グラム染色画像からの白血球貪食の分類と検出

影本 幸哉[†1,a)]　中城 龍之介[†1,b)]　平田 耕一[†2,c)]

概要：グラム染色とは細菌の推定手法の一種であり，染色液を用いて菌を染め上げ，顕微鏡を用いてその染色性や形状などから菌種を推定する手法である．本論文では，白血球が異物を取り込み分解しようとする作用である貪食に着目し，貪食，擬貪食，非貪食の 3 つの状態を分類，検出する．そして，各貪食像に対してアノテーションを適用したデータをもとに，画像分類器である DenseNet，MobileNet，VGG，Vision Transformer でそれらを分類すると共に，物体検出器である YOLOv5 と YOLOv8 でそれらを検出する．

Classifying and Detecting Phagocytotic Activity of Leukocytes from Gram Stained Smear Images

Abstract: Gram staining is a bacterial classification method that involves staining bacteria using a dye solution and estimating bacterial species based on their staining characteristics and morphology under a microscope. In this paper, we focus on the phagocytotic activity of leukocytes, a process in which leukocytes engulf and degrade foreign substances, and classify and detect three states of phagocytosis, pseudo-phagocytosis, and non-phagocytosis. Then, based on the annotated images of the three states, we classify them by image classifiers of DenseNet, MobileNet, VGG and Vision Transformer, and detect them by object detectors of YOLOv5 and YOLOv8.

1. はじめに

グラム染色 [1] とは，1884 年に Christian Gram によって開発された細菌染色手法である．この手法では血液や尿，喀痰，便といった検体中に存在する細菌を，染色液に応じてグラム陽性とグラム陰性に染め分け，この染色性（藍色・赤色）や細菌の形態（球菌・桿菌）から感染症の起炎菌を推定する．

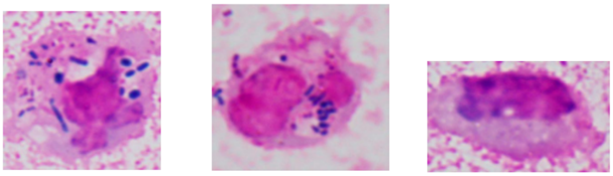
グラム染色には高額な医療機器の導入が不要であり，また染色から検査までが 30 分程度で行えるため安価かつ迅速な検査が可能である．その一方で，推定には熟練が必要であり，臨床の現場ではグラム染色による微生物検査を行える検査技師が不足しているという問題がある．そこで，顕微鏡検査の補助や自動化を目指し，深層学習を利用したグラム染色画像からの画像分類や物体検出の研究が進められている．

グラム染色画像には，菌だけではなく白血球も映っている．白血球は，免疫系において重要な役割を果たす細胞群であり，体内に侵入した病原体や異物を排除する機能を持つ．白血球にはいくつかの種類があり，特に好中球やマクロファージは貪食と呼ばれる作用を持つ．貪食とは，好中球などが異物を取り込み，内部で分解する作用のことを指す．グラム染色画像における白血球像には，**図 1** のように，白血球が異物を完全に取り込んでいる貪食像，白血球と菌が観測者から見て前後に並んで位置するため貪食状態のように見える擬貪食像，明らかに貪食していない白血球像である非貪食像の３つがある．このうち，貪食像と擬貪食像は差異が少なく判別が難しい．

グラム染色画像からの白血球貪食の分類では，Yoshihara と Hirata [12] は，VGG16 [10] とそれを改良した分類器を用いて，白血球の貪食活性を分類した．ここでは，分類対象の映る画像全体から白血球の映る領域のみを矩形として切り出し，その矩形を画像分類器に入力することで画像分類を行っている．

一方, グラム染色画像からの白血球貪食の検出では, Yoshi-

†1 現在，九州工業大学大学院情報工学府
Presently with Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology
†2 現在，九州工業大学情報工学研究院
Presently with Department of Artificial Intelligence, Kyushu Institute of Technology
a) kagemoto.kouya998@mail.kyutech.jp
b) nakajo.ryunosuke428@mail.kyutech.jp
c) hirata@ai.kyutech.ac.jp



(a) 貪食像　(b) 擬貪食像　(c) 非貪食像

図 1: 3 種の貪食像のグラム染色画像

Fig. 1 Gram stained smear images of three types of leukocyte phagocytosis

hara と Hirata [13] は，Faster R-CNN [9]，RetinaNet [7]，YOLOv5 [5] の物体検出器を用いて，白血球貪食を検出した．その結果，YOLOv5 が最も高い検出性能であった．

そこで本論文では，貪食像，擬貪食像，非貪食像の 3 つの白血球像を対象として，画像分類器として VGG，MobileNet [3]，DenseNet [4]，Vision Transformer (ViT) [2]，物体検出器として YOLOv5 と YOLOv8 [6] を用いて，グラム染色画像における白血球貪食の分類と検出を行う．

2. 画像分類器

2.1 VGGNet

VGGNet [10] は，3 × 3 の畳み込みフィルターを備えた畳み込みニューラルネットワーク（CNN）であり，複数の CNN の後ろに全結合層を繋げたシンプルな構成を持つ．

本論文では, VGG16 および VGG19 を使用する．VGG16 は 13 層の CNN と 3 層の全結合層からなる合計 16 層の VGGNet であり，VGG19 は，16 層の CNN と 3 層の全結合層からなる合計 19 層の VGGNet である．

2.2 MobileNet

MobileNet [3] は，通常の畳み込み層に加えて Depthwise Separable Convolution と呼ばれる層から構成された CNN である．通常の CNN では畳み込み処理を一度に処理するが，Depthwise Separable Convolution では空間方向の畳み込みとチャンネル方向の畳み込みを分離し，それぞれの畳み込みで学習を行う．MobileNet ではすべての層の後に，バッチ正規化と活性化関数として ReLU による処理が行われる．

本論文では，MobileNetV2 (MNV2)，MobileNetV3-Small (MNV3-S)，MobileNetV3-Large (MNV3-L) を使用する．V2 では Bottleneck 構造を応用した Inverted Residual と呼ばれる構造の採用が，V3 では SENet の Squeeze-and-Excitation と呼ばれる構造の採用と一部活性化関数の Hardswish への変更が行われている．

2.3 DenseNet

DenseNet [4] は，Dense ブロックと呼ばれる，すべてのブロック間がスキップ接続で繋がれた構造を持つブロックから構成された CNN である．これによりパラメータ数を抑えたまま高い複雑性を持つネットワークが構成でき，勾配消失問題の緩和にも寄与する．

本論文では，層の数が 121 層，161 層，169 層，201 層ある DenseNet121 (DN121)，DenseNet161 (DN161)，DenseNet169 (DN169)，DenseNet201 (DN201) を使用する．

2.4 Vision Transformer

Vision Transformer (ViT) [2] は，CNN とは異なり，おもに自然言語処理で使用される Transformer [11] を画像処理に応用したものである．画像を小さなパッチに分割した後にベクトル化させ，それに元の位置情報やクラス情報を付加したトークンを Transformer エンコーダーで処理する．Transformer エンコーダーは Multi-Head Attention と多層パーセプトロン層からなり，Self-Attention を活用して，トークン同士の関連度やトークンの文脈を学習する．

ViT には，Base, Large, Huge のモデルがあり，それぞれ層の数とハイパーパラメータが異なっている．また，パッチに分割する際のパッチサイズが 16 または 32 のモデルがある．本論文では，ViT-B/16，ViT-L/16，ViT-B/32，ViT-L/32 を使用する．

2.5 画像分類の評価方法

本論文では，画像分類における評価指標として，正解率，適合率，再現率，F 値を使用する．これらの値はそれぞれ以下のように定義される．

$$\text{正解率} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{適合率} = \frac{TP}{TP + FP}$$

$$\text{再現率} = \frac{TP}{TP + FN}$$

$$\text{F 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

ここで，TP（真陽性）は，モデルが真であると予測した画像のうち，実際に真である画像の数，TN（真陰性）は，モデルが偽であると予測した画像のうち，実際に偽である画像の数，FP（偽陽性）は，モデルが真であると予測した画像のうち，実際には偽である画像の数，FN（偽陰性）は，モデルが偽であると予測した画像のうち，実際には真である画像の数である．

3. 物体検出器

3.1 YOLOv5

Redmon ら [8] によって開発された YOLO（You Only Look Once）は，その名の通り物体の位置と種類を同時に予測する検出器であり，従来の物体検出器よりも高速に処理を行うことができる．

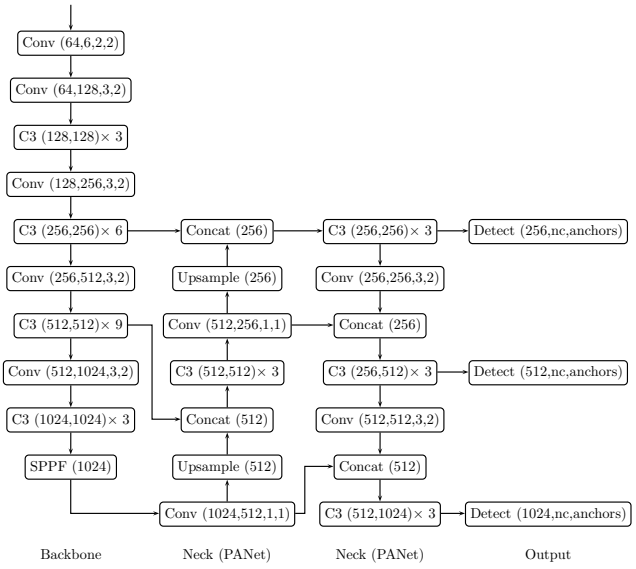


図 2: YOLOv5 のアーキテクチャ
Fig. 2 Architecture of YOLOv5

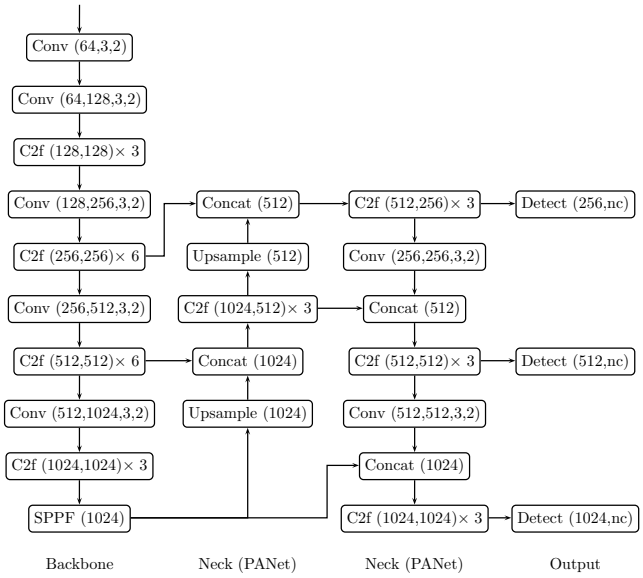


図 3: YOLOv8 のアーキテクチャ
Fig. 3 Architecture of YOLOv8

YOLOv5 [5] は，Ultralytics 社によって提案された物体検出器であり，**図 2** のアーキテクチャが示す通り，Backbone，Neck，Output によって構成されている．ここで，Backbone と Neck には C3 を採用している．

YOLOv5 では，アンカーボックスと呼ばれる仕組みが採用されている．YOLOv5 は入力画像を正方形にリサイズしたの後グリッドに分割する．ここで，各グリッドの中心をアンカーとしてアンカーボックスを生成し，このアンカーボックスに対して予測を行う．YOLOv5 では生成するアンカーボックスのサイズを学習することにより，学習時間の短縮と検出性能の向上を図っている．

YOLOv5 には，深さ乗数と幅乗数に応じて，n，s，m，l，x の 5 種類の事前モデルが準備されている．**表 1** に，YOLOv5 の各事前モデルにおける深さ乗数および幅乗数を示している．本論文では，5 種類の事前モデルをすべて使用する．

3.2 YOLOv8

YOLOv8 [6] は，YOLOv5 と同様に Ultralytics 社によって提案された物体検出器であり，**図 3** のアーキテクチャが示す通り，YOLOv5 と同じく Backbone，Neck，Output によって構成されている．ここで，Backbone と Neck には C2f を採用している．

YOLOv8 では，事前にアンカーボックスを生成する必要のないアンカーフリーの手法が採用されている．この手法によりネットワークがアンカーの位置に依存しない自由な予測が可能になっており，違いの大きな物体の検出などに対する検出性能の向上が期待される．

YOLOv5 と同様に YOLOv8 にも，深さ乗数と幅乗数に応じて，n，s，m，l，x の 5 種類の事前モデルが準備されている．表 1 に，YOLOv8 の各事前モデルにおける深さ乗数および幅乗数を示している．本論文では，5 種類の事前モデルをすべて使用する．

表 1: YOLOv5 と YOLOv8 の事前モデルの深さ乗数と幅乗数
Table 1 Depth Multiple and Width Multiple for Each Prepared Model in YOLOv5 and YOLOv8

	YOLOv5		YOLOv8	
モデル	深さ乗数	幅乗数	深さ乗数	幅乗数
n	0.33	0.25	0.33	0.25
s	0.33	0.50	0.33	0.50
m	0.67	0.75	0.67	0.75
l	1.00	1.00	1.00	1.00
x	1.33	1.25	1.00	1.25

3.3 物体検出の評価方法

物体検出における予測の正誤判定には IoU（Intersection over Union）を用いる．実際に対象を囲んだ矩形を正解矩形，モデルが予測した対象を囲んだ矩形を予測矩形とするとき，IoU は以下のように定義される．

$$\mathrm{IoU} = \frac{\text{正解矩形と予測矩形の共通面積}}{\text{正解矩形と予測矩形の面積和}}$$

そして，IoU が一定の閾値を超えた場合に，モデルは予測に成功した (TP) と判定する．一方で，本来は検出対象の物体が入力画像に映っているにも関わらずモデルが予測しなかった場合 (FP)，および，本来は検出対象の物体ではないのにモデルが誤って検出した場合 (FN) は誤検出と判定する．物体検出における適合率，再現率，F 値では，IoU の閾値は 0.5 とする．

さらに，平均適合率（Average Precision，AP）は，同一

表 2: 画像分類の実験環境
Table 2 Computer Environment for image classification

OS	Windows 10 Home
CPU	11th Gen Intel(R) Core(TM) i9-11900 @ 2.50GHz
RAM	16.0 GB
GPU	NVIDIA GeForce RTX 3060

表 3: 物体検出の実験環境
Table 3 Computer Environment for object detection

OS	Windows 11 Home
CPU	AMD Ryzen(TM) 5 4500 6-Core Processor @ 3.60GHz
RAM	16.0 GB
GPU	NVIDIA GeForce RTX 3060

再現率の画像の適合率の平均をすべての再現率に対して平均した値である．複数のクラスがある場合は，各クラスごとに AP を算出する．mAP（mean AP）は，閾値を 0.5 から 0.95 まで 0.05 間隔で求めた AP の平均値である．

4. 白血球貪食の分類と検出

4.1 計算機環境と交差検証

表 2 と**表 3** は，画像分類および物体検出における計算機環境である．また，分類では学習回数を 100 エポック，バッチサイズを 8 に，検出では学習回数を 500 エポック，バッチサイズを 8 に設定している．

本論文では，分類では 5 分割，検出では 3 分割の交差検証を適用する．5 分割交差検証では，すべてのデータをランダムに振り分け 5 等分し，3 つを訓練データ，1 つを検証データ，1 つをテストデータとする．3 分割交差検証では，すべてのデータをランダムに振り分け 3 等分し，1 つを x 軸反転と y 軸反転を用いて 3 倍にデータ拡張した後に訓練データに，1 つを検証データ，1 つをテストデータとする．モデルは，訓練データによって訓練し，モデルのパラメータを調整する．また，訓練中 1 はエポックごとに検証データを使って過学習が起きていないかなどを検証する．最後に訓練が終了した後，テストデータを使用してモデルを評価する．これを 1 セットとし，訓練データ，検証データ，テストデータの分け方を変えて分類では 5 セット，検出では 3 セット行って各セットの平均を評価指標とする．

4.2 白血球貪食の分類

白血球貪食の分類では，分類対象の映る画像全体から白血球の映る領域のみを矩形として切り出し，その矩形を画像分類器に入力することで画像を分類する．ここでは，貪食，擬貪食，非貪食の 3 クラス，および，擬貪食像と非貪食像を合わせて 1 つの非貪食クラスとした貪食と非貪食の 2 クラス分類を行う．

表 4 は，画像分類において使用する画像枚数である．

表 4: 画像分類における画像枚数
Table 4 Number of Images in Image Classification

	対象	画像枚数
貪食像	貪食	243
非貪食像	擬貪食	249
	非貪食	800

表 5: 3 クラス分類の結果・貪食像と擬貪食像
Table 5 Results of Three-Class Classification – Phagocytic and Pseudo-Phagocytic Images

	貪食像			擬貪食像		
画像分類器	適合率	再現率	F 値	適合率	再現率	F 値
VGG16	**0.645**	**0.729**	**0.676**	0.504	0.493	0.484
VGG19	0.596	0.687	0.605	0.488	0.366	0.405
MNV2	0.426	0.655	0.515	0.350	0.024	0.044
MNV3-S	0.497	0.618	0.531	0.443	0.088	0.131
MNV3-L	0.505	0.547	0.522	0.377	0.060	0.103
DN121	0.518	0.708	0.590	0.552	0.301	0.349
DN161	0.557	0.623	0.574	**0.557**	0.301	0.362
DN169	0.576	0.545	0.532	0.529	0.280	0.323
DN201	0.532	0.626	0.563	0.542	0.276	0.332
ViT-B/16	0.587	0.638	0.591	0.532	**0.498**	**0.498**
ViT-L/16	0.572	0.692	0.611	0.533	0.467	0.485
ViT-B/32	0.572	0.527	0.524	0.505	0.458	0.461
ViT-L/32	0.627	0.509	0.529	0.421	0.409	0.412

表 6: 3 クラス分類の結果・非貪食像と 3 クラス平均
Table 6 Results of Three-Class Classification – Non-Phagocytic Images and Three-Class Average

	非貪食像			3 クラス平均			
画像分類器	適合率	再現率	F 値	適合率	再現率	F 値	正解率
VGG16	**0.858**	0.813	0.834	**0.669**	**0.678**	**0.664**	**0.735**
VGG19	0.846	0.843	**0.842**	0.643	0.632	0.617	0.721
MNV2	0.766	0.861	0.810	0.514	0.514	0.456	0.661
MNV3-S	0.756	0.864	0.804	0.487	0.523	0.401	0.668
MNV3-L	0.745	**0.926**	0.826	0.543	0.511	0.484	0.688
DN121	0.830	0.816	0.819	0.633	0.608	0.586	0.697
DN161	0.805	0.871	0.833	0.640	0.598	0.590	0.714
DN169	0.785	0.881	0.828	0.630	0.569	0.561	0.702
DN201	0.790	0.848	0.816	0.621	0.583	0.570	0.696
ViT-B/16	0.830	0.796	0.812	0.650	0.644	0.634	0.709
ViT-L/16	0.832	0.786	0.804	0.646	0.648	0.633	0.707
ViT-B/32	0.803	0.813	0.804	0.627	0.599	0.596	0.690
ViT-L/32	0.777	0.799	0.784	0.608	0.572	0.575	0.669

表 5 と**表 6** は，それぞれの画像分類器における 3 クラス分類での評価指標の値であり，**表 7** と**表 8** は，それぞれの画像分類器における 2 クラス分類での評価指標の値である．また，表中の太字はそれぞれの指標における最大値を表している．

表 7: 2 クラス分類の結果・貪食像と非貪食像
Table 7 Results of Two-Class Classification – Phagocytic and Non-Phagocytic Images

	貪食像			非貪食像		
画像分類器	適合率	再現率	F 値	適合率	再現率	F 値
VGG16	**0.752**	0.589	**0.646**	0.910	0.948	**0.928**
VGG19	0.649	**0.666**	0.645	**0.923**	0.906	0.913
MNV2	0.675	0.328	0.395	0.860	0.944	0.899
MNV3-S	0.580	0.280	0.369	0.852	0.954	0.900
MNV3-L	0.630	0.371	0.454	0.867	0.947	0.905
DN121	0.661	0.588	0.604	0.908	0.922	0.914
DN161	0.713	0.516	0.549	0.897	0.925	0.906
DN169	0.677	0.490	0.557	0.889	0.939	0.913
DN201	0.743	0.490	0.587	0.891	**0.960**	0.924
ViT-B/16	0.690	0.568	0.610	0.904	0.931	0.916
ViT-L/16	0.574	0.605	0.577	0.908	0.889	0.897
ViT-B/32	0.569	0.540	0.529	0.895	0.889	0.890
ViT-L/32	0.577	0.544	0.500	0.898	0.876	0.881

表 8: 2 クラス分類の結果・2 クラス平均
Table 8 Results of Two-Class Classification – Two-Class Average

	2 クラス平均			
画像分類器	適合率	再現率	F 値	正解率
VGG16	**0.831**	0.769	**0.787**	**0.881**
VGG19	0.786	**0.786**	0.779	0.861
MNV2	0.768	0.636	0.647	0.828
MNV3-S	0.716	0.617	0.635	0.827
MNV3-L	0.749	0.659	0.679	0.838
DN121	0.785	0.755	0.759	0.859
DN161	0.805	0.721	0.728	0.848
DN169	0.783	0.715	0.735	0.854
DN201	0.817	0.725	0.755	0.872
ViT-B/16	0.797	0.750	0.763	0.863
ViT-L/16	0.741	0.747	0.737	0.836
ViT-B/32	0.732	0.715	0.710	0.824
ViT-L/32	0.737	0.710	0.691	0.813

表 5 から表 8 より，全体的に VGG16 および VGG19 の評価指標の値が大きい．特に，表 6 と表 8 より，クラス平均の結果では 2 クラス平均の再現率を除いてすべて VGG16 が最大値であり，正解率も 2 クラス分類と 3 クラス分類共に VGG16 が最大である．ただし，3 クラス分類における擬貪食では，ViT-B/16 の再現率と F 値が最大である．また，ほとんどの場合において VGG16 の評価指標の値の方が VGG19 の評価指標の値よりも大きいなど，層数を多くしても必ずしも評価指標の値が大きくなるとは限らないことが分かる．

表 9: 物体検出における矩形画像の枚数
Table 9 Number of Rectangles in Object Detection

	対象	矩形数
貪食像	貪食	243
非貪食像	擬貪食	249
	非貪食	801

表 10: 物体検出の結果
Table 10 Results of Object Detection

	3 クラス検出			2 クラス検出	
物体検出器	貪食	擬貪食	非貪食	貪食	非貪食
YOLOv5n	0.190	0.152	0.292	0.201	0.405
YOLOv5s	**0.434**	**0.294**	**0.463**	0.340	0.484
YOLOv5m	0.236	0.177	0.352	0.244	0.447
YOLOv5l	0.256	0.168	0.356	0.262	0.447
YOLOv5x	0.241	0.179	0.351	0.267	0.458
YOLOv8n	0.314	0.211	0.385	0.321	0.476
YOLOv8s	0.354	0.224	0.413	**0.388**	**0.519**
YOLOv8m	0.346	0.235	0.421	0.358	0.515
YOLOv8l	0.276	0.203	0.405	0.306	0.499
YOLOv8x	0.264	0.202	0.402	0.296	0.486

4.3 白血球貪食の検出

白血球貪食の物体検出でも画像分類と同様，貪食，擬貪食，非貪食の 3 クラス検出を行った後，擬貪食像と非貪食像を合わせて 1 つの非貪食クラスとし，貪食と非貪食の 2 クラス検出を行う．**表 9** は，物体検出において使用する矩形画像枚数である．1 枚の画像に各クラスの矩形が混在しているためクラスごとの画像枚数は記載していないが，データセット全体では 202 枚の画像を用いている．**表 10** は，それぞれの物体検出器における 3 クラス検出と 2 クラス検出検出における mAP である．また，表中の太字はそれぞれのクラスにおける mAP の最大値を表している．

表 10 より，3 クラス検出では YOLOv5s が，2 クラス検出では YOLOv8s がすべてのクラスの mAP を最大とする．また，YOLOv5s を除くすべての検出器では，2 クラス検出における貪食クラスの mAP が 3 クラス検出における貪食クラスの mAP よりも大きいが，YOLOv5s では逆に小さい．さらに，画像分類と同様に，単純に事前モデルを大きくしても必ずしも mAP が大きくなるとは限らないことが分かる．

5. まとめと考察，今後の課題

本論文では，グラム染色画像からの白血球貪食の分類と検出に取り組んだ．その結果，白血球貪食の分類では VGG16 が最も検出性能が高い画像分類器であった．また，白血球貪食の検出では，3 クラス検出で YOLOv5s が，2 クラス検出で YOLOv8s が最も検出性能が高い物体検出器であった．なお，菌の検出では検出性能が低い事前モデルで

ある YOLOv5s が白血球貪食の 3 クラス検出では検出性能が高かったのは，菌の画像と比較して白血球像が大きいことが理由だと考えられる．

今後の課題としては，ConvNeXt や YOLO11 といった新たな画像分類器や物体検出器の適用，よりエポック数を増やした分類と検出，データ拡張を利用した分類と検出などが挙げられる．

参考文献

[1] J. W. Bartholomew, T. Mittwer: *The Gram stain*, Bacteriol. Rev. **16**, 1–29 (1952). https://doi.org/10.1128/br.16.1.1- 29.1952.

[2] A. Dosoviskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby: *An image is worth 16 × 16 words: Transformers for image recognition at scale*, Proc. ICLR'21 (2021). https://openreview.net/forum?id=YicbFdNTTy.

[3] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M Andreetto, H. Adams: *MobileNets: Efficient convolution neural networks for mobile vision applications*, arXiv:1704.04861 (2017).

[4] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger: *Densely connected convolution networks*, Proc. CVPR'17, 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243.

[5] G. Jocher: *YOLOv5*, https://github.com/ultralytics/yolov5 (2020).

[6] G. Jocher, A Chaurasia, J. Qui: *YOLOv8 by Ultralytics*, https://github.com/ultralytics/ultralytics (2023).

[7] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár: *Focal loss for dense object detection*, Proc. ICCV'17, 2980–2988 (2017). https://doi.org/10.1109/ICCV.2017.324.

[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi: *You only look once: Unified, real-time object detection*, Proc. CVPR'16, 778-788 (2016). https://doi.org/10.1109/CVPR.2016.91.

[9] S. Ren, K. He, R. Girshick, J. Sun: *Faster R-CNN: Towards real-time object detection with region proposal networks*, IEEE Trans. Pattern Anal. Mach. Intell. **39**, 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031.

[10] K. Simonyan, A. Zisserman: *Very deep convolution networks for large-scale image recognition*, Proc. ICLR'15 (2015).

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin: *Attention Is All You Need*, Proc. NIPS'17, 6000–6010 (2017).

[12] K. Yoshihara, K. Hirata: *Detecting Campylobacter bacteria and phagocytotic activity of leukocytes from Gram stained smears images*, Proc. ESKM'21, 10–15 (2021). https://doi.org/10.1109/IIAI-AAI53430.2021.00002.

[13] K. Yoshihara, K. Hirata: *Object detection as Campylobacter bacteria and phagocytotic activity of leukocytes from Gram stained smears images*, Proc. ICPRAM'22, 534–541 (2022). https://doi.org/10.5220/0000155500003122.