グラム染色画像の菌の矩形領域分類における矩形領域数の影響について

黒石 陽夢[1,a)]　平田 耕一[1,b)]

概要：本研究では，グラム染色画像に含まれる 14 種類の細菌を対象に，細菌を囲む矩形領域の画像分類を行う．各菌種について，互いに素な 300 領域の集合 A, B, C を構成し，その和 $A \cup B, A \cup C, B \cup C$（$|A \cup B| = |A \cup C| = |B \cup C| = 600$）および $A \cup B \cup C$（$|A \cup B \cup C| = 900$）を用いて領域数の影響を評価した．VGG，MobileNet，DenseNet，ViT を ImageNet 事前学習により転移学習し，5 分割交差検証で正解率・適合率・再現率・F 値を評価した．多くの条件で ViT が最高となり，領域数の増加は F 値の単調な向上を必ずしももたらさないことを確認した．

キーワード：2130502 分類学習，2150205 画像認識・理解，2170301 医療・福祉支援

On the Effect of the Number of Regions in Classifying Rectangle Regions of Bacteria in Gram Stained Smears Images

Hiromu Kuroishi[1,a)]　Kouichi Hirata[1,b)]

Abstract: In this study, we classify rectangle regions of 14 bacteria in Gram-stained smear images. For each bacterium, we construct three disjoint sets A, B, C consisting of 300 regions, and use their unions $A \cup B$, $A \cup C$, $B \cup C$ (with $|A \cup B| = |A \cup C| = |B \cup C| = 600$) and $A \cup B \cup C$ (with $|A \cup B \cup C| = 900$) to evaluate the effect of the number of regions. We train VGG, MobileNet, DenseNet, and ViT with ImageNet pretraining, and evaluate accuracy, precision, recall, and F1 by 5-fold cross validation. ViT achieves the best results in most settings. Then, we confirm that increasing the number of regions does not necessarily improve F1 monotonically.

Keywords: 2130502 Classification, 2150205 Image recognition/understanding, 2170301 Medicine and welfare

1. はじめに

臨床微生物学における検査は，主に顕微鏡検査，培養，同定に分かれる．グラム染色 [1] は，1884 年に Hans Christian Gram によって導入された顕微鏡検査の一種であり，染色性と形状に基づいて細菌を分類する手法である．これはクリスタルバイオレット染色と赤色サフラニン対比染色に対する細胞壁の保持性の違いに基づく．紫色に染まる細菌はグラム陽性菌とよばれ，ピンク色に染まる細菌はグラム陰性菌とよばれる．グラム染色後，細菌は顕微鏡下で球状または桿状として観察される．球状の細菌は球菌，桿状の細菌は桿菌とよばれる．したがって，細菌は，グラム陽性球菌，グラム陽性桿菌，グラム陰性球菌，グラム陰性桿菌の 4 種類に分類できる．

グラム染色は 30 分以内で実施でき，かつ高価な装置を必要としないため，培養や同定と比べて，感染症の初期診療において非常に重要な役割を果たす．しかし，グラム染色では細菌以外にも白血球，埃，油滴，結晶などが同じ染料で染まるため，細菌を正確に識別するには熟練した技術

[1] 九州工業大学
Kyushu Institute of Technology
[a)] kuroishi.hiromu361@mail.kyutech.jp
[b)] hirata@ai.kyutech.ac.jp

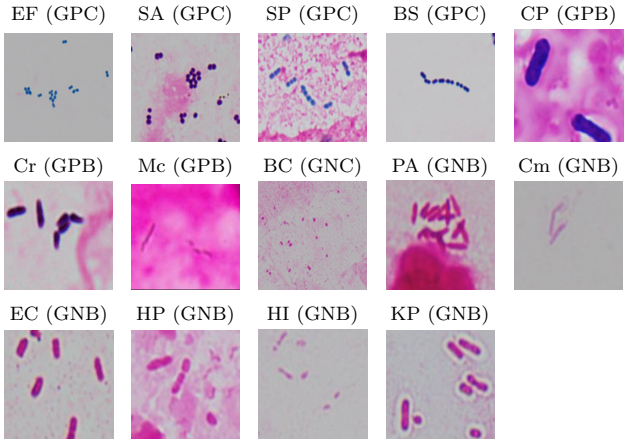


図 1 対象とする細菌の拡大画像
Fig. 1 The images of the target bacteria.

が必要となる．

本研究では，腸球菌（EF），黄色ブドウ球菌（SA），肺炎球菌（SP），B 群レンサ球菌（BS）の 4 種のグラム陽性球菌，クロストロジウム・パーフリンゲンス（CP），コリネバクテリウム（Cr）の 3 種のグラム陽性桿菌，ブランハメラ・カタラーリス（BC）の 1 種のグラム陰性球菌，緑膿菌（PA），キャンピロバクター（Cm），大腸菌（EC），ピロリ菌（HP），インフルエンザ菌（HI），肺炎桿菌（KP）の 6 種のグラム陰性桿菌からなる 14 種の細菌を分類の対象とする．**図 1** は各細菌の拡大画像である．これらの細菌を矩形で囲った領域でアノテーションを行い，これらを画像分類器に入力する．グラム染色塗抹画像における菌の分類に関して，Smith ら [13] は，血液検体を対象に，固定サイズ（146 × 146 画素）に切り出した領域を用いて，グラム陰性桿菌（GNB），クラスタ状のグラム陽性球菌（GPC），および双対状と連鎖状の GPC の 3 クラスを CNN で分類した．Satoto ら [11] は，グラム染色画像からグラム陰性菌を VGG16 [12] で分類した．さらに，Kawano ら [8] は，13 種類の細菌の矩形領域を対象に，VGG16/19，MobileNet [6]，DenseNet [7] により分類した．Kuroishi と Hirata [9] は，VGG，MobileNet，DenseNet，RegNet [16]，ConvNeXt [10]，ViT [2]，EfficientNet [14]，EfficientNetV2 [15] を用いて，14 種の細菌の矩形領域を分類した．その際，各細菌あたりの領域数を 300（ただしクロストリジウムは 144）に揃えていた．

本研究では，この先行研究を発展させ，各細菌の矩形領域数を 300，600 および 900 に増やし，VGG，MobileNet，DenseNet，ViT を用いて，領域数が分類性能に与える影響を調べる．

2. 画像分類器

本研究では，画像分類器として VGGNet，MobileNet，DenseNet，ViT(Vision Transformer) を用いる．以下では，各分類器の概要を説明する．

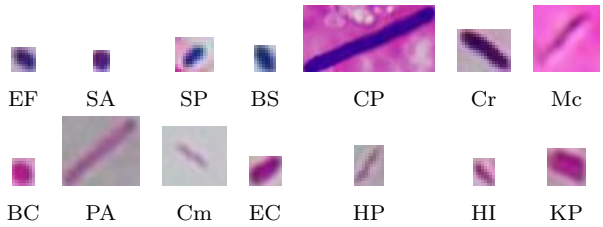


図 2 細菌の矩形領域（CP は 0.5 倍で表示）
Fig. 2 The rectangle regions of bacteria (CP is scaled by 0.5)

2.1 VGGNet

VGGNet [12] は，3 × 3 の小さな畳み込みフィルタを多数積み重ねることでネットワークを深層化した畳み込みニューラルネットワーク (CNN) である．入力画像 (224 × 224) に対して，畳み込み層と最大プーリング層を交互に配置した特徴抽出部の後に，3 層の全結合層と softmax 層を接続する構成を持つ．本研究では VGG16，VGG19 を用いる．VGG16 は 13 層の畳み込み層と 3 層の全結合層からなる合計 16 層のネットワークであり，VGG19 は 16 層の畳み込み層と 3 層の全結合層からなる合計 19 層のネットワークである．

2.2 MobileNet

MobileNet [6] は，Depthwise Separable Convolution を用いて計算量とパラメータ数を大幅に削減した軽量 CNN である．通常の畳み込みが空間方向のフィルタリングとチャネル方向の結合を同時に行うのに対し，Depthwise Separable Convolution では，まず各入力チャネルごとに 3 × 3 の畳み込み（depthwise）を行い，その後 1 × 1 の畳み込み（pointwise）でチャネルを結合する．これにより，計算量とパラメータ数が大幅に削減される．また，最初の層を除き，各層の後段に Batch Normalization と ReLU を適用し，最終の全結合層の出力を softmax 層に入力して分類する．本研究では MobileNet-Small，MobileNet-Large を用いる．

2.3 DenseNet

DenseNet [7] は，各層がそれ以前のすべての層からの特徴マップを受け取る CNN である．これにより，勾配消失問題を緩和，特徴の再利用の促進，パラメータ数の削減を可能にする．本研究では DenseNet-169，DenseNet-201 を用いる．

2.4 ViT

Vision Transformer（ViT）[2] は，画像分類に Transformer アーキテクチャを適用したものである．入力画像を固定サイズのパッチに分割し，各パッチを線形埋め込みに変換する．これらのパッチ埋め込みに位置エンコーディングを加え，それらを Transformer エンコーダに入力する．

表 1 各細菌の顕微鏡画像枚数（#im）と矩形領域数（#rec）
Table 1 The number (#im) of images of bacteria, the number (#rec) of rectangle regions of bacteria.

菌	#im	#rec	菌	#im	#rec
EF	21	2,687	BC	37	3,553
SA	75	13,832	PA	94	6,418
SP	91	1,007	Cm	272	20,096
BS	19	2,332	EC	97	5,322
CP	12	144	HP	49	1,783
Cr	57	3,093	HI	31	6,189
Mc	157	2,120	KP	39	3,012

Transformer エンコーダは，マルチヘッド自己注意機構と位置ごとのフィードフォワードネットワークから構成される．最終的な分類には，クラス埋め込みトークンの出力を用いる．本研究では ViT-B/32，ViT-L/32 を用いる．

3. 実験方法

本節では，本研究で実施した分類実験の手順を述べる．

3.1 データセット

本研究では，14 種類の細菌についてグラム染色顕微鏡画像を収集し，各画像中の細菌を矩形で囲んでアノテーションした．1 枚の顕微鏡画像から複数の細菌領域が抽出されるため，顕微鏡画像数と矩形領域数は一致しない．

これらのグラム染色顕微鏡画像から抽出した細菌の矩形領域をデータセットとする．

表 1 は，元となる顕微鏡画像枚数（#im）およびそこから抽出された全矩形領域数（#rec）である．

これらの全矩形領域集合から互いに重ならない 3 つの集合 A, B, C をそれぞれ 300 領域ずつ無作為に選択する．次に，それらを組み合わせて $A \cup B$, $A \cup C$, $B \cup C$ の 3 つの集合を構成し，それぞれ 600 領域からなる集合と，$A \cup B \cup C$ からなる 900 領域の集合を構成する．なお，CP は全体で 144 領域しか存在しないため，すべての条件で 144 領域を用いる．

矩形領域は元画像から切り出され，サイズは細菌の大きさに応じて異なる．このデータのサイズ特性は，実寸比の例（**図 2**）と統計量（**表 2**：最大・最小・平均）で示す．なお，実寸比の図では視認性のため CP のみ 0.5 倍で表示している．全体画像と矩形領域の例を**図 3** に示す．

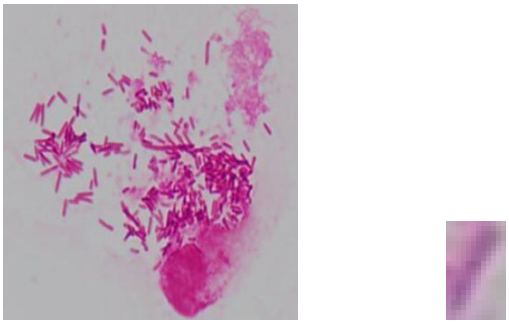
3.2 前処理

本研究では，すべての矩形領域を各分類器の入力サイズに合わせて 224×224 画素にリサイズする．ここで，表 2 のように，CP の矩形領域は他の細菌と比較して極端に大きい．

すべての矩形領域は，RGB の 3 チャネル画像として扱い，画素値は $[0, 255]$ から $[0, 1]$ の範囲に正規化する．これらの前処理を施した画像を，画像分類器への入力として用いる．

表 2 各細菌における矩形領域の横方向および縦方向の画素数の最大値・最小値・平均値
Table 2 The maximum (max.), the minimum (min.) and the average (ave.) of pixels for the horizontal side and the vertical side in all the regions for every bacterium.

	横方向（画素）			縦方向（画素）		
細菌	最大	最小	平均	最大	最小	平均
EF	15	4	6.94	15	4	6.91
SA	10	5	6.45	13	4	7.08
SP	23	9	15.30	26	7	13.66
BS	11	4	6.30	11	4	6.66
CP	191	10	39.12	232	13	50.55
Cr	24	6	12.80	18	5	9.02
Mc	75	15	30.59	66	9	24.64
BC	10	4	6.72	15	4	7.62
PA	60	4	14.08	48	5	14.01
Cm	70	9	18.68	52	8	17.21
EC	29	6	13.89	25	6	10.77
HP	30	8	14.10	26	5	11.24
HI	24	3	7.70	20	4	7.71
KP	27	9	13.99	25	9	12.73



(a) 全体画像　(b) 矩形領域の例
図 3 緑膿菌（PA）の全体画像（a）と矩形領域（b）
Fig. 3 Whole image (a) and rectangle region (b) of Pseudomonas aeruginosa (PA).

3.3 評価方法

分類性能の評価には，正解率，適合率，再現率，および F 値を用いる．真陽性（TP），偽陽性（FP），真陰性（TN），偽陰性（FN）を用いて，以下のように定義する．

$$\text{正解率} = \frac{TP+TN}{TP+TN+FP+FN}$$
$$\text{適合率} = \frac{TP}{TP+FP}$$
$$\text{再現率} = \frac{TP}{TP+FN}$$
$$\text{F 値} = \frac{2\,\text{適合率}\,\text{再現率}}{\text{適合率}+\text{再現率}}$$

3.4 学習設定

すべての画像分類器は，ImageNet [5] で事前学習されたモデルを初期値として用いる転移学習 [17] により学習される．これにより，限られた枚数の画像データに対しても安定した特徴表現を獲得できる．

評価には 5 分割交差検証 (5-fold cross validation) を用いる．全矩形領域を無作為に 5 つの互いに重ならない部分集合に分割し，そのうち 3 つを学習，1 つを検証，1 つをテストデータとして分類器を学習・評価する．この手順を 5 回繰り返し，すべての部分集合が 1 回ずつテスト用になるようにする．最終的な性能は，5 回の結果の平均値とする．

本研究におけるすべての実験は，Ubuntu 22.04.4 LTS を搭載した計算機上で実行する．CPU には Intel Xeon プロセッサ (2.20 GHz)，メモリは 16 GB を使用し，GPU には NVIDIA L4 を用いる．すべての分類器は同一の計算環境および学習条件の下で学習，評価する．学習時のエポック数は 300，バッチサイズは 32 とする．

4. 実験結果

本節では，分類に関する実験結果を示す．**表 3** は 300 領域からなる集合 A, B, C，**表 4** は 600 領域からなる集合 $A \cup B$, $A \cup C$, $B \cup C$，および**表 5** は 900 領域からなる集合 $A \cup B \cup C$ に対する各分類器の分類性能を示す．なお，太字は各評価指標における最大値を表す．

表 3, 表 4, 表 5 より，集合 B を除くすべての集合において，ViT-B/32 が最高の分類性能を示したことが分かる．一方，集合 B に対しては ViT-L/32 が最高の分類性能を示した．したがって，ViT は本研究の分類において最も有効な分類器であると考えられる．また，集合 A は，すべての分類器において F 値が最大であり，集合 C は F 値が最小であった．特に集合 C では，$A \cup C$ および $B \cup C$ の結果からわかるように，領域数が増加するにつれて F 値が大きくなる．しかし，この性質は集合 A および B には見られない．

5. おわりに

本研究では，グラム染色画像における細菌の矩形領域を対象として，各細菌について 300，600，900 の領域を用いて分類した．その結果，ViT がすべての分類器の中で最も有効な分類器であること，および領域数を増やしても F 値は必ずしも増加しないことを確認した．

本研究では，300，600，900 領域の設定において，CP の領域数が他の細菌と揃えられなかったため，すべての条件で同じ領域数を用いることができなかった．そのため，今後の課題として，CP を分類対象から除外する，あるいはデータ拡張を適用して領域数を増やすことにより，すべてのクラスの領域数を揃えた上で，領域数と分類性能の関係を詳細に分析することが挙げられる．

また，矩形領域のリサイズに伴う情報損失を避けるため元のサイズのまま適用可能な分類器の設計と，Vision Mamba [3] や VSSM（Vision State Space Model） [4] といっ State Space Model 系分類器の適用による CNN 系や Transformer 系との比較評価も今後の課題である．

表 3 300 領域（集合 A, B, C）に対する分類結果
Table 3 The evaluation values of the classification for A, B and C.

300 領域	A			
分類期	正解率	適合率	再現率	F 値
VGG16	0.924	0.927	0.925	0.925
VGG19	0.916	0.920	0.918	0.918
MobileNet-Small	0.908	0.918	0.911	0.911
MobileNet-Large	0.920	0.927	0.923	0.923
DenseNet-169	0.938	0.943	0.940	0.940
DenseNet-201	0.937	0.940	0.940	0.939
ViT-B/32	**0.942**	**0.945**	**0.944**	**0.944**
ViT-L/32	0.929	0.933	0.931	0.931

300 領域	B			
分類期	正解率	適合率	再現率	F 値
VGG16	0.827	0.835	0.833	0.832
VGG19	0.841	0.847	0.846	0.846
MobileNet-Small	0.856	0.864	0.862	0.861
MobileNet-Large	0.863	0.870	0.868	0.867
DenseNet-169	0.820	0.828	0.825	0.824
DenseNet-201	0.845	0.853	0.850	0.850
ViT-B/32	0.871	0.877	0.876	0.875
ViT-L/32	**0.906**	**0.911**	**0.911**	**0.910**

300 領域	C			
分類期	正解率	適合率	再現率	F 値
VGG16	0.826	0.834	0.832	0.831
VGG19	0.832	0.838	0.838	0.837
MobileNet-Small	0.859	0.866	0.864	0.863
MobileNet-Large	0.847	0.856	0.853	0.852
DenseNet-169	0.823	0.831	0.827	0.827
DenseNet-201	0.843	0.851	0.848	0.848
ViT-B/32	**0.863**	**0.869**	**0.867**	**0.867**
ViT-L/32	0.848	0.854	0.853	0.852

表 4 600 領域（集合 $A \cup B$, $A \cup C$, $B \cup C$）に対する分類結果
Table 4 The evaluation values of the classification for $A \cup B$, $A \cup C$ and $B \cup C$.

600 領域	$A \cup B$			
分類期	正解率	適合率	再現率	F 値
VGG16	0.887	0.894	0.891	0.892
VGG19	0.883	0.891	0.888	0.890
MobileNet-Small	0.867	0.875	0.872	0.873
MobileNet-Large	0.895	0.901	0.900	0.900
DenseNet-169	0.902	0.908	0.907	0.907
DenseNet-201	0.901	0.907	0.906	0.906
ViT-B/32	**0.910**	**0.916**	**0.915**	**0.915**
ViT-L/32	0.906	0.911	0.911	0.910

600 領域	$A \cup C$			
分類期	正解率	適合率	再現率	F 値
VGG16	0.878	0.886	0.883	0.883
VGG19	0.876	0.883	0.881	0.882
MobileNet-Small	0.897	0.903	0.903	0.902
MobileNet-Large	0.847	0.856	0.853	0.852
DenseNet-169	0.869	0.877	0.873	0.875
DenseNet-201	0.888	0.895	0.894	0.894
ViT-B/32	**0.910**	**0.915**	**0.915**	**0.915**
ViT-L/32	0.896	0.901	0.901	0.901

600 領域	$B \cup C$			
分類期	正解率	適合率	再現率	F 値
VGG16	0.856	0.865	0.864	0.863
VGG19	0.847	0.857	0.854	0.854
MobileNet-Small	0.883	0.891	0.889	0.889
MobileNet-Large	0.847	0.856	0.853	0.852
DenseNet-169	0.848	0.858	0.852	0.854
DenseNet-201	0.872	0.881	0.879	0.879
ViT-B/32	**0.894**	**0.900**	**0.899**	**0.899**
ViT-L/32	0.882	0.889	0.889	0.888

表 5 900 領域（集合 $A \cup B \cup C$）に対する分類結果
Table 5 The evaluation values of the classification for $A \cup B \cup C$.

900 領域	$A \cup B \cup C$			
分類期	正解率	適合率	再現率	F 値
VGG16	0.877	0.886	0.883	0.884
VGG19	0.882	0.890	0.889	0.888
MobileNet-Small	0.897	0.904	0.903	0.903
MobileNet-Large	0.882	0.890	0.889	0.888
DenseNet-169	0.873	0.882	0.879	0.880
DenseNet-201	0.892	0.899	0.898	0.898
ViT-B/32	**0.913**	**0.920**	**0.919**	**0.919**
ViT-L/32	0.901	0.907	0.906	0.906

参考文献

[1] J. W. Bartholomew, T. Mittwer: *The Gram stain*, Bacteriol. Rev. **16**, 1–29 (1952).
[2] L. Beyer, M. Dehghani, A. Dosovitskiy, S. Gelly, G. Heigold, N. Houlsby, A. Kolesnikov, M. Minderer, T. Unterthiner, J. Uszkoreit, D. Weissenborn, X. Zhai: *An image is worth 16×16 words: Transformers for image recognition at scale*, Proc. ICLR'21 (2021).
[3] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, X. Wang: *Vision Mamba: Efficient visual representation learning with bidirectional state space model*, arXiv:2401.09417 (2024).
[4] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, J. Jiao, Y. Liu: *VMamba: Visual State Space Model*, arXiv:2401.10166 (2024).
[5] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. Fei-Fei: *ImageNet : A large-scale hierarchical image database*, Proc. CVPR'09, 248–255 (2009).
[6] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam: *MobileNets: Efficient convolutional neural networks for mobile vision applications*, arXiv:1704.04861 (2017).
[7] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger: *Densely connected convolutional networks*, Proc. CVPR'17, 2261–2269 (2017).
[8] I. Kawano, E. Kurumida, S. Terada, K. Hirata: *Classifying Gram positive cocci and Gram negative bacilli in Gram stained smear images*, Proc. ESKM'22, 55–60 (2022).
[9] H. Kuroishi, K. Hirata: *Classifying rectangle regions of bacteria in Gram stained smears images*, Proc. ESKM'25, 43–48 (2025).
[10] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, S. Xie: *A ConvNet for the 2020s*, Proc. CVPR'22, 11976–11986 (2022).
[11] B. D. Satoto, I. Utoyo, R. Rulaningtyas, E. B. Khoendori: *An improvement of Gram-negative bacteria identification using convolution neural network with fine tuning* Telekomnika **18**, 1397–1405 (2020).
[12] K. Simonyan, A. Zisserman: *Very deep convolutional networks for large-scale image recognition*, Proc. ICLR'15 (2015).
[13] K. P. Smith, A. D. Kang, J. E. Kirby: *Automated interpretation of blood culture Gram stains by use of a deep convolutional neural network*, J. Clin. Microbiol. **56**, e01521-17 (2018).
[14] M. Tan, Q. V. Le: *EfficientNet: Rethinking model scaling for convolutional neural networks*, Proc. ICML'19, 6105–6114 (2019).
[15] M. Tan, Q. V. Le: *EfficientNetV2: Smaller models and faster training*, Proc. ICML'21, 10096–10106 (2021).
[16] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, Z. Xu: *RegNet: Self-regulated network for image classification*, IEEE Trans. Neural Netw. Learn. Syst. **34**, 9562–9567 (2023).
[17] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He: *A comprehensive survey on transfer learning*, Proc. IEEE **109**, 43–76 (2021).