

病理画像における癌種領域分割のための Adaptively Weighting Multi-scale FCNの提案

徳永 宏樹^{1,a)} 寺本 祐記^{2,b)} 吉澤 明彦^{2,c)} 備瀬 竜馬^{1,3,d)}

概要：様々ながんの診断において、組織を採取して、顕微鏡で拡大観察する病理診断が行われている。しかし、診断数に対して病理医の数は不足しており、医師の負担は大きい。そこで、本研究では、深層学習を用いた病理画像における癌種ごとの自動領域分割を行う。病理医は実際の診断において、画像をスケール変化させ低解像度で広視野画像と高解像度で狭視野画像の双方を観察して判断する。しかし、深層学習の自動領域分割手法である Fully Convolutional Network(FCN) では、単一視野の入力しか扱えず識別に有用な情報が含まれない可能性がある。そこで、本研究では、複数の視野情報を入力として、画像に応じて視野領域を適応的に組み合わせて利用する Adaptively Weighting Multi-scale FCN を提案する。実験では、提案手法が従来の FCN と比較して、高精度な領域分割を実現することを確認した。

1. はじめに

自動病理画像解析は、病理医の腫瘍および癌のサブタイプ診断を助ける重要なタスクである。これまで、デジタル病理画像から2つの領域（正常および腫瘍）を分割する課題は数多く取り組まれており、高精度の領域分割が実現されてきている [1][2][3]。近年、癌のサブタイプによって患者の予後を予測できる可能性が示されてきており [4]、複数のサブタイプの領域分割を行うことは、病理学的画像解析における重要な課題となってきた。

このような領域分割問題に対して、畳み込みニューラルネットワーク (CNN) [5] や全層畳み込みネットワーク (FCN) [6][8] が広く使われており、様々なタスクにおいて従来の手法よりすぐれた結果を実現している。このようなネットワークにおいては、計算機のメモリサイズの制限により、ネットワークへの入力画像のサイズは制限される。一般には、 256×256 画素程度の画像が用いられることが多く、通常、入力画像はネットワークの入力画像サイズにダウンサンプリングされる。しかし、病理画像診断によく用いられる Whole Slide Tissue Image (WSI) のサイズは、一般の画像と比較して非常に大きく、ダウンサンプリングなしに WSI のパッチ画像が入力画像として利用される場

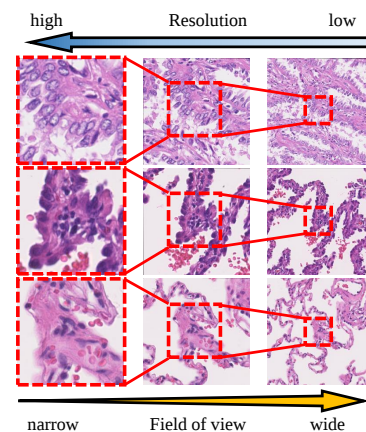


図 1 複数のスケール画像例。上段：Papillary, 中段：Lepidic, 下段：Normal. スケールは左から右にそれぞれ 1, 0.5, 0.25.

合、視野が狭すぎて癌領域のサブタイプを区別できない。一方、視野を広げるために、入力画像をダウンサンプリングすると、図 1 に示すように、空間解像度が低下する。この視野と解像度のトレードオフは、単一の入力画像からサブタイプごとに領域分割することを困難にしている。

一方、病理医は通常、WSI のスケールを変更しながら病理画像を観察し、癌領域を特定する。つまり、狭視野の高解像度画像と広視野の低解像度画像の両方を診断に使用していると言える。さらに、入力画像に応じて識別に重要なスケールが異なると考える。例えば、図 1 では、Papillary (上段) および Lepidic (中段) を区別するためには、広い視野情報が有効である。一方、Lepidic (中段) と Normal (下段) を区別するためには、広視野・低解像度画像ではパターンが類似しているため、狭視野・高解像度画像で、よ

¹ 九州大学大学院 システム情報科学府
² 京都大学医学部附属病院 病理診断科
³ 国立情報学研究所 医療ビッグデータ研究センター
a) hiroki.tokunaga@human.ait.kyushu-u.ac.jp
b) tera1980@kuhp.kyoto-u.ac.jp
c) akyoshi@kuhp.kyoto-u.ac.jp
d) bise@ait.kyushu-u.ac.jp

り詳細部分の情報が重要となる。

Contribution : 本研究では、複数の視野情報を入力として、画像に応じて視野領域を適応的に組み合わせて利用する Adaptively Weighting Multi-scale FCN を提案する。提案手法は複数スケールの画像を適応的に統合するため、識別に高解像度画像と広い視野領域の両方が必要なサブタイプの分類に有効に働くと考えられる。実験では、提案手法を病理学において重要なタスクである肺腺癌のサブタイプ領域分割へ適用し、他の最新の手法と比較してよい性能を示すことを確認した。

関連技術 : FCN は、領域分割を行うために開発された手法で、局所の特徴を抽出するためのダウンサンプリング部と、全体的位置情報を復元するためのアップサンプリング部から構成される [6]。FCN の中でも、U-net[8] は、医用画像解析に広く使用され、高精度な領域分割を実現している。しかし、U-net は単一画像入力であるため、空間解像度と視野の広さのトレードオフは課題として残る。

他の分野において複数のネットワークを統合する手法が提案されている。一般画像内のテキスト領域を抽出するために、スケールの異なる 3 種類の FCN を統合して特徴を抽出するマルチスケールの FCN が提案されている [9]。この手法では、すべての入力画像に対して複数ネットワークを統合する際の重みが固定される。また、群集の計測のために、複数の異なるスケールに特化して訓練された CNN の中から単一の最適なネットワークを選択する Switch-CNN が提案されている [10]。本論文では、複数の癌サブタイプ領域を領域分割するために、これら既存の手法とは異なり、入力画像に応じて複数スケールの画像の特徴を適応的に組み合わせて利用する新規ネットワークを提案する。

2. Adaptively Weighting Multi-scale FCN

図 2 に Adaptively Weighting Multi-scale FCN の構造を示す。空間解像度と視野の広さのトレードオフに対処するために、提案手法では、 i 番目の画像の入力として 3 つの異なるスケール画像 (X_i^1, X_i^2, X_i^3) を用いる。ここで、 X_i^1 は領域分割を行う対象領域の画像パッチであり、最も狭い視野で最も高い解像度を有する。 X_i^2, X_i^3 は X_i^1 のダウンサンプリング画像であり、スケールはそれぞれ 0.5 倍および 0.25 倍である。 X_i^1 を領域分割する際に X_i^2, X_i^3 における対象領域の周辺画像特徴が利用される。提案手法では、特定の画像に応じて 3 つの異なるスケールの重みを適応的に推定し、その重みを用いてこれらの画像特徴を統合し、サブタイプ別に領域分割する。

2.1 ネットワーク構成

図 2 に示すように、Adaptively Weighting Multi-scale FCN は、Expert FCN (f_{E1}, f_{E2}, f_{E3}), Weighting CNN f_W , Integrating FCN f_I の 3 種類のネットワークから構成される。 (f_{E1}, f_{E2}, f_{E3}) は、スケールの異なる画像を領

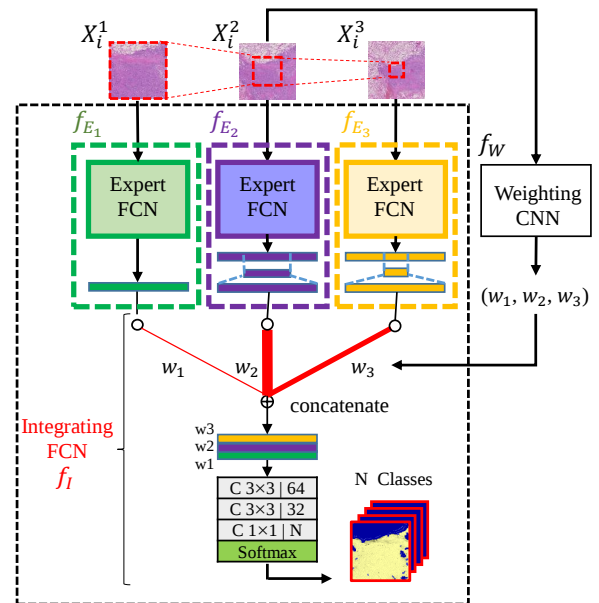


図 2 Adaptively Weighting Multi-scale FCN 構成概要図

域分割するために特化して訓練されており、対象領域における各サブタイプの尤度マップを出力する。 X_i^2, X_i^3 の視野は対象領域と異なるため、各出力尤度マップの X_i^1 に対応する領域が切り取られ、 f_{E1} の出力と同じサイズにアップサンプリングされる。各 Expert FCN の出力は以下のように統合される。

入力画像に応じて識別に重要なスケールの組み合わせが異なることを仮定しており、特定の入力画像に対する Expert FCN の重みを適応的に推定する Weighting CNN を作成する。Weighting CNN には、Google が開発した Xception モデル [11] の出力層を 3 クラスの全結合層に置き換え、解析対象画像を用いて fine-tuning する。

Integrating FCN では、Weighting CNN で推定された重みを使用して、Expert FCN の出力を重み付け結合する。このネットワークは、対象視野領域の最終的な領域分割結果を出力する。ここで、出力チャンネル数はサブタイプのクラス数である。この FCN は、3 つの畳み込み層と softmax 層で構成される単純な構成とする。

2.2 学習法

Algorithm 1 に、学習手順の概要を示す。学習には、 N 枚の訓練パッチ画像 $\{X_i^1, X_i^2, X_i^3\}_{i=1}^N$ と正解マスク画像 $\{D_{i,c}^{GT} | i = 1, \dots, N, c = 1, \dots, M\}$ を各クラスごとに使用する。ここで、 M はサブタイプ数である。上述したように、画像 $\{X_i^1, X_i^2, X_i^3\}$ は、同じ対象領域を含む異なるスケールの画像の組である。異なるスケールの画像についての 3 つの Expert FCN $\{f_{E_k}\}_{k=1}^3$ は、各サブタイプの尤度マップを予測するためにあらかじめ訓練画像を用いて独立に訓練される。それぞれの Expert FCN を訓練するための損失関数として、クラス内の画像数によって重み付けされた categorical cross entropy を使用し、Adam を用いて最適化され、ネットワークパラメータ $\{\Theta_k\}_{k=1}^3$ が更新される。こ

Algorithm 1 Adaptively Weighting Multi-scale FCN training algorithm

```

1: Input:  $N$  training image patches  $\{X_i^1, X_i^2, X_i^3\}_{i=1}^N$  with
   ground truth segmentation map  $\{D_{i,c}^{GT}\}_{i=1, \dots, N, c=1, \dots, M}$  for each class to train all networks
2: % Initialization: Pre-training for expert FCNs  $f_{E_1}, f_{E_2}, f_{E_3}$ 
   independently
3: Backpropagation to train  $\{\Theta_k\}_{k=1}^3$  using  $\{D_{i,c}^{GT}\}_{k=1}^3$  respectively
4: % Training for  $T_d$  epochs
5: for  $t = 1$  to  $T_d$  do
6:   % Generate training data to train Weighting CNN
7:   for  $i = 1$  to  $N$  do
8:      $D_{i,c}^k = f_{E_k}(X_i^k; \Theta_k)$  % output of  $f_{E_k}$  with input  $X_i^k$ 
9:      $w_i^k = 2 |D_{i,c}^k \cap D_{i,c}^{GT}| / (|D_{i,c}^k| + |D_{i,c}^{GT}|)$ ,
        $\vec{w}_i = [w_i^1, w_i^2, w_i^3]$ 
10:   end for
11:    $S_{train} = \{\vec{X}_i, \vec{w}_i\}_{i=1}^N$ 
12:   % training soft switch for 1 epoch
13:   Train Weighting CNN with  $S_{train}$  and update  $\Theta_W$ 
14:   % train the integrated shared networks ( $\{f_{E_k}\}_{k=1}^3, f_I$ )
   with Weighting CNN
15:   for  $i = 1$  to  $N$  do
16:     % estimate the weights  $\vec{w}_i^W$  by Weighting CNN  $f_W$ 
     with the current  $\Theta_W$ 
17:      $\vec{w}_i^W = f_W(X_i^2; \Theta_W)$ 
18:     % train  $\{\Theta_k\}_{k=1}^3$  and  $\Theta_I$  with  $\vec{w}_i^W$ 
19:     Backpropagation to train the integrated shared networks
     ( $\{f_{E_k}\}_{k=1}^3, f_I$ ) with  $\vec{w}_i^W$  and update  $\{\Theta_k\}_{k=1}^3$ 
     and  $\Theta_I$ 
20:   end for
21: end for
22: Output: trained parameters  $\{\Theta_k\}_{k=1}^3$  for  $f_{E_k}$ ,  $\Theta_I$  for the
   Integrating FCN layer  $f_I$  and  $\Theta_W$  for Weighting CNN  $f_W$ 

```

の事前学習によって、各 Expert FCN はそれぞれのスケールの画像に特化される。

これらの初期化されたパラメータを使用して、2種類のネットワーク ($\{f_{E_k}\}_{k=1}^3, f_I$) (図2 黒点線枠) と Weighting CNN f_W が交互に最適化される。はじめに、Weighting CNN を訓練するために、訓練画像 $\{\{X_i^k\}_{k=1}^3, \{D_{i,c}^{GT}\}_{c=1}^M\}_{i=1}^N$ を使用して重みの組の訓練データを生成する。特定のスケール画像が識別に重要である場合、そのスケールに対応する Expert FCN は、他の Expert FCN と比較して高精度な予測を行うと考えられる。そこで、予測結果 $D_{i,c}^k$ と正解マスク画像 $D_{i,c}^{GT}$ との間のダイス係数を Expert FCN の重みとして以下のように定義する。

$$w_i^k = 2 * |D_{i,c}^k \cap D_{i,c}^{GT}| / (|D_{i,c}^k| + |D_{i,c}^{GT}|),$$

$$(k = 1, 2, 3), \quad \vec{w}_i = [w_i^1, w_i^2, w_i^3].$$

訓練画像と重みの組 $S_{train} = \{X_i^2, \vec{w}_i\}_{i=1}^N$ を使用して、Weighting CNN f_W が訓練され、ネットワークパラメータ Θ_W が更新される。最適化には確率的勾配降下 (SGD)、損失関数として平均二乗誤差 (MSE) を使用した。

次に、統合された共有ネットワーク ($\{f_{E_k}\}_{k=1}^3, f_I$) を、

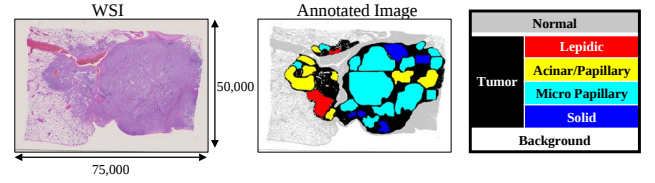


図3 WSI 画像の例. 左: WSI オリジナル画像, 中: 正解マスク画像, 右: ラベル情報

Weighting CNN によって推定された重みを用いて訓練する。中間スケールの訓練画像 X_i^2 に対して、Weighting CNN $f_W(X_i^2; \Theta_W)$ によって、各 Expert FCN の重み $\vec{w}_i^W$ を推定する。推定された重みを使用して統合された共有ネットワークは、損失関数 categorical cross entropy を SGD を用いて最適化することで、End-to-End で訓練される。この学習プロセスを、すべての訓練データ $\{X_i^1, X_i^2, X_i^3\}_{i=1}^N$ に対して反復的に適用する。

この交互学習アルゴリズムは、最大エポックまたは損失関数が収束するまで実行される。この学習アルゴリズムを介して、各 Expert FCN は、複数のサブタイプを領域分割するための対応するスケール画像に特化され、Weighting CNN は、特定の入力画像の Expert FCN の重みを推定するように訓練されることになる。Integrating FCN は、推定された重みを用いて各サブタイプの最終的な領域分割結果を推定するように訓練される。

3. 実験結果

WSI における腫瘍および正常領域の2クラス分割と、肺腺癌における複数のサブタイプ領域分割の2つの領域分割課題による評価実験を行った。領域分割の精度を、提案手法と関連が深い以下の3つの最新の手法と比較した。医療画像の領域分割に広く用いられている U-net[8]、複数ネットワークから最適ネットワークを選択する Hard Switch FCN (HS) [10]、複数スケールのネットワークの出力を固定ウェイトで統合する Multi-scale Shared Net (MSN) [9] との比較実験を行った。

実験では29枚のWSIを使用した。WSIは、スライスされた肺腺癌組織を染色し、最大倍率40倍のバーチャルスライドスキャナによって撮影することによって作成される。画像サイズは、最大で $54,000 \times 108,000$ 画素である。訓練およびテスト画像を作成するために、病理医によって、癌サブタイプ (1. Normal, 2. Lepidic, 3. Acinar/Papillary, 4. Micro-papillary, 5. Solid) の領域にアノテーションが行われた。ここで、Normal は腫瘍外領域を示し、他の4つのクラスは腫瘍のサブタイプを示す。複数のサブタイプ領域を領域分割するために、はじめに腫瘍領域を2クラス領域分割し、次に、腫瘍領域と推定された領域からサブタイプ領域分割を行った。

図3に、WSI および対応する正解マスク画像を示す。腫瘍領域には、病理医が特定のサブタイプに分類することが

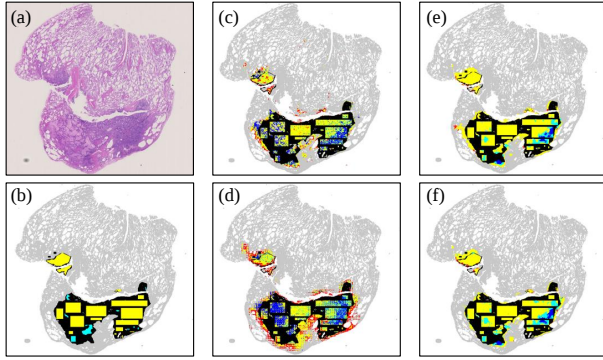


図 4 テスト画像の領域分割結果例 (a) オリジナル画像, (b) 正解マスク画像, (c) Unet[8], (d) HS[10], (e) MSN[9], (f) 提案手法.

表 1 比較手法の精度評価

	2 class			
	Unet[8]	HS[10]	MSN[9]	Proposed
OP	0.939	0.937	0.942	0.959
JI	0.852	0.857	0.860	0.900
DC	0.917	0.922	0.923	0.947
	5 class			
	Unet[8]	HS[10]	MSN[9]	Proposed
OP	0.917	0.882	0.934	0.936
JI	0.358	0.330	0.441	0.448
DC	0.456	0.424	0.558	0.563

難しい曖昧な領域が存在する。これらの領域は黒で示されており、マルチクラス領域分割の精度を評価するにはこれらの領域を対象外とした。

Adaptively Weighting Multi-scale FCN を訓練するために、図 1 に示すように、同じ領域に対応する 3 つの異なるスケールのパッチ画像 $\{X_i^1, X_i^2, X_i^3\}_{i=1}^N$ の組を抽出した。画像サイズは 256×256 画素、スライディングウィンドウの移動幅は 256 画素、スケールは $(0.5, 0.25, 0.125)$ とした。対応する正解マスク画像 $\{D_{i,c}^{GT} | i = 1, \dots, N, c = 1, \dots, M\}$ は 1 画素単位でラベルデータとして使用される。ここで N は 108,842 枚のパッチ画像である。訓練データ用に 25 枚の WSI をランダムに選択し、テストデータ用に 4 枚の WSI を使用した。データセットは、すべての比較方法の評価で共通したものを使用した。

図 4 に、サブタイプ領域分割結果の例を示す。U-net と HS[10] の結果では、図 4(c)(d) に示すように断片化した領域が多い。これらの結果と比較して、MSN[9] (図 4(e)) および提案手法 (図 4(f)) がよい結果を示している。

また、Overall Pixel (OP) 精度、Jaccard Index (JI) の平均、および Dice Coefficient (DC) の平均の 3 つの指標を用いて定量的に評価した。これらの評価指標は、次のように定義される。

$$OP = \frac{\sum_c TP_c}{\sum_c (TP_c + FP_c)}, JI = \frac{1}{M} \sum_c \frac{TP_c}{(TP_c + FP_c + FN_c)},$$

$$DC = \frac{1}{M} \sum_c \frac{2TP_c}{((TP_c + FP_c) + (TP_c + FN_c))}. \quad (1)$$

ここで、 TP_c , FP_c , FN_c はそれぞれクラス c の真陽性、

偽陽性、および偽陰性の数である。単一のネットワークを使用する Unet および HS[10] と比較すると、MSN[9] および提案手法はすべての評価指標において上回る結果となった。この結果から、複数スケールの画像と比較して、単一画像では複数のサブタイプを識別するための情報が不足していると考えられる。また、2 クラスと複数サブタイプの領域分割で全ての評価指標について、提案手法が最もよい精度となった。これは、Weighting CNN が入力画像に応じて異なるスケールの画像から画像特徴を適応的に用いることで、他の方法に比べて精度が向上したためと考えられる。

4. まとめ

本論文では、肺病理画像を癌サブタイプ毎に領域分割するために、入力画像に応じて異なるスケールの画像から画像特徴を適応的に組み合わせて使用する Adaptively Weighting Multi-scale FCN を新規に提案した。提案手法は、最新の領域分割手法と比較して精度を改善した。今後の研究では、画像スケールの重要度とサブタイプの種類との関係を調べる。

参考文献

- [1] Xu, Y., Jia, Z., and Wang, L.B., *et al.*; Large Scale Tissue Histopathology Image Classification, Segmentation, and Visualization via Deep Convolutional Activation Features, BMC Bioinformatics, 18(1):281, DOI 10.1186/s12859-017-1685-x (2017).
- [2] Xu, Y., Li, Y., and Liu, M., *et al.*; Gland Instance Segmentation by Deep Multichannel Side Supervision, MICCAI, vol. 9901, pp. 496-504 (2016).
- [3] Chen, H., Qi, X., Y, L., and Heng, P.A.; DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation, CVPR, pp.2487-2496 (2016).
- [4] Yoshizawa, A., Motoi, N., and Riely, G.J., *et al.*; Impact of Proposed IASLC/ATS/ERS Classification of Lung Adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases, Modern Pathology, 24, pp.653-664 (2011).
- [5] Krizhevsky, A., Sutskever, I., and Hinton, G.E.; ImageNet Classification with Deep Convolutional Neural Networks, NIPS, vol.1, pp.1097-1105 (2012).
- [6] Long, J., Shelhamer, E., Darrell, T.; Fully convolutional networks for semantic segmentation. CVPR. pp. 3431-3440 (2015).
- [7] Hou, L. Samaras, D., and Kurc, T.M., *et al.*; Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification, CVPR, pp.2424-2433 (2016)
- [8] Ronneberger, O., Fischer, P., and Brox, T.; U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI, 9351, pp.234-241 (2015).
- [9] He, D., Yang, D., and Liang C., *et al.*; Multi-scale FCN with Cascaded Instance Aware Segmentation for Arbitrary Oriented Word Spotting In the Wild, CVPR, pp.3519-3528 (2017).
- [10] Sam, D.B., Surya, S., and Babu, R.V.; Switching Convolutional Neural Network for Crowd Counting, CVPR, pp.5744-5752 (2017).
- [11] Chollet, F.; Xception: Deep learning with depthwise separable convolutions, CVPR, pp.1251-1258 (2016).