

畳み込みニューラルネットワークによるチンパンジーの個体識別

Identifying Individual Chimpanzee with Convolution Neural Networks

池田 宥一郎*¹ 飯塚 博幸*¹ 山本 雅人*¹
 Yuuichiro Ikeda Hiroyuki Iizuka Masahito Yamamoto

*¹北海道大学 大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

The development of information science in recent years has greatly contributed to studies of animal behavior. We are collaborating with Maruyama Zoo in Sapporo to reduce the load on zookeepers taking care of animals using artificial intelligence. One of our targets is to automatically create chimpanzee's Ethogram for health management and maintenance of breeding environments. Ethogram is an inventory of all behavior patterns of specific individuals and species. In order to create an Ethogram, it is necessary to identify individuals. Therefore, in this research, we examined whether each individual chimpanzee can be recognized using Convolutional Neural Networks, which has high accuracy in image recognition field. Experimental results showed that it is possible for our system to identify individual chimpanzees.

1. はじめに

近年の情報科学分野の発展は動物行動学分野の研究に大きく寄与しており、動物の行動に関するデータを情報科学の技術を持って自動的に計測、分析を行っていく”Computational Ethology” [Anderson and Perona 14] とよばれる新領域の発展を促進している [福永と岩崎 15]. Computational Ethology においては画像や動画データにおける動物の個体識別および行動分類が課題として挙げられている。この分野の発展は情報科学分野の発展と連動しており、近年めざましい成果を上げている畳み込みニューラルネットワーク (CNN) を代表とした深層学習による画像認識技術の応用が求められている。

Computational Ethology ならびに深層学習による画像認識技術は実社会応用も求められている。札幌市円山動物園では動物の健康管理や飼育環境の整備のために、人工知能によるチンパンジーの管理システムの開発を試みている。従来は飼育員が膨大な時間と労力をかけて飼育動物に異常がないか録画したビデオを早送りして観察していた。さらに、動物の健康状態や動物同士の関係を分析するのは熟練の飼育員の経験に頼っていた。この状況を改善するため人工知能によるエソグラムの自動作成を行う。エソグラムとはある特定の個体や種の全行動パターンの目録であり、動物の行動を研究するうえでもっとも基本的な記録である。これによって飼育動物がいつ、どこで、何をしていたのか自動的に収集され、飼育動物の24時間監視システムの開発や、最適な環境エンリッチメント [Maple and Perdue 13] の提供ができると考えられる。自動エソグラム作成の先行研究として、加速度センサを用いて野外で行動する鳥類のエソグラムを自動生成する試みはあったが [Sakamoto et al. 09], 本研究においてチンパンジーが活動するのは閉鎖的な環境であるため、動画によってより仔細な行動の分析が可能であるとともに、センサ類を個体に取り付ける必要がないためほぼストレスを与えることなく分析できる。エソグラムを作成するためには個体の識別・トラッキング・行動分類が必要であり、特に個体の識別はエソグラムを作成する上で初めの手順となるため識別可能性を検証する必要がある。

したがって本研究では、エソグラム自動作成にむけた初期段



図 1: 識別する 9 頭のチンパンジー

階としてチンパンジーの各個体の識別が可能であるかを検証する。個体識別の手法としては深層学習の一種である CNN を用いる。また、個体識別の精度と汎化性を高めるために求められるデータの数や性質についても検証する。

2. データセットの作成および加工

本研究では札幌市円山動物園のチンパンジー館で飼育されている図 1 に示す 9 頭のチンパンジーの画像データを用いる。撮影環境は朝 11 時頃のチンパンジー館清掃終了後に餌が撒かれた状態で、チンパンジーがケージの中に入れられた時から撮影開始した。撮影者は 9 人であり、それぞれ異なる個体を追跡しつつ数秒おきに撮影を行い、10 分ごとに撮影する個体を交代した。これによって各個体について全撮影者の写真が含まれるため、撮影者の癖により個体分類されることを防ぐことがで

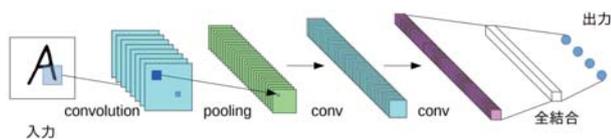


図 2: CNN の処理の流れ

表 1: CNN の構造

	縦×横×チャンネル数	フィルタ
入力	128×128×3	-
C1	128×128×32	3×3
C2	128×128×32	3×3
P1	64×64×32	2×2
C3	64×64×64	3×3
C4	64×64×64	3×3
P2	32×32×64	2×2
C5	32×32×128	3×3
C6	32×32×128	3×3
P3	16×16×128	2×2
C7	16×16×256	3×3
C8	16×16×256	3×3
P4	8×8×256	2×2
FC	1×1×1024	-
出力	1×1×9	-

きると考えられる。

撮影された写真は各個体の全身が大きく映るように手で正方形に切り抜き、128 × 128 ピクセルにリサイズした。また、身体のごく一部しか写っていないものや、ぶれが大きいもの、フォーカスが大幅に合っていないものなど、識別困難であると考えられる写真は取り除いた。また、ほぼ同様の写り方をしている画像は正答率に対して信頼性を欠くため取り除いた。以上の方法によって総数 4,986 枚の画像データを取得した。

3. チンパンジーの個体識別実験

3.1 実験設定

本研究では様々な画像認識タスクにおいて高い精度を示した CNN[LeCun et al. 98] を用いてチンパンジーの個体分類を行う。CNN とは図 2 のように画像の特徴を抽出する畳み込み層と画像をダウンサンプリングしていくプーリング層を繰り返すことで、入力画像から抽象化された特徴を抜き出すネットワークであり、画像認識分野の幅広いタスクに応用されている。なお、CNN の学習は誤差逆伝搬法により行われる。本研究では Tensorflow をバックグラウンドとする Keras をライブラリとして用いた。本研究で用いる CNN の構成は表 1 に示す通り 15 層のネットワークであり、128 × 128 ピクセルのある個体の画像を入力とし、Softmax 関数を用いて 9 頭の個体のそれぞれである確率を出力とする。誤差関数は多クラス交差エントロピーであり、Adam オプティマイザを用いた。バッチサイズは 32 とした。

実験では 2 章にて作成した 4,986 枚の画像のうち各個体ごとに訓練データとテストデータを 4 : 1 の比率でランダムに分

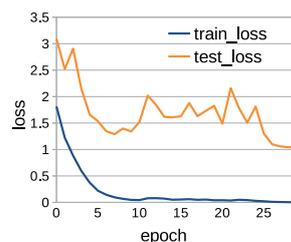


図 3: 誤差の推移

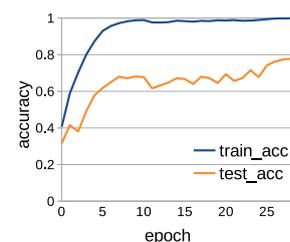


図 4: 正答率の推移

表 2: 各個体ごとのデータ数と正答率

個体名	ジェーン	コユキ	ガチャ	ハル	チャコ
データ数	257	442	659	828	520
正答率	68.2	70.2	75.4	84.4	83.8

個体名	テス	アッキー	スージー	レディ
データ数	425	671	478	706
正答率	73.4	75.4	80.6	81.8

け、訓練データ 3,982 枚、テストデータ 994 枚で 30epoch 学習する。学習は 5 回行い、そのたびに訓練データとテストデータをランダムに選び直し、正答率と誤差の推移について平均をとった。

3.2 実験結果

図 3, 図 4 から、訓練データに対する平均誤差の収束が確認でき、訓練データに対する平均正答率も 1.0 近くとなっていることがわかる。テストデータに関しては平均誤差は不安定な結果となったが、平均正答率は 0.78 程度と、9 頭をランダムに分類したときの正答率の期待値である 0.11 と比べて大幅に上回る結果となった。

次に、個体ごとの分類結果の平均を示す。表 2 から、データ数の最も少ないジェーンの正答率が最も低いことがわかる。学習 5 回分の各個体のデータ数と正答率の相関係数は $r=0.53$ ($p<0.001$) であり、一般的な基準で正の相関が認められた。したがって、データ数が少ないほど正答率は低くなる傾向にあると考えられ、データセットの見直しが求められる。しかし、高い正の相関は認められなかったため、正答率および汎化性を高めるためにはデータの数のみではなく質も考慮する必要があると考えられる。

4. 撮影者および撮影時間帯の違いに応じた汎化性の検証実験

4.1 実験設定

24 時間トラッキングする上で各個体の状況は多様に変化していくと考えられ、いかなる状況でも信頼できる正答率が出せるのか検証する必要がある。本実験で用いた画像データは撮影者の違いや撮影した時間帯の違いによって各個体の状況が異なっていると考えられる。したがって、加工後の画像データが全個体について一定枚数揃っている特定の撮影者 A による画像 400 枚すべてをテストデータとし、それ以外の撮影者による画像 4,586 枚を訓練データとして同様に学習を行った。

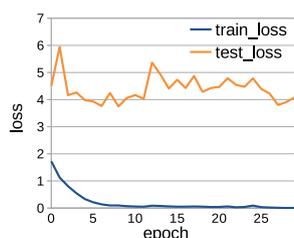


図 5: 撮影者 A 以外の画像を用いて学習した誤差の推移

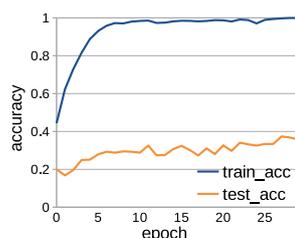


図 6: 撮影者 A 以外の画像を用いて学習した正答率の推移



図 7: 撮影者 A のガチャと他の画像の違い

4.2 実験結果

テストデータの誤差は図 5 のように推移し、初めの epoch の学習を終えたあとテストデータの誤差がほとんど下がっていないことがわかる。正答率は図 6 のように推移し、テストデータの正答率は 38 % 程度となった。

3 章における学習と比較して訓練データが増加しているにも関わらず正答率が下がってしまった原因として、やはり撮影された個体が写る角度や光量の違いが考えられる。実際のデータ例として図 7 を見てみると、コユキに誤分類された撮影者 A のガチャは逆光のある高所におり、他の撮影者のガチャとは写る角度も光量も異なることがわかる。一方で他の撮影者のコユキの中に高所にいるものが含まれ姿勢も似ていたため、コユキに誤って分類されたと考えられる。したがって、全個体について多様な時間帯の多様な写り方のデータがあれば汎化性が高まり、さらに正答率の向上が見込めると考えられる。

5. おわりに

本研究では、CNN を用いてチンパンジーの個体識別が可能であることを確認した。また、各個体のデータ数やデータの多様性に応じて正答率は変化するため、多数のデータを集めていくことが汎化性を高める上で必要であることも確認した。

今後は自動エソグラム作成にむけ、動画データにおいてチンパンジーの各個体をトラッキングし、行動の分類を行う。なお、自動エソグラム作成においては画像ではなく動画を用いるため、時系列情報を扱える。よって、動画のある特定のフレームのみで識別した結果が誤っていたとしても、過去のフレームの情報によって補完することで正しく識別させることが可能であると考えられる。したがって、時系列情報を扱える物体検出

モデルの検討、および多様な動画データの収集とそれらに対する正確なラベル作成が今後の課題である。

6. 謝辞

本研究を進めるにあたり、札幌市経済観光局様、札幌市環境局札幌市円山動物園様、一般財団法人さっぽろ産業振興財団様に様々なご助言、ご援助を戴いた。ここに深謝の意を表す。

参考文献

- [Anderson and Perona 14] Anderson, D. J. and Perona, P. : Toward a science of computational ethology. *Neuron*, 84, 18-31 (2014),
- [福永と岩崎 15] 福永津嵩, 岩崎渉: Computational ethology: バイオインフォマティクスと動物行動学の融合. *領域融合レビュー*, 4, e003(2015)
- [Sakamoto et al. 09] Sakamoto KQ, Sato K, Ishizuka M, Watanuki Y, Takahashi A, et al.: Can Ethograms Be Automatically Generated Using Body Acceleration Data from Free-Ranging Birds?. *PLOS ONE*, 4(4), e5379, (2009),
- [Maple and Perdue 13] Maple, Terry and Perdue, Bonnie: *Environmental Enrichment*. 95-117 (2013),
- [LeCun et al. 98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner: Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11),2278-2324 (1998)