

U-Net による CT 画像における脊椎の自動検出

鎌田 理詩[†] 菊池 眞之[†] 庄野 逸^{††} 林 勲^{†††} 福島 邦彦^{††††}

[†] 東京工科大学 コンピュータサイエンス学部 〒192-0982 東京都八王子市片倉町 1404-1

^{††} 電気通信大学 大学院情報理工学研究科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

^{†††} 関西大学 大学院 総合情報学研究科 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

^{††††} ファジィシステム研究所 〒820-0067 福岡県飯塚市川津 680-41

E-mail: [†] {c0114161c5@edu, kikuchi@stf}.teu.ac.jp, [‡] shouno@uec.ac.jp,

[‡] ihaya@cbii.kutc.kansai-u.ac.jp, [‡] fukushima@m.ieice.org

あらまし 本研究では、患者の CT 画像から脊椎の 3D メッシュモデルを自動生成するための足がかりとして、深層学習を用いた CT 画像における脊椎部分の自動抽出を目指した。具体的には、10 本分の脊椎データのうち、学習用に 5 本、テスト用に 5 本を使用し、U-Net と呼ばれるニューラルネットワークで学習とテストを行った。学習させた脊椎をテストしたときの Dice 係数による教師画像とセグメンテーション結果との一致度の評価値(以下、一致度とする)は 98.5%、学習させていない脊椎をテストしたときの一致度は 82.7%であった。この一致度を向上させるためには、U-Net のパラメータを調整する他、三次元的に学習を行えるようにすることが有効であると考えられる。

キーワード U-Net, 医用画像処理, セグメンテーション

Application of U-Net to spine image extraction in CT image

Mikoto KAMATA[†] Masayuki KIKUCHI[‡] Hayaru SHOUNO[‡] Isao HAYASHI[‡] and
Kunihiko FUKUSHIMA[‡]

[†] School of Computer Science, Tokyo University of Technology 1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan

[‡] Graduate School of Informatics and Engineering, University of Electro-Communications 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

[‡] Faculty of Informatics, Kansai University 2-1-1 Kamisenji, Takatsuki, Osaka 569-1095, Japan

[‡] Fuzzy Logic Systems Institute 680-41 Kawatsu, Iizuka, Fukuoka 820-0067, Japan

E-mail: [†] c0114161c5@edu.teu.ac.jp, [‡] kikuchi@stf.teu.ac.jp, [‡] shouno@uec.ac.jp,

[‡] ihaya@cbii.kutc.kansai-u.ac.jp, [‡] fukushima@m.ieice.org

Abstract In this study, we aimed at automatic extraction of spinal parts in CT images using deep learning as a foothold for automatically generating 3D mesh model of spine from CT image of patient. We used 10 spinal data which were learned and tested on U-Net for the segmentation of spine region. The percentage of correct answers that the learned spine was tested was 98.5%, and the similarity between teacher image and segmented spine image measure by Dice coefficient of having tested the unedited spinal column was 82.7%. In order to improve this similarity value, it will be effective to adjust the parameters of U-Net and to conduct 3D learning as well.

Keywords U-Net, Medical image processing, segmentation

1. はじめに

近年、深層学習の登場により AI 分野が活性化され、様々な業界で実際に活用されている。一方で、手作業による仕事は依然として多い。具体例として、医療分野では、患者の CT 画像から脊椎の 3D メッシュモデルを作成する際に、CT 画像一枚一枚から脊椎部分のみを手作業で切り出している状況であることが挙げられる。この作業は、非常に時間と労力がかかり、効率的とは言いがたい。本研究では、この問題を解決するための前段階として、CT 画像の脊椎部分を自動的に抽出することを目指す。具体的には、U-Net [1] と呼ばれるニューラルネットワークを用いる。

2. 分析方法

2.1. U-Net

本研究では、U-Net と呼ばれる Encoder-Decoder 構造を持つ U 字型のニューラルネットワークを用いる。このニューラルネットワークは、下向きのパスで畳み込みとプーリングを行い、上向きのパスで畳み込みとアップサンプリングを行うものである。ここで、本研究で U-Net を用いる理由として、このニューラルネットワークが医用画像処理のために考案されたものであることが挙げられる。医用画像のセグメンテーションにおいては、個人情報保護などのため、医用画像の入手が困難であるため、少ない画像枚数で処理が出来なければならない他、各組織の局所的な特徴と全体的な位置の情報の両方を元画像上で特定しなければならない。加えて各組織同士の境界をはっきりと判別出来なければならない。U-Net は、このような問題を解決するために考案されたものであることから、本研究においても高い精度を得ることが出来るのではないかと考えた。

本研究で構築する U-Net では、一般的な U-Net とはやや異なり、出力関数をシグモイド関数からハイパボリックタンジェント関数に変更している。また、過学習を防止するために、アーリーストッピングと呼ばれるテクニックを用いて、50 回学習の精度が落ちた場合に学習を停止させるようにした他、0.5 の確率でドロップアウトを 2 回行うようにした。なお、誤差逆伝搬法と一致度の算出には Dice 係数を用いている。

$$\frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (1)$$

ここで式(1)において、X は教師データの画素値の総和、Y は予測結果の画素値の総和である。

2.2. 具体的な手順

本研究では、前述した U-Net を用いて CT 画像から脊椎部分の学習とテストを行う。具体的には、SpineWeb [2] が提供している 10 本分の脊椎を CT スキャンした画像(枚数は脊椎それぞれで不揃いであるが、大体 500~600 枚前後。サイズは 512×512 pixel)を用いる。なお、それぞれの CT 画像における CT 値の基準を統一させるため、入力画像に以下のような正規化を予め行ってから、U-Net で学習とテストを行う。

$$i_{new}(x) = \frac{i(x) - i_{avr} + 3\sigma}{6\sigma} \quad (2)$$

ここで、 $i(x)$ 、 $i_{new}(x)$ 、 i_{avr} は、正規化前の画素値、正規化後の画素値、および画像全体の画素値の平均であり、シグマは標準偏差である。

脊椎 10 本分のデータのうち、5 本分(2852 枚)を学習させ(その際、予めシャッフルを行う)、残り 5 本分をテストさせる。なお本研究では、バッチサイズを 10、閾値を 0.9 としている。

最後に、テストで得られた結果を 3D Slicer(医用画像の可視化と解析のためのオープンソースソフトウェアの一つ)[3]に読み込ませて、脊椎の 3D メッシュモデルが構築出来ているかどうかを目視で確認する。

3. 結果

学習を行った結果、87 回で学習が停止し、学習させた脊椎をテストしたときの一致度は 98.5%、学習させていない脊椎をテストしたときの一致度は 82.7%であった。ここで、具体的にどのように学習が進んだのかをグラフ化した図を図 3.1 に示す。図 3.1 では、1 エポックごとに一致度をプロットしており、青線のグラフ「dice_coef」は学習させた脊椎画像をネットワークに入力したときの一致度、橙線のグラフ「val_dice_coef」は学習させていない脊椎画像をネットワークに入力したときの一致度を表している。

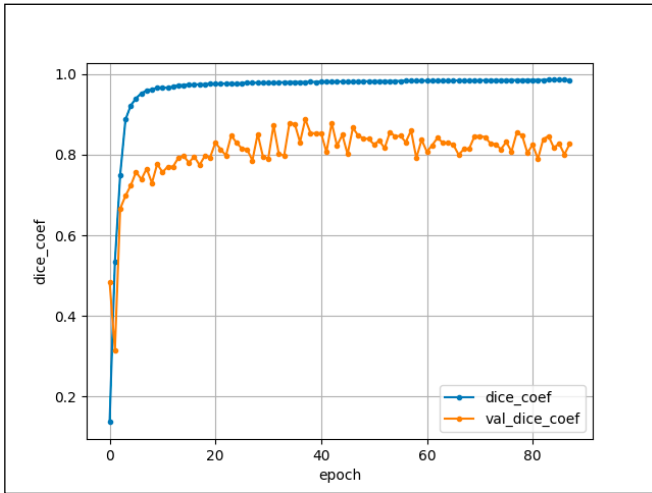


図 3.1 脊椎抽出結果の一致度

図 3.1 より、橙線のグラフを見ると、過学習を起こすこと無く順調に学習できていることが見て取れる。

図 3.2 は学習させた脊椎の予測結果，図 3.3，図 3.4 は学習させていない脊椎の予測結果である．ここで，図 3.2，図 3.3，図 3.4 は，それぞれ 3 行 5 列で構成されており，上から生データ，教師データ，予測結果を表している．なお，図 3.3 は学習させていない脊椎の予測結果のうち，特に一致していない部分を示しており，図 3.4 はそれ以外の部分(ある程度一致している部分)を示している(図 3.3 は脊椎の上部，図 3.4 は脊椎の中部の CT 画像である)．

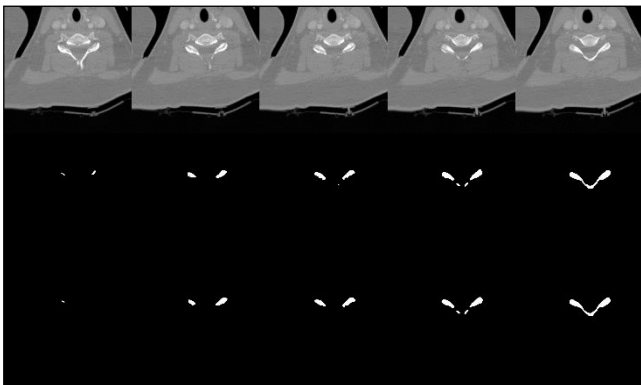


図 3.2 学習させた脊椎の予測結果

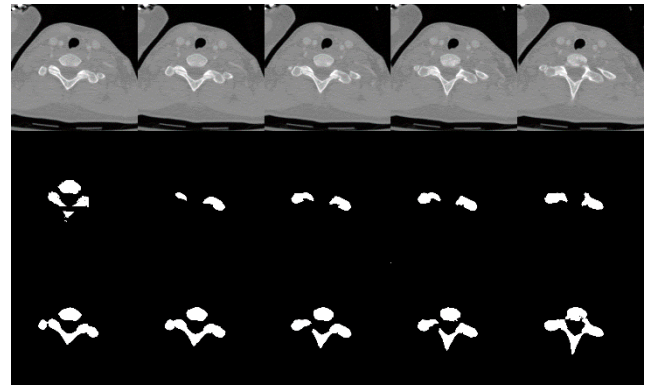


図 3.3 学習させていない脊椎の予測結果(1)

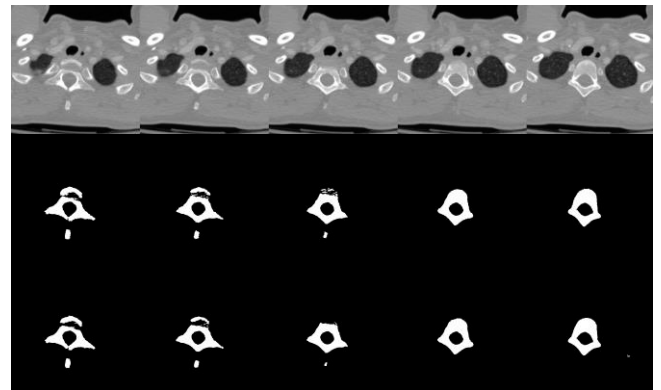


図 3.4 学習させていない脊椎の予測結果(2)

図 3.2 より、脊椎の学習が確かに行えていることが見て取れる．一方で，図 3.3，図 3.4 を見ると，図 3.3 では，教師データと予測結果があまり一致しておらず，図 3.4 では，大方一致していることが分かる．ここで，図 3.3 の一行目を見ると，脊椎が予測結果のような形で存在しているように見えるが，教師データとは異なる形であることが分かる．しかし，図 3.4 では，図 3.3 のような大きな違いは見られない．これは，CT 画像を生成する際に，脊椎を上からスキャンすると，最も表面の層よりも奥の層まで見えるケースがあるためで (CT 画像上で脊椎の影も見ることがある)，図 3.3 はそのケースに当たる．ただし，全体としては，図 3.4 のような結果が大半であった．

図 3.5 は，学習させていない脊椎の予測結果を 3D Slicer に読み込ませ，自動的に 3D メッシュモデル化させたものである．

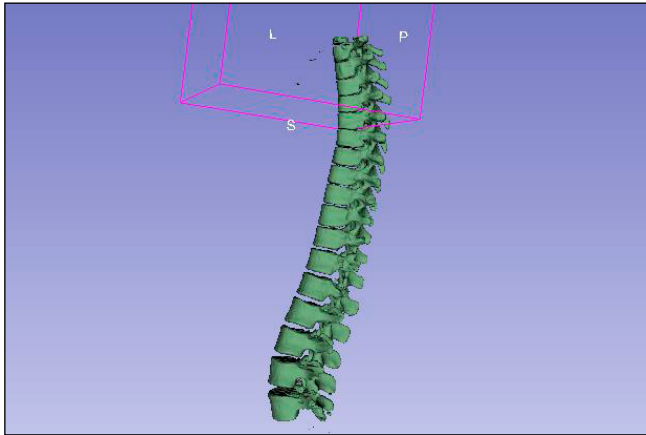


図 3.5 学習させていない脊椎の 3D メッシュモデル

図 3.5 より、目視で脊椎のような形状をしていることが分かる。脊椎の 3D メッシュモデルの周囲には、わずかに粒のようなものが散っているが、3DSlicer の Islands 機能(図 3.5 に示されるような小さな粒は取り除き、小さな穴は埋める機能)などを用いるか、本研究では実装していないが、コードにそういった機能を実装することで十分に対応出来るため、概ね期待通りの成果が得られたと言える。一方で、正確な脊椎の形状と構築された 3D メッシュモデルの形状を比較すると、本研究で構築したニューラルネットワークが図 3.3 のように影を誤識別してしまっているために、構築された 3D メッシュモデルがやや分厚くなり、棘突起の部分も垂直気味になるなど、100%の一致度ではないことも見て取れる。

4. 考察

実験結果より、本研究で構築したニューラルネットワークでは、脊椎の影の部分まで誤って抽出してしまっていたものの、82.7%と概ね正確な脊椎の形状に近い 3D メッシュモデルが構築出来たことが分かった。ここで、何故このような結果が得られたのかを考えてみる。今回、事前に入力画像に対し正規化を行い、CT 値の基準値を統一したり、U-Net の出力層を敢えてシグモイド関数からハイパボリックタンジェント関数に変更したりするなどの工夫を施した。予測結果の大半は、図 3.4 のようにしっかりと脊椎の抽出が行っていたが、中には図 3.3 のように脊椎の影まで抽出してしまっているものもあった。これは、学習に用いた CT 画像が上からスキャンしたものだけであったことが原因ではないかと予想される。単純にチューニングして精度を高めることで対処する方法も有効ではあるが、より高い精度を得るには、横からスキャンした CT 画像も同時に学習させる、即ち三次元的な学習を行うことが適しているのではないかと予想される。また、上か

らと横からの二種類の画像を用意せずとも、上からスキャンした CT 画像の画素値を配列化し、並び替えを行うことで擬似的に横からの CT 画像の画素値を用意出来るため、その値を用いて学習を行うことでもより高い精度が得られるのではないかと考えられる。また、本研究では、入力画像中の脊椎に関するクラス分類までは行っていないため、脊椎が男性のものであるか女性のものであるか、大人のものであるか子供のものであるか、骨折をしているかどうかなどの判別がつかない。よって、クラス分類を行えるようにすることで、より本研究の有用性を高めることが出来ると考えられる。

5. まとめ

本研究では、患者の CT 画像から脊椎の 3D メッシュモデルを自動生成するための足がかりとして、深層学習を用いた CT 画像における脊椎部分の自動抽出を目指した。具体的には、10 本分の脊椎データに対し、U-Net で学習とテストを行った。学習させた脊椎をテストしたときの一致度は 98.5%、学習させていない脊椎をテストしたときの一致度は 82.7%であった。この一致度を向上させるためには、U-Net のパラメータを調整する他、三次元的に学習を行えるようにすることが有効であると考えられる。また、脊椎のクラス分類を行えるようにすることで、更に本研究の有用性を高めることが出来ると考えられる。

文 献

- [1] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention, no.9351, pp.234-241, Munich, Germany, Oct.2015.
- [2] SpineWeb, <http://spineweb.digitalimaginggroup.ca/spineweb/index.php?action=home>, 2017.1.1.
- [3] 3DSlicer, <https://www.slicer.org/>, 2017.12.1.