

# 視線データ分析による内視鏡医熟練度検出の検討

梅田 幹大<sup>1</sup> 長松 隆<sup>2,a)</sup> 廣江 葵<sup>2,3,4,b)</sup>

**概要：**近年、視線データを用いた技能伝承方法が検討されている。本研究では、熟練度の可視化・形式化を目的として、内視鏡医の専門スキルの共通部分を視線データから探索した。実験では、視鏡画像10枚と内視鏡画像ではない画像10枚計20枚を刺激として、内視鏡医2名と一般学生3名の計5名の視線を計測した。得られた視線データのヒートマップに対して、画像構造の比較を行う指標であるSSIMを用いることで、内視鏡医2名が内視鏡医画像を見ている時の視線分布の類似度が高いことを示した。

## 1. はじめに

近年、少子高齢化が日本のみでなく多くの先進国で大きな問題になっている。こうした状況の中で、多くの職業において人材が不足し、高度な技能や知識の伝承は重要な課題となっている。この課題に対応するため、熟練者の技能を効率的に伝え、さらには指導者がいなくても伝承ができるようにする技術の開発が求められる。特に熟練者が直感的に身につけてきた「カン」や「コツ」と呼ばれる暗黙知の伝承は、数値や言語による明示的な表現が難しく、その可視化や伝承には依然として多くの課題が残されている。

こうした課題に対して、これまでの研究では、身体の動き・触覚・視覚・筋電位などを対象に、モーションキャプチャ、慣性センサ、視線計測装置などを用いた計測と分析が行われてきた。その中でも視覚は、多くの作業上で最初に用いられることの多い、極めて重要な感覚である。したがって、熟練者が作業中に注目している重要な領域や特徴を反映した視線データは、暗黙知を可視化・伝達するうえで大きな役割を果たすと考えられる。

本研究の最終目的は、視線計測データと視覚情報を組み合わせ、熟練者の視線パターンを分析し、熟練者の視線を可視化・推定可能とする手法を新たに確立することである。この目的を達成するための初期段階として、本研究では、作業の大部分をモニター画面を注視しながら行う内視鏡医の視線に着目し、内視鏡医と一般学生の視線データから視線パターンの違いを定量的に示せるか探索した。

先行研究では、内視鏡検査における熟練度差が視線に表

れ得ることを示す一方で、画面分割や領域(AOI)に基づく指標に依存する解析が中心であり、注視分布の「形状」や「配置関係」といった空間構造について、比較する枠組みは十分に確立されていない。そこで本研究では、内視鏡画像提示下で取得した視線分布から視線ヒートマップを生成し、ヒートマップ間の類似度を用いて熟練者・非熟練者の視線パターン差を定量化する手法を探索する。

本研究の成果は、内視鏡診療のみならず、他の専門的技術の教育や評価にも応用可能である。医療分野だけでなく、スポーツや製造業における熟練技能の伝承にも活用できると考えられ、少子高齢社会における技術継承の課題解決に大きく寄与すると考えられる。

## 2. 関連研究

### 2.1 熟練度と視線計測に関する研究

Brunyćら<sup>1</sup>が乳腺の生検標本をデジタル化した病理画像(スライド)の所見の判定時において、熟練の病理医と研修医の視線を計測し、経験の違いが視線行動に影響するか、その視線行動が診断精度と関係するかについて、初期閲覧(ズーム前)と後期閲覧(ズーム後)に分けて分析を行っている[1]。初期閲覧(ズーム前)において、熟練の病理医と研修医の両群ともに、早期に重要領域を見つける確率は同程度だったが、熟練の病理医は重要領域での注視がより多く、注視時間も長い傾向が示された。また研修医については研修年次が上がるほど初期閲覧で重要領域へ注視する確率が高まった。一方で、視線行動と診断精度の間に有意な関連は見られなかった。なお、熟練度と視線行動の関係は医療分野に限らず、ゲームなどの一般領域でも報告されている。Ahmadiら<sup>2</sup>はクラウドゲーミングを対象に、プレイヤーのスキルレベル(初心者・中級者・上級者)と視線パターンの関係に着目し、スキルに基づいた視覚的注意モデ

<sup>1</sup> 神戸大学海洋政策科学部

<sup>2</sup> 神戸大学大学院海事科学研究科

<sup>3</sup> 神戸大学大学院医学研究科

<sup>4</sup> 大阪成蹊大学データサイエンス学部

a) nagamatu@kobe-u.ac.jp

b) hiroe@ieee.org

ル (Skill-Based Visual Attention Model) を提案した [2]. 135 名の視線データを解析した結果, 初心者は画面全体を広く見る傾向があるのに対し, 上級者はゲーム上の重要要素に視線を集中させる傾向が明らかになった.

## 2.2 内視鏡画像を対象とした視線計測に関する研究

Troya らは大腸内視鏡動画をコンピュータ支援によるポリープ検出システム (CADe) のあり/なしで提示し, 経験差のある参加者の反応時間と視線指標を用いて比較した [3]. この時, CADe 自体はヒトより速くポリープを検出したが, CADe 使用時の参加者の反応時間は非使用時と同程度で改善がみられなかった. さらに CADe の利用は正常粘膜の誤認を増やし, 視線移動距離を短縮させるなど, AI 導入が探索戦略に与える影響について指摘した.

## 2.3 内視鏡画像を対象とした視線計測を用いた熟練度に関する研究

Lami らは 18 名の内視鏡医を対象として, 内視鏡を盲腸から 2 分間引き抜く間の視線を, 視線追跡用メガネを用いて計測し, ポリープ検出率 (PDR) の高低と視線パターンの関係を検討した [4]. ここで高いポリープ検出率を示した内視鏡医は画面を  $3 \times 3$  グリッドで分割した際の外周部をより長く観察し, 画面下部の特定領域への注視分布も PDR と相関することを示した. さらに隠れマルコフモデル (HMM) 解析により PDR に対応する複数の視線パターンを特定し, 視線の可視化がトレーニングや評価指標になりうると述べている. Karamchandani らは研修医の大腸内視鏡検査時の視線を眼鏡型アイトラッカーで記録し, 手技スコア (TWPS) と視線指標の関連を検証を行っている [5]. TWPS が高いほど腸管粘膜の周辺部への注視 (注視時間/注視回数) が増え, 反対に腸管内腔中心への注視は減少するという相関が報告されており, 視線パターンを用いた熟練度推定・可視化の妥当性を支持する知見として位置づけられる.

## 2.4 視線類似度の評価に関する研究

Bylinskii らは, サリエンスモデルの性能評価に用いられる複数の指標を整理し, 評価指標の選択が結論 (モデル順位や解釈) を変えることを体系的に示した [6]. Kullback-Leibler divergence (KLD) を含む複数の評価指標について, 偽陽性・偽陰性の扱い, 中心バイアスの補正, 空間ずれの許容など, 指標が暗黙に仮定している性質を比較している. この研究から視線ヒートマップの比較においても, KLD が「分布の差」をよく捉える一方で, 解析目的に応じて他指標 (構造的類似度など) との併用が必要になり得ることを示唆している. Zang らは, 緑内障 OCT レポートを読む眼科医の視線を計測し, 得られた注視点から注視ヒートマップを作成したうえで, 深層学習モデルによ

る注視領域予測を行った [7]. 予測されたサリエンスマップ (注視予測マップ) と, 実測の注視ヒートマップの一致度を評価するために KLD に加えて SSIM 分析を行い, 両者の画像としての類似性を示している. このように SSIM は, 視線分布を画像化したヒートマップ同士を比較する際に, 画素単位の誤差ではなく「構造の近さ」を指標化できるため, 視線パターンの類似度評価に利用されている.

## 3. 内視鏡検査の熟練度定量化手法探索のためのデータ取得

データ取得の目的は, 医師がディスプレイに投影される内視鏡画像を観察し, 病変の有無を判断する過程において, 病変のある部位および病変のない部位の静止画像を用いて, 熟練した医師と一般学生が腫瘍を観察している際の視線データを取得し, 医師独特の注視パターンや習熟度が反映されるような注視方法を探索するためのデータセットを作成することである. また, 視線計測と同時に病変の発見率を記録する. 得られた眼球運動の各種パラメータ (瞳孔径, 注視時間, 眼球運動の種類など) と静止画像とを組み合わせて機械学習を行うことで, 視線行動と熟練度の関係を定量的に評価することを目指す.

### 3.1 実験システムの構成

視線の記録については, Tobii AB 製の Tobii pro fusion (60Hz) を使用した. また, 視線計測の際に被験者に課題を表示するために用いたディスプレイは 15.6 インチ ( $1920 \times 1080$ ) であった.

実験用ソフトの開発には, オープンソースで Python 言語をベースとした心理実験環境構築用 Psychopy [8] と, Tobii のデータを取得するための Titta[9] を利用した.

### 3.2 刺激

#### 3.2.1 内視鏡画像

内視鏡医が臓器や病変の種類によって, 視線の位置が変わってくるのを調べるために, 食道の病変 3 種類, 胃の病変 2 種類計 5 種類の病変の画像をそれぞれ 2 枚ずつの 10 枚を GastroVision データセット [10] より使用した. これらの画像の内訳は, バレット食道 (Barrett's esophagus), 食道静脈瘤 (Esophageal varices), 食道炎 (Esophagitis), 胃ポリープ (Gastric polyps), 胃食道接合部正常 Z 線 (Gastroesophageal-junction-normal z-line) である. さらに, これらの病変に対応する正常な臓器の画像が胃の画像 4 枚と食道の画像 6 枚の合わせて 10 枚を用意した. 画像は, 不規則な順番にできるよう設定した. 今回はバレット食道, 食道炎と胃ポリープの画像と, 正常な食道と胃の画像それぞれ一種類ずつ示す (図 1).

#### 3.2.2 内視鏡画像以外の画像

実験には, 内視鏡医とそうでない人とで内視鏡画像以外

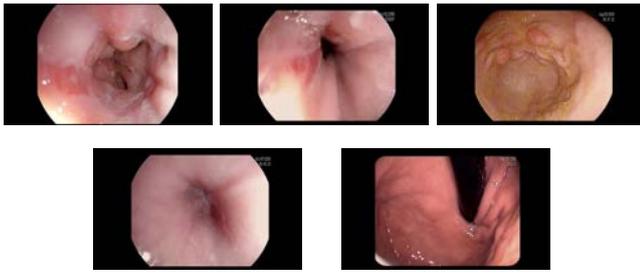


図 1: 内視鏡画像の例（左上がバレット食道，中央が食道炎，左上が胃ポリープ，左下が正常な食道，右下が正常な胃）

の画像を見た時に視線を送る場所に違いがあるのかも観察するために CAT2000 という顕著性マップ用データセット [11] の画像を使用した。これらの画像も内視鏡画像と同様に 20 枚用意した。図 2 に例（4 種類）を示す。



図 2: 内視鏡以外の画像の例

## 4. 分析

取得した内視鏡医 2 名 (D1, D2) と工学系の一般学生 3 名 (S1, S2, S3) の視線データについて、内視鏡画像や内視鏡に関係のない画像について、画像の見方の違いがあるかどうかを明らかにすることを目的として、評価を行った。

### 4.1 分析方法

画像を見ているときの視線計測データから視線ヒートマップを作成した。また、ヒートマップから SSIM (Structural Similarity Index Measure) [12] による比較をおこなった。

#### 4.1.1 視線ヒートマップ作成

取得した視線データに対してガウスクERNELのカーネル密度推定を行い視線ヒートマップを作成した。例として、胃ポリープを示した刺激画像 13 (図 3b) と内視鏡以外の画像である刺激画像 26 (図 3b) について被験者ごとに作成した視線ヒートマップをそれぞれ示す (図 4, 図 5)。青から赤にかけてその領域を見ていた時間が長いことを示している。示した図の色尺度は、可視化のためそれぞれ正規化したものである。

#### 4.1.2 SSIM の計算

視線密度分布の差異を評価する方法としては、KLD を使った確率分布距離の計算が一般的である。情報量の差



(a) 刺激画像 13

(b) 刺激画像 26

図 3: ヒートマップを作成した刺激画像

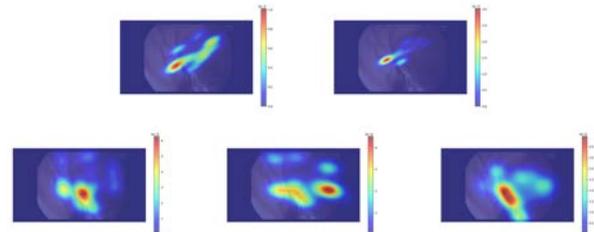


図 4: 内視鏡画像でのヒートマップ（上段左：D1, 上段右：D2, 下段左：S1, 下段中央：S2, 下段右：S3）

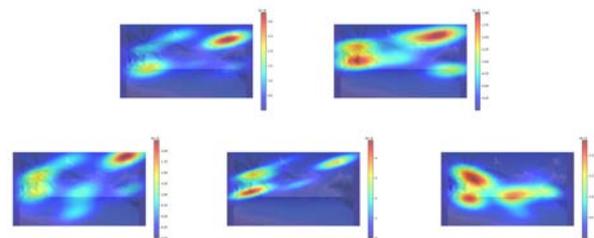


図 5: 内視鏡画像ではない画像でのヒートマップ（上段左：D1, 上段右：D2, 下段左：S1, 下段中央：S2, 下段右：S3）

を定量化する点で有用である一方、分布の空間的な位置ずれや局所的な変動に対して敏感であり、些細な位置ずれで評価がかわってしまうため、注視領域の配置や形状といった構造的な類似性を直接的に評価することは難しい。

一方、今回の研究で着目したいのは、注視が集中する領域の位置関係や分布形状といった空間的構造の類似性である。視線ヒートマップは、視線の空間分布を画像として表現したものであり、観察者がどの領域にどの程度注目していたかを可視化する手法である。

このようなヒートマップ間の類似性を定量的に評価するためには、単純な画素差や相関係数だけでなく、局所的な空間構造を考慮した指標を用いることが望ましい。そこで本研究では、カーネル密度推定によって得られた視線密度分布を画像として扱い、その空間的構造の類似性を評価するために SSIM を計算した。視線ヒートマップを画像として扱い、SSIM を用いることで、視線分布のわずかな位置ずれに対して、過度に敏感になることなく、注視領域の相対的な配置や分布形状といった空間構造の類似性を評価することを目指した。

SSIM は輝度、コントラスト、構造に関する局所統計量

に基づいて算出される。具体的には、2枚の画像  $x, y$  の局所統計量（平均  $\mu_x, \mu_y$ 、分散  $\sigma_x^2, \sigma_y^2$ 、共分散  $\sigma_{xy}$ ）を用いて局所窓ごとに SSIM を計算し、その平均を画像全体の類似度として定義する。SSIM は一般に  $[-1, 1]$  をとり、実際の画像では  $[0, 1]$  に収まることが多い。

局所窓における SSIM は式 1 で定義される ([12])。

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

ここで  $C_1, C_2$  は分母が 0 に近づくことを防ぐための定数であり、画像のダイナミックレンジを  $L$  として一般に

$$C_1 = (K_1L)^2, \quad C_2 = (K_2L)^2$$

と置く（例： $K_1 = 0.01, K_2 = 0.03$ ）。

## 4.2 分析結果

視線ヒートマップから SSIM を計算し、視線分布の類似度を示すとともに、検定によりその結果の有意差を確かめた。

### 4.2.1 SSIM

計 40 枚の刺激画像を閲覧しているときの被験者 5 名の視線から視線ヒートマップを作成し、各刺激画像を見ている時の 5 名の被験者の全通りの組み合わせで SSIM を算出した。

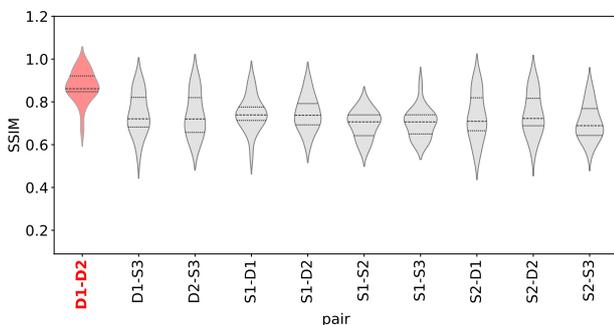


図 6: 内視鏡画像のみのペアごとの SSIM

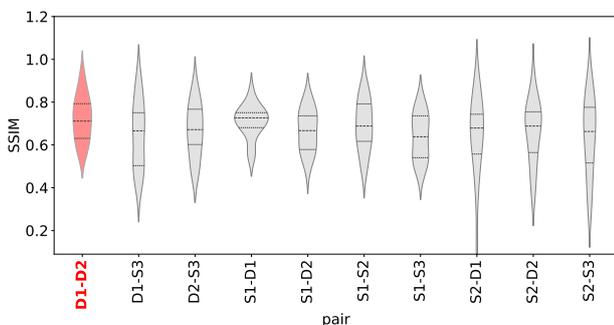


図 7: 内視鏡画像ではない画像のみのペアごとの SSIM

図 6 と図 7 は各被験者ペアにおける SSIM の分布を示し

ている。縦軸が SSIM で値域が 0~1 であり、値が大きいくほど視線ヒートマップの構造が似ていることを示し、横軸は被験者ペアを示している。各バイオリン図は、複数刺激に対して算出された SSIM 値の分布を表している。バイオリンが横に膨れるほど、そこに多くのデータが含まれていることを示している。図 6 から、内視鏡画像を見ている際には、赤色で示した D1-D2（内視鏡医同士）ペアは他の被験者ペアと比較して SSIM が高い傾向を示しており、視線分布の類似性が高いことが示唆される。内視鏡画像ではない画像を見ている時の SSIM は、内視鏡医同士も他の組み合わせと分布に違いは見られなかった（図 7）。

次に、内視鏡画像の種類ごとの SSIM に差ができるかを可視化するため、縦軸を SSIM、横軸を刺激画像番号として散布図を作成し図 8 として示す。

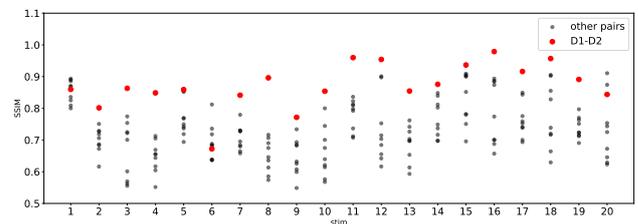


図 8: 内視鏡画像での SSIM の散布図

黒点は内視鏡医-学生および学生-学生の組み合わせを示し、赤点は内視鏡医同士（D1-D2）の SSIM 値を示す。多くの刺激において、D1-D2 の SSIM は他のペアよりも高い値を示しており、刺激に依存せず安定した視線類似性が示唆された。

### 4.2.2 検定

SSIM を用いた解析により、内視鏡医が内視鏡画像を閲覧している際の視線分布に類似性が存在する可能性が示唆された。ペア単位で算出した SSIM 値を用いた探索的解析として統計的検定を行い、内視鏡医間で観察された視線分布の類似性が有意なものであるかを検証した。

#### 4.2.2.1 Welch ANOVA

被験者ペアによる差を検証するため、等分散を仮定しない統計手法である Welch の一元配置分散分析（Welch one-way analysis of variance: Welch ANOVA）を用いた。Welch ANOVA は 3 群以上の平均値の間に差があるかを同時に検定する手法である。一般的な一元配置分散分析（one-way ANOVA）は、各群における分散の等質性を仮定するが、本研究で扱う SSIM 値は値域が 0~1 に制限されており、被験者ペアごとに分散およびサンプルサイズが異なる可能性がある。そのため、分散不均一性およびサンプルサイズの不均一性に強い Welch ANOVA を採用した。Welch ANOVA において算出される F 値は、群間（例:D1-D2 と D1-S2 の間）の平均値の差が、各群内（例:D1-D2 内）のばらつきに対してどの程度大きいかを示す指標であり、分散

およびサンプルサイズの不均一性を考慮した重み付けが行われている。F 値が大きいほど、群間の平均値の差が群内変動に比べて相対的に大きいことを意味する。

また、p 値は、すべての群の母平均が等しいという帰無仮説のもとで、観測された F 値以上の値が偶然によって得られる確率を表す。p 値が有意水準（本研究では  $\alpha=0.001$ ）を下回る場合、群間の平均値に偶然では説明しにくい差が存在すると判断し、帰無仮説を棄却した。

医師同士 (D1-D2) を基準となる群として検定すると、内視鏡画像のみに絞ったとき、F 値は 8.91 で p 値  $4.33 \times 10^{-9}$  であり、内視鏡以外の画像に絞ったとき、F 値は 1.27 で p 値 0.266 となった。

Welch ANOVA により群間の全体差の検定を行い、有意な結果が認められた。しかし、Welch ANOVA の結果が有意であっても、どの群とどの群の間に差があるかを特定することはできない。そこで、各被験者ペア間の比較を Welch の t 検定で行った。また得られた p 値に対して Holm 法による多重比較補正を適用することとした。

#### 4.2.2.2 Welch の t 検定と Holm 法

Welch の t 検定は、比較する 2 群の分散が等しいことを仮定せず、各群の分散およびサンプルサイズを個別に考慮して平均値の差を検定する方法であり、分散不均一性が想定されるデータに強い手法である。本研究では、被験者ペア間の SSIM の差を評価するため、各群間の比較に Welch の t 検定を適用した。また、複数の群間比較を行うことによる第一種過誤の増加を抑制するため、Welch の t 検定によって得られた未補正の p 値に対して Holm 法による比較補正を行った。Holm 法は、少なくとも 1 つの偽陽性が生じる確率（家族誤差率; Family-Wise Error Rate）を制御する多重比較補正法である。本研究では、Holm 法により補正された p 値を用いて統計的有意性の判断を行うこととし、有意水準を  $\alpha=0.001$  とした。

以下の表 1 は、Welch の t 検定で算出した t 値と Holm 法で多重比較した p 値の結果である。t 値は平均差の相対的な大きさを示し、p 値は偶然にその結果が得られうる確率を示している。内視鏡画像では、t 値はすべてのペアと比較して、4 より大きくなり、補正した p 値はすべてにおいて 0.001 より低くなった。逆に、内視以外の画像では、t 値は 2 より小さくなり、補正した p 値はすべてにおいて 0.001 よりはるかに大きくなった。これにより、医師同士 (D1-D2) の内視鏡画像の見方は、ほかのペアと異なり、視線分布の類似性を確認することができた。

### 4.3 考察

結果として、内視鏡画像を見ている時の内視鏡医 2 名の視線の分布は、類似していることが示された。専門性のある人間が、その専門性のフィールドに関する刺激を観察する際に一定のパターンがあるという関連研究での知見が内

表 1: 多重比較の結果

被験者ペア	内視鏡画像		内視鏡画像以外	
	t 値	p 値	t 値	p 値
S1-S2	8.09	<0.001	0.90	0.75
S1-S3	7.53	<0.001	2.49	0.16
S1-D1	5.22	<0.001	0.50	0.75
S1-D2	5.20	<0.001	1.74	0.68
S2-S3	6.88	<0.001	1.69	0.68
S2-D1	4.85	<0.001	1.61	0.68
S2-D2	4.71	<0.001	1.49	0.68
D1-S3	4.71	<0.001	1.76	0.68
D2-S3	4.72	<0.001	1.55	0.68

視鏡医のスキルでも同様であることが示唆された。

ここで内視鏡医 2 名の視線が他のペアと比べどのような違いがあるかを確認するため、図 8 より、医師-医師の組み合わせで 1 番 SSIM の値が大きくなった刺激画像 16 閲覧時のヒートマップ (図 9) を示す。この画像は、病変のない食道を示している。上段のヒートマップでは医師 2 名は部位の極めて限られた領域にしか見ていないのに対して、下段の一般学生のヒートマップでは医師と比較してももう少し広い領域を見ていることがわかる。ここから、熟達した内視鏡医は着目点を絞って画像を観察していることが考えられる。

また、病変のある画像と病変のない画像でどのような違いがあるかを確認するため、図 3a の画像について考察する。この画像は胃ポリープを表し、手前の先端が透明に見える部分は、内視鏡の処置具（スネア/ループ/注入針など）を出し入れするためのカテーテルであり、その隣の丸みを帯びたふくらみがポリープである。この画像から作成した視線ヒートマップ (図 4) を見ると、上段の内視鏡医 2 名の視線は胃ポリープに集中しているのに対し、下段の一般学生 3 名の視線はカテーテルを中心にそのほかのところにも視線が移ってしまっていることが確認できる。これは、内視鏡 2 名はカテーテルから胃ポリープをすぐ確認できた可能性が考えられる。逆に一般学生 3 名に関しては明らかに色が違うカテーテルに視線が集中してしまった可能性が考えられる。このように内視鏡医 2 名に関して病変がある画像での視線データはその病変に視線が集中しやすいことが考えられる。そのほかの画像について SSIM の結果を確認すると、病変のある画像に関してすべての画像で SSIM が他のペアと比べて 1 番大きい値を示し、食道炎の一例を除きすべての SSIM で 0.8 を上回っていた。このように病変のある場合は、その病変に視線が集中することが示唆された。

以上より、内視鏡画像閲覧時において医師間で類似した視線分布が形成される可能性が示されたが、本研究にはいくつかの制約が存在する。まず、本研究で用いた刺激数および被験者数が限られているため、特定の病変種類や病変

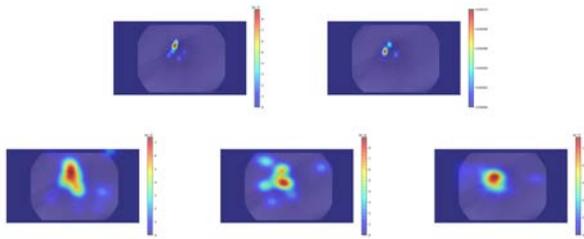


図 9: 内視鏡画像 (16) での SSIM の散布図 (上段左: D1, 上段右: D2, 下段左: S1, 下段中央: S2, 下段右: S3)

の有無ごとに視線行動の差異を詳細に検討するには不十分である。

また、本研究では視線データをヒートマップとして集約し、SSIM を用いて空間的な類似度を評価したが、この手法では視線の時系列的な動きや注視順序、注視時間といった動的特徴を捉えることはできない。そのため、医師間で類似した SSIM 値が得られた場合であっても、必ずしも同一の視線戦略や探索過程を共有しているとは限らない。

今後は被験者数および症例数を拡充するとともに、時系列解析や注視行動指標を組み合わせた多面的な解析を行うことで、内視鏡診断における視線行動の特徴をより詳細に明らかにする必要がある。

## 5. おわりに

本研究では、視線データを用いた熟練度の可視化を目的に、作業のほとんどを画面を見て行う内視鏡検査に着目し、研究を行った。具体的には、内視鏡医 2 名と工学系の一般学生 3 名に対し、内視鏡画像 (病変の有無あり) を 20 枚、内視鏡画像ではない画像 20 枚を見せ、その時の視線を視線計測装置を用いて記録した。

分析にはヒートマップを用い、作成したヒートマップの SSIM を比較することで内視鏡医ペアとその他のペアで視線の比較を行った。ここで内視鏡医ペアとほかペアで差が見られたため、Welch ANOVA 検定で内視鏡医ペアの群が他と異なるかどうかという全体的な差の有無を確認した。加えて、Welch の  $t$  検定と多重比較方法である Holm 法を持ちいることで、内視鏡医ペアの SSIM の結果は有意的に生じたことが確認できた。

今後は、被験者数および症例数を拡充するとともに、時系列解析や注視行動指標を組み合わせた多面的な解析を行う予定である。それに加え最終目的である熟練度の可視化に向け、研修医と熟練医での視線をとり、両者の違いを明らかにしていきたい。

**謝辞** 本研究は JSPS 科研費 23K16932 の助成を受けたものである。研究の計画段階より多くのご助言とご協力をいただいている神戸大学医学部附属病院 国際がん医療・研究センター 消化器内科の森田 圭紀先生と鷹尾 まど佳先生

に深くお礼申し上げます。また、実験に協力していただいた皆様に深くお礼申し上げます。

## 参考文献

- [1] Brunyé, T. T., Drew, T., Kerr, K. F., Shucard, H., Weaver, D. L. and Elmore, J. G.: Eye tracking reveals expertise-related differences in the time-course of medical image inspection and diagnosis, Vol. 7, No. 5, p. 051203, , DOI: 10.1117/1.JMI.7.5.051203 (2020).
- [2] Ahmadi, H., Zadtootaghaj, S., Pakdaman, F., Hashemi, M. R. and Shirmohammadi, S.: A Skill-Based Visual Attention Model for Cloud Gaming, Vol. 9, pp. 12332–12347, , DOI: 10.1109/ACCESS.2021.3050489 (2021).
- [3] Troya, J., Fitting, D., Brand, M., Sudarevic, B., Kather, J. N., Meining, A. and Hann, A.: The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze, Vol. 54, No. 10, pp. 1009–1014, , DOI: 10.1055/a-1770-7353 (2022).
- [4] Lami, M., Singh, H., Dilley, J. H., Ashraf, H., Edmondson, M., Orihuela-Espina, F., Hoare, J., Darzi, A. and Sodergren, M. H.: Gaze patterns hold key to unlocking successful search strategies and increasing polyp detection rate in colonoscopy, Vol. 50, No. 7, pp. 701–707, , DOI: 10.1055/s-0044-101026 (2018).
- [5] Karamchandani, U., Erridge, S., Evans-Harvey, K., Darzi, A., Hoare, J. and Sodergren, M. H.: Visual gaze patterns in trainee endoscopists - a novel assessment tool, Vol. 57, No. 9, pp. 1138–1146, , DOI: 10.1080/00365521.2022.2064723 (2022).
- [6] Bylinskii, Z., Judd, T., Oliva, A., Torralba, A. and Durand, F.: What Do Different Evaluation Metrics Tell Us About Saliency Models?, Vol. 41, No. 3, pp. 740–757, , DOI: 10.1109/TPAMI.2018.2815601 (2019).
- [7] Zang, M., Mukund, P., Forsyth, B., Laine, A. F. and Thakoor, K. A.: Predicting Clinician Fixations on Glaucoma OCT Reports via CNN-Based Saliency Prediction Methods, Vol. 5, pp. 191–197, , DOI: 10.1109/OJEMB.2024.3367492 (2024).
- [8] Peirce, J., MacAskill, M. and Hirst, R.: *Building experiments in PsychoPy*, SAGE Publications Ltd (2022).
- [9] Niehorster, D. C., Andersson, R. and Nyström, M.: Titta: A toolbox for creating PsychToolbox and Psychopy experiments with Tobii eye trackers, Vol. 52, No. 5, p. 1970–1979, , DOI: 10.3758/s13428-020-01358-8 (2020).
- [10] Jha, D., Sharma, V., Dasu, N., Tomar, N. K., Hicks, S., Bhuyan, M. K., Das, P. K., Riegler, M. A., Halvorsen, P., Bagci, U. et al.: Gastrovision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection, *Workshop on machine learning for multimodal healthcare data*, Springer, pp. 125–140 (2023).
- [11] Borji, A. and Itti, L.: CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research, *ArXiv*, Vol. abs/1505.03581 (2015).
- [12] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612 (2004).