで撮影される画像から、周辺の物体を瞬時に正しく認識・抽出できる処理手法が開発されている。これらの画像処理手法の設計に用いられている「深層学習」とよばれるアプローチ [1] が医用画像からの臓器抽出にも利用でき、そのような事例が少なからず報告されている。そこで、本稿では、深層

\*<sup>1</sup> 岐阜大学工学部電気電子・情報工学科〔〒501-1194 岐阜市柳戸 1-1〕  
e-mail: zxr goo@gmail.com

論文受付：2017年6月14日

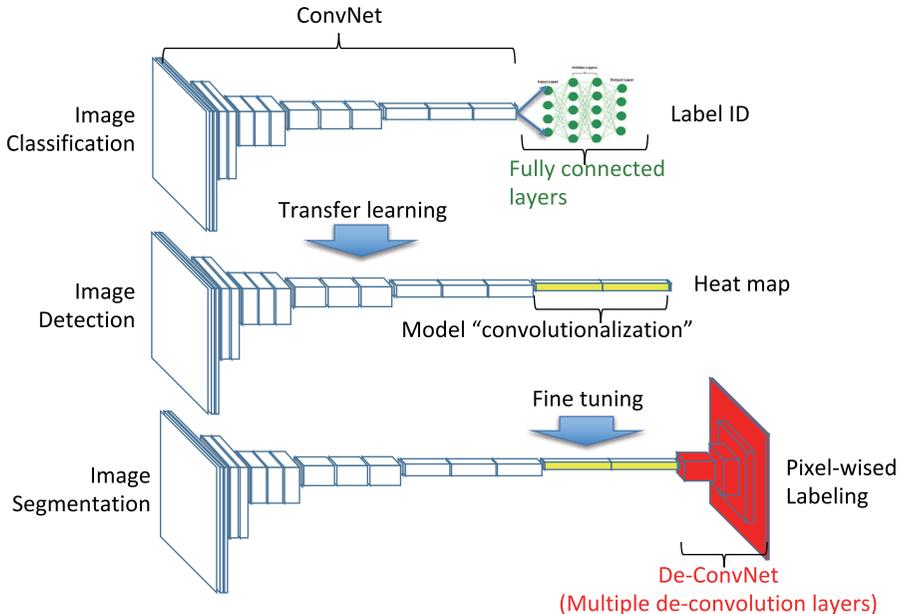


図1 深層学習の技術の進化。(画像の分類処理から、物体の位置検出に、さらにセグメンテーション処理までの発展)

学習に基づく三次元 CT 画像の解剖構造の自動抽出に関する最新の研究成果を紹介する。

## 2. 深層学習に基づく自然画像のセグメンテーションの方法

深層学習に基づく画像のセグメンテーションは、よくセマンティック・セグメンテーション (semantic segmentation) とよばれる。具体的な処理は、画像内に存在している複数の物体の“意味”を理解 (認識) し、“存在範囲”を示す物体の輪郭を高精度で決定すること (抽出) に分割することができる。また、認識処理は物体の存在位置 (例えば、画像上で物体を囲むボックス (bounding box) の座標) の検出 (detection) と物体の分類 (classification) から構成される。歴史的には深層学習の研究は、最初に画像上での物体の分類問題に着目し、次に物体の位置検出問題に移り、最後に画像のセグメンテーション問題の解決を目指し、それぞれ成功してきた。すなわち、画像のセグメンテーション手法が、物体分類と位置検出の研究成果を吸収して、発展してきたといえる。その典型例として、全畳み込みネットワーク (fully convolutional networks, 以下、FCN と略称) の方法を紹介する。

深層学習の研究分野で代表的なセグメンテーションの方法の一つとして、この FCN がよく知られている [2]。FCN は、画像分類のニューラルネットワークから発展してきた (図 1 参照)。具体的には、FCN は自然画像認識のコンテスト ILSVRC<sup>1)</sup> において、これまで優勝した AlexNet [3], VGG [4], GoogLeNet [5] のネットワーク (図 1 の上段を参照) をベースとしている。基本的な考え方は、まず、図 1 の上段の画像特徴の抽出部分である畳み込みネットワーク部 (ConvNet) の構造と学習結果は残して、結果の判定処理を行う全連結層 (fully connected layers) 部を改造する (全畳み込み化, convolutionalization とよぶ)。これによって、図 1 の中段に示すように、各対象を識別するために学習された画像特徴量の活性化を表すヒートマップ (heat-map) を直接出力する。このヒートマップは医用画像処理分野でよく用いる確率的臓器アトラスと類似するものであり、画像上に存在している対象領域を識別するとともに、存在位置と範囲を同時に出力できる。すなわち、画像の物体分類の学習成果 (画像特徴量) をう

<sup>1)</sup> <http://www.image-net.org/>

まく位置検出に転用している。しかし、出力ヒートマップの画像解像度が低く、物体の輪郭をはっきり識別できないため、画像解像度を向上させる De-ConvNet (図 1 下段の赤色部分、超解像処理に類似) を加える。実際には、複数の De-convolution 処理層を多段式で連結することによって徐々に結果画像の解像度を向上させる。

FCN の学習にも特徴がある。FCN の前半部分 (ConvNet) は、従来の画像分類のネットワークの学習結果から調整 (fine-tuning) する (transfer learning, 転移学習ともよぶ)。これは、物体認識処理に適切な画像特徴がセグメンテーションにも有効であると考えられるためであり、これにより早く学習曲線 (誤差値) が収束される。一方、FCN の後半である De-ConvNet については、新たに学習するしかない。具体的には複数のデコンボリューション (De-convolution) 層を逐次的に FCN の後に付け加えて訓練を繰り返す方法を用いる。狙いとしては、学習法である誤差逆伝搬法の特徴 (教師信号に近いほど学習効果がよい) を生かして、毎回新しく訓練するパラメータの数を少なめにし、かつ教師信号に近い位置に配置することで、よい学習効果を得ることと考えられる。

上述のように、FCN は自然画像からの物体セグメンテーションで大きな成功を取めた。最近の多くの研究では、FCN の考え方を生かして、抽出精度をさらに改善し続けている。例えば、画像解像度を向上させるときに用いられるアッププーリング (Up-pooling) 処理を工夫した SegNet [6]、セグメンテーションによく使用されている条件付き確率場 (CRF-RNN) を加えた FCN+CRF-RNN 法 [7] などが挙げられる。これらの技術は医用画像の解剖学的構造の認識・抽出にも応用可能と考える。

### 3. FCN に基づく CT 画像からの解剖構造の自動認識・抽出

三次元 CT 画像から複数の臓器・組織を自動的に認識・抽出する処理は、自然画像における semantic segmentation と類似している。自然画像と比較して、CT 画像や MR 画像などに代表される医用画像の場合には、(1) 処理対象が三次元配列であり、情報量が多く計算コストが高い、(2) 撮影原理が自然画像と異なり、画像上で解剖構

造が明確に識別できない場合がある、(3) 深層学習に使える症例の数が少なく、かつ (4) 三次元医用画像から解剖学的構造の教師信号を獲得することが困難、などの問題点がある。これらの問題点を解決するために、さまざまなアプローチが提案されてきている。以下では、いくつかの代表的な考え方を紹介する。

#### 1) 2D-FCN の結果の三次元投票 [8]

このアプローチは、三次元 CT 画像における二次元の断面画像から解剖学的構造を部分的に抽出して、それらの抽出結果 (三次元解剖構造の断片) を三次元画像空間に蓄積 (多数決による投票) することによって、三次元である解剖学的構造を決定する (図 2 参照)。この方針は、医師による CT 画像の読影 (複数の二次元スライス画像で臓器領域を確認・記憶して、頭の中で最終的に三次元の解剖構造を構築する方式) を真似ている。技術的には、三次元 CT 画像をさまざまな方向性をもつ断面画像にサンプリングすることは、深層学習に用いられる訓練標本の不足問題を解決し、冗長化した抽出結果群を三次元空間へ投票することによって、抽出処理の可用性 (availability) や信頼性 (reliability) を確保する利点が挙げられる。

実現する際には、転移学習を利用した。すなわち、CT 画像を用いて、自然画像におけるセグメンテーションの学習結果を調整する。これは、異なる画像の種類でも基礎的な画像特徴量は共通であると考え、数の少ない CT 画像を用いてもよい学習効果が得られるためである。また、輪郭抽出の役割を果たすデコンボリューション層の数を増やして、CT 画像上の曖昧な臓器輪郭の抽出処理を強化する。

この方法の弱点は、形状が細長い臓器 (血管、消化器官) の抽出精度が悪い点が挙げられる。その原因は、二次元断面画像群における血管領域は面積が小さく、形状のばらつきが大きいため、FCN での抽出が困難であるためである。

#### 2) 走査ウィンドウに基づく 3D-FCN [9, 10]

この方法では、二次元画像を対象にして作られた FCN の三次元拡張版を開発している。具体的には、畳み込みなどの層は全部三次元に拡張する。これらの研究の代表例として、3D U-Net [9] が挙げられる。このアプローチには、三次元画像に適用できるストレートな構造の利用に

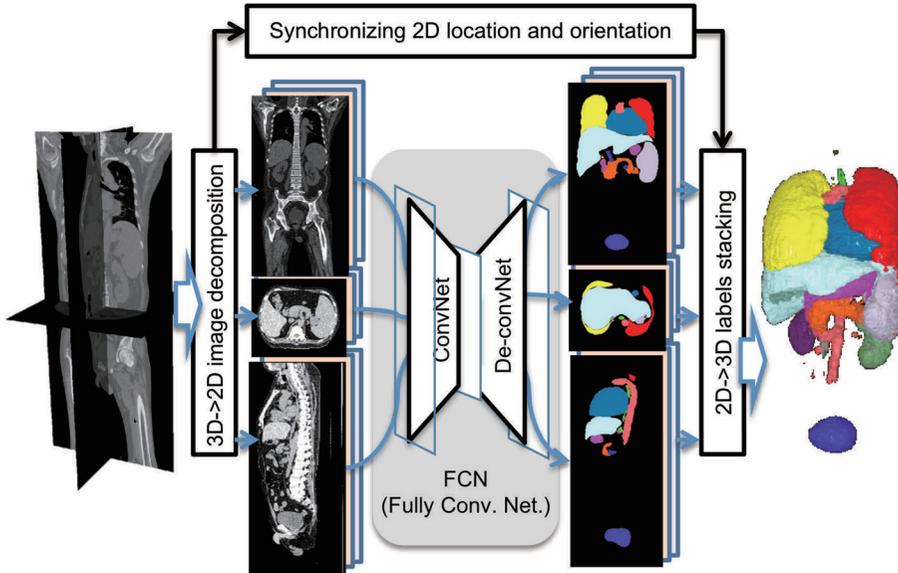


図2 2D-FCNの結果の三次元投票による解剖学的構造の抽出処理の概略 [8].

よって、豊富な三次元画像情報を漏れなく学習に使える利点があり、性能のよい抽出手順が見つけれることが期待される。しかし、FCNにおけるパラメーターの数が指数関数的に増え、問題点 (3) に対して学習用のCT画像の症例数が比較的少ないため、有効な学習結果を得ることが困難であると考えられる。また、人体の解剖学的構造を表す三次元CT画像のデータ量(典型的な画像サイズが $512^3$ 画素)が大きく、3D U-Netが直接的に処理することが時間的にも計算機のメモリー容量においても、現状では非現実的である。

以上の問題を解決するために、CT画像をサイズの小さい領域に分割・走査 (sliding window) しながら、三次元パッチ単位でFCNの学習とテストを行うことが提案されている。しかし、CT画像をそのままパッチに分割する場合には、解剖構造の大局的な情報が失われ、各パッチで表現される解剖構造の分散がさらに大きく、解剖構造を高精度に認識することが難しい。逆に各パッチに含まれる解剖学的構造の大局的な情報を増やす場合には、CT画像の解像度を落とす必要があり、臓器輪郭の抽出精度が低下する問題がある。最近では、複数の3D U-Netsを階段 (cascade) 状に接続して、CT画像からの解剖学的構造の認識と抽出を複数の画像解像度のレベ

ル (coarse-to-fine) で進めることなどの工夫が提案されている [10].

### 3) 臓器のバウンディングボックスに基づく3D-FCNs [11, 12]

計算機の処理能力を考慮して、3D-FCNの処理範囲をCT画像における臓器・組織の存在範囲に限定する方法も報告されている [11, 12]. このアプローチでは、各解剖構造の自動認識と抽出処理を分離して個別に対応する。すなわち、CT画像から各臓器・組織を認識してその領域を囲むバウンディングボックスを、従来の機械学習の方法 [11] によって自動的に決める。求めたそれぞれのボックス領域を対象領域と限定して、3D-FCNの訓練とテストを行う [12].

この研究では、各臓器を処理する専用の3D-FCNを臓器ごとに個別に訓練することになる [12]. これは、異なる臓器の大きさに応じて、3D-FCNの性能を最大限に引き出すためである。また、CT画像における臓器抽出の精度評価の基準である一一致度 (dice index) をそのまま3D-FCNの損失 (Loss) 値として逆伝搬することによって、ネットワークのパラメーターを順次更新する方法などの工夫を加えている [12]. このアプローチでは、三次元の画像情報を用いて、臓器ごとに画像特徴量と判定パラメーターを学習することで、よい抽出精度を得ることを狙っ

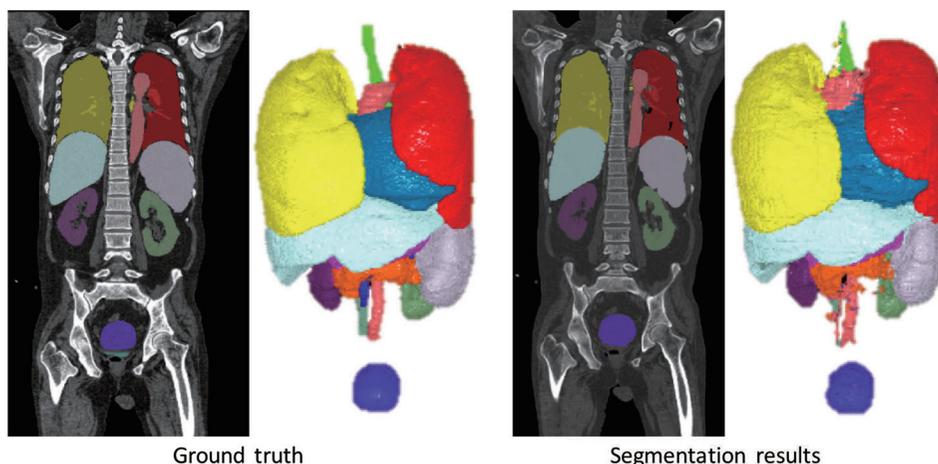


図3 体幹部 CT 画像からの複数の解剖学的構造（色分けで表示）の抽出結果 [8]. 原画像の 1 スライスでの二次元表示法（左半分）とサーフェスレンダリング法による三次元表示法（右半分）を用いた. Ground truth: 手動での抽出結果. Segmentation results: 自動抽出の結果.

ているが、各臓器抽出に共通で使える特徴の集約は、今後の課題である。

#### 4. 実験

CT 画像における解剖学的構造の自動認識・抽出について、以上に述べた 3 つのアプローチをわれわれは最近検証した。深層学習に基づく解剖学構造の自動抽出手順は、2D-FCN の結果の三次元投票と臓器のバウンディングボックスに基づく 3D-FCNs について、筆者らが開発したシステムを利用した。スライディングウィンドウに基づく 3D-FCN は、公開されている 3D U-Net システム [9] を用いた。また、従来法と比較するために、代表的な手法として「臓器の確率アトラスと形状モデル」に基づいて開発されたシステム [13] (筑波大学の岡田らより提供) を利用した。

これらのシステムを文科省科学研究費補助金新学術領域研究 (2009~2013 年度) 計算解剖学の研究プロジェクト<sup>2)</sup> で開発されたデータベース (DB) に適用し、性能比較を行った。この DB はさまざまな人体範囲を撮影した CT 画像 240 例 (画像サイズが  $512 \times 512 \times 99 \sim 1141$  (voxels), 解像度が  $0.625 \sim 1.148 \times 0.625 \sim 1.148 \times 1.0$  (mm<sup>3</sup>), 155 例が造影撮影) と画像内の解剖学的構造 (左右の肺, 心臓, 大動脈, 食道, 肝臓, 胆嚢, 胃, 脾臓, 左右の腎臓, 大静脈, 門脈・脾静脈・上腸間膜静脈, 膀胱, 膀胱, 前立腺, 子宮,

計 17 種類の臓器領域) と食道内腔および胃の内容物の 2 種類の関心領域を表す正解画像 (ラベル画像) で構成されている。ここで紹介する実験では、全症例の 95% (228 例) を用いてシステムを訓練し、その結果を残りの 12 例 (全症例の 5%) でテストした。抽出結果の例を図 3 に示す [8]。各臓器における抽出精度は、抽出結果と正解領域との一致度で評価した。また、CT 画像に含まれる全臓器領域の一致度の加重平均を、画像単位での評価基準とした。実験では、Graphics Processing Unit (NVIDIA 社製, GeForce GTX Titan X 12 GB) を装着した計算機を使用した。解剖学構造の自動抽出処理は、深層学習ソフトウェアパッケージ Caffe Framework<sup>3)</sup> の上に実装した。

#### 5. 考察

深層学習による手法の性能を従来法 [13] と比較した。その結果、従来法で抽出対象となっている 7 種類の臓器で比較したところ、深層学習による結果の精度 (一致度の平均値) が 8.8% 良かった。また、処理手順の汎用性、頑丈性、計算速度の面においても、深層学習による手法が従来法よりよい結果を示した。

2D-FCN の結果の三次元投票と 3D U-Net の比較では、3D U-Net は CT 画像をそのまま処理で

<sup>2)</sup> <http://www.comp-anatomy.org/wiki/>

<sup>3)</sup> <http://caffe.berkeleyvision.org>

きないために、CT画像を複数の三次元パッチに分割してからスライディングウィンドウに基づく3D U-Netの方式を利用した。19種類の対象領域における抽出精度の比較結果から、2D-FCNの処理結果の一致度(平均値)が13.2%高かった。その原因は、CT画像の分割による三次元解剖学的構造の断片化、学習症例が少ない、かつ転移学習の利用ができない、などの要因で、3D U-Netの十分な学習が困難であったためと推測される。

そこで、CT画像から各臓器の存在範囲を検出して、臓器別に3D-FCNを学習する臓器のバウンディングボックスに基づく3D-FCNsの方式に切り替えて比較すると、3D-FCNの抽出結果のほうが2D-FCNよりよいことが示され、16種類の臓器の一致度(平均値)が15.1%改善された。特に消化器官と血管において、結果の抽出精度が大幅に向上した。しかし、この方式では、臓器ごとに専用の3D-FCNを学習してCT画像に適用する必要がある。よって、全臓器を1つの2D-FCNで対応するわれわれの方法と比較すると、使用の利便性と計算の効率性では不利であるといえる。

## 6. まとめ

本稿では、CT画像における解剖学的構造の自動認識・抽出問題について、深層学習を用いた解決法を解説した。従来法と比較した結果から、深層学習のアプローチがCT画像から広範囲・多種類の解剖学的構造の自動認識・自動抽出の解決に寄与できることを示した [14]。また、深層学習の実現の過程において、二次元と三次元FCNに関する最新の研究結果を紹介し、それぞれの利点と弱点を考察した。今後の課題として、二次元と三次元のFCNネットワークの融合、異なる画像モダリティからの複数の臓器の認識と抽出処理の融合、教師ありの学習と教師なしの学習の融合、解剖構造の抽出と病変検出機能の融合などが挙げられる。

## 謝 辞

実験に協力していただきました岐阜大学藤田研究室の高山良介氏と飯盛広規氏、並びに筑波大学の岡田俊之先生と名古屋大学のHolger Roth先生に感謝いたします。本研究の一部は、文部

科学省科学研究費補助金新学術領域研究(課題番号26108005)およびJSPS科研費(基盤研究C26330134)によって行われました。

## 文 献

- [1] LeCun Y, Bottou L, Bengio Y, et al.: Gradient-based learning applied to document recognition. Proc. of the IEEE **86**: 2278-2324, 1998
- [2] Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, 2015, pp3431-3440
- [3] Krizhevsky A, Sutskever I, Hinton GE, et al.: Image Net classification with deep convolutional neural networks. In proceedings of Advances in Neural Information Processing Systems 25, Nevada, 2012, pp1097-1105
- [4] Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition, Proc. ICLR, <https://arXiv:1409.1556v6> (accessed 2017. 6. 14)
- [5] Szegedy C, Liu W, Jia Y, et al.: Going deeper with convolutions. In proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, 2015, No. 15523970
- [6] Badrinarayanan V, Kendall A, Cipolla R: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Patt Anal Mach Intell, 2017 (in press)
- [7] Zheng S, Jayasumana S, Romera-Paredes B, et al.: Conditional random fields as recurrent neural networks. In proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, No. 15801852
- [8] Zhou X, Ito T, Takayama R, et al.: First trial and evaluation of anatomical structure segmentations in 3D CT images based only on deep learning. Med Image Inf Sci **33**: 69-74, 2016
- [9] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Ourselin S, Joskowicz L, Sabuncu ML, et al. eds.: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016, Lecture Notes in Computer Science Vol. 9901, Springer, Cham, pp424-432
- [10] Roth HR, Oda H, Hayashi Y, et al.: Hierarchical 3D fully convolutional networks for multi-organ segmentation. <https://arxiv.org/abs/1704.06382v1> (accessed 2017.6.14)
- [11] Zhou X, Morita S, Zhou X, et al.: Automatic anatomy partitioning of the torso region on CT images by using multiple organ localizations with a group-wise calibration technique. Proc SPIE **9414**: 94143K-1-94143K-6, 2015
- [12] 周 向榮, 山田一馬, 飯盛広規, 他: 3D-Deep CNNを用いたCT画像からの多臓器の自動抽出と性能評価. 第36回日本医用画像工学会大会予稿集, 岐阜, 2017, OPI-7
- [13] Okada T, Linguraru MG, Hori M, et al.: Abdominal

multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. *Med Image Anal* **26**: 1-18, 2015

[14] Zhou X, Takayama R, Wang S, et al.: Deep learning

of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med Phys*, 2017 (in press)

## Simultaneous Recognition and Segmentation of Multiple Anatomical Structures on CT Images by Using Deep Learning Approach

Xiangrong ZHOU <sup>\*1</sup>, Hiroshi FUJITA <sup>\*1</sup>

<sup>\*1</sup> *Gifu University*

This paper introduces research works that apply deep learning approaches based on ConvNet to solve automatic multi-organ segmentations on CT images that cover a wide range of human body. In particular, we describe our recent research work as an example to show multiple-organ segmentation methods on CT images by using ConvNets. We discuss strength and weakness of the ConvNet that is majorly used for 2D image processing and its extension for 3D images with the latest research progresses. Finally, we compare the deep learning approaches to the conventional approach that is designed by the processing procedures based on human experience and shows an advantage and potential possibility of ConvNets to address the issue of automatic multi-organ segmentations on CT images covering a wide range of human body.

**Key words:** Deep learning, Convolutional neural network, 3D CT images, Anatomical structures recognition and extraction

*Med Imag Tech* **35**(4): 187-193, 2017



周 向荣 (しゅう こうえい)

1993年中国ハルビン工業大・工・電気卒。2000年名古屋大学工学研究科情報工学専攻博士後期課程修了。同年岐阜大学バーチャルシステムラボラトリー研究員, 2002年同大学院医学系研究科助手, 2006年同助教, 2017年同工学部助教, 工学博士。医用画像における画像処理および画像認識に関する研究に従事。電子情報通信学会, 日本医用画像工学会, 日本生体医工学会, 日本放射線技術学会などの会員。



藤田広志 (ふじた ひろし)

1976年岐阜大・工・電気卒。1978年同大学院修士課程修了。同年岐阜高専助手, 1986年同助教。この間, 1983~1986年シカゴ大学ロスマン放射線像研究所客員研究員。1991年岐阜大工学部助教, 1995年同教授, 2002年同大学院医学系研究科再生医科学専攻主任教授。2017年同工学部教授, 工学博士。著書: 医用画像ハンドブックおよび実践医用画像解析ハンドブック(監・編著, オーム社)など多数。医用画像情報学会(名誉会長), 日本医用画像工学会(2017年大会長), 電子情報通信学会(フェロー, 東海支部長), IEEE, SPIEなどの会員。

\* \* \*