

2019 年度 修士論文概要

主査	舟橋 健司	副査	本谷 秀堅	研究室	舟橋研究室
入学年度	2018 年度	学籍番号	30414114	氏名	満安 佑亮
論文題目	2 種類の CNN モデルを用いた細胞核抽出と分類 Cell Nucleus Extraction and Classification Using Two CNN Models				

1 はじめに

近年、医療の進歩には目覚ましいものがあるが、依然として癌の脅威は未だに去っていない。癌の早期発見は今後の日本の国民の寿命の延長には欠かせないものである。そのため、細胞や組織を直接診察する細胞診・組織診は癌の早期発見することにおいてとても重要な役割を担っている。しかし、これらの診断の大部分は病理専門医の経験と技術に依存しているため、様々な課題がある。そこで本研究では、癌の客観的指標の作成及び病理専門医の負担軽減のための癌診断システムの構築を目的とする。その一環として細胞核の抽出と分類を行う。ヘマトキシリン・エオジン (HE) 染色された病理画像から細胞核の抽出および分類を行った手法として文献 [1][2] がある。しかし、文献 [1] ではスコア画像を作成するが、その画像はあくまで細胞核が存在する確率が高い領域を示しているため誤検出や検出漏れが多くなってしまふ。文献 [2] では良性、悪性、転移の 3 種類を判定しており、分類精度は全体では約 8 割だが、個々の分類率を見ると悪性に関しては約 6 割の分類しかできておらず信憑性に欠けてしまふ。以上の課題をふまえ本稿では同様に細胞核に着目し、抽出および分類に Deep Learning のアルゴリズムの一種である Convolutional Neural Network(CNN) モデルを 2 種類用いて従来の手法よりも精度の向上と高速化を図る。

2 細胞核の抽出および分類における提案手法

2.1 細胞核抽出における提案手法

本手法の細胞核抽出には、医療画像のセグメンテーション用のモデルである U-Net[3] を用いた。本手法で用いる U-Net の構造を図 1 に示す。細胞画像およびマスク画像よりパッチという小さい画像を作成し、学習用データセットとしている。パッチを作成する前処理としてグレースケール化、ヒストグラム平坦化、ガンマ補正を加えている。パッチ画像の例を図 2 に示す。本手法の抽出方法として画素ごとではなく検出ウィンドウを設けてそれを細胞画像に対してラスタスキャンを行い、検出ウィンドウから取得した画像を入力画像とする。モデルに通した入力画像は細胞核が抽出された状態で出力され、全ての画素に対して終わるまで繰り返し、細胞画像の抽出結果画像を作成する。また、作

成された画像は膨張処理や収縮処理を行うモルフォロジー処理を加えノイズの低減を行う。

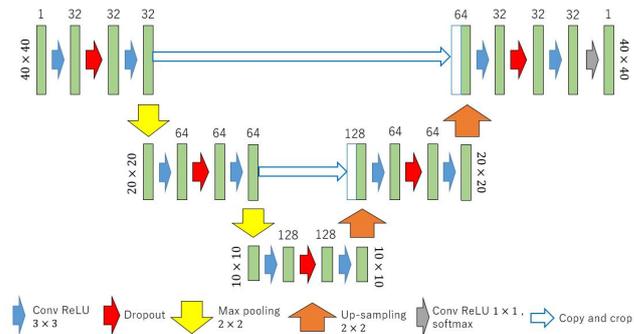
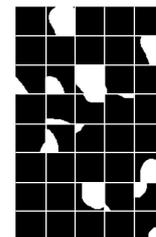
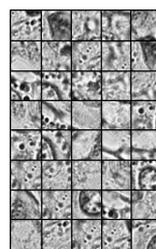


図 1: 抽出用モデルの構造



(a) 細胞画像のパッチ

(b) マスク画像のパッチ

図 2: 作成するパッチ画像の例

2.2 細胞核の分類における提案手法

本手法の細胞核分類には、CNN モデルである VGG16[4] に学習済みネットワークを利用する Fine-tuning を行い、細胞核分類用に拡張する。本手法で用いるモデルの構造を図 3 に示す。細胞核分類の処理手順として抽出された細胞核にラベリングを行い、細胞核領域を矩形化する。矩形化された細胞核を入力画像とし学習させ、正常細胞か癌細胞に分類する。

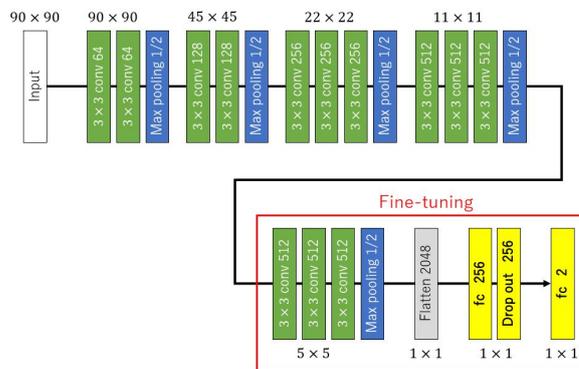
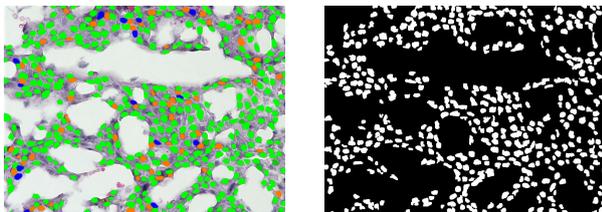


図 3: 分類用モデルの構造

3 細胞核の抽出および分類実験

3.1 細胞核抽出実験

提案手法の有効性を確かめるため実際に HE 染色された遺伝子改変マウスに発症したメラノーマ組織画像より細胞核の抽出を行った。学習用データセットとして 40×40 px のパッチ画像 190000 枚作成した。図 4 にはそれぞれの真値と抽出結果を示す。また、精度評価には Precision, Recall, F-measure を用いた。結果として、Recall, F-measure は提案手法の方が高精度な結果となった (表 1)。特に、提案手法の Recall の値に関しては 80%以上と高い値であることが確認でき、抽出漏れが少ないことを表している。また、抽出画像を作成する時間は文献 [1] が約 18 時間、提案手法が約 6 時間であった。細胞核の抽出において、抽出漏れを少なくすることが重要であると考えられる。細胞核の分類をすることを考えると、細胞の見逃しが原因で患者に不利益な状況につながってしまうことも考えられるため、抽出漏れが少ない本手法の有効性が確認できた。



(a) 真値 (b) 提案手法の抽出結果

図 4: 真値と抽出結果

表 1: 精度評価 [%]

細胞画像 1	Precision	Recall	F-measure
文献 [1]	83.5	63.0	71.8
提案手法	75.2	84.5	79.6

3.2 細胞核分類実験

提案手法の有効性を確かめるため、病理医がメラノーマか否かを診断し患者から採取した HE 染色された皮膚組織の病理画像に対して実験を行った。メラノーマの判定は良性腫瘍と診断された「良性」、悪性腫瘍と診断された「悪性」、メラノーマが他の組織に転移したもので悪性と診断された「転移」の 3 種類としている。分類基準として 1 枚の画像に対し正常細胞が多ければ正常、癌細胞が多ければ悪性と判断する。また、病理画像には細胞核が数千個ほど点在しているので細胞核を一つ一つ分類していくと時間がかかってしまうので、病理画像と病理画像を 16 等分した切片画像の 2 種類に対して分類を行う。VGG16 での Fine-tuning 向け学習用データセットとして正常細胞画像 77 枚、癌細胞画像 98 枚を構築した。良性の切片画像 10 枚、悪性の切片画像 20 枚、転移の切片画像 10 枚の計 40 枚の切片画像と良性病理画像 2 枚、悪性病理画像 2 枚、転移病理画像 2 枚の計 6 枚の病理画像全体に対して分類を行った。

結果として、分類全体の正答率が 8 割を超える結果となった (表 2)。特に、良性と転移は正確に分類できている。しかし、悪性に関しては正常と分類してしまっているものもある。原因としては、悪性は正常細胞と癌細胞が入り混じっているものであるため癌細胞を正常細胞として誤分類してしまっていることが考えられる。また、切片画像の実験と同様に良性と転移は分類結果が明確になっていることが確認できる (表 3)。悪性に関しては切片画像での実験で誤分類してしまっているものがあつたが病理画像全体で見るとしっかりと分類できている。以上より本手法の有効性が確認できた。

表 2: 切片画像に対する分類結果 (提案手法)

ラベル	正常分類数	悪性分類数	正答率	全体の正答率
良性	10	0	1.0	0.88
悪性	5	15	0.75	
転移	0	10	1.0	

表 3: 病理画像に対する分類結果

ラベル	番号	正常細胞核数	癌細胞核数	分類結果
良性	1	1807	1038	正常
	2	2264	1723	正常
悪性	1	1414	1422	癌
	2	1886	1979	癌
転移	1	1785	2475	癌
	2	1588	1761	癌

4 おわりに

本稿では、2 種類の CNN モデルを用いた細胞核の抽出と分類を行う手法を提案した。実験では、実際の HE 染色された病理画像を用いて細胞核の抽出および分類を行い本手法の有効性を確認した。今後の課題として、抽出の分野では重複している細胞を分離する手法の提案と更なる精度向上、分類の分野では癌のステージを考慮した分類手法の提案やメラノーマ以外の癌に対しての分類手法の提案などが挙げられる。

参考文献

- [1] Yuya Tsukada, et al, "Extraction of Cell Nuclei using CNN Features", Procedia Computer Science, Elsevier, Vol.112, pp.1633-1640, 2017.
- [2] Yuji Iwahori, Yuya Tsukada, et al, "Classification of Cell Nuclei using CNN Features", International Conference on Intelligence Science, Springer, pp.195-208, 2019.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", International Conference on Medical image computing and computer-assisted intervention, Springer, pp. 234-241, 2015.
- [4] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.