

令和三年度卒業論文

プールベース能動学習を用いた  
牛の顔データセット構築の効率化の検討

宮崎大学 工学部 情報システム工学科

小八重 悠作

指導教員 椋木 雅之 教授

# 目次

1	はじめに	1
2	データセット構築	2
2.1	データセット構築の必要性	2
2.2	能動学習によるデータセット構築の効率化	3
2.2.1	受動学習と能動学習	3
2.2.2	プールベース能動学習	4
2.2.3	データセット構築の効率化	6
3	プールベース能動学習による牛の顔データセット構築	7
3.1	処理の流れ	7
3.2	牛画像の準備	8
3.3	牛の顔学習	9
3.4	追加データの選択	10
3.5	再学習とデータセット構築	11
4	実験	12
4.1	実験の目的	12
4.2	実験方法	13
4.3	実験結果	15
4.4	データセット構築実験	16
4.5	考察	17
5	おわりに	18
	謝辞	19
	参考文献	20

# 1 はじめに

近年、深層学習は目覚ましい進歩を遂げている。深層学習は機械学習の一種で、画像認識や音声認識といった多数の分野で活用されている。利点としては、データの特徴をより深めて学習させることができるので、性能が良いことが挙げられる。一方、欠点としては、多くの学習データが必要であることが挙げられる。この学習データには、入力データに対する正しい認識結果を人手で与えるアノテーションが必要である。多くの学習データを必要とする深層学習では、アノテーションのコストが高くなってしまう。

椋木研究室では、牛の顔画像からの個体識別の研究を行っている。ネット上には、さまざまな公開データセットが存在するが、個体別の牛の顔画像の大規模な公開データセットは存在しない。そのため、個体別の牛の顔画像の大規模な公開データセットを構築する必要があるが、その場合でも、アノテーションの手間が問題となる。

機械学習において、学習データのアノテーションの手間を減らす手段として能動学習がある。能動学習は、学習効果の高いデータを選択的にアノテーションするというもので、これを使うことで、アノテーションの手間を減らし、少ないデータで十分な学習を行うことができるという利点がある。

本研究では、能動学習の一種であるプールベース能動学習を使って、個体別の牛の顔画像データセット構築の効率化を検討する。具体的には、牛の顔検出器の学習と学習データの作成、個体識別用牛の顔データセットの構築を同時に進める枠組みを検討する。その中で、牛の顔検出器で検出された信頼度の高い検出結果をそのまま学習データに追加することで、検出器の性能向上が見られるか、実験により比較する。個体別牛の顔データセットが構築できれば、従来、人間の顔に対する研究が主流だった個体識別研究に、新たな可能性を提供できる。

以下、2章ではデータセット構築、3章ではプールベース能動学習による牛の顔データセット構築、4章では本研究で行った実験について述べる。5章では本論文を総括する。

## 2 データセット構築

### 2.1 データセット構築の必要性

深層学習では、学習用の大規模なデータを元に学習を行う。その際、データの量や多様性は、学習の精度にかかわってくるため、多くのデータが必要になる。

ネット上には、公開データセットが多数存在する。例えば、ImageNet [1] は、1400 万枚以上の画像と、2 万種類以上の物体名（クラス名）が収録されている、他と比べて桁違いに大きなデータセットである。VGG16、VGG19、ResNet といった著名な深層学習ネットワークについて、ImageNet を使って学習済みのモデルが公開され、広く引用されている。このように、ImageNet は、深層学習の研究の進展に大きく貢献している。しかし、牛の個体識別に利用できる大規模牛顔画像データセットは存在しない。よって、大規模牛顔画像データセットが構築できれば、牛の個体識別の研究の進展に大きく貢献できると考えた。

## 2.2 能動学習によるデータセット構築の効率化

### 2.2.1 受動学習と能動学習

機械学習の学習方法の1つに、教師あり学習というものがある。教師あり学習は、学習データに、ラベルという正解を与えた状態で学習させる方法である。これは、ラベルを与えられたデータを受動的に学習することから、受動学習と呼ばれる。教師あり学習では、大量のデータを使用してモデルを学習させることで、高い性能を得ることができる一方、そうした大量のデータに対して、アノテーションと呼ばれるラベル付け作業のコストが高くなり、人間の手間がかかりすぎてしまうことが問題となる。また、類似するデータがデータセットに含まれる場合、それらは学習データとして冗長なものになる可能性がある。つまり、大量のラベル付きデータだからといって、必ずしも全てのデータが性能の向上に寄与するとは限らない。このように、ランダムに収集したデータでデータセットを作成した場合、モデルの学習に有用なデータは、大量のラベル付きデータの内の一部に過ぎないという可能性がある。

このアノテーションの手間を減らすために、能動学習というものがある。能動学習は、教師あり学習の一種で、比較的少数のラベル付きデータで学習された機械学習モデルを使って、ラベル付けされていないデータから、学習効果の高いデータを選んで、アノテーションを行うというものである。これにより、あらかじめ大量のラベル付きデータを作成する必要性がなくなり、冗長性を排除した最低限の数の学習データによる学習が可能になることが期待される。

## 2.2.2 プールベース能動学習

プールベース能動学習は、能動学習の一種で、pool というラベルなしデータの集合から、学習データに追加すべきデータを選択するという方法である。プールベース能動学習の構成要素と処理の流れは、以下の通りである（図1）。

### 構成要素

- 初期ラベル付きデータ集合
- 機械学習モデル
- ラベルなしデータの集合（pool）
- オラクル（人間）

### 処理の流れ

1. 初期ラベル付きデータ集合を学習データとする
2. 学習データで機械学習モデルを学習する
3. 学習済みモデルを使って pool からデータを選択する
4. 選択したデータをオラクルに問い合わせる
5. オラクルがアノテーションを行い、学習データに追加する
6. 学習済みモデルが十分なデータを学習するまで、2～5 を繰り返す

学習データに追加すべきデータの選択方法には、様々な方法がある。また、数多くあるデータ選択方法から、最適なデータ選択方法を実行するという研究 [2] も存在するが、いずれも、オラクルによるアノテーションを必要とする方法である。本研究では、オラクルによるアノテーションを行うことなく、アノテーションの手間を減らす方法を検討する。

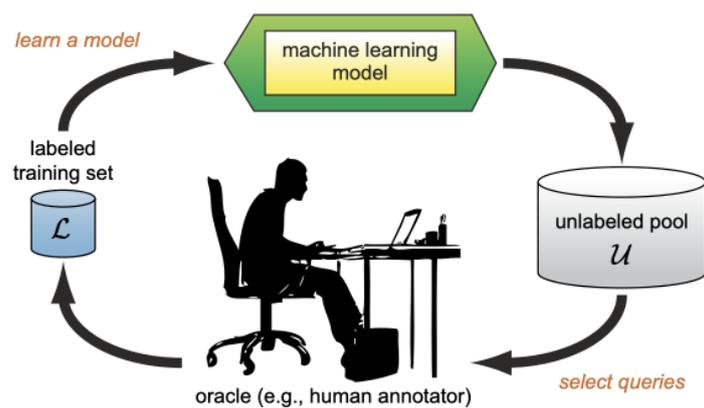


図 1 プールベース能動学習における学習サイクル [3]

### 2.2.3 データセット構築の効率化

本研究では、データセット構築の効率化のために、牛の個体毎に動画像を撮影する。次に、この動画像に対して牛の顔検出を適用する。これにより、得られた動画像毎の検出領域が、牛個体の顔データセットに対応する。

牛の顔検出を行うためには、牛の顔検出器の構築を行わなければならないため、牛の顔検出器の構築にプールベース能動学習を用いる。プールベース能動学習において、学習データに追加するデータは、人手によってアノテーションが行われたデータであるが、本研究では、牛の顔検出器で検出された顔を、牛の顔学習データに追加する。この場合、人手によるアノテーションが不要であるという利点があるが、検出できたデータを学習データに追加しても性能向上につながらないという可能性がある。この点については、実験によって評価する。

## 3 プールベース能動学習による牛の顔データ セット構築

### 3.1 処理の流れ

プールベース能動学習による牛の顔データセット構築の処理の流れは以下の通りである。

1. 牛画像の準備
2. 牛の顔学習
3. 追加データの選択
4. 再学習とデータセット構築

以下、それぞれの処理について述べる。

## 3.2 牛画像の準備

牛画像の準備のために、みらいグローバルファーム [4] にて、2日間牛の動画撮影を行った。この動画は、牛を個体別に、正面方向や横方向から撮影したものである。撮影した動画の本数は1509本である。それらをフレームレート（1秒間の画像の枚数）10として画像に変換した結果、画像の枚数は280536枚となった。図2に取得した牛画像の例を示す。こうして得られた牛画像の一部を、プールベース能動学習におけるpoolとして利用する。また、別の一部に対して、人手でアノテーションを行い、初期ラベル付きデータ集合とする。



図2 牛画像の例

### 3.3 牛の顔学習

牛の顔学習には、機械学習モデルである YOLO [5] を用いる。YOLO は、物体検出に使用される代表的な機械学習モデルで、従来の物体検出法と比べて、高速かつ正確に物体を検出できるというメリットがある。2015 年に論文で発表されて以降、新たなバージョンが発表され、最新のバージョンは YOLOv5 である。

YOLOv5 を使った学習では、事前学習済みモデルを利用する。事前学習済みモデルは、牛の顔は学習されていないので、牛の顔を追加学習させることで、牛の顔を検出できるようにする。YOLOv5 は、ネットワークの規模により、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x の 4 種類あり、学習の精度は、YOLOv5s < YOLOv5m < YOLOv5l < YOLOv5x、学習の速度は、YOLOv5x < YOLOv5l < YOLOv5m < YOLOv5s となっている (図 3)。学習データ、事前学習済みモデル、画像サイズ、バッチサイズ、エポック数を指定することで YOLOv5 の牛の顔学習が行われる。事前学習済みモデルとしては、YOLOv5s を COCO dataset で学習したものをを用いる。バッチサイズは、学習データをいくつかに分けたそれぞれのサブセットに含まれるデータの数である。また、エポック数は、データセット全体を学習する回数である。

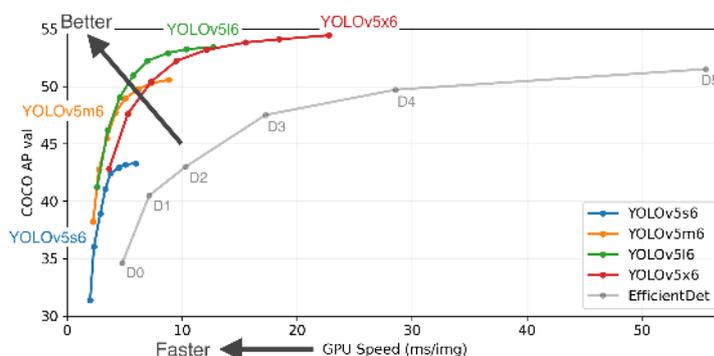


図 3 YOLOv5 の各ネットワークの性能 [6]

### 3.4 追加データの選択

牛の顔を学習した牛の顔検出器を使って、pool に対して牛の顔検出を行い、検出結果から、学習データに追加するデータを選択する。YOLO による牛の顔検出を行うと、検出結果として、牛の顔を囲う矩形とその信頼度 (0~1 の値) が出力される (図 4)。信頼度が十分高い検出結果は、牛の顔をしっかりと囲うことができているため、ほぼ正解とみなして良いとする。信頼度が十分高いデータから、一部をランダムに選択して、学習データに追加する。



図 4 YOLO による牛の顔検出結果画像

## 3.5 再学習とデータセット構築

学習データに追加データを加えた、新たな学習データで牛の顔を再学習する。牛の顔を再学習した牛の顔検出器を使い、pool に対して牛の顔検出を行い、信頼度が十分高いデータを学習データに加え、再び牛の顔を学習する。これらの処理を繰り返し、十分高い性能の検出器が構築できたら、プールベース能動学習は終了となる。

このプールベース能動学習によって構築した検出器を、個体毎に撮影した牛動画像に適用することで、個体毎の牛の顔データセットを作成する。

## 4 実験

### 4.1 実験の目的

この実験では、プールベース能動学習による牛の顔データセット構築において、信頼度 0.9 以上の検出結果を学習データに加える提案手法 (conf0.9) により、牛の顔検出の性能が向上するかを評価することを目的としている。比較手法として、オラクルによるアノテーションが行われた信頼度 0.4 以下の牛のデータを学習データに加えたとき (conf0.4) と、オラクルによるアノテーションが行われたランダムな牛のデータを学習データに加えたとき (random) を用いて実験する。

## 4.2 実験方法

本実験では、以下の設定でプールベース能動学習を行った。

- 初期ラベル付きデータ集合：初期牛の顔データセット（train40枚、validation10枚）
- 機械学習モデル：YOLOv5
- pool：牛画像10000枚、正解ラベルあり
- test データ：牛画像1000枚、正解ラベルあり

初期牛の顔データセットは、3.2節で示した280536枚の牛画像から、ランダムに選択した50枚に対して、牛の顔部分に手でアノテーションを行ったものである。また、poolとtestデータは、3.2節とは別に用意したtrainデータセット2077枚、validationデータセット510枚の牛画像で学習したYOLOv5を使って、3.2節の画像の内、56084枚に対して牛の顔検出を行うことで作成した。信頼度0.9以上であり、正確に検出できていると思われる10000枚をpool、1000枚をtestデータとした。また、この牛の顔検出による検出領域をそれぞれの正解ラベルとする。

プールベース能動学習において、poolから学習データに追加する牛のデータの選択基準には、YOLOv5の牛の顔検出による信頼度を用いている。conf0.9では、その時点の学習済みYOLOv5をpoolに適用し、信頼度が0.9以上の検出結果の中から、ランダムに8枚選択し、そのまま追加する。conf0.4では、信頼度0.4以下の牛画像から、ランダムに8枚選択し、予め与えておいた正解ラベルを追加する。randomでは、pool内の全ての牛画像から、ランダムに8枚選択し、同様に正解ラベルを追加する。YOLOv5の学習は、事前学習済みモデルをYOLOv5s、画像サイズを640、バッチサイズを32、エポック数を300として行った。また、プールベース能動学習では、conf0.4において、追加される牛のデータの数が8枚未満となるまでの5回、再学習を繰り返した。

検出の性能の評価指標としては、IoU（図5）を用いた。IoUは、2つの領域がどれくらい重なっているかを表す指標で、2つの領域の共通部分を和集合で割ったものになる。正解領域と予測領域の重なりが大きいほどIoUの値は大きくなり、予測領域がズレていたり、正解領域内でも小さすぎたり、正解領域を含んでいても大きすぎるとIoUの値は低くなる。つまり、IoUの値が大きいほど、物体検出がうまくできていると言える。

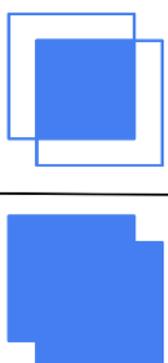
$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$


図5 IoU [7]

### 4.3 実験結果

実験では、それぞれの手法で、test データの正解ラベルと検出結果ラベルによる IoU を計算し、平均 IoU の変化を図 6 に示した。図 6 より、データを追加する毎に、平均 IoU が大きくなり、 $\text{conf}0.9 < \text{conf}0.4 < \text{random}$  の順に検出の性能がよいことが分かった。

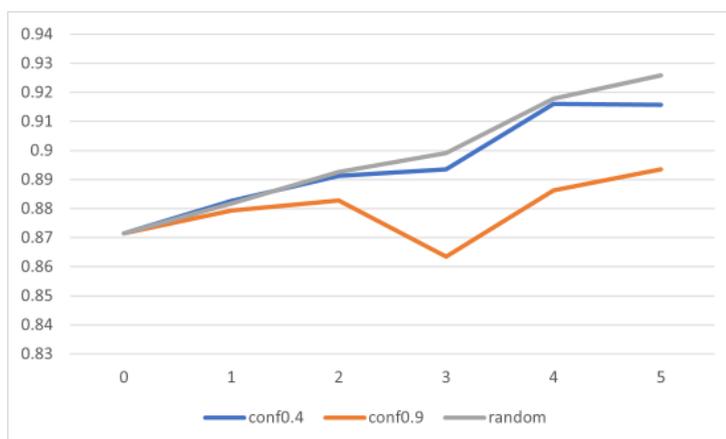


図 6 プールベース能動学習による平均 IoU の変化

## 4.4 データセット構築実験

4.2 節のそれぞれの手法で、再学習を 5 回繰り返したことによって構築した牛の顔検出器による牛の顔検出を、4.2 節の test データに適用する。この牛の顔検出による信頼度 0.8 以上の検出結果を、牛の顔データセットとする。このとき、牛の顔データセット内のデータが、正しくアノテーションされているか評価する。牛の顔データセットの正解ラベルと検出結果ラベルによる IoU を計算し、IoU が 0.8 以上の場合、そのデータは正しくアノテーションされているとする。4.2 節の初期ラベル付きデータ集合だけを学習したことによって構築した牛の顔検出器 (seed) でも、同様の実験を行う。

実験結果を表 1 に示した。信頼度 0.8 以上のデータ数は、牛画像 1000 枚の内、どのくらいの数を牛の顔データセットに登録できるかを示している。また、IoU0.8 以上のデータ数は、牛の顔データセットに登録したデータの内、正しいデータはどのくらいかを示している。表 1 より、信頼度 0.8 以上のデータ数は、seed < conf0.4 < conf0.9 < random の順に多く、それらのデータの内、IoU が 0.8 以上のデータの割合は、seed < conf0.9 < random < conf0.4 の順に大きかった。

表 1 牛の顔データセットの正確さ

手法	信頼度 0.8 以上のデータ数	IoU0.8 以上のデータ数	正しいデータの割合
seed	822	790	96.1 %
conf0.4	916	908	99.1 %
conf0.9	940	914	97.2 %
random	972	955	98.3 %

## 4.5 考察

4.3 節より、実験結果から、人手でアノテーションを行ったデータを学習データに追加した方が、検出の性能の向上に適していることが分かる。しかし、提案手法でも、まったく手をかけずに検出の性能の向上につながるということが分かった。

4.4 節より、50 枚の初期ラベル付きデータ集合のみを使って学習した場合に比べ、提案手法は人手をかけず 118 枚の画像をデータセットに追加できた。また、アノテーションの正確性も多少向上した。比較手法では正確性がさらに向上するが、人手による追加のアノテーションが必要である。アノテーションの手間を考えると、提案手法でデータセット構築した後に、誤って登録されたデータを削除する方が手間が少ない。

したがって、検出の性能の向上につながる点、人手によるアノテーションの手間を減らすことができる点から、提案手法は、牛の顔データセット構築の効率化に寄与すると考えられる。

## 5 おわりに

本研究では、プールベース能動学習を用いた牛の顔データセット構築の効率化について検討した。プールベース能動学習により牛の顔検出器を作成し、それを利用して牛の顔データセットを構築するアプローチをとった。プールベース能動学習では、データを順次追加する際に、追加するデータへの正解の付与が必要である。本研究では、その時点での学習済みの牛の顔検出器を使って信頼度高く検出された結果をそのまま追加データとすることで、人手をかけずに学習データを増やすことの有効性について調査した。

実験により、プールベース能動学習において、信頼度が十分高い検出結果を学習データに追加する方法は、信頼度の低いデータやランダムに選んだデータに対して人手でアノテーションを行いデータを追加する方法に比べて、検出精度の向上が低かった。また、学習後の検出器を使ったデータセット構築でも、正確性がやや劣った。しかし、初期ラベル付きデータ集合のみを学習したものよりは、検出精度は向上しており、データセット構築の正確性やデータセットに登録されるデータ数も向上した。人手をかけずにこれらの性能向上が得られることから、提案手法は牛の顔データセット構築の効率化に寄与することが分かった。

本研究の問題点は、実験回数が少ないこと、プールベース能動学習での性能の上限を調査できなかったことである。今後の課題としては、追加データの数を減らし、実験回数を多くすること、poolの正解データ全部を使って学習を行うこと、信頼度0.9以上の検出結果全部を使って学習を行うこと等、さらなる実験と検討が必要である。

## 謝辞

本研究を進めるにあたり、指導教員である棕木雅之教授に大変感謝致します。お忙しい中、研究に対してアドバイスを頂きありがとうございました。また、技術職員の高塚佳代子さん、そして棕木研究室の皆様にも、研究室での活動をサポートして頂いたことを感謝致します。

## 参考文献

- [1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge”, IJCV, doi:10.1007/s11263-015-0816-y, 2015
- [2] Yao-Yuan Yang, Shao-Chuan Lee, Yu-An Chung, Tung-En Wu, Si-An Chen, Hsuan-Tien Lin, “libact: Pool-based Active Learning in Python”, arXiv:1710.00379v1, 2017
- [3] Burr Settles, “Active Learning Literature Survey”, Computer Sciences Technical Report 1648, University of Wisconsin-Madison, doi:10.1.1.167.4245, 2010
- [4] みらいグローバルファーム (株) , <http://miraifg.co.jp/info/mirai-global-farm.html>
- [5] Dweepna Garg, Parth Goel, Sharnil Pandya, Amit Ganatra, Ketan Kotecha, “A Deep Learning Approach for Face Detection using YOLO”, IEEE Punecon, doi:10.1109/PUNECON.2018.8745376, 2018
- [6] v5.0 - YOLOv5-P6 1280 models, <https://github.com/ultralytics/yolov5/releases>
- [7] 物体検出の評価指標 IoU の計算方法, <https://qiita.com/shokupan/items/35eae224c59989957623>