

脂肪カテゴリ導入によるがん診断支援システムの性能向上の検討

馬淵 凌[†] 近藤 日向子[†] 早津 祐真[†] 浅生田 良[†] 大同 海良[†] 多田 賢一郎[†] 堀内 晴貴[†] 金道 敏樹[†]

[†]金沢工業大学 工学部 情報工学科

1. はじめに

近年,病理専門医の人手不足が深刻化している.その要因として,がん患者とその検査数の増加が挙げられ,病理専門医の負担軽減へ向けた取り組みが検討されている.その代表例として,CNN(畳み込みニューラルネットワーク)などの人工知能を使った病理画像診断技術の研究がある.時に医師と同等かそれ以上の診断精度を示すといった成果を上げている.その一方で,がん評価の判定根拠と教師データのアノテーション精度が臨床において十分でないという課題も見られ,医師に求められるレベルには至っていない.

本研究では,"医師によるがん診断"を支援する技術の開発に CAMELYON16[1]のデータ(教師データ)を使用して取り組んでいる.これまでに,「がんらしさ」を情報量の密度の多寡で可視化する情報密度法[2,3,4]を提案し,がん領域と正常領域を区別可能であるという結果を得ている.一方,正常領域内の脂肪領域や血管を含む領域を「がんらし」とする技術的課題も見られた.

本研究では,課題のひとつである脂肪領域の輝度が細胞領域よりも高いことに注目した.輝度分布を使用した新たな脂肪検出アルゴリズムを作成することで,情報密度法の課題のひとつを間接的に解決するとともに,このアルゴリズムが診断効率の向上に貢献できることを示す.

2. アプローチ手法

我々は,脂肪領域と細胞領域を見分ける判断材料として,明るさの違いに注目した.そこで,各色の明るさを表す輝度を用いて研究を行った.以下の研究では,病理画像から250×600枚に切り取られたマス画像(384×360px)を使用する.

手法として,マス画像を100(10×10)分割する.100分割された小片ごとにピクセル単位で輝度を取得し,平均値を算出する.すると,ひとつのマス画像に対し,100個の輝度からなるデータを作成できる.この輝度データ群をもとにヒストグラム(Fig.1)を作成した.すると,脂肪領域と細胞領域で輝度分布に偏りが見られることがわかった.

これにより,輝度による脂肪領域と細胞領域の分類は可能であると考え,研究に取り組んだ.

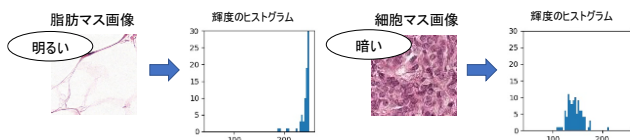


Fig.1: マス画像の輝度分布(左:脂肪, 右:細胞)

3. 輝度分布でのクラスタ分析

ここでは,輝度分布が脂肪の分類に有効であることを示す.

マス画像から得た輝度データを使用してクラスタ分析を行った.使用したマス画像は,連続した範囲の10,934枚に正常領域とがん領域の教師データ各1,000枚を加えた合計12,934枚である.

クラスタ分析で得られた結果のデンドログラムとその内訳を示す.(Fig.2, Table1)赤色の枠内には脂肪領域のマス画像が分類されている.赤色の枠に分類されたマス画像には,がん領域の教師データは含まれていない.これにより,輝度分布を使用して脂肪を識別することは可能であることがわかる.

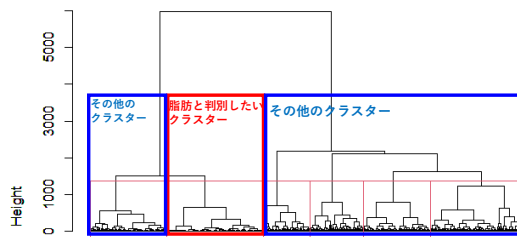


Fig.2 クラスタ分析結果(12,934枚)

Table1 クラスタ分析結果の内訳

	ガン	合計
赤枠のクラスタ	0	2,868
青枠のクラスタ	1,000	10,066

4. 脂肪検出パラメータの決定

これまで,輝度分布の違いが脂肪領域とがん領域を分類するための基準になり得ることを示した.ここでは,最初の着眼点である脂肪領域の輝度が細胞領域よりも高いという特徴に立ち戻る.

がんと正常の教師データである各1,000枚のマス画像からなる輝度データ群のうち,輝度230以上の小片が75個以上あるマス画像にがん領域は含まれていなかった.つまり,輝度と小片の個数のふたつをパラメータとして適切な値を算出できれば,脂肪を識別するためのロジックになると考えた.脂肪かどうかの評価基準を決めるパラメータの作成手順を以下のように算出した.

対象のマス画像内で輝度 θ_y 以上の小片の数を数える.
ある小片を

$$F_{ab} = \theta(Y_{ab} - \theta_y)$$

とする.ここで,

$$\theta(x) = \begin{cases} 1 & x \geq 1 \\ 0 & x < 0 \end{cases}$$

$$Y_{ab} = \int_{x_a}^{x_a+\Delta x} \int_{y_b}^{y_b+\Delta y} Y_{ab}(x, y) dx dy / \int_{x_a}^{x_a+\Delta x} \int_{y_b}^{y_b+\Delta y} dx dy$$

Introduction of the third category, fat, improves the accuracy of the cancer diagnosis support system

[†]Ryo Mabuchi, Hinako Kondo, Yuma Hayatsu, Ryo Asoda, Kaira Daido, Kenitirou Tada, Haruki Horiuchi, Toshiki Kindo

[†]Department of Information and Computer Science, College of Engineering, Kanazawa Institute of Technology

となり,マス画像内の 100 個の小片を 0,または 1 で表現できる.すべての小片を足し合わせると

$$N = \sum_{a=1}^{10} \sum_{b=1}^{10} \theta(Y_{ab} - \theta_y)$$

と表すことが出来る.ここで,ある輝度 θ_y 以上の小片の和 Y_{ab} を以下レートと呼ぶ.

上記の計算を脂肪領域,がん領域の各 1,000 枚の合計 2,000 枚で行い,輝度 0~255,レート 0~100 のすべての組み合わせで脂肪とがんの検出数のデータを取得した.このデータをもとに Confusion Matrix (Fig.3)を作成した.作成した Confusion Matrix をもとに,評価手法を以下のように算出した.

$$\begin{aligned} \text{適合率: Precision} &= \frac{TP}{TP+FP} \\ \text{再現率: Recall} &= \frac{TP}{TP+FN} \\ \text{F 値: F} &= \frac{2RP}{R+P} \end{aligned}$$

本研究では,病理専門医が効率よくがんの有無を判定できることを目的としている.そのため,脂肪の未検出とがんの誤検出をバランスよく減らせる F 値を使用し,取得数のデータ内で F 値が最大となった輝度 239,レート 43 を以下の研究のパラメータとした.

		予測	
		脂肪 (Positive)	ガン (Negative)
実際	脂肪 (Positive)	脂肪を脂肪と判定 TP 1000	脂肪をガンと判定 FN 0
	ガン (Negative)	ガンを脂肪と判定 FP 6	ガンをガンと判定 TN 994

Fig.3 脂肪とがんの Confusion Matrix

5. がんと誤認識される脂肪領域の除去

決定したパラメータを昨年度までの研究結果に加える.昨年までの研究結果にてがんと誤認識されていた脂肪領域の色付けがどのように変化したかを示す.(Fig.4) 3 枚の画像のうち,左は元画像であり青色の枠内が教師データ上のがん領域である.中央は昨年度までの研究結果のひとつである「がんらしさ」を色付けした画像である.正常領域であるはずの脂肪領域を「がんらしい」と色付けしている.右は中央の画像に加えて,作成した脂肪検出アルゴリズムに基づいて脂肪と検出された領域を灰色に色付けしたものである.色付けされた画像を比較すると,今まで「がんらしい」とされていた脂肪領域の大部分を取り除けることがわかる.

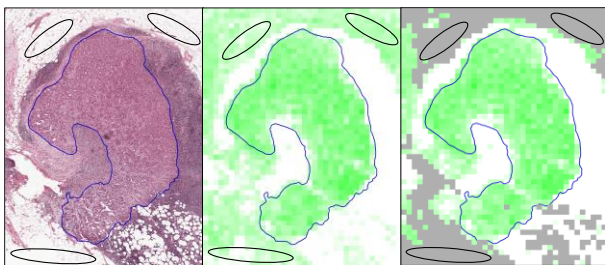


Fig.4 脂肪領域の除去(CAMELEYON16[1])

6. 病理スライドでの脂肪領域の除去

実際に診断に用いる病理スライドで脂肪を除去した場合の色付け結果とその内訳を示す.(Fig.5, Table2)左が元画像で,右が脂肪と検出された領域を灰色に色付けした画像である.150,000 (250×600) 枚のマス画像からなる元画像のうち,輝度 255,レート 100 に該当する領域を脂肪と細胞のどちらでもない余白部分とし,診断領域には含めないこととした.余白を除いた 63,649 枚の診断領域のうち 49,107 枚を脂肪として検出しており,約 77%の脂肪を除去できることになる.つまり,病理専門医の診断範囲は残りの約 20%となる.また,他の病理スライド画像でも同様に,77%前後の脂肪を除去できることがわかっている.

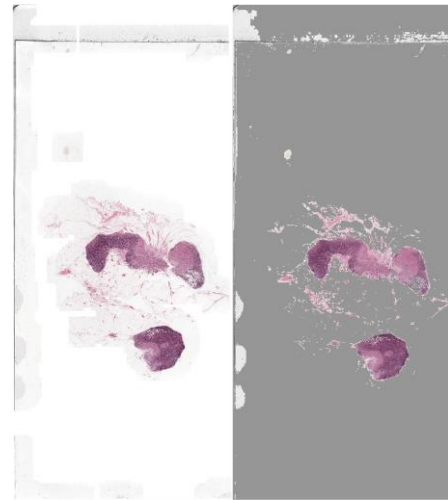


Fig.5 全体に対する脂肪領域の除去 (CAMELYON16[1])

Table2 病理スライド上での脂肪除去結果の内訳

	脂肪	診断領域	割合(脂肪/診断領域)
F値最大:(239,43)	49,107	63,649	77.2

7. 総括

本研究では,輝度を用いて脂肪検出アルゴリズムを作成し,正常領域内の脂肪領域を「がんらしい」とする情報密度法の技術課題のひとつを解決できることを示した.同時に,病理スライド上の診断範囲を約 20%に軽減できることを確認した.

一方で,血管を含む領域での識別精度が十分でないという課題も残されている.この領域に対しても,脂肪領域ほどではないが輝度分布の偏りが見られることがわかっている.今後も本研究で開発した手法を応用し,さらなる精度向上に努めていく.

参考文献

- [1]Ginneken van B. et al., “CAMELYON16”, 2016.
- [2]田村一希, 田島和征, 金道敏樹: “情報量を用いた画像特徴量解析とそのガン画像診断への応用” 情報処理学会第 83 回全国大会
- [3]陸田駿弥, 鎌田千晶, 佐藤大樹, 他: “情報量を用いたがん画像診断—特徴量の特性解析による性能向上の検討—” 情報処理学会第 84 回全国大会
- [4]金道敏樹, 陸田駿弥, 山田壮亮: “病理画像のがんらしさと正常らしさを情報量で評価する情報密度法”, MEDICAL IMAGING TECHNOLOGY(in press)