

局所特徴量が担う情報量に基づくがん領域の高解像度可視化

陸田 駿弥[†] 山田 壮亮[‡] 金道 敏樹[†]金沢工業大学 工学部 情報工学科[†] 金沢医科大学 臨床病理学[‡]

1. 研究の背景と目的

国立がん研究センターによると、近年、日本では高齢者の人口増加とともにがんの患者とその検査数が増加している。医療診断の現場では、そのすべての確定診断を約2,500人の病理専門医が行っており、人手不足が深刻化している。そこで我々は、“医師によるがん診断”を支援する病理画像診断技術の開発に取り組んでいる。

現在、AIによる病理画像診断は注目を集めている。2016年に行われた乳がんのリンパ節転移の検出精度を競う国際コンペティション (CAMELYON16[1]) で上位の成績を収めた手法は、いずれも Deep Learning を用いたモデルであり、Deep Learning はがん領域の検出において一定の有効性があることが示されている。しかし、我々は、Deep Learning を用いた手法は、医師が結果を信頼するには以下の二つの問題があると考えている。一つ目の問題は、がん領域の評価とその判断根拠となる特徴量を明確に示せない点である。CNN の判断根拠の可視化として、Ramprasaathら[2]によって Grad-CAM が考案されているが、根拠となる特徴自体を示せていない。二つ目は、教師データの問題である。Deep Learning を用いた手法では、大量の画像とともに、正確にアノテーションされた情報が必要である。しかし、そのようなデータを用意するのは困難で、CAMELYON16で提供された学習用データのアノテーション精度は、がん細胞が顕著な領域に限定されていることが我々の先行研究[3, 4]によって指摘された。

したがって、本研究では、境界線情報を必要とせず、判断根拠となる特徴量が明らかな情報密度法 [3, 4]に基づく病理画像診断技術を提案する。また、本手法で、CAMELYON16の病理画像データを用いて、がん領域を高解像度で可視化できることを示す。

2. がん領域の高解像度可視化に向けた提案手法

我々が提案する情報密度法を用いた病理画像診断は、局所特徴量が担う情報量を算出することで、個々の特徴量のがんらしさを定量化し判断根拠としている。そこで、がん領域を高解像度に可視化するには、病理画像小片から高密度に局所特徴量を検出し、それら全てのがんらしさを評価する必要があると考えた。

2.1. 高密度な SIFT 特徴検出

本研究では、約 100 μm 四方の病理画像小片からおおよそ 4,000個の SIFT 特徴量を検出して、がんらしさの定量化・判断根拠となる特徴量の可視化に用いた。なお、検出には OpenCV-v4.5.4 の SIFT 特徴検出アルゴリズム、標準パラメータ (特徴検出数: 最大値) を用いた。

がん領域と正常領域それぞれの病理画像小片に対して、SIFT 検出を行った結果を Fig. 1 に示す。図より、病理画像中に局所特徴量が高密度に分布していることが分かり、高解像度な可視化が可能であることが示唆される。

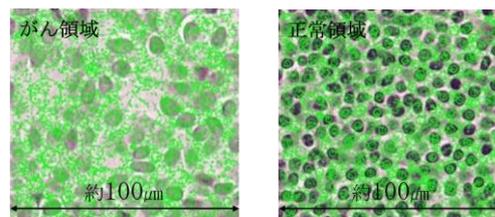


Fig. 1 高密度に SIFT 特徴検出した結果 (左: がん, 右: 正常)

2.2. 特徴量のクラス分割による計算量削減

前章の、高密度な SIFT 特徴検出により、評価対象の画像特徴が大幅に増え、計算量を削減するための工夫が必要となった。そこで、特徴マッチングのパターン削減を見込み、特徴量のサイズの違いに注目した。

がん領域と正常領域の病理画像小片を各 100 枚 (標本画像) 用意し、それぞれの領域で検出される特徴量のサイズによる偏りがあるか確認したところ、各領域から検出される特徴量のサイズには偏りが確認できた (Fig. 2)。これは、がん細胞と正常細胞は形状が異なり、細胞核などの大きさが違うという医学的知見と一致している。

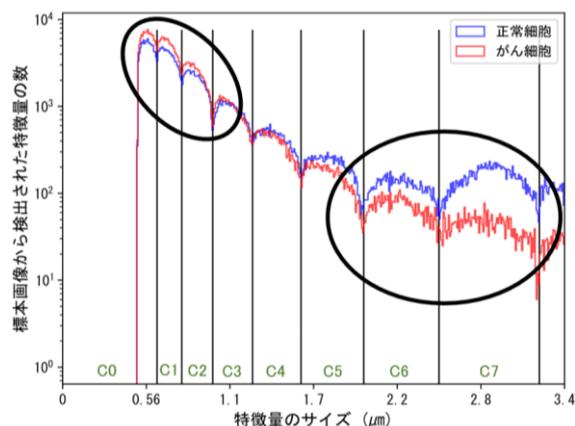


Fig. 2 検出された特徴量のサイズ分布 (赤: がん, 青: 正常)

そこで、本来、SIFT 特徴量はスケール変化に不変であるが、特徴量のサイズを考慮して、Fig. 2 のように 10 個のクラス (C0, C1, ...) を作成し、特徴マッチングの計算対象を同一クラス内に限定することで、特徴マッチングのパターンを 1/100 程度まで削減した。なお、これらの局所特徴量が担う情報量の算出は、別途報告予定の識別情報量評価手法を使って、がんらしさを定量化した。

こうして、高密度な SIFT 特徴検出と特徴量のサイズを考慮したクラス分割による計算量削減で、検出した全て局所特徴量が担うがんらしさを定量化し、可視化を行う。

Visualization of cancer areas in high resolution based on the information density of local feature

[†]Shunya Mutsuda, [‡]Sohsuke Yamada, [†]Toshiki Kindo

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

[‡]Department of Clinpath Pathology, Kanazawa Medical University

3. CAMELYON16 病理画像データへの適用実験

2章では、情報密度法に基づく、がん領域の高解像度可視化を実現する具体的な手法を提案した。ここでは、本手法を CAMELYON16 の病理画像データに適用させ、その有効性を検証する。

Fig. 3 中のブロックマップ（緑色）可視化は、病理画像、特徴量のサイズ、標本の識別情報量を用いて行った。まず、対象画像を病理画像小片に分割し、高密度で SIFT 特徴量を検出する。次に、ある画像特徴 f と最も類似度の高い代表的な画像特徴を、特徴量サイズと特徴マッチング計算により決定し、マッチングした代表的な画像特徴の識別情報量をその画像特徴 f のがんらしさとする。これを検出されたすべての画像特徴に対して行い、病理画像小片単位で識別情報量の総和を算出し、その結果を色付けの濃淡（がんらしさの判断根拠の強さ）として表現した。

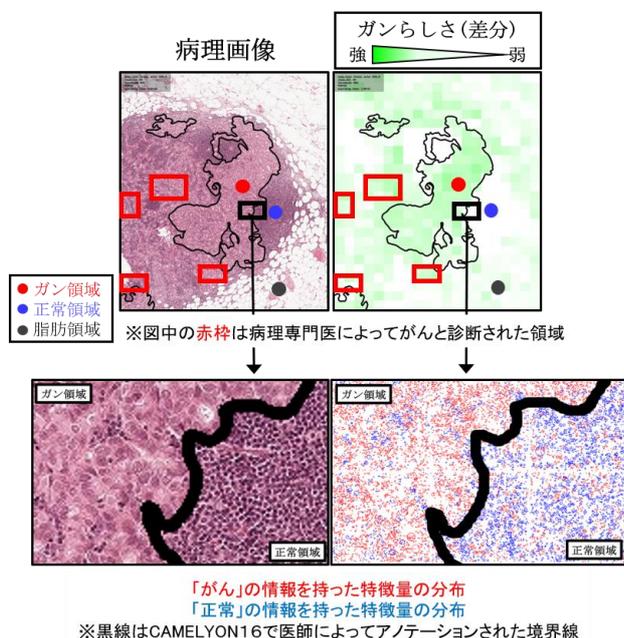
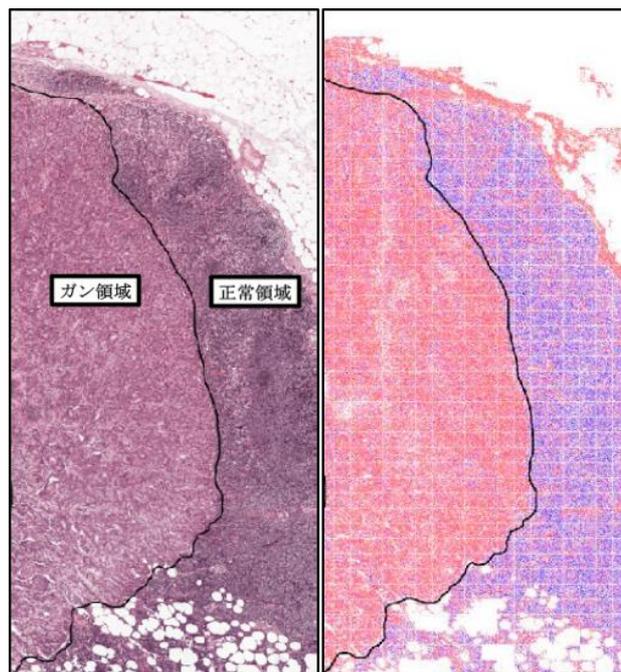


Fig. 3 ブロックマップの色付け結果および判断根拠となる特徴量を可視化した拡大像（赤: がん, 青: 正常）

こうして作成した図（Fig. 3 右上）は、情報密度法が個々の画像特徴 f が担う情報量の算出に、がんと正常を分ける境界線情報を用いない方法であるにも関わらず[3, 4], CAMELYON16 のがん領域のアノテーションを良く再現できていることを示している。それに加えて、アノテーション外であるが、病理専門医によってがんと診断された赤枠領域（がん細胞と正常細胞が混在している領域）に対しても本手法は、がんらしいと判断することができている。そして、本手法の判断根拠である局所特徴量を色付け（赤: がん特徴, 青: 正常特徴）した拡大像を Fig. 3 右下に示した。図より、がん領域・正常領域の判断根拠となる特徴量の分布（密度）には明確な違いがみられ、その境界線は CAMELYON16 のアノテーションに概ね一致していることがみてとれる。また、Fig. 3 とは別の領域で、より広範囲に判断根拠となる特徴量の分布を可視化した結果（Fig. 4 右）でも同様に、CAMELYON16 で医師によってアノテーションされた教師データの境界線を再現できていることが確認できる。



「がん」の情報を持った特徴量の分布
「正常」の情報を持った特徴量の分布

※黒線はCAMELYON16で医師によってアノテーションされた境界線

Fig. 4 別領域かつより広範囲で特徴量を可視化した結果

以上のことから、高密度に検出した非常に多くの SIFT 特徴量を用いることで、がん領域を高解像度で可視化できることが確認できた。これによって、局所特徴量はそれ単独でがん細胞と正常細胞を識別する情報を意外にも多く持っていることが示唆される。

4. まとめと今後の展望

本研究では、境界線情報を必要とせず、判断根拠となる特徴量が明らかな情報密度法に基づく病理画像診断技術の開発に取り組んだ。高密度な SIFT 特徴検出と特徴量のサイズを考慮して計算量削減の工夫をした本手法は、CAMELYON16 の病理画像データにおいて、教師データの境界線を高い精度で再現でき、がん領域と正常領域を高い解像度で見分けることに成功した。このことから、我々が提案する情報密度法に基づく病理画像診断の有効性を確認することができた。

今後は、医療現場から提供される画像データを用いて、様々な症例へ適用させることで、我々が提案する情報密度法に基づく病理画像診断技術の汎用性と臨床診断レベルの有効性を確認し、本手法の確立を目指す。

参考文献

[1] Ginneken van B. *et al.*, “CAMELYON16”, 2016.
 [2] Ramprasaath R. *et al.*, “Visual Explanations from Deep Networks via Gradient-based Localization”, ICCV’17, 2017.
 [3] 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”, IPSJ-Z83, 2021.
 [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, 2022.