

令和 2 年度 修士論文

出芽酵母の細胞形態からその染色体構成を予測する

Identification and classification of chromosome
compositions from cell morphology in budding yeast

高知工科大学大学院
工学研究科 基盤工学専攻 生命科学コース
学籍番号 1235041 清水公貴 Koki Shimizu

指導教員 田中誠司

令和 3 年 3 月 16 日 提出

目次

1. 要旨	5
2. 緒言	7
3. 材料と研究方法	8
3-1. 実験材料	8
3-1-1. 細胞株	8
3-1-2. プラスミド	9
3-2. 培地	10
3-3. 実験方法	11
3-3-1. 出芽酵母形質転換 (Yeast transformation)	11
3-3-2. パルスフィールドゲル電気泳動のためのプラグ作製方法	12
3-3-3. パルスフィールドゲル泳動法 (PFGE: Pulsed-Field Gel Electrophoresis)	13
3-3-4. 細胞の顕微鏡観察、撮影	13
3-4. 異数性細胞の単離方法	14
3-4-1. 過剰複製による異数性細胞の単離方法	14
3-4-2. 動原体機能阻害による異数性細胞の単離方法	14
3-5. 使用した計算機と構築した環境	16
3-6. 深層学習のモデルとその構造	16
3-6-1. 畳み込みニューラルネットワーク	16
3-6-2. 畳み込み層	16
3-6-3. プーリング層	17
3-6-4. 全結合層	17
3-6-5. 出力層	17
3-6-6. 損失関数	17
3-6-7. 勾配降下法	17
3-6-9. Xception	17

3-6-10. ミニバッチ学習	18
3-6-11. ファインチューニング	18
3-6-12. ドロップアウト	18
3-6-13. 本研究で利用したモデルの構造	18
3-6-14. モデルの評価方法	19
3-6-15. 入力画像の前処理	19
3-7. YeastSpotter	20
3-8. 細胞の検出	20
3-8-1. 細胞検出手法の検討	20
3-8-2. 利用した、Python ライブラリとその仕様について	20
3-8-3. 二値化処理	21
3-8-4. Non-Maximum Suppression	21
4. 結果	22
4-1. 撮影画像からの酵母細胞検出に際し、NMS 処理がある程度有効に機能した	22
4-2. YeastSpotter が生成するマスク画像を利用することで、酵母細胞をより効率よく検出 することができた	23
4-3. 検出した酵母細胞を 82.3%の正解率で細胞周期ごとに分類することができた	23
4-4. 短い染色体では安定な形態を示す 2n-1 細胞を単離することができた	25
4-5. 第 I, III, VI, IX 番染色体モノソミー細胞、野生型細胞を学習させたところ、III 番染色 体を除き、画像撮影日の方が結果に与える影響が大きかった	28
4-6. 細胞を混ぜて写真を撮ることで、野生型細胞と第 I, III 番染色体モノソミー細胞を 66.7%の正解率で分類できた	30
5. 考察	32
5-1. YeastSpotter を用いた細胞検出システムがうまく機能した	32
5-2. 異数体細胞の単離にはそれぞれ適した方法がある	32
5-3. H2A-GFP/mCherry を用いた系の評価	33
5-4. 第 III 番染色体では、細胞周期に依存しない特徴がある	33
5-5. 細胞形態から染色体構成を予測するにあたっての深層学習の有用性	34
6. 文献	36

7. 謝辞	38
-------	----

1. 要旨

出芽酵母 *Saccharomyces cerevisiae* は真核生物のモデル細胞であり、約 6000 個の全遺伝子のうち約 4700 個の非必須遺伝子を個別に破壊した遺伝子破壊ライブラリーが早くから整備されたことで、それを利用した多数の網羅的な解析が行われてきた。その一つに、細胞形態と遺伝子機能との関係性を解析したものがある¹。そこでは、各遺伝子破壊株について、CalMorph と名付けられた新規プログラムを用いて、細胞形態・アクチン骨格・核の形態の蛍光顕微鏡画像データから「細胞の大きさ」や「楕円への近さ」等、501 個の形態表現型の定量的な抽出が行われた。その結果、同一あるいは類似した生物学的プロセスに属する各遺伝子機能の欠損が類似した細胞形態を生じること、すなわち、細胞の「形」はその「生理機能」を反映し、ひいては細胞の形から未解析遺伝子の機能すら予測可能であることも示された。さらに、本手法は、薬剤の標的経路の予測や酵母細胞の生理状態の把握等、様々な局面で使える強力な手法であることも示されている。このような背景のもと、私は細胞形態から、異数性のような染色体構成の異常を予測・同定することが可能なのではないかと考えた。一般的に、染色体異常はそのゲノム構成が変化してゆく不安定化状態を経て生じる。CalMorph に加え、新規な予測手法として深層学習を適用することができれば、様々な染色体異常を持つ酵母細胞のより迅速な同定や、細胞分裂に伴い染色体構成が変化している状態の細胞レベルでの追跡も可能となることが期待できる。

本研究では、様々な異数性細胞の解析を試みたが、特に解析が進んだ、二倍体 ($2n$) 細胞から個々の染色体について 2 本のうち 1 本を失った ($2n-1$) 細胞株の解析を代表例として以下に述べる。

深層学習においては、教師データによる事前のトレーニングが必須である。そこでまず、コントロールとなる野生型 $2n$ 細胞を用いて、効率的な細胞認識手法の確立を行った。まず細胞を透過光撮影した画像から YeastSpotter² を用いて酵母細胞を抽出し、 100×100 pixel の背景画像と合成したものを教師データとした。その際、出芽酵母はその細胞周期を通じて形態が変化するため、より同じ特徴を備えた画像を学習させることが望ましいと考え、細胞周期ごとに分類することとした。生産性の高い CNN モデル Xception を用いて細胞を 6 種類 [G1, S, G2/M 期, 例外3種 (Contamination, 複数細胞, 見切れた細胞)] に分類するよう学習を行ったところ、82%の精度で分類するモデルの作製に成功した。

$2n-1$ 細胞株の解析にあたっては、まず、片方の染色体のセントロメア機能を人工的に阻害することで染色体の不分離に起因する欠損を誘導^{3,4}して $2n-1$ 候補細胞を取得し、それらの染色体構成をパルスフィールドゲル電気泳動法で確認した。本手法で 16 本ある染色体のうち、サイズの小さい染色体(第1,3,6,9番)については、安定な $2n-1$ 株が得られた。その他の染色体では、候補株は増殖遅延を示し、そこから(おそらく $2n$ への復帰による)増殖速度が回復したものが一定頻度で

出現した。そのため、第1,3,6,9番染色体の $2n-1$ 株について、コントロール $2n$ 細胞と同様のモデルを用い、染色体構成の学習を行った。複数回の画像取得と学習を行ったところ、学習精度にばらつきが認められたため、写真の撮り方の誤差を認識している可能性を考えた。この可能性を排除するべく、第1,3番染色体の $2n-1$ 株を蛍光タンパクの発現と組み合わせることで「答え合わせ」ができるように工夫し、学習させた結果を検証したところ、 $2n$ 野生型細胞と第1,3番染色体の $2n-1$ 細胞とを混合したものについて、70%の精度で分類できていることがわかった。

現在は、CalMorph で解析に用いる画像と同じく、核・アクチン・細胞壁を蛍光染色した画像を撮影し、学習精度がどうなるかを比較解析している。

文献

1. Ohya Y et al. (2005) PNAS USA 102: 19015–19020.
2. Lu AX et al. (2019) Bioinformatics 35: 4525–4527.
3. Reid RJ et al. (2008) Genetics 180: 1799–1808.
4. Hill A & Bloom K. (1987) Mol Cell Biol 7: 2397–2405.

2. 緒言

生物のゲノム DNA に変異が起こった場合、細胞の死やがん化といった、さまざまな変化が現れると同時に、細胞形態も変化することがある。このような細胞形態の変化と遺伝子の変異を関連付けることは遺伝子の働きを理解するうえで重要である。出芽酵母 *Saccharomyces cerevisiae* は真核生物の良いモデル細胞であり、様々なレベルで最も解析が進んだ真核生物の一つである。そのゲノムは約 6000 個の遺伝子をコードするが、全遺伝子レベルでそれらの機能や様々な表現型との関係を網羅的に調べた多くの解析がある。そのような網羅的な解析の一つとして、出芽酵母における細胞形態・アクチン細胞骨格・核の形態を顕微鏡観察することで、「細胞の大きさ」、「細胞の細長さ」、「出芽の方向」、「楕円への近さ」、等、501 個の形態表現型を解析したものがある。そこでは、上記 501 個の表現型の定量解析を可能とするために、細胞形態定量化プログラム CalMorph (カルモルフ) が開発され、約 4700 の非必須遺伝子を単独破壊した細胞について、網羅的な解析が行われた¹。その結果、類似した遺伝子機能の欠損は比較的同様の細胞形態を示すことが統計的に示され、「細胞の形」と「細胞の生理的状態」には関連があること、さらには、「細胞の形」を知ることで、機能未知遺伝子であってもその「機能」が予測可能であることが示された¹。大矢研究室のその後の解析から、CalMorph を用いた細胞形態解析は薬剤の標的分子の予想、清酒酵母の生理状態の把握等、様々な局面で使える強力な手法であることも示されてきた^{5,6}。

このような背景のもと、私は細胞形態から、個々の遺伝子レベルを超えた染色体レベルでのゲノムの状態、例えば、異数性のような染色体構成の異常を予測・同定することが可能なのではないかと考えた。一般的に、染色体異常はそのゲノム構成が変化してゆく不安定化状態を経て生じる。CalMorph に加え、新規な予測手法として深層学習を適用することができれば、様々な染色体異常を持つ酵母細胞のより迅速な同定や、細胞分裂に伴い染色体構成が変化している状態での細胞レベル・世代レベルでの追跡も可能となることが期待できる。

そこで私は、出芽酵母を材料として用いて、その細胞形態から深層学習で染色体構成を予測することが可能であるかどうか、研究を行なった。この目的で、まず様々な染色体構成をもつ細胞の単離を行い、単離した様々な細胞の顕微鏡撮影を行った。この画像データを用いて、効率的な細胞の検知、学習データの作製を行った。このようにして得た学習データを用いて深層学習を行い、予測精度の検証を行った。

3. 材料と研究方法

3-1. 実験材料

3-1-1. 細胞株

本研究で用いた株を表1に示す。

表 3.1 酵母細胞株

Name	Genotype	Reference
YST3321 (BY4741/ BY4742 diploid)	<i>MATa/MATalpha his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 met15Δ0/met15Δ0 ura3Δ0/ura3Δ0</i>	Lab. stock
YKS001	YST3321 <i>CEN1/CEN1::GAL7p-URA3</i>	This study
YKS002	YKS001 <i>Chromosome I/-</i>	This study
YKS005	YST3321 <i>CEN2/CEN2::GAL7p-URA3</i>	This study
YKS007	YST3321 <i>CEN3/CEN3::GAL7p-URA3</i>	This study
YKS009	YST3321 <i>CEN4/CEN4::GAL7p-URA3</i>	This study
YKS011	YST3321 <i>CEN5/CEN5::GAL7p-URA3</i>	This study
YKS013	YST3321 <i>CEN6/CEN6::GAL7p-URA3</i>	This study
YKS015	YST3321 <i>CEN7/CEN7::GAL7p-URA3</i>	This study
YKS017	YST3321 <i>CEN8/CEN8::GAL7p-URA3</i>	This study
YKS019	YST3321 <i>CEN9/CEN9::GAL7p-URA3</i>	This study
YKS021	YST3321 <i>CEN10/CEN10::GAL7p-URA3</i>	This study
YKS023	YST3321 <i>CEN11/CEN11::GAL7p-URA3</i>	This study
YKS025	YST3321 <i>CEN12/CEN12::GAL7p-URA3</i>	This study
YKS027	YST3321 <i>CEN13/CEN13::GAL7p-URA3</i>	This study
YKS029	YST3321 <i>CEN14/CEN14::GAL7p-URA3</i>	This study
YKS031	YST3321 <i>CEN15/CEN15::GAL7p-URA3</i>	This study
YKS033	YST3321 <i>CEN16/CEN16::GAL7p-URA3</i>	This study
YKS035	YST3321 <i>Chromosome III/-</i>	This study
YKS037	YST3321 <i>Chromosome VI/-</i>	This study

Name	Genotype	Reference
YKS039	YST3321 Chromosome IX/-	This study
YKS071	YKS001 <i>HTA1-EGFP-kanMX</i>	This study
YKS073	YKS007 <i>HTA1-mCherry-kanMX</i>	This study
YKS078	YKS071 Chromosome I/-	This study
YKS080	YKS073 Chromosome III/-	This study

3-1-2. プラスミド

Name	Genotype	Reference
pST2370	YCplac33 4xGALps	Lab Stock
pST2374	YEplac195 4xGALps	Lab Stock
pKS001	pST2374 CDC6	This study
pKS002	pKS001 DPB11	This study
pKS003	pST2370 CDC6	This study
pKS004	pKS003 DPB11	This study
pKS005	pKS004 SLD2-11D	This study
pKS006	pKS002 SLD2-11D	This study

3-2. 培地

・完全培地 YPAD (per Litter)

Yeast Extract	10g
Peptone	20g
Adenine	40 mg
(Agar	20 g)
H ₂ O	900mL

オートクレーブ後(120°C, 15 min)、別に調整・滅菌した 20% Glucose を 100 mL 加えた.

・合成培地 Synthetic Complete: SC (per Litter) plate

以下の Flask1, 2 を別々にオートクレーブ滅菌(120°C, 15 min)

Flask1	
Glucose	20 g
Agar	20 g
H ₂ O	800 mL
Flask 2	
aa mix (SC – His, Leu, Trp, Ura)	1.394 g
Adenine	40 mg
H ₂ O	100 mL

オートクレーブ後、Flask1, 2 を混合、さらに別に調整・フィルター滅菌した 10xYeast nitrogen base (YNB, 67g/L) を 100 mL 加え、1 L とした。

SC-Ura 培地作製時には、別にオートクレーブしておいた 10 mg/mL Histidine を 4.0 mL(per L)、10 mg/mL Leucine を 10.0mL(per L)、2 mg/mL Tryptophan を 20 mL(per L)加えた。

3-3. 実験方法

3-3-1. 出芽酵母形質転換 (Yeast transformation)

1. 前培養: シングルコロニーから少量の細胞を 2 mL の YPAD へ懸濁。震盪培養 30°C o/n。
2. 本培養: 前培養液を OD=0.1 となるよう YPAD を用いて希釈。震盪培養 (30°C, 3hrs)。
3. 集菌: 50 mL コニカルチューブに培養液を移し、遠心 (4000 rpm, 30 sec. (Kubota 遠心機))。
4. Wash: 細胞を 1 mL の滅菌水で懸濁後、1.5 mL tube に移し、遠心 (10,000 rpm, 5 sec. (Eppendorf 遠心機))。
5. 上清を捨て、細胞を適量の Trafo mix に懸濁し、予め DNA を入れておいた 0.2 mL tube に細胞入り Trafo mix を入れ、よく混合した。
2 μ L DNA + 25 μ L Trafo Mix
直鎖断片: 4-5 μ L DNA + 50 μ L Trafo Mix
6. 42°C で 40 分間保温 (PCR 機)
7. 軽く遠心し、細胞を回収。10,000 rpm, 5 sec. 程度 (Eppendorf)
8. 薬剤耐性 (Kan) の場合: 細胞を 1 mL の YPAD に再懸濁、30°C, 2hrs. 培養、その後、細胞を回収し、適当な培地にガラスビーズを使って均一に拡げた。
栄養要求性 (ウラシル) の場合: 細胞を適量の滅菌水に懸濁後、SC-Ura 培地にガラスビーズを使って均一に拡げた。
9. 30°C で 2o/n 培養 → 形質転換後のコロニーの出現。
10. コロニーPCR により形質転換の確認

3-3-2. パルスフィールドゲル電気泳動のためのプラグ作製方法

1. シングルコロニーから少量の細胞を 2 mL の YPAD へ懸濁。震盪培養 30°C o/n
2. OD 測定。(1 plug あたり必要な細胞は OD=0.4 とし、必要量を算出)。必要量を 1.5mL tube に取り、集菌 (cfg: 10000 ~15000 rpm, ~30 sec) 後、上清を完全に除去。
3. 細胞ペレット(~ 10 μ L)に 106 μ L の Lysis buffer (w/ Sorbitol, + Enz)と 74 μ L の 2% Low Melt Agarose (Bio-Rad's certified Low Melt Agarose #161-3111, melt in H₂O and kept on 60°C block) を加え、固まる前にピペッティングでよく混合し、plug mold に注ぐ。

Lysis Buffer (w/ Sorbitol)	per 50 ml	(final conc.)
0.5M EDTA	10.0 mL	(100 mM)
1 M Tris-Cl (7.5)	0.5 mL	(10 mM)
Sorbitol	5.47 g	(0.6 M)
2-Mercaptethanol	100 μ L	(0.20%)
H ₂ O	35.5 mL	(to 50 mL)

4. cell-agarose mix を注いだプラグを固まらせた (4°C, >30 min)。
5. plugs を 500 μ L の Lysis buffer (w/o sorbitol) 内でインキューベート (37°C, 4 hrs ~ o/n)。

Lysis Buffer (w/o Sorbitol)	per 50 ml	(final conc.)
0.5M EDTA	10.0 mL	(100 mM)
1 M Tris-Cl (7.5)	0.5 mL	(10 mM)
2-Mercaptethanol	100 μ L	(0.20%)
H ₂ O	39.4 mL	(to 50 mL)

6. Lysis buffer を抜いて、500 μ L の ProK buffer (containing 0.5 mg/ml Proteinase K)に置換し、インキューベート (50°C, o/n)。

ProK Buffer (w/ Sorbitol)	per 1.0 ml	(final conc.)
0.5M EDTA	200 μ L	(100 mM)
1 M Tris-Cl (8.0)	10 μ L	(10 mM)
10% Na-N-Lauroyl Sarcosine	100 μ L	(1.0%)
20 mg/ml Proteinase K	25 μ L	(0.5 mg/ml)
H ₂ O	665 μ L	(to 1.0 mL)

7. 冷蔵庫 (4°C) で保管

サイズの大きい染色体におけるモノソミー細胞の解析時は、上記 1 の代わりに、以下の方法で行った。

- 1: OD=0.04 となるよう集菌し YPAD プレートにガラスビーズを用いて広げ、培養 (30°C 2o/n)

3-3-3. パルスフィールドゲル泳動法 (PFGE: Pulsed-Field Gel Electrophoresis)

1. 0.5x TBE を 2.0 L 作製.
2. そのうち 120 mL を用いて 0.8% Gel (PFGE certified agarose, Bio-Rad) を作製、残りは泳動用バッファとして使用.
3. 1/3 に切ったサンプルプラグをウェルに挿入し、隙間を 0.8 %アガロースで固める.
4. 泳動 (Bio-Rad CHEF-DR-II, 設定は 6 V/cm, 50-90 sec, 22 hrs, 14°C)
5. 染色: 0.3 x GelRed in 200 mL used 0.5x TBE. 染めすぎた場合、脱イオン水で脱染.

3-3-4. 細胞の顕微鏡観察、撮影

1. 前培養: シングルコロニーから少量の細胞を 2 mL の YPAD へ懸濁。震盪培養 (30°C 18hrs ~ 20hrs)。
2. 本培養: 前培養液を OD=0.1 となるよう YPAD を用いて希釈。震盪培養 (30°C, 3hrs ~ 4.5hrs)。
3. 1.5mL tube に移し、超音波処理 (125watt 20kHz 6second)
4. 遠心にて集菌 (12000 rpm, 1 min)
5. 0.7 μ L の 99% Glycerol をスライドグラスに乗せチップを使って突いた少量の細胞を懸濁
6. カバーガラスを乗せ、裏返した後、湿度を保った状態で静置 (10min)
7. EVOS FL Auto2 (Thermo Scientific 社) 顕微鏡にて、40 倍、透過光 (0.1) {蛍光撮影時は GFP 緑光、mCherry 赤光}、Z-stack モード*で細胞を撮影。

*Z-stack 条件は以下の通り。0.25x Depth of Field, Step Size: 0.3 μ M, Number of planes: 9.

3-4. 異数性細胞の単離方法

3-4-1. 過剰複製による異数性細胞の単離方法

出芽酵母細胞では、Sld2-D と Cdc45^{JET1} を同時に発現させると過剰複製が起こり、ほとんどの細胞は死ぬが、生き残ってくるサバイバーには高頻度にゲノムが不安定化した細胞が含まれる^{7,8}。この系を利用するために、ガラクトース誘導性プロモーター (*GALp*) 下流に sld2-11D と *CDC45-JET1* を組み込んだ Yep/YCp タイプの plasmid (pKS01-06) を作製・酵母細胞に導入し、非同調、あるいは G1 期停止させた同調環境下で目的タンパク誘導を行った。同時に培養液の一部をプレートに撒き、細胞生存率をモニターしたが、生存率が大きく下がるような条件は見出せなかったため、この方法による単離は一時停止した。

3-4-2. 動原体機能障害による異数性細胞の単離方法

図 3.1 に本方法の原理を模式的に示す。ガラクトース誘導性プロモーター (*GALp*) 下流にセントロメアを配置し、転写誘導をかけると、セントロメアの機能が損なわれることが知られている^{3,4}。そこで、各染色体のセントロメア近傍に、URA3-GAL7p を挿入できるような PCR 産物を作製し、2倍体細胞 (YST3321) に導入し、図 3.1 上側に示したような酵母株を作製し、その後、*GALp* 発現誘導によるセントロメアの無効化処理を行った。PCR 産物を組み込むためのオリゴヌクレオチド配列は、基本的には Reid et al., (2008)³ と同じものを用いた。

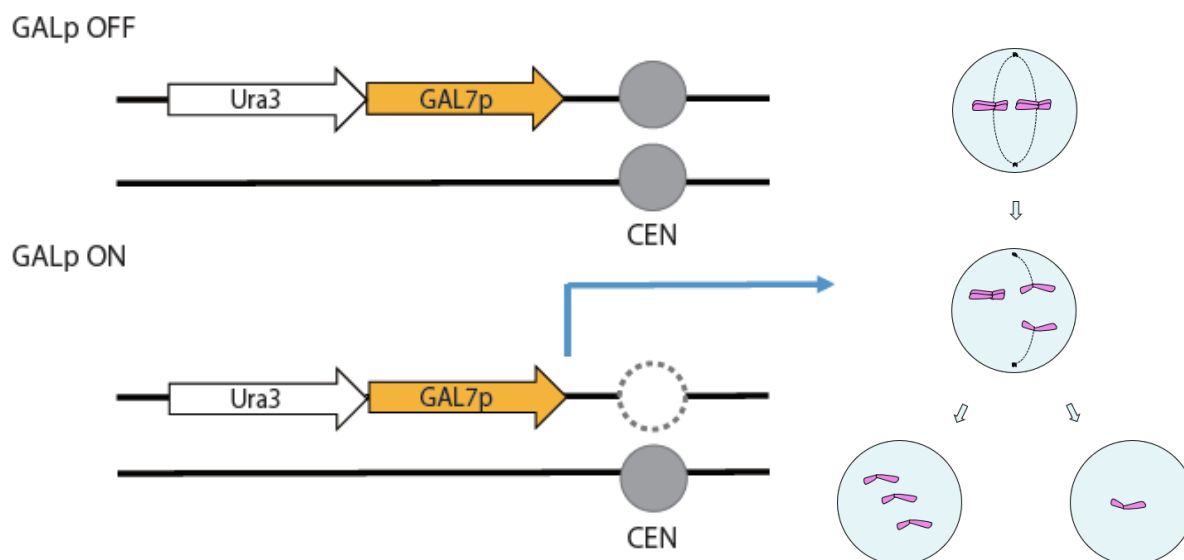


図 3.1 染色体欠損誘導の原理

下記に、実験の方法を示す。

1. PCR により、各染色体 Centromere 領域 の左側*の近傍領域の中央部に (800~1200 bp)に URA3-GAL7p を持つ DNA 断片を overlapping PCR にて作製し、アガロースゲルを用いて分離・精製した。PCR に用いたオリゴヌクレオチドは、研究室のストック番号で#4665-4668, 4703-4789.
2. 作製した、直鎖断片を 3-3-1 に記載の方法により導入。Ura+となった形質転換体について、URA3-GAL7p を持つことを、コロニーPCR で確認した。
3. 前培養: シングルコロニーから少量の細胞を 2 mL の YPA-raffinose へ懸濁。震盪培養 (30°C o/n)
4. 本培養: 前培養液を OD=0.1 となるよう YPA-raffinose を用いて希釈。震盪培養 (30°C, 3hrs~4.5hrs)
5. 2%となるようガラクトースを加え、震盪培養 (30°C, 4hrs~o/n)
6. YPAD プレートにガラスビーズを用いて広げ、培養 (30°C o/n)
7. SC-URA, SC+5FOA プレートにレプリカを行い、培養 (30°C 2o/n)
8. SC+5FOA プレートにのみ生えてきたコロニーを再培養し、パルスフィールドゲル電気泳動を用いて、解析。

サイズの大きな染色体については、上記操作 5-8 の代わりに、以下の方法で行った。

1. 2%となるようガラクトースを加え震盪培養(30°C, 6hrs)
2. YPAD プレートにガラスビーズを用いて広げ、培養(30°C 2~3o/n)
3. 生えてきたコロニーのうち、生育の遅いコロニーを再培養(30°C 2~3o/n)
4. 生育のスピードが回復したクローンが高頻度に出現してくるコロニーを選択し、その中の生育の遅いクローンを、パルスフィールドゲル電気泳動を用いて、解析。

*CEN14 に関しては、Centromere DNA の左側に ORF が存在するため右側(800~1200 bp)を増幅した。

3-5. 使用した計算機と構築した環境

パソコン環境: OS: Windows、CPU: Core i7-8750H、GPU: GeForce RTX 2080 with Max-Q design、

言語: Python3.7.9、フレームワーク: Tensorflow1.14.0, Keras2.2.4

3-6. 深層学習のモデルとその構造

3-6-1. 畳み込みニューラルネットワーク

本研究では、深層学習を行うにあたり、畳み込みニューラルネットワーク(Convolutional Neural Network, 以下 CNN)を用いて酵母細胞の分類を行っている。CNN は画像認識に用いられる深層学習モデルであり、通常の Neural Network に加えて、主に複数の畳み込み(Convolution)層とプーリング(Pooling)層から構成されている(図 3.2)。

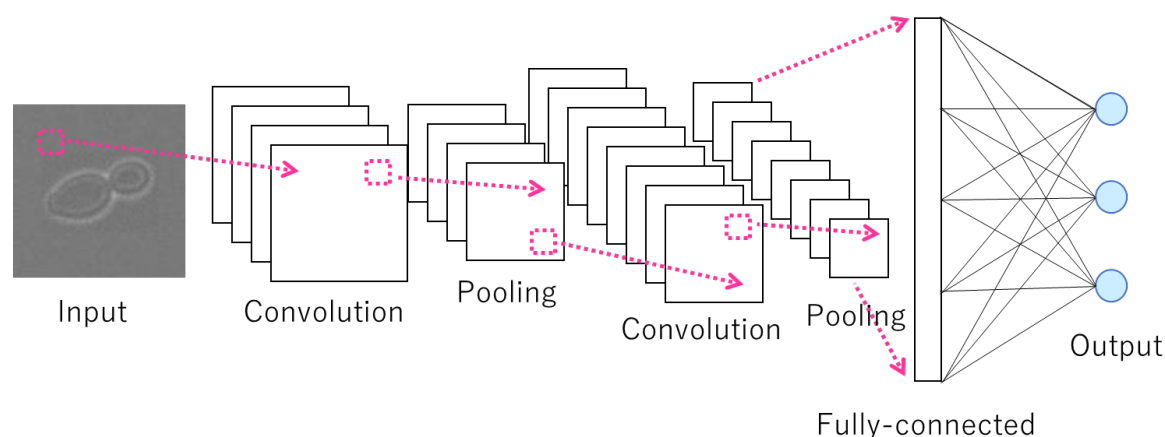


図 3.2 畳み込みニューラルネットワーク (CNN)

3-6-2. 畳み込み層

畳み込み (Convolution) 層とは画像に対して、フィルタを掛け合わせる畳み込み演算を行い、特徴量を抽出する層のことである。畳み込み演算とはマスごとにフィルタの数字を掛け合わせた後に足し合わせる処理のことを言う。

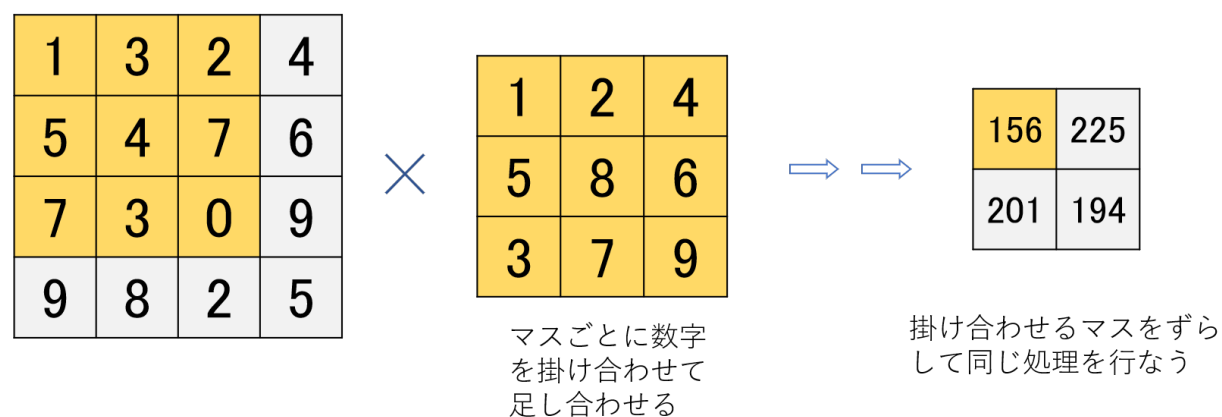


図 3.3 畳み込み演算

畳み込み層では、図 3.3 のように畳み込みを演算行なうことによって出力される画像サイズが小さくなる。加えて、この方法では画像の端の特徴を捉えきれ可能性がある。それらを考慮し、畳み込みを行なう前に、あらかじめ出力画像のサイズが入力画像のサイズと同じになるように、入力画像の周りを 0 で

覆うゼロパディングという手法を用いる場合が多い。また、より特徴を抽出できるようにするため、異なった数値のフィルタを何枚使用するかといったチューニングが必要である。

3-6-3. プーリング層

プーリング (Pooling) 層とは主に畳み込み層の後につけられる特徴を圧縮する層のことである。本研究ではマックスプーリング (Max Pooling) という手法を使用した。図 3.4 のように範囲内で最大の数値をその範囲内で数値とし特徴の圧縮処理を行う。マス大きさ、マスのずらし方によって特徴の圧縮度を決めることができる。

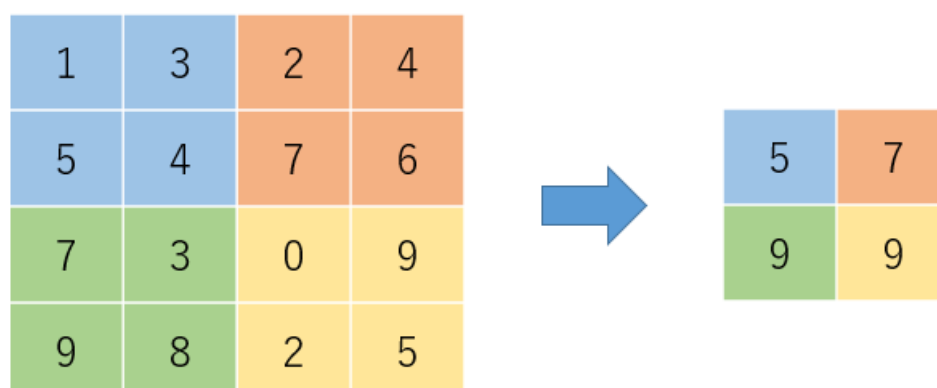


図 3.4 マックスプーリング

3-6-4. 全結合層

全結合 (Fully-connected) 層は、その名の通り全ての入力値を結合し出力する層である。CNN では、出力 (Output) 層と共に、畳み込み層と、プーリング層により抽出された特徴をもとに、画像を識別し分類する役割を担っている。

3-6-5. 出力層

出力層は、モデルの最後に配置され、全結合層からの出力をもとにして、分類を行う層である。CNN ではソフトマックス関数 (Softmax function) により、出力層における出力を確率に変換する。このようにして出力された確率をもとに学習を行う。

3-6-6. 損失関数

損失関数 (Loss Function) とは、出力層における出力と、正解データとの誤差を表す関数である。正解データは、正解ラベルのみ1であり、その他は0となっている。CNN では一般的に、損失関数は Categorical Cross Entropy が用いられることが多い。本研究においても Categorical Cross Entropy を使用しており、以下の式 3-1 で表される。

$$E = -\sum_k t_k \log y_k \quad (\text{式 3-1})$$

(t_k : 正解データ、 y_k : 出力層の出力、 k : クラス数)

3-6-7. 勾配降下法

CNN における学習では、損失関数を最小化するように学習を行う。最小化する手法として勾配降下法が用いられる。このような勾配降下法はいくつか存在し、最適化関数 (Optimizer) と呼ばれる。本研究では、Adam (Adaptive Moment Estimation) を採用している⁹。ハイパーパラメーターとして、学習率 (Learning rate: Lr)、学習率減衰 (Decay) などがある。これらのパラメーターを上手くチューニングすることで、学習の精度が向上する。

3-6-9. Xception

Xception¹⁰ とは、畳み込み層においてチャンネル方向の畳み込みと空間方向の畳み込みは完全に分けて考えることができると仮定して作成された、CNN モデルである。空間方向の畳み込み、チャンネル

方向の畳み込みを合わせ、合計 40 の畳み込み層を使用する。チャンネル方向の畳み込みとは 3-6-4 で示した畳み込みと同様であるのに対して、空間方向の畳み込みは 1×1 マスのフィルタをフィルタの枚数分用いて行なわれる。本研究では、Keras で使用可能な CNN モデルにおいて、学習時間と精度に関して最も生産性が良いことから Xception を採用している。図にその構造を示す。

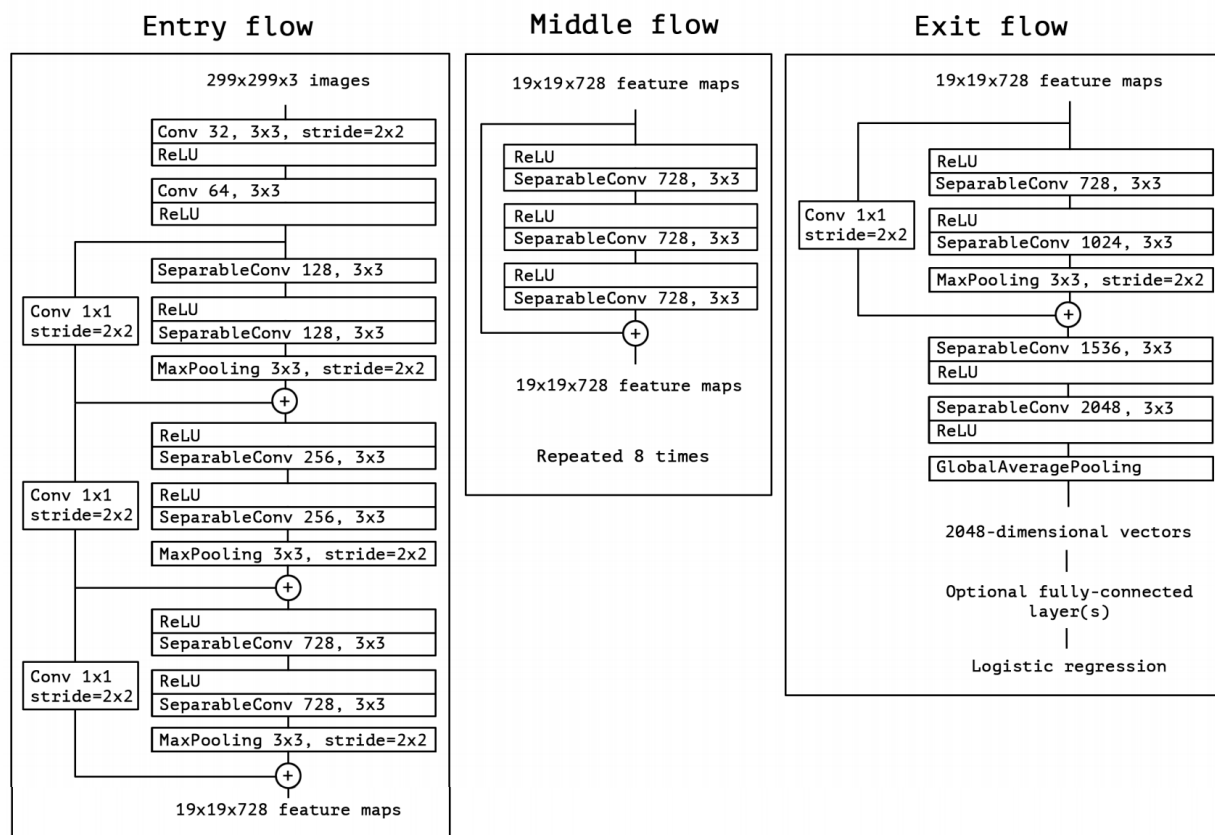


図 3.5 Xception の構造 (Chollet, 2016¹⁰)

3-6-10. ミニバッチ学習

ミニバッチ学習とは、少数のサンプル(バッチサイズ)における学習誤差を平均化してから学習を行う方法である。これにより、計算の高速化が可能である。さらに本研究で採用した Xception では畳み込み層の後にバッチノーマライゼーション (Batch Normalization) 処理が行われており、畳み込み層の出力をバッチサイズごとに標準化している。

3-6-11. ファインチューニング

ファインチューニング (Fine Tuning) とは、学習済みモデルを初期値として使用し、モデルの再学習を行う手法である。本研究では、ImageNet 学習済みモデルを初期値として使用した。

3-6-12. ドロップアウト

ニューラルネットにおいて、学習用データに対して過度に適合し、未知のデータに対しての汎用性がなくなってしまうことは過学習 (Overfitting) と呼ばれる。ドロップアウト (Dropout) はニューラルネット各層のつながりをランダムで不活化することで、このような過学習を改善する手法である¹¹。学習時に不活化するユニット数を指定することができる。本研究では、デフォルト値の 0.5 を使用した。

3-6-13. 本研究で使したモデルの構造

本研究で使したモデルの構造を下記に示す。

表 3.2 本研究で使したモデルの構造

各層	出力サイズ	関数
Input	100 × 100 × 3	–
Xception structure*	3 × 3 × 2048	–
flatten	18432	–
Fully-connected	1024	Relu
Dropout**	1024	–
Output(Fully-connected)	Number of classes***	Softmax

Loss Function: Categorical Cross Entropy

Optimizer: Adam (Lr=0.0001, Decay=0.00001)、但し、Adam の学習率の減衰に加え、validation データに対する Loss Function の値が 3 epochs 改善されない場合は Lr を 0.1 倍する。

Number of Epoch : 50epochs, 但し、Early Stopping を採用しているため validation データに対する Loss Function の値が 10 epochs 改善されない場合は学習が停止する。

* Xception structure は ImageNet を初期値として Fine-tuning を行っている。

** Dropout はデフォルトの 0.5 を採用した。

*** Number of classes は分類のクラス数、本研究では 3, 5 or 6 である。

3-6-14. モデルの評価方法

本研究では、正解率 (Accuracy)、または適合率 (Precision) を用いてモデルの評価を行っている。正解率は、全予測に対する正答率を表しており、適合率は各クラスの正答率を表している。

3-6-15. 入力画像の前処理

CNN を用いて学習させるには、入力画像が同じ大きさである必要がある。本研究では、撮影した画像中における、細胞のない領域 (100 × 100 ピクセル) を背景画像として、検出した細胞画像と合成し入力画像とした。100 × 100 ピクセルを超えるサイズの細胞画像は、そのほとんどが複数の細胞を含んでいるため、入力画像作製前に除外した。

入力画像は、あらかじめ、訓練データ、検証データ、テストデータに分けられる。訓練データは学習に使用したデータ、検証データは Adam のパラメーターの更新に使用したデータ、テストデータは学習に使用していないデータである。本研究では、主に入力画像の 1/5 をテストデータとし、残りの 4/5 をさらに 5 分割し、1/5 を検証データに、残りを訓練データとして使用した。

本研究では、訓練データは Keras の Image Data Generator を用いて、ランダムに画像を回転 (−180 ~ 180 度)、左右の反転により画像の水増しを行い、加えて、正規化 (1/255) 処理を行った。

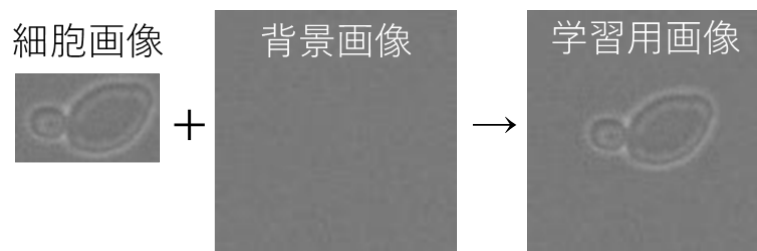


図 3.6 入力画像の合成

3-7. YeastSpotter

YeastSpotter²とは、酵母のセグメンテーションを行うツールの一つである。ヒトの核の画像を染めるために訓練した深層学習モデル MASK-RNN を出芽酵母に用いたところ、酵母細胞で訓練されていないにもかかわらず、モデルの微調整なしで酵母のセグメンテーションが可能であったツールである。本研究では、細胞を検出するために YeastSpotter における出力途中のマスク画像を使用した。また、実装には GitHub (github.com) にて公開されているプログラムを使用した。

3-8. 細胞の検出

3-8-1. 細胞検出手法の検討

近年、物体検出手法には、物体の検出と分類を同時に行う YOLO¹² (You Only Look Once) のような深層学習モデルが多く使われている。YOLO は、従来の、物体検出の後に分類を行うモデルと比較して、処理速度が速いモデルとして知られている。本研究では、研究開始当初、YOLO を使って細胞の検出と染色体構成の分類を同時に行うことを考えた。しかしながら、「While it can quickly identify objects in images it struggles to precisely localize some objects, especially small ones.」(Redmon, 2015)¹² にもあるように YOLO は小さいオブジェクトの検出が得意ではなく精度が低くなってしまう。そのため、本研究では採用しなかった。

また、物体検出の深層学習モデルには、物体の検出の後に分類を行う Faster R-CNN¹⁴ のようなモデルも存在している。本研究では細胞形態のように非常に似た画像を分類するタスク、詳細な画像分類 (Fine-Grained Visual Categorization, FGVC) を扱っている。このようなタスクではより詳細な特徴抽出が必要になるため、CNN の構造が大きくなってしまいがちが多い。そのため、Faster R-CNN のようなモデルを使用する場合、学習に時間がかかってしまい効率が悪くなってしまうことから、本研究では採用しなかった。このような理由から、細胞の検出、染色体構成の分類を分けて行う方法を採用した。

3-8-2. 使用した、Python ライブラリとその仕様について

本研究では、細胞を検出するにあたり、主に Opencv-Python、Numpy を使用した。細胞検出の検証に使用したモジュールと使用目的を下記の表に示す。

表 3.2 使用したモジュールとその用途

ライブラリ	モジュール	使用目的
OpenCV	threshold	二値化処理
	findContours	領域抽出
	contourArea	領域の面積を計算
	boundingRect	外接矩形の計算
	rectangle	外接矩形の出力

3-8-3. 二値化処理

二値化処理とは、グレイ画像において、設定した閾値をもとにして黒と白の二値に変換する処理のことである。画像を二値に変換することで、様々な画像分析が行えるようになる。本研究では、閾値は 130 とし、処理を行った。

3-8-4. Non-Maximum Suppression

物体検出では、1つの物体に対して、複数回の検出、領域の重複がよく生じる。これを解決するために検出結果を統合し1つとする Non-Maximum Suppression (NMS) が、検出後の後処理によく用いられる。本研究では、Numpy を用いて実装を行った。以下にその処理を示す。

1. 出力した矩形をリスト化する。
2. 重複した矩形の重なり(Overlap Ratio)を計算する。
3. 設定した閾値以上の Overlap Ratio を示した矩形を削除する。

4. 結果

4-1. 撮影画像からの酵母細胞検出に際し、NMS 処理がある程度有効に機能した

深層学習を行うには、撮影した画像から酵母細胞を検出する必要がある。深層学習を用いず、定量解析を行った Calmorph では、3種類の蛍光染色を用いて、細胞形態の定量化を行なっている¹が、単純な透過光画像を用いた深層学習が機能すれば、撮影が簡便になること、生細胞での解析が可能になる等、メリットが多い。そこでまず野生型 2n 細胞を透過光撮影し、その画像から酵母細胞を検出するシステムを開発することとした。透過光画像では、撮影時の焦点深度の位置の違いが画像の見かけに大きな影響を及ぼすため、z-stack モードを用いた撮影を行い、焦点位置による影響が最小となるようにした。このようにして撮影した画像を入力画像(図 4.1)として用い、二値化処理を行うことで細胞の輪郭抽出を行った。この輪郭をもとに細胞の外接矩形を検出した(図 4.2)。

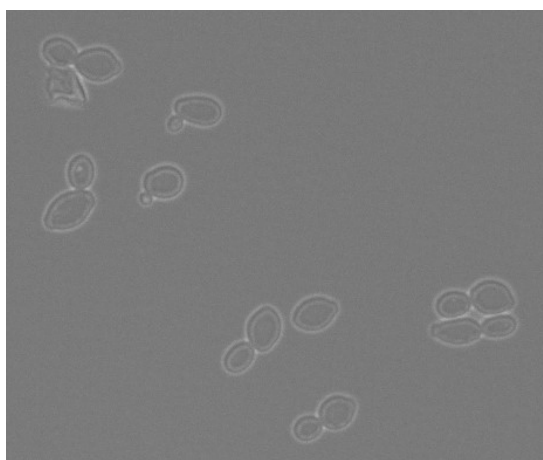


図 4.1 入力画像

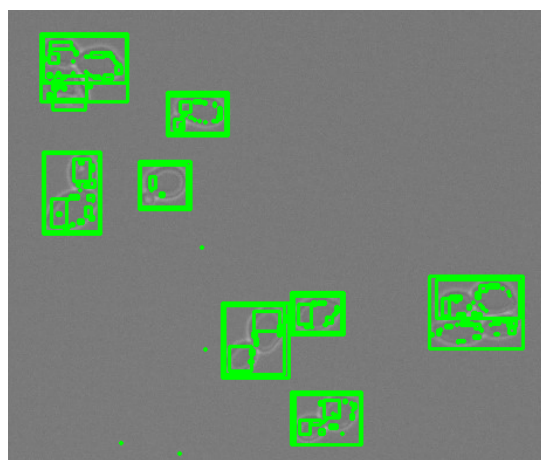


図 4.2 二値化画像による外接矩形の検出結果

このようにして検出した外接矩形はノイズが大きく、細胞をうまく検出できていなかった(図 4.2)。そこで、検出精度を改善するために、小さすぎる領域と、大きすぎる領域を、ノイズとして除外することとした。このようなノイズ除去処理を行うことで、明らかな検出精度の向上が見られた(図 4.2 と 4.3 を比較)。しかし、ノイズ除去処理後も検出領域の重複が多くみられた(図 4.3)ことから、重複した外接矩形を除去するアルゴリズムである NMS(方法: 3-8-3)を用いた。これにより、より重複の少ない細胞の検出が可能となった(図 4.4)。

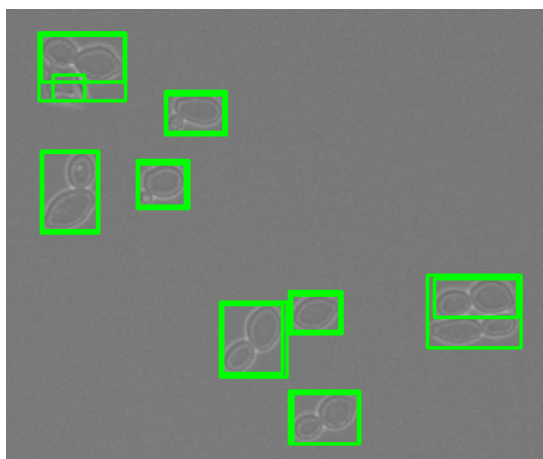


図 4.3 ノイズ除去の結果

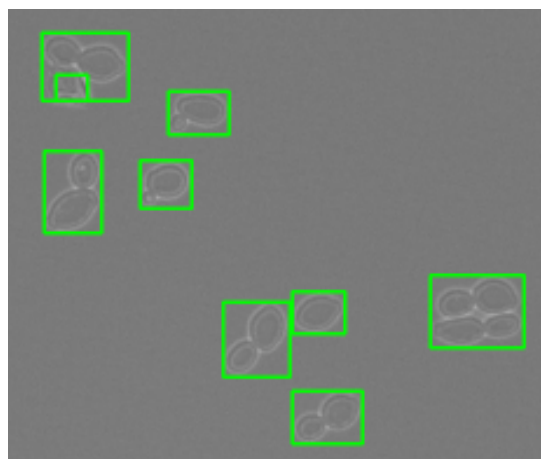


図 4.4 NMS 使用後の結果

4-2. YeastSpotter が生成するマスク画像を使用することで、酵母細胞をより効率よく検出することができた

上記 4-1 の処理により、重複の少ない酵母細胞の検出が可能となったが、依然として、検出のできていない領域や、重複が残っていた(図 4.4)。検出精度をさらに改善するために、YeastSpotter を使用してみることにした。YeastSpotter は、深層学習を使って作られた出芽酵母細胞専用のセグメンテーション(認識・分離)ツールである²。実際に web 上にて本ツールを使用したところ、単一の酵母細胞を複数の細胞として認識してしまった(図 4.5)。また、深層学習を行うには、外接矩形を検出することが望ましい。そこで、GitHub (github.com)に公開されているプログラムを用いて、YeastSpotter がその処理に生成するマスク画像を二値化処理し、酵母細胞の検出に利用してみたところ、二値化画像(マスクなし)を使用し、NMS 処理した時(図 4.4)に比べ、細胞検出精度が向上した(図 4.6)。

以上の結果より、YeastSpotter 出力途中のマスク画像を使用して酵母細胞を検出するシステムを作製し、以下の解析に使用した。

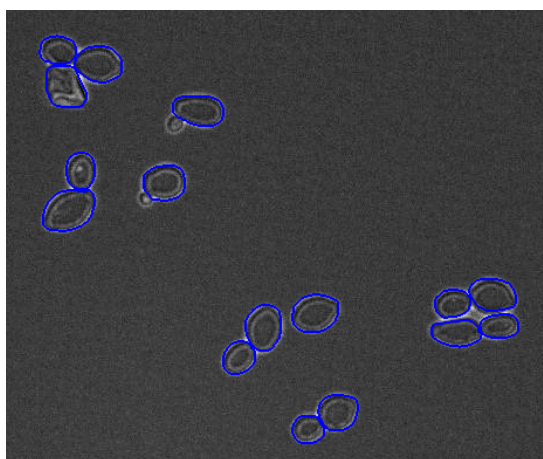


図 4.5 YeastSpotter を使用時の細胞検出結果

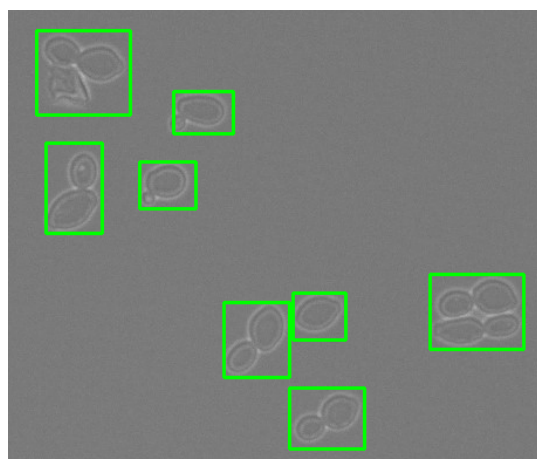


図 4.6 マスク画像使用後の細胞検出結果

4-3. 検出した酵母細胞を 82.3%の正解率で細胞周期ごとに分類することができた

CNN を用いて、機械に染色体構成の異なる画像を学習させる際、より同じ特徴を持った細胞を学習させることが望ましい。出芽酵母は、細胞周期の進行に伴って出芽し、その芽が成長して行くため、細胞周期の時期によって細胞の形が異なる。そのため、上記 4-1 で作製したシステムを用いて検出した細胞を、芽の有無とそのサイズをもとに分類した細胞周期3種類に加え、検出ミスの細胞3種類を加えた6種類に分類することとした。G1 期細胞は出芽前で芽を持たないもの、S 期、G2/M 期細胞は、芽をもつ細胞で、母細胞と娘細胞(芽)の面積比がそれぞれ 1/2 未満、1/2 以上のものとした(図 4.7)。

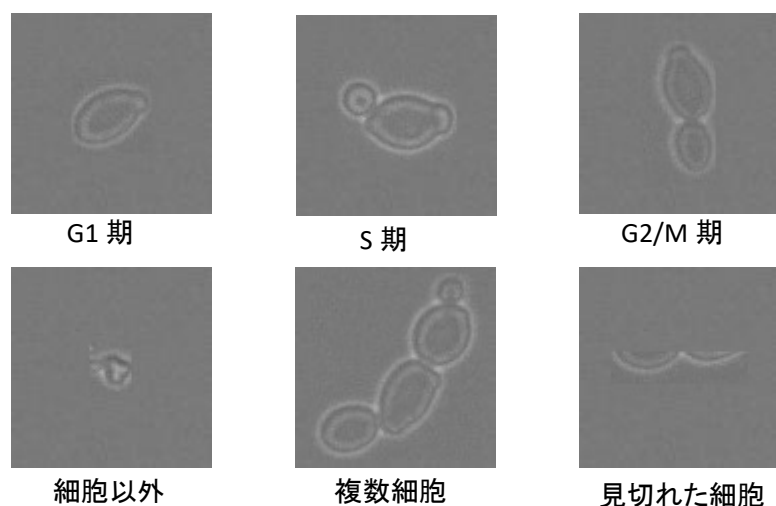


図 4.7 図 4.6 に示した方法で検出した細胞画像を6種類に分類した結果の例

37 枚の撮影画像より、1113 個の細胞を検出し、分類を行なった。次にこの画像データを用いて 3-6-13 に記載したモデルを使用し、表 4.1 に示したような画像データの振り分けで学習を行った結果、テストデータに対し、83.2%の正解率で分類することができた(表 4.2)。

表 4.1 細胞周期分類用データ

クラス	訓練(枚)	検証(枚)	テスト(枚)
G1 期	188	46	58
S 期	144	36	45
G2/M 期	188	46	58
細胞以外	32	8	9
複数細胞	80	19	24
見切れた細胞	85	21	26
合計	717	176	220

表 4.2 細胞周期学習結果

クラス	正解率[訓練]	正解率[検証]	正解率[テスト]
G1 期	1	0.957	0.914
S 期	1	0.891	0.862
G2/M 期	1	0.694	0.822
細胞以外	1	0.625	0.667
複数細胞	1	0.684	0.667
見切れた細胞	1	0.81	0.808
合計	1	0.824	0.832

この結果より、細胞の検出と大まかな分類が可能なステムが構築できた。すなわち、異なる染色体構成を持つ細胞を学習させるための準備が整ったと判断した。

4-4. 短い染色体では安定な形態を示す $2n-1$ 細胞を単離することができた

本研究では染色体構成の異なる様々な出芽酵母細胞画像を深層学習させ、分類させることを目的としているが、学習に適した既存の「染色体構成の異なる様々な出芽酵母細胞画像」はないため、異数性細胞の単離を行った(具体的な手法とその原理は 3-4 を参照)。予備的な解析より、2倍体($2n$)細胞から出発し、特定の染色体を1本失った $2n-1$ モノソミー細胞が再現性良く単離できることが分かったため、3-4-2 に示した方法でモノソミー細胞の単離を行った。このようにして単離した候補クローンについては、パルスフィールドゲル電気泳動法を用いた解析を行い、期待通りにモノソミーとなっているかどうかを決定した。この方法で、サイズの小さい染色体である第 I, VI, III, IX 番染色体について、モノソミー状態を安定に維持できる細胞を単離することができた(図 4.8)。

中～大サイズの染色体モノソミー候補クローンは、*URA3* マーカー欠損による 5-FOA 含有培地上での生育が安定しない、あるいはそもそもコロニー形成が不良となる傾向が現れた。そこで、ガラクトースによる転写誘導後、選択をかけずに YPAD プレートで培養したところ、明らかに生育の遅い細胞が現れた。これらのクローンを再培養したところ、生育のスピードが回復したクローンが高頻度に出現してくることがわかった(図 4.9)。生育の遅いコロニーより細胞を集めてパルスフィールドゲル電気泳動での解析に供したところ、バンド濃度より、それらは目的の染色体を1本欠損した細胞であることが分かった(data not shown)。第 II, IV, V, X, XI, XII, XIII, XIV, XV, XVI 番染色体モノソミー株はこのような性質を示した。

第 V 番染色体では、生育速度の異なる染色体が確認され、異数性細胞だけではなく異常な染色体構成を示した細胞が見られた(図 4.10)。

第 VIII 番染色体では、生育速度の異なるコロニーは確認できたが、再培養後、スピードが回復したクローンは確認できず、パルスフィールドゲル電気泳動法での解析結果、モノソミー細胞ではなかった(data not shown)。

第 VII 染色体については、ガラクトースでの転写誘導後に YPAD に生育の遅い候補クローンが出現しなかった。このことは、第 VII 染色体については、 $2n-1$ モノソミー株が致死となり、単離できないことを示唆する。

以上の結果より、本研究では、第 I, VI, III, IX 番染色体モノソミー細胞を撮影し、学習させることとした。

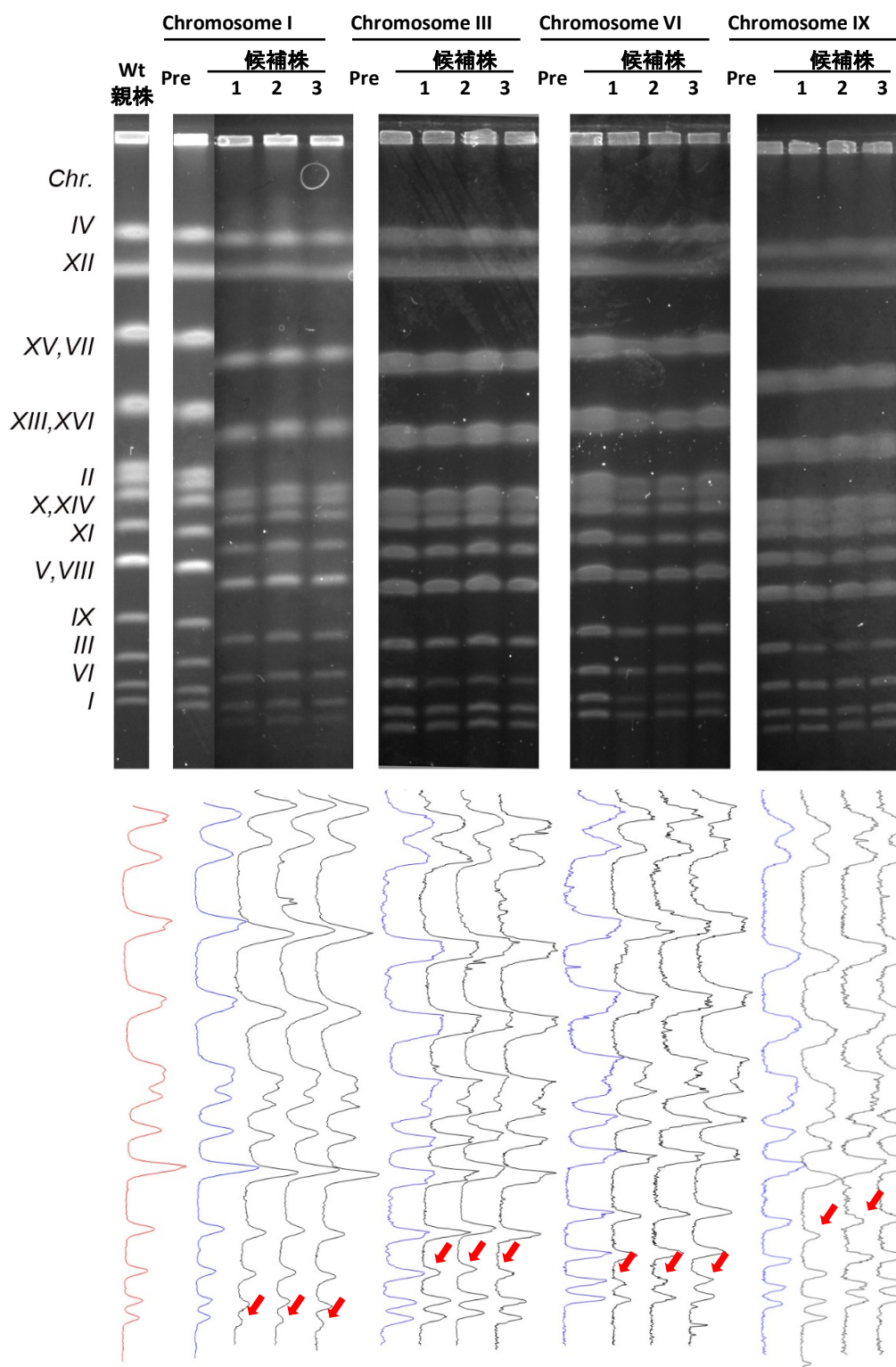


図 4.8 第 I, VI, III, IX 番染色体モノソミー株の単離

Wt (YST3321)、セントロメア機能喪失処理(ガラクトース添加処理)前の細胞(Pre: YKS001(Chr. I), YKS007 (Chr. III), YKS013 (Chr. VI), YKS019 (Chr. IX))ならびに処理後に現れたモノソミー候補クローン(候補株: 各3クローン)の染色体をパルスフィールドゲル電気泳動法で解析した結果(上: 電気泳動像、下: 泳動像より各バンドの濃さを定量したもの)を示す。ガラクトース添加処理前(Pre)と比べてバンドのシグナルが低くなった、すなわち、モノソミーとなったものを矢印で示す。

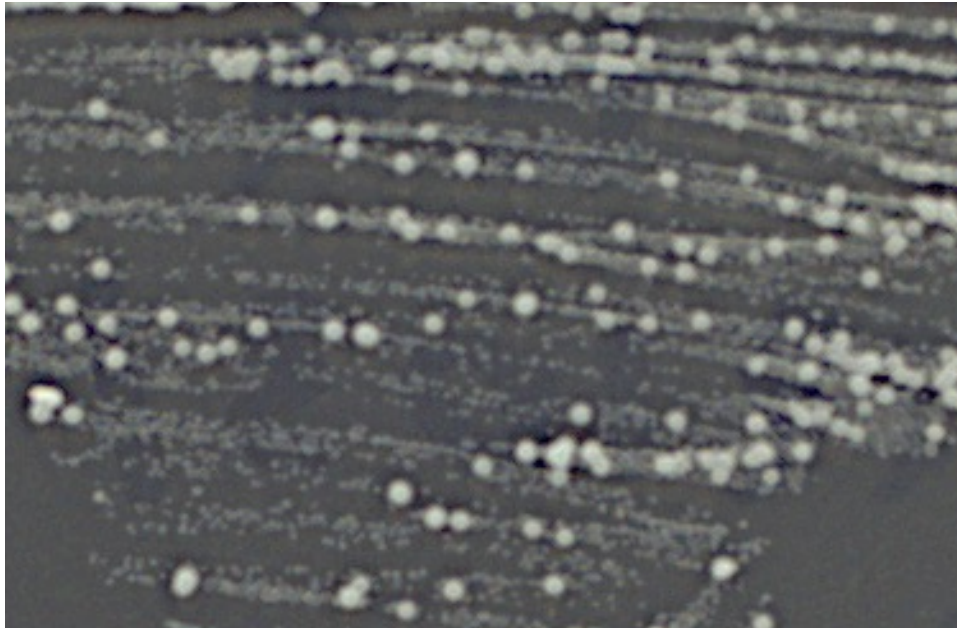


図 4.9 $2n-1$ モノソミー候補である生育の遅いコロニーをみると、生育の早いクローンが高頻度に出現する

YKS023 (CEN11::GAL7p-URA3)細胞をガラクトース培地で6時間培養した後に、YPAD プレートで2日間培養し、出現した生育の遅いコロニー(モノソミー候補株)を再び YPAD プレートで2日間再培養した結果を示す。

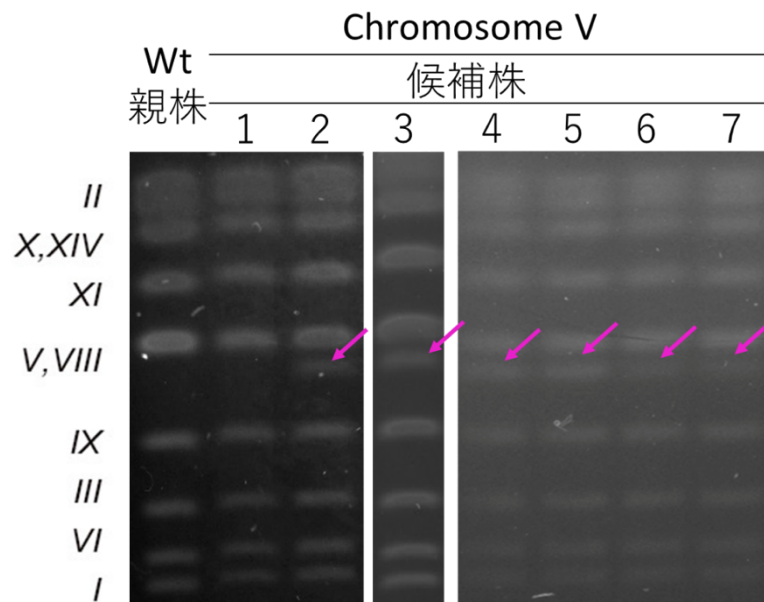


図 4.10 第 V 番染色体の $2n-1$ モノソミー候補では、バンドの短くなった株が出現した

Wt (YST3321)、YKS011(CEN5::GAL7p-URA3)細胞をガラクトース培地で12時間培養した後に、YPAD プレートで 1 日、SC+5FOA プレートで 2 日培養し、出現したコロニー(候補株 1-3)、YKS011(CEN 5::GAL7p-URA3)細胞をガラクトース培地で6時間後、YPAD プレートで2日間培養し、出現した生育の遅いコロニー(モノソミー候補株 4~7)をパルスフィールドゲル電気泳動法で解析した結果。異常な染色体のバンドを矢印で示す。

4-5. 第 I, III, VI, IX 番染色体モノソミー細胞、野生型細胞を学習させたところ、III 番染色体を除き、画像撮影日の方が結果に与える影響が大きかった

学習・評価用の細胞画像を得るため、まず、安定に増殖することが可能な、第 I, III, VI, IX 番染色体 $2n-1$ 細胞と野生型 $2n$ 細胞を撮影した。この実験では、細胞を 2 日に分け撮影した。予備解析で確立した方法を用いて撮影した画像から細胞を検出し、図 4.1 同様に、細胞周期の区別を含む 6 種類に分類した。分類した細胞データについて分類誤りの修正を手動で行った後、同日に撮影された細胞を、訓練データと検証データに分け、別日に撮影した細胞をテストデータとして(表 4.3)、細胞周期ごとに 3-6-13 に示した条件で学習を行った(表 4.4)。

表 4.3.1 モノソミー細胞データセット(G1 期)

クラス	訓練(枚)	検証(枚)	テスト(枚)
Chr1	356	88	200
Chr3	371	92	200
Chr6	168	42	200
Chr9	62	15	200
Wt	234	58	200
合計	1191	295	1000

表 4.3.2 モノソミー細胞データセット(S 期)

クラス	訓練(枚)	検証(枚)	テスト(枚)
Chr1	288	71	200
Chr3	116	28	200
Chr6	85	21	200
Chr9	58	14	200
Wt	182	45	200
合計	729	179	1000

表 4.3.3 モノソミー細胞データセット(G2/M 期)

クラス	訓練(枚)	検証(枚)	テスト(枚)
Chr1	394	98	200
Chr3	192	48	200
Chr6	106	26	200
Chr9	73	18	200
Wt	200	49	200
合計	965	239	1000

表 4.4.1 モノソミー細胞学習結果(G1 期)

クラス	正答率[訓練]	正答率[検証]	正答率[テスト]
Chr1	0.997	0.864	0.47
Chr3	0.981	0.935	0.76
Chr6	0.988	0.81	0.26
Chr9	0.984	0.867	0.015
Wt	0.987	0.931	0.355
合計	0.988	0.892	0.372

表 4.4.2 モノソミー細胞学習結果(S 期)

クラス	正解率[訓練]	正解率[検証]	正解率[テスト]
Chr1	0.992	0.939	0.42
Chr3	0.974	0.938	0.635
Chr6	0.991	0.577	0.4
Chr9	1	0.611	0.075
Wt	0.99	0.98	0.46
合計	0.989	0.883	0.398

表 4.4.3 モノソミー細胞学習結果(G2/M 期)

クラス	正解率[訓練]	正解率[検証]	正解率[テスト]
Chr1	1	0.93	0.75
Chr3	1	1	0.78
Chr6	1	0.667	0.405
Chr9	1	0.857	0.05
Wt	1	0.911	0.46
合計	1	0.899	0.489

学習の結果、同日に撮影した細胞を用いた「検証」の正解率はおしなべて高かった(G1期: 89.2%、S期: 88.3%、G2期: 89.9%)ものの、「テスト」の正解率(G1期: 37.2%、S期: 39.8%、G2期: 48.9%)は非常に低いという結果になった。その時期により大きく細胞の形が異なる細胞周期に比べ、各細胞周期ごとに分けた細胞間では目視でわかるほどの特徴がないことから、この検証とテストに現れた大きな違いは、撮影日が異なっていたこと(訓練用データと検証用データは同日取得)から、細胞撮影時が異なることで生じた誤差を学習してしまっている可能性があると考えた。そこで、この可能性を検証するべく、細胞周期を分類せずに学習を行った(表 4.5、表 4.6)。

その結果、細胞周期未分類時においても、検証正解率が高いがテスト正解率は低いという、細胞周期ごとに分類して学習を行なった時と同様の結果（検証正解率: 93.4%、テスト正解率: 45.8%）となった。この結果から、細胞を撮影時の誤差を学習している可能性が考えられた。

学習全体としては上記のような結果であったが、第 III 番染色体の結果のみに注目すると、細胞周期ごとの分類の有無にかかわらず、テスト正答率が高かった（G1 期: 76.0%、S 期: 63.5%、G2 期: 78.0%、細胞周期分類なし: 84.0%）。このことは、第 III 番染色体が細胞周期に依存しない何らかの特徴を持っており、深層学習モデルがその違いを認識したことを示唆している。

第 I 染色体については G2/M 期の検証正解率とテスト正解率が共に高いという結果（検証正解率: 93.0%、テスト正解率: 75.0%）あったため、G2/M 期に何らかの特徴を持っていることが示唆された。

一方、第 IX 染色体については、検証正答率が高いがテスト正答率は非常に低いという結果であったため、学習枚数が少ないため十分に学習できていないか、撮影時の誤差を学習している可能性が考えられた。

表 4.5 モノソミー細胞データセット（細胞周期未分類）

クラス	訓練(枚)	検証(枚)	テスト(枚)
Chr1	1038	257	600
Chr3	679	168	600
Chr6	359	89	600
Chr9	193	47	600
Wt	616	152	600
合計	2885	713	3000

表 4.6 モノソミー細胞学習結果（細胞周期未分類）

クラス	正解率[訓練]	正解率[検証]	正解率[テスト]
Chr1	0.998	0.942	0.423
Chr3	0.996	0.976	0.84
Chr6	1	0.775	0.4
Chr9	1	0.957	0.033
Wt	0.99	0.961	0.593
合計	0.996	0.934	0.458

4-6. 細胞を混ぜて写真を撮ることで、野生型細胞と第 I, III 番染色体モノソミー細胞を 66.7% の正解率で分類できた

各モノソミー細胞をそれぞれ個別に撮影することによる、撮影の誤差を学習している可能性を排除するために、細胞を混ぜて撮影することを考えた。しかし、この場合、目視、あるいは何らかの方法で正確に細胞の正誤判定を行える必要がある。そこで、モノソミー株の同一遺伝子を株ごとに異なる色の蛍光タンパクでラベルし、「答え合わせ」ができるようにすることとした。そこで、ヒストン H2A タンパクをそれぞれ mCherry、GFP でラベルした第 I, III 番染色体 2n-1 モノソミー細胞を作製し、野生型細胞と共に学

習させることとした。学習には透過光画像のみを用いるが、細胞撮影時に同時に mCherry, GFP を検出した蛍光画像を撮影しておくことで、学習後に簡便かつ正確に「答え合わせ」をすることが可能となる(図 4.11)。このように作製した細胞を用いて、細胞周期を分類せず学習を行った(表 4.7、表 4.8)。この実験では、同日に、二回に分けて細胞を撮影し、一方を訓練データと検証データとして、もう一方をテストデータとして用いた。

その結果、第Ⅲ番染色体では、検証正解率、テスト正解率が共に高い(検証正解率: 77.5%、テスト正解率: 77.0%)という結果となった。これらの結果からも、第Ⅲ番染色体では細胞周期に依存しない何らかの特徴があることが示唆された。

一方、野生型2倍体細胞(Wt)では、検証正解率に比べ、テスト正解率が低いという結果(検証正解率: 77.5%、テスト正解率: 50.0%)であったことからうまく特徴を捉え切れていない可能性が考えられた。

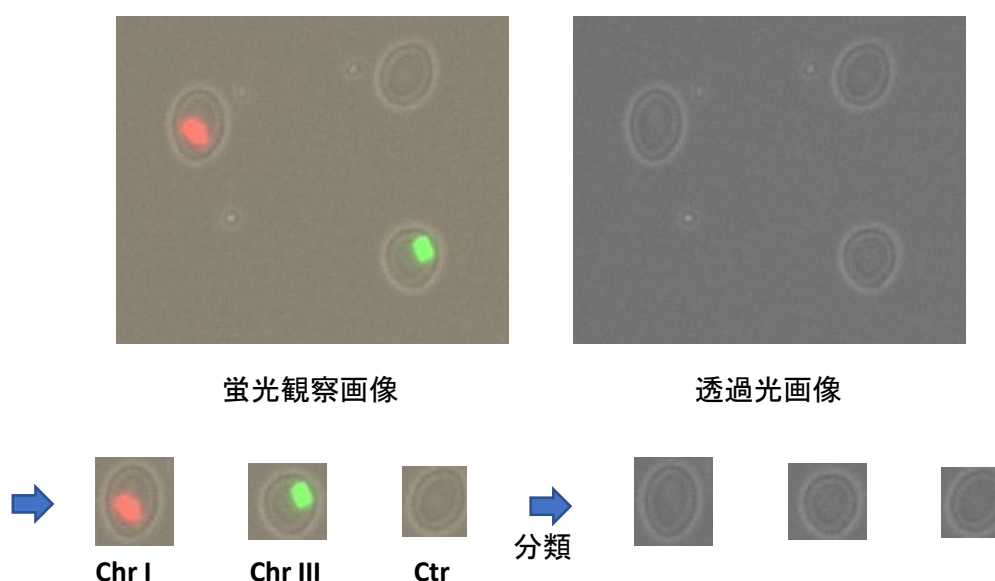


図 4.11 蛍光画像を利用した、 $2n-1$ モノソミー細胞の分類

表 4.8 モノソミー細胞データセット(同一プレート撮影、細胞周期未分類)

クラス	訓練(枚)	検証(枚)	テスト(枚)
Chr1	160	40	100
Chr3	160	40	100
Wt	160	40	100
合計	480	120	300

表 4.9 モノソミー細胞学習結果(同一プレート撮影、細胞周期未分類)

クラス	正解率[訓練]	正解率[検証]	正解率[テスト]
Chr1	0.983	0.55	0.56
Chr3	1	0.775	0.77
Wt	1	0.775	0.5
合計	0.994	0.7	0.61

5. 考察

本研究では、細胞形態から異数性のような染色体構成の異常を予測・同定することが可能なのではないかと考え、新規な予測手法としての深層学習を適用し、細胞形態と染色体構成の関連付けができるシステムの構築を目指した。

5-1. YeastSpotter を用いた細胞検出システムがうまく機能した

深層学習を行うには、撮影した画像から酵母細胞を検出し、大きさを揃えた画像を用意する必要がある。酵母細胞は楕円形がひとつあるいは2つつながった形態であるため、単純な画像抽出ではうまく認識できない。酵母細胞用の画像抽出ツールである YeastSpotter を用いても、単一細胞の中で、母細胞とそこから出芽した芽を別のものとして誤った認識をしていた。しかし、YeastSpotter がその抽出過程で生成するマスク画像を物体検出に用い、NMS を使用することで、明らかに単一細胞を複数細胞と認識することや、細胞を重複して検出してしまうことが減少した。したがって、本研究で開発したシステムは、酵母細胞を抽出する上で、有効かつ効率的なものであると考えられる。

本研究では、新たに深層学習モデルの学習を行うことによる、細胞の検出は行っていない。本研究での目的は染色体構成を予測することであったので、既存のツールを組み合わせ細胞の検出を行った。しかしながら、更に検出精度を向上させるのであれば、Faster R-CNN のような深層学習モデルを用いることで検出精度が向上する可能性がある。

また、本研究では、画像の二値化の閾値を 256 階調のうち 130 の固定値とした。この条件で、細胞境界がよく検出できただけでなく、画像のノイズも低くおさえることができた。酵母細胞撮影時には光の強さの数値を固定して撮影を行っている。そのため、向上する可能性は低いと考え本研究では行わなかったが、二値化の閾値を画像一枚ずつ個別に計算することにより YeastSpotter を用いずとも検出精度が向上する可能性が考えられる。

5-2. 異数体細胞の単離にはそれぞれ適した方法がある

異数体細胞と一口に言っても、様々なもの、例えば、一倍体細胞由来であれば、特定の染色体数が増えた $n+1$ (時には $n+2$, $n+3$...) 細胞、二倍体細胞由来であれば、特定の染色体数が増減した $2n\pm 1$ 細胞など、がある。本研究では、このようなものを取得するための方法を一通り試した。 $n+1$ 細胞を得るために過剰複製を利用した実験では(単離方法:3-4-1)、過剰複製誘導による細胞生存率の低下が顕著ではなかったため、今回は、候補株の選出は行わなかった。また、過剰複製による単離方法では、特定の染色体異数性細胞を狙って取得することが難しいという問題点はあるが、この系は、うまく働けば一度に様々な種類の異数隊を取得できるため、条件設定さえできれば将来的に有用な系であると思える。

今回行なったセントロメア機能の条件的阻害による方法(3-8-2)では、理論上、 $2n-1$ モノソミー細胞に加え、 $2n+1$ トリソミー細胞も出現しているはずである。今回の学習と組み合わせた解析には用いなかったものの、実際にいくつかのトリソミー細胞の単離を行うことに成功している。トリソミー細胞は培地による選択ができないため、単離効率は悪くなるものの、 $2n-1$ モノソミーとのセットで解析を行うことができれば、より興味深い結果が得られることも期待できる。

本研究で用いたモノソミー細胞の単離法は、大変最も効率が良かったほとんどの染色体でセントロメア不活化処理を行なった後に生育速度の異なるコロニーが形成した。生育速度は異なったが、サイズの小さい染色体ではモノソミー状態を安定に維持できることが分かった。しかし、中～第サイズの染色体では、生育速度の遅いコロニーを再培養したところ、生育が遅いものに加え、生育が野生型株と同等のレベルまで回復したクローンが出現した。これらのうち、生育の遅いものがモノソミー細胞であった。

細胞における染色体の欠損は細胞に非常に大きな影響を与える。さらに、「Interestingly, most $2n-1$ strains endoduplicate and become $2n$.」(Reid, 2008)³にもあるように、ほとんどのモノソミー株では、 $2n$ 細胞に戻る傾向がある。このような現象はモノソミーレスキューとして知られている。モノソミーレスキューの原理は解明されていないが、「染色体不分離によって残存の相同染色体が倍加して復帰する機構」(松崎,1993)¹⁴として考えられている。本研究では、図として示してはいないが、SC+5FOA プレートで生えたものの中には $2n$ 細胞が数多く存在した。そのため、このようなことが、私が作製した細胞においても高頻度にかかるものと考えられる。

また、新たな発見として第 V 番染色体における動原体機能阻害によって、染色体の長さが再現性良く変化した異数体が発現した。おそらく、第 V 番染色体にある何らかの特定の染色体構造と関係があると思われるが、その理由を本研究で解明することは、時間の関係もあり、行わなかったものの、非常に興味深い現象と言える。

5-3. H2A-GFP/mCherry を用いた系の評価

本研究では、染色体構成を学習させるにあたり、細胞を撮影する際に生じた何らかの誤差が大きな問題となることがわかった。この問題を解決するために、異なる色の蛍光タンパクである GFP と mCherry を用いて個別にラベルした異なるモノソミー細胞を混ぜて撮影を行った。この系を用いることで、少なくとも三種類の細胞を同時に撮影することが可能となった。本研究では、さらに多くの異数性細胞を学習・予測させることを考えている。したがって、多種類の異数性細胞を学習させるには、このような方法だけでは限界があるかもしれない。本研究では、このような方法で撮影時誤差の克服を行ったが、そのほかの方法として、複数回・複数日に分けて細胞を撮影し、その細胞を混ぜ合わせて学習させることで、撮影時の誤差を誤って学習してしまうという問題を改善できる可能性がある。

5-4. 第Ⅲ番染色体では、細胞周期に依存しない特徴がある

深層学習によるモノソミー細胞の予測の結果、第 I 番染色体では、G2/M 期において予測精度が高いように見られた(表 4.4.3)。しかし、蛍光タンパクを用いた系では撮影した枚数が少なくその検証ができなかった。第 VI, IX 番染色体では、撮影時の誤差を考慮した学習が行えていないため、本研究においてはその特徴を深層学習によって捉え予測することはできなかった。しかし、第Ⅲ番染色体では、特に、その予測精度が高いという結果が得られた。第Ⅲ番染色体では、細胞周期を未分類とした場合にも、その精度が高く、細胞周期に依存しない特徴があると考えられた。しかしながら、その特徴は目視でもわからないためどのような特徴をもとに判断しているのかはわからなかった。

特徴を検出する方法の一つに、Generative Adversarial Networks(GAN)¹⁵を用いる方法がある。GAN は画像を生成する深層学習モデルであり、学習した特徴をもとに実在しないデータを生成することができる。これを利用することで、細胞の平均的な特徴を持った画像を生成することにより、機械側で抽出した特徴を目で見える形で抽出できる可能性がある。この方法を使って仮に「第Ⅲ番染色体モノソミーが備える特徴」が目視で認識できるレベルで抽出できれば、それは大変興味深い結果となることが期待できる。

5-5. 細胞形態から染色体構成を予測するにあたっての深層学習の有用性

表 4.2 の結果から、Xception を用いることである程度、細胞形態の違いを学習できることが示された。そして、表 4.9 のように、蛍光たんぱく質を使うことで、その精度にばらつきはあったが、染色体異数性細胞を学習することでその予測がある程度可能であることが示された。

本研究では深層学習時にファインチューニングを使用しているため、画像枚数が少ないとしても、ある程度の結果を得ることができている。しかし、図 5.1 (渡辺、2004)¹⁶ にもあるように一倍体と二倍体ではその細胞形態に大きな違いはあるものの、二倍体であっても一倍体の特徴を示す細胞が出現する。これらのことから、画像の枚数を増やすことで本来の特徴を持った細胞を多く学習することができ、より精度が向上する可能性がある。

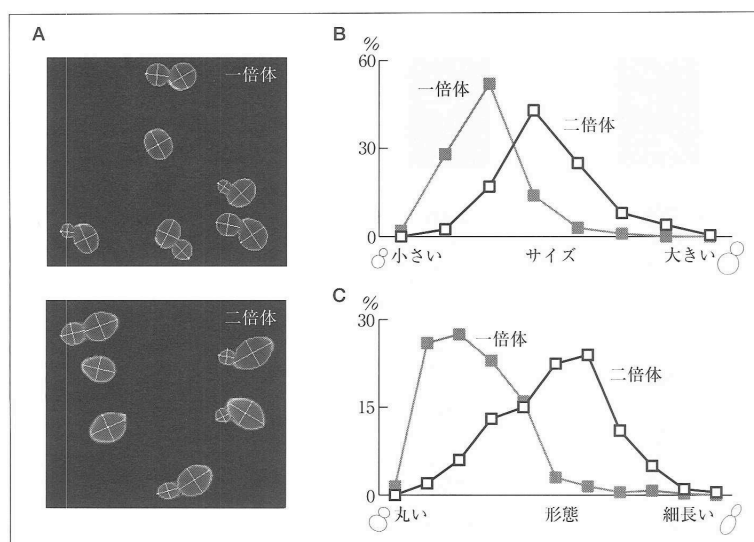


図 5.1 CalMorph システムを用いた倍数性による細胞形態の違いの検出 (渡辺、2004)¹⁶

A CalMorph を用いた細胞を定量解析、B 一倍体と二倍体の細胞形態の違い

更に、本研究では、3-6-13 に示したモデルを用いて深層学習を行っているが、このようなモデルのほかにも学習の精度が高い方法がある。マルチタスク学習はそのような方法の一つで、CNN により抽出された特徴ベクトルに加えて、画像を処理することにより測定した特徴などを加えて学習する。そのため、この方法ではデータの作製に膨大な時間が必要となる。本研究では、時間の関係でこの方法は採用しなかったが、本法を採用することでさらに精度が向上する可能性がある。

深層学習はどのような細胞形態を認識し、判断しているのかわからない。しかし、近年深層学習がどこを見て判断しているかを可視化する方法が研究されている。その一つとして、Grad-CAM¹⁷ という方法がある。

このような方法を用いることで、深層学習がどこを見て判断しているかが可視化できる可能性がある (図 5.2)。

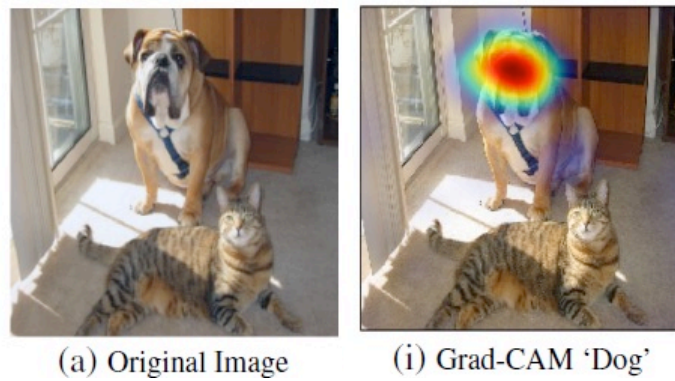


図 5.2 Grad-CAM による深層学習の視覚化 (Selvaraju, 2016¹⁷ を改変)

しかしながら、実際には、どのような特徴があるかを明確に示すことは難しい。そのため、特徴を明確に示したい場合には、CalMorph を用いた定量解析は非常に有効であると考えられる。一方、深層学習では細胞の染色を行わずに分類することが可能であるため、非常にスピーディーである。両者の長所を組み合わせる、例えば、あらかじめ染色体構成を CalMorph で解析し、特徴を明確にした状態で深層学習を用いて染色体構成やその機能を予測させる等の手法を採用することで、より強力な解析ツールとなることが期待できる。

6. 文献

- 1 Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci U S A* **102**, 19015-19020, doi:10.1073/pnas.0509436102 (2005).
- 2 Lu, A. X., Zarin, T., Hsu, I. S. & Moses, A. M. YeastSpotter: accurate and parameter-free web segmentation for microscopy images of yeast cells. *Bioinformatics* **35**, 4525-4527, doi:10.1093/bioinformatics/btz402 (2019).
- 3 Reid, R. J. *et al.* Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomyces cerevisiae* chromosomes. *Genetics* **180**, 1799-1808, doi:10.1534/genetics.108.087999 (2008).
- 4 Hill, A. & Bloom, K. Genetic manipulation of centromere function. *Mol Cell Biol* **7**, 2397-2405, doi:10.1128/mcb.7.7.2397 (1987).
- 5 Okada, H., Ohnuki, S. & Ohya, Y. Quantification of cell, actin, and nuclear DNA morphology with high-throughput microscopy and CalMorph. *Cold Spring Harb Protoc* **2015**, 408-412, doi:10.1101/pdb.prot078667 (2015).
- 6 Ohnuki, S. *et al.* Analysis of the biological activity of a novel 24-membered macrolide JBIR-19 in *Saccharomyces cerevisiae* by the morphological imaging program CalMorph. *FEMS Yeast Res* **12**, 293-304, doi:10.1111/j.1567-1364.2011.00770.x (2012).
- 7 Tanaka, S. & Araki, H. Multiple regulatory mechanisms to inhibit untimely initiation of DNA replication are important for stable genome maintenance. *PLoS genetics* **7**, e1002136, doi:10.1371/journal.pgen.1002136 (2011).
- 8 Tanaka, S. *et al.* CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* **445**, 328-332, doi:nature05465 10.1038/nature05465 (2007).
- 9 Kingma, D. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv*, 1412.6980 (2014).
- 10 Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv*, 1610.02357 (2016).
- 11 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting *J Machine Learn Res* **15**, 1929-1958 (2014).
- 12 Redmon, Joseph. *et al.* You Only Look Once: Unified, Real-Time Object Detection. *arXiv*, 1506.02357(2015).
- 13 Ren, Shaoqing. *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv*, 1506.01497
- 14 松崎 浩明、酵母異数体から正常細胞への復帰機構の解明。科学研究費補助金研究成果報告書(1993)
- 15 Goodfellow, J.S. Generative Adversarial Networks. *arXiv*, 1406.2661 (2014)

- 16 渡辺大輔 他、出芽酵母における網羅的形態解析. 化学と生物、2004 年 42 巻 4 号 p. 240-248.
- 17 Selvaraju, R. *et.al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. arXiv, 1610.02391(2016)

7. 謝辞

本研究を遂行するに当たり、指導教員である高知工科大学環境理工学群生命科学専攻の田中誠司教授には研究遂行にあたって大変お世話になり、また、多大なご指導を頂きました。同じく小川助教授には様々な助言を頂きました。また、共同研究者である東京大学大矢禎一教授、同じく大貫慎輔助教には細胞を撮影するにあたり様々な助言をいただきました。さらに、ICI Award@KUT で、私をご支援くださいました前田建設工業株式会社様に心から感謝申し上げます。