

# 少数画像からの生物個体識別手法に関する研究

## ゴリラの個体識別を題材として

岡 幸秀<sup>†</sup>

岡留 剛<sup>‡</sup>

Sachiho Oka

Takeshi Okadome

### 1. はじめに

研究背景として絶滅危惧種の増加が問題になっている。ある生き物が絶滅することで生態系のバランスがくずれ、自然環境に大きな影響を与えてしまう可能性がある。国際自然保護連合によれば2019年時点で約三万種がレッドリストに挙げられており、世界規模で絶滅危惧種を守る運動が活発になっている。その中でも本研究ではゴリラに対して注目することにする。ゴリラの個体減少の原因として、密猟やゴリラ自身のストレスが考えられる。そこで、本研究ではゴリラのストレスにフォーカスを当てることにする。例えば、ゴリラがストレスを感じて餌を食べない日があった際、同じ個体が次の日も餌を食べないとなると、そのゴリラの体調が良くないのではないかなどの予想をたてることができる。体調悪化などの問題の早期発見を行うことは、ゴリラの保護に繋がると考えられる。日本で飼育されているゴリラは1つの動物園につき、1家族程度なため、個体識別の必要性はあまりないと考えられるが、海外のサファリパークや野生に生息している個体に着目すると、個体識別は必要である。一例を挙げると、ウガンダのブウィンディ国立公園のマウンテンゴリラの個体数は2012年時点で、合計400頭のマウンテンゴリラが生息していることが報告されており、このことから個体識別の必要性は充分にあると考えられる。本研究ではゴリラの保護を目的としてゴリラの顔画像を用いて微細な差を検出する識別器の作成を行う。日本の動物園で飼育されているゴリラのほとんどがニシローランドゴリラであるため、本研究ではニシローランドゴリラのデータセットに限定して研究を進める。

また絶滅危惧種は個体数が少なく、かつ得られる画像には限りがあるため少数データに限って研究を行う。人間の神経細胞 (Neuron) と神経回路網の繋がり仕組みを模したシステムをベースとして、そのシステムを数式的にモデル化したものを Neural Network とよぶ。Neural Network を多層にして用いることで、データに含まれる特徴を段階的に、より深く学習することが可能となる。Deep Neural Network に大量のデータを入力することでモデルはデータに含まれる特徴量を自動的に学習する。しかし入力が少数データの場合、Deep Neural Network

を用いると一般的に精度が芳しくないことが知られており、そのため、Data Augment や fine-tuning といわれる、関係するデータを大量に集めそのデータセットを用いてモデルを訓練し、元々保持していたデータセットを用いて、事前学習した Neural Network を微調整する研究があり、少数データしか得られない状況下において推定精度を向上させる研究が注目されている。

### 2. 関連研究

#### 2.1 少数データに対する個体識別に関する研究

先行研究にパンダの顔画像を用いて個体識別を行う研究 [Matkowski et al. 2019] がある。この研究のアルゴリズムはまず2枚の画像のうち1枚を非剛体変換し、2枚の画像に映るパンダの顔の位置合わせを行ったのち、Local Binary Pattern 特徴量 (LBP 特徴量) と gabor Filter, 2種類の特徴量抽出を行う。gabor filter とは画像中にどの向きの線が含まれているかを抽出できるフィルタである。Judson P. Jones と Larry A. Palmer によってネコの視覚野にある単純型細胞の形が2次元の Gabor Filter で表せることが示された [Jones et al. 1987]。このフィルタは虹彩認識や指紋認識などの生体認証に必要なパターン認識技術にも利用されている。gabor filter は、「正弦波」と「ガウス関数」の積として定義される。

$$g(t) = ke^{j\theta}w(at)s(t) \quad (1)$$

ここで  $w(t)$  がガウス関数、 $s(t)$  が正弦波を表す (指数表現されている)。

$$w(t) = e^{-\pi t^2} \quad (2)$$

$$s(t) = e^{j(2\pi f_0 t)} \quad (3)$$

直感的なイメージはこのようになる。

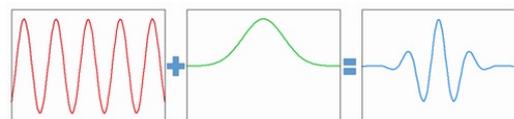


図 1: 正弦波とガウス関数の積の直感的イメージ。

最後に部分最小二乗回帰を用いて識別を行う (図 2)。

<sup>‡</sup> 関西学院大学, Kwansei Gakuin University

<sup>†</sup> 関西学院大学, Kwansei Gakuin University

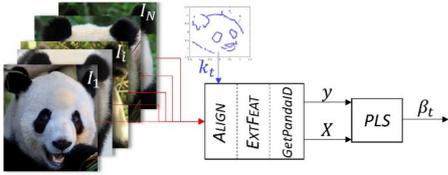


図 2: [Matkowski et al. 2019] による個体識別アルゴリズムの概略図.

## 2.2 Convolutional Neural Network

Convolutional Neural Network(以下, CNN)は通常のNeural NetworkにConvolution層とPooling層を追加したもので, 画像認識に使われることが多い. 活性化関数は, 通常のNeural Networkにおけるシグモイド関数に対して, CNNではReLU (rectified linear unit) 関数が利用される. Convolution層では, 前層のノードとKernelで畳み込み演算を行って特徴マップを得る. Pooling層で, convolution層から出力された特徴マップを縮小して新たな特徴マップを得る. これは, 画像の特徴を維持しながら画像の持つ情報量を大幅に圧縮していることができる

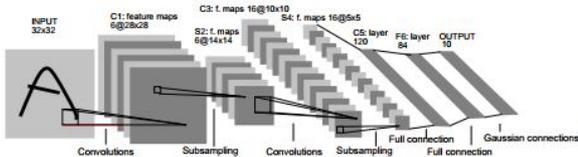


Fig. 1. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

図 3: CNN の流れ.[LeCun et al. 1999] より.

. CNNは特徴を抽出するための検出器である, フィルタのパラメータ(重み)を自動で学習していくが, データ数やその質が芳しくないとき, パラメータをうまく学習することができず, 一般的に過学習が発生する. この問題を解決するために, パラメータに関して新たな初期化戦略を行う必要があると指摘されており [Glorot et al. 2010, Zhang et al. 2015, ], 解決すべき問題の一つとなっている. 同様に, ハイパパラメータの最適化に関しても考慮すべき問題で, 最適化序盤において重要度の高い変数を見つけ出す必要があることに加えて, さらなる性能向上のために重要度の低い変数と変数間依存関係を考慮した探索を行う必要がある [Watanabe et al. 2020].

## 2.3 Data Augmentation

機械学習で学習する際, 十分に訓練データが存在しない, もしくは手持ちの訓練データが偏っていることがある. そのような場合, 過学習を防ぎ汎化性能を向上させるために, 手持ちの訓練データに何らかの処理を施してデータを拡張する. これをData Augmentationという. 基本的な手法として, 回転や平行移動, 反転, ノイズ付与, 画像のマスクなどが挙げられる [Krizhevsky et al. 2012, Zhong et al. 2020]. しかし, Data Augmentationに関してデータを増やせばいいというものではなく, データを考慮して拡張する必要がある. 例えば, 手書き数字データで回転処理を拡張手法として選択した場合, 画像の重要な意味が失われる可能性がある. さらに, 最適なデータ拡張方法を見つけるAuto Augmentの研究 [Hataya et al. 2019, Cubuk et al. 2020] や新たなデータ拡張手法に関する研究 [He et al. 2019] などがある.

## 3. 提案手法

大まかな処理の流れを図4に示す. 元画像の顔部分のトリミングを行ったのち, すべての画像に対して $32 \times 24$ ピクセルにリサイズ, その後データ拡張の処理を行った. データ拡張では, 左右反転と上下反転の2種類の処理を施した. それらのデータをゴリラ識別器に入力し, 結果, そのゴリラ個体はどのゴリラなのかを出力する. ゴリラ識別器にはCNNは2つの畳み込み層と3つの全結合層を持つCNNを用いる. それぞれユニット数は1つめの畳み込み層が6, 2つめの畳み込み層が16, 1つめの全結合層が120, 2つめの全結合層が84, 1つめの全結合層が6になるように設定した(図5). Pooling層にはMax-Pooling, Kernel Sizeは $3 \times 3$ を採用している. 最後の全結合層のユニット数は識別する個体が今回6個体だったため, この値になっている.

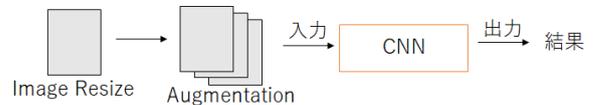


図 4: 処理の流れ.

## 4. 実験

### 4.1 実験データ

京都市動物園に依頼, 協力を得て実験データを入手した. データ数は以下の表1の通り, ゴリラ個体計6体, 元データ総数は317枚である. さらに, 実験に使用した個体の顔画像の一例を図6に挙げる. 特徴を記載している

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 6, 28, 28]	456
MaxPool2d-2	[-1, 6, 14, 10]	0
Conv2d-3	[-1, 16, 10, 6]	2,416
MaxPool2d-4	[-1, 16, 5, 3]	0
Linear-5	[-1, 120]	28,920
Linear-6	[-1, 84]	10,164
Linear-7	[-1, 7]	595

Total params: 42,551  
 Trainable params: 42,551  
 Non-trainable params: 0

Input size (MB): 0.01  
 Forward/backward pass size (MB): 0.04  
 Params size (MB): 0.16  
 Estimated Total Size (MB): 0.21

図 5: CNN 内のパラメータ数.

個体は血縁関係のある家族で現在も京都市動物園で飼育されている個体だが、記載していない個体に関しては現在、京都市動物園で飼育されていないため、新たなデータを入手するのは困難である。

表 1: 実験データ

個体名	データ数	特徴
キンタロウ	74	子供 (弟)
ゲンタロウ	69	子供 (兄)
モモタロウ	73	父
ゲンキ	76	母
ゴン	12	
ヒロミ	13	

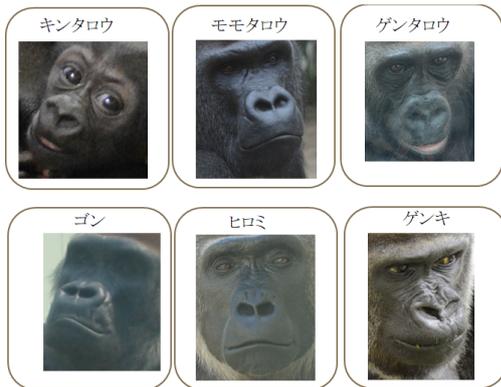


図 6: 各ゴリラ個体の顔画像の例.

## 4.2 実験結果

実験結果は以下の表 2, 表 3 の通りである. 全体の accuracy は 52% で, 6 体をランダムに分類したときの正答率の期待値, 0.17 と比べると上回る結果となった.

表 2: 実験結果 1

個体名	precision	recall	f1-score	support
キンタロウ	0.59	0.87	0.70	23
ゲンタロウ	0.43	0.29	0.34	21
モモタロウ	0.48	0.55	0.51	22
ゲンキ	0.50	0.52	0.51	23
ゴン	0.00	0.00	0.00	4
ヒロミ	0.00	0.00	0.00	4

表 3: 実験結果 2

	precision	recall	f1-score	support
accuracy			0.52	97
macro avg	0.33	0.37	0.34	97
weighted avg	0.46	0.52	0.48	97

## 4.3 考察

個体ごとに考察した結果, ゴンとヒロミに関しては極端にデータ数が少ないため, 学習が十分にできず, precision, recall, f1-score がそれぞれ低くなつたと予想する. また, キンタロウに関しては唯一子供の個体で他の個体との差が顕著なため, その他の個体と比較すると高い値が出たと考察する. データ数が少なく, 全体として学習がうまくできていないため, CNN を用いるだけでは精度があまり向上していないと考えられる.

## 5. 今後の展望

この研究によって, 動物園やサファリパークでは個体の健康管理, 野生では行動把握が可能になると考えている. しかし, この研究にはまだ再考すべき点がいくつかある. 専門家は”鼻”に着目してゴリラの個体を識別することを動物園職員の方から伺ったため, 鼻の特徴を捉えるような Kernel(Filter) を作成し, CNN に組み込みたい. また, 鼻の特徴を取得するためにゴリラの顔のどこに鼻があるのかを知る必要があるため, あらかじめゴリラの顔の向きを推定し, 鼻の位置を明確にする必要がある. 顔の向きの検出には gabor 特徴量を取得し, 複数個体の顔画像の平均を求め, 対象画像の特徴量とのマッチングを行って顔の向きを推定する手法 [Srimuang et al. 2004] を試してみたい. さらに, CNN のハイパーパラメーターについても再考が必要であると考えている. また, データが現在極端に少ないため, 個体数と一個体に対するデータ数, とともに増やして実験を行う予定である.

## 6. まとめ

本研究では絶滅危惧種であるゴリラの保護を目的として, ゴリラの個体識別を行うために CNN を用いて実験

を行った。実験結果は accuracy52% で、ランダムに分類を行うときの期待値よりも高い結果を得た。各個体のデータ数に応じて正答率が変動しているため、汎化性能を高める上でデータ収集が必要である。今後は、鼻の特徴を捉えるような Kernel(Filter) を作成し CNN に組み込むこと、データを増やすこと、CNN のハイパーパラメータの再考を行い、より精度の高い識別器の作成に取り組む。

## 参考文献

- Cubuk, E. D., B. Zoph, J. Shlens, and Q. V. Le (2020). “Randaugment: Practical automated data augmentation with a reduced search space”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 702–703.
- Glorot, X. and Y. Bengio (2010). “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256.
- Hataya, R., J. Zdenek, K. Yoshizoe, and H. Nakayama (2019). “Faster autoaugment: Learning augmentation strategies using backpropagation”. In: *arXiv preprint arXiv:1911.06987*.
- He, Z., L. Xie, X. Chen, Y. Zhang, Y. Wang, and Q. Tian (2019). “Data Augmentation Revisited: Rethinking the Distribution Gap between Clean and Augmented Data”. In: *arXiv preprint arXiv:1909.09148*.
- Jones, J. P. and L. A. Palmer (1987). “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex”. In: *Journal of neurophysiology* 58.6, pp. 1233–1258.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*, pp. 1097–1105.
- LeCun, Y., P. Haffner, L. Bottou, and Y. Bengio (1999). “Object recognition with gradient-based learning”. In: *Shape, contour and grouping in computer vision*. Springer, pp. 319–345.
- Matkowski, W. M., A. W. K. Kong, H. Su, P. Chen, R. Hou, and Z. Zhang (2019). “Giant panda face recognition using small dataset”. In: pp. 1680–1684.
- Srimuang, K. and Y. N. Munetoshi Iwakiri (2004). “Detection of a Person Face and Head Pose Estimation using the Gabor wavelet”. In: *IPSIJ SIG Technical Report* 2004.119 (2004-FI-077), pp. 1–6.
- Watanabe, s., M. Nomura, and O. Masaki (2020). “The Characteristics Required in Hyperparameter Optimization of Deep Learning Algorithms”. In: *The 34th Annual Conference of the Japanese Society for Artificial Intelligence(2020)*. Information and Media Technologies, 1J3OS1002–1J3OS1002.
- Zhang, G., J. Kato, Y. Wang, and K. Mase (2015). “How to initialize the CNN for small datasets: Extracting discriminative filters from pre-trained model”. In: *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, pp. 479–483.
- Zhong, Z., L. Zheng, G. Kang, S. Li, and Y. Yang (2020). “Random Erasing Data Augmentation.” In: *AAAI*, pp. 13001–13008.