

情報理論を用いたガン画像診断

-特徴量の特性解析による性能向上の検討-

陸田 駿弥[†] 鎌田 千晶[†] 佐藤 大紀[†] 米谷 吏[†] 近藤 日向子[†] 早津 祐真[†] 馬淵 凌[†] 金道 敏樹[†]

[†]金沢工業大学 工学部 情報工学科

1. 研究の背景と目的

近年、ガン患者の数とその検査数の増加に伴い、ガン診断医の負担が増加している。そこで本研究では、“医師によるガン診断”の支援を目的としたガン画像診断支援システムの開発を目指している。

昨年の報告では、各画像特徴が担う情報量に注目する情報密度法を提案し、ガン画像診断コンペティション (CAMELYON16 [1]) で用意された画像データと SIFT 特徴量を用いて、識別空間のみで特徴量を選択 (識別空間選択) し、ガンと疑わしい領域を概ね特定できることを示した [2]。しかし、識別空間選択では、画像の一定の領域のガンらしさは示せても、ガンの典型的な特徴量は何かについて答えることはできない。

そこで本研究では、特徴量空間における特徴量選択 (特徴量空間選択) により、ガン特有の特徴量領域を見出すことで、色付けの高精度化を試みるとともに、その結果に対する医師による見解について述べる。

2. ガン画像診断支援システムの概要

本研究で提案するガン画像診断支援システムは、STEP1: 情報密度法による画像特徴の定量的な評価、STEP2: 標本画像特徴群の選択、STEP3: ガンらしさの色付けで構成されている (Fig. 1)。

STEP1: 情報密度法

まず、ガン・正常領域の標本画像を準備し、各画像から SIFT 特徴量 (画像特徴) を抽出する。次に、各画像特徴とガン領域の画像特徴群および正常領域の画像特徴群の両者をマッチングし、各標本画像とのマッチング数を求める。さらに、それぞれのマッチング数とカルバック・ライブラー情報量の概念 [3] を用いて、各画像特徴の情報量 (SKC 値) を算出する。SKC 値は、各画像特徴がガン・正常を識別する情報量をどれだけ持っているかを表現している。

STEP2: 標本画像特徴群の選択

我々は、ガン特有の特徴量領域を特定するために、画像特徴の類似性を考慮する必要があると考え、クラスタ分析を行った。なお、クラスタの結合にはウォード法、距離計算にはユークリッド距離を用いた。その後、各クラスタを、識別空間においてガンらしい画像特徴群と正常らしい画像特徴群に分け、標本画像特徴群を選択した。

STEP3: ガンらしさの色付け

色付けは、対象画像、選択した画像特徴群とその SKC

値を用いて行う。まず、対象画像をマス画像に切り出し、各マス画像の画像特徴を抽出する。次に、抽出した画像特徴と標本画像特徴群をマッチングし、各マスの合計 SKC 値を算出する。その合計値は、“ガンらしさ”、“正常らしさ”の強度を表し、多寡に応じて色付けの濃淡を変えることで色付けマップを作成する。さらに、“ガンらしさ”から“正常らしさ”を除いた情報差分を計算することで、ガンらしさ (差分) マップを作成する。

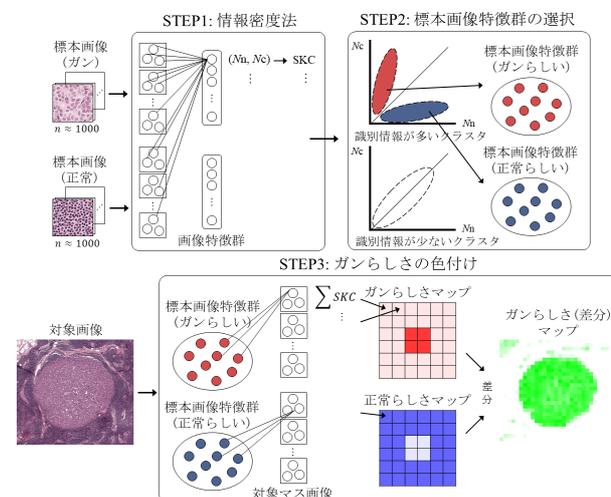


Fig. 1: ガン画像診断支援システムの概要

3. 特徴量空間におけるクラスタ分析

クラスタ分析を行い、ガン領域・正常領域の画像特徴群 (各 40 万個) をそれぞれ 1600 個のクラスタに分類した。この内、代表的なクラスタの識別空間における画像特徴分布を Fig. 2 に示す。ここで、横軸は正常領域の標本画像とのマッチング数 (N_n)、縦軸はガン領域の標本画像とのマッチング数 (N_c) である。各クラスタは、ガンらしい画像特徴群 (cluster 1157)、正常らしい画像特徴群 (cluster 199)、識別情報が少ない画像特徴群 (cluster 296, cluster 638) の代表例である。

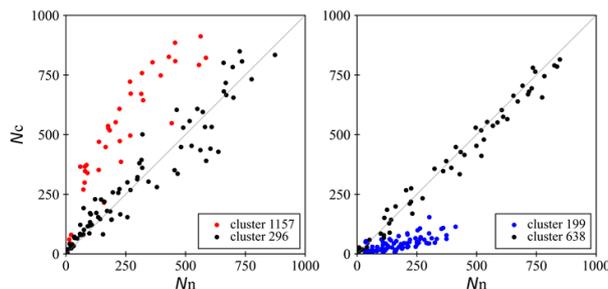


Fig. 2: クラスタに含まれる画像特徴群の分布 (一例) (左: ガン, 右: 正常)

Diagnose cancer images using information theory
-Examine performance improvement by feature analysis-
[†]Shunya Mutsuda, Chiaki Kamada, Daiki Sato, Tsukasa Yonetani,
Hinako Kondo, Yuma Hayatsu, Ryo Mabuchi, Toshiki Kindo
Department of Information and Computer Science,
College of Engineering, Kanazawa Institute of Technology

ガン・正常を識別する情報が顕著なクラスタを見積もるために、横軸にクラスタ内の画像特徴の共分散 (Cov)、縦軸に原点とクラスタの中心を結んだ直線のなす角度 (Phi) を取った散布図 (Fig. 3) を作成した。ガン・正常を識別する情報が顕著なクラスタとして、 $\Phi \geq 60$ 度および $Cov \leq 20,000$ のクラスタ (赤点群) に含まれる画像特徴をガンらしい標本画像特徴群、 $\Phi \leq 25$ 度および $Cov \leq 20,000$ のクラスタ (青点群) に含まれる画像特徴を正常らしい標本画像特徴群として選択した。

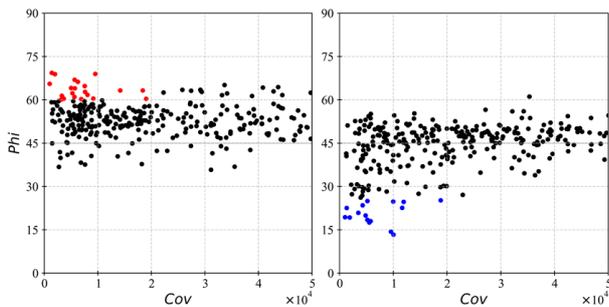


Fig. 3: クラスタの分布 (一部)
(左: ガン, 右: 正常)

4. 選択した標本画像特徴群の比較

今年 (2021) と昨年 (2020) で、識別空間において選択した標本画像特徴群を比較した結果を Tab. 1 および Fig. 4 に示す。これらより、画像特徴の類似性を考慮したことで、昨年選択した画像特徴を概ね包含する標本画像特徴群を選択できている。しかし、性能向上のための $\Phi=45$ 度付近の特徴を減らす絞込みで課題が残る。その原因は、特徴量空間内の分布のいびつさであることが分かりつつある。

Tab. 1: 標本画像特徴群の比較 (数)

	Cancer		Normal	
	Cluster	Feature	Cluster	Feature
2021	107	24314	72	19894
2020	-	17270	-	8020

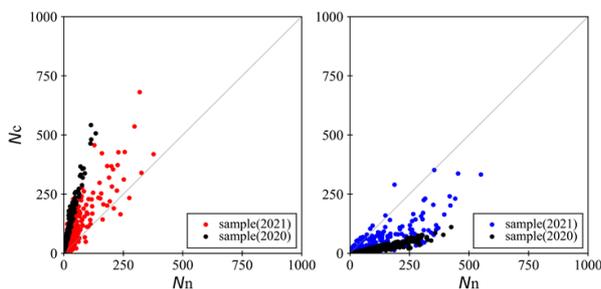


Fig. 4: 標本画像特徴群の比較 (分布)
(左: ガン, 右: 正常)

5. 色付け結果の比較

今年 (2021) と昨年 (2020) の色付けマップの比較を Fig. 5 に示す。図より、今年の標本画像特徴群から、昨年の結果と同程度の色付け結果が得られており、特徴量空間での特徴量選択が可能であることが分かった。また、正常らしさの色付けにおいては、ガン領域や非細胞領域

での色付け強度が低下していることが分かった。これは、正常細胞の画像特徴がガンのそれに比べて顕著であり、特徴量空間で正常特有の領域を概ね特定できたことを示唆している。

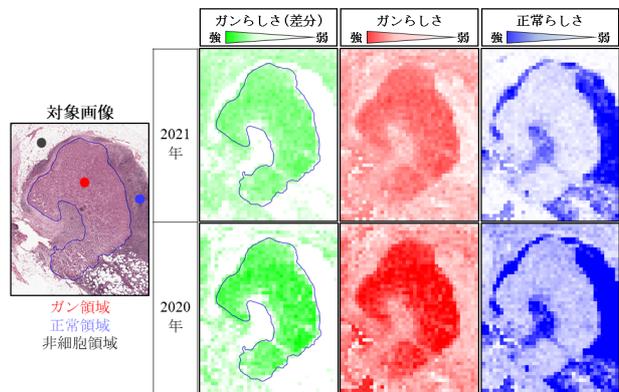


Fig. 5: ガン・正常らしさマップの比較

6. 医師による診断結果と整合性

本研究では、教師ラベルが“正常領域”としている領域で色付けマップによりガンらしいと識別されている領域 (Fig. 6 内の黒枠 4 領域) に対して、ガンの確定診断を専門とする臨床医に診断協力を仰いだ。その結果、下図の 4 領域すべてにおいて“ガン領域である”と診断を得た。このことから、本システムは教師ラベルの誤りを特定する色付け精度およびガン領域の誤診を軽減する可能性を有することが示唆される。

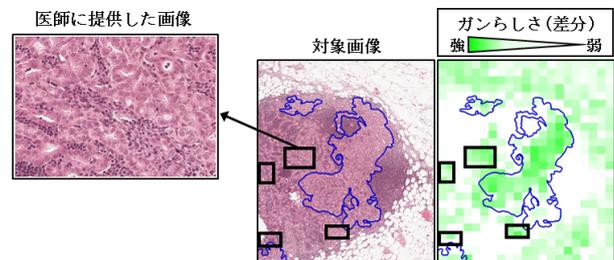


Fig. 6: 教師ラベルと医師による診断が異なる領域 (一部)
*青枠は CAMELYON16 [1] の教師ラベル (ガン領域)

7. まとめと今後の展望

本研究では、特徴量空間で特徴量選択を行い、情報密度法を用いて、ガンと疑わしい領域を概ね特定できることを確認した。今後は、さらなる特徴量空間の解析を行い、画像特徴の共起を考慮したクラスタ分析により、ガン特有の特徴量領域を特定し、色付け精度を向上させる予定である。

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”, IPSJ-Z83, 2020.
 [3] Kindo T. et al., “Adaptive Information Filtering System that organizes personal profiles automatically”, Proc. of 15th IJCAI97, pp.716-721, 1997.