

深層学習を用いた府大猫識別手法の推定

1 はじめに

近年、機械学習を用いた画像認識の技術は大幅に進歩している。例えば、普段からよく使われているスマートフォンでは顔認証や網膜認証などがすでに十分実用レベルに達している。こうした技術は一部動物の識別タスクにも適用されており、例えばペンギンであれば胸の模様、サメであれば背びれの形などそれぞれに個体認識ができる部位の識別に応用されている。

私は現在大阪府立大学において、ひと☆ねこという部活に所属している。大学内において多数の猫が存在するが、糞尿被害や研究施設への入り込みなどの問題があったため、猫の管理を目的として設立された部活である。この部活の活動の際に重要なのが個体識別である。現時点では、猫の管理を人が目で見て行っているため、カメラと機械学習を用いることでより効率的な個体識別が可能となれば大幅に作業を効率化できる。

そこで本研究では、深層学習を用いた府大猫識別手法の推定を行う。

2 要素技術

2.1 CNN

CNN(Convolutional neural network) は日本語で畳み込みニューラルネットワークと呼ばれる。CNN は、主に Convolution 層と Pooling 層からできており、特徴として、それぞれの層の間に生物の脳の視覚野に関する脳科学の知見にヒントを得た、「局所受容野」「重み共有」という結合を持っている。Convolution 層では、入力データにフィルタ（カーネル）を利用して畳み込み演算を行う。Pooling 層では、カーネルを画像全体に対し適用し統計的な処理を行う。頻繁に使われる Pooling 層は Average Pooling 層と Max Pooling 層の二つである。Max Pooling 層は適用されたカーネルの最大値を取り、Average Pooling 層は適用されたカーネル内の値の平均値をとる。

2.2 YOLO

一般物体認識 YOLO は物体検出のアルゴリズムである。You only look once の頭文字からとられており、一度読みこんだだけで識別できるというところから由来があり、その実行速度は非常に早く動画ですら、リアルタイム検出が可能となっている。YOLO では検出と識別を同時にを行うことで、処理時間の遅延を解消しようとしている。YOLO のアルゴリズムは以下の通りである。

1. 入力画像を $S \times S$ のグリッドセルと呼ばれる領域に分割する。
2. (a) それぞれのグリッドセルについて B 個の Bounding Box と信頼度スコアを推測する。信頼度スコアとは、どれくらい Box が物体を含んでおり、どのくらいの精度で Box がそれを予測したかというモデルのことである。ここで信頼度スコアをと定義する。上記の式では物体がグリッドセルに存在しなければ信頼度スコアは 0 となる。
(b) それと同時にそれぞれのグリッドセルは C 個の物体クラスそれぞれの条件付きクラス確率である Pr を表す。
3. 条件付きクラス確率と個々の Box の信頼度スコアをかけ合わせて、それぞれの BoundingBox のクラス固有の信頼度スコアを得られる。

以上のようにして YOLO は検出と識別を同時にに行っている。

2.3 mAP

mAP とは、物体検出の精度を比較するための指標である。

2.4 IoU

IoU とは Intersection over Union の頭文字であり、2つの領域がどれだけ重なっているかを $[0,1]$ の値で表す。

3 関連研究

関連研究として、中坪らの生成型学習における害獣検出の検討がある。これは、害獣による農作物への被害の増加を受け、害獣検出及び撃退システムを構築し、これを害獣対策に活用することを目指すものである。生成型学習とは

4 府大猫

5 提案手法

先に述べたように、本研究では府大猫の個体識別を行うが、まずは研究の第一段階として、個体識別をする前にカメラに写っている対象物が猫であるかを識別する必要がある。猫を識別するためには当然多量の学習データが必要となるが、猫が写っている写真の中から猫を取り出すことができれば次回以降訓練する際に自動でデータを集めることができる。そのため、猫の写っている写真の中から、猫をトリミングする仕組みを作る。まず、ImageNetで訓練されたVGG16の層を全部凍結して転移学習を行い、出力は猫か猫じゃないかの2値分類とした。

表1に転移学習時のパラメータを示す。実験的に 5×5 でのトリミングを行い、25箇所のうち、最も猫であると判断した所だけを抜き出した。図1、図2に訓練時のaccuracy, lossをそれぞれ示す。はじめから、train, validationともaccuracyが高く上手く学習が進んでいない。図3、図5、図7、図9をそれぞれトリミングした結果が図4、図6、図8、図10である。この方法では、図6のように見切れてしまうものが出現したり、また図10のように猫じゃないものを捉えてしまうものも存在した。猫がいるかどうかは判断できているものの、猫がどこにいるかを上手く判定できていない。図10に関しては、とがっているはっぱを猫の耳ととらえて探索を終了てしまっていると思われるため、猫らしさを耳の形で識別てしまっていると思われる。

表 1: モデル

クラス	2 クラス分類
訓練データ数	500 (各 250)
input	image($224 \times 224 \times 3$)
output	class(2)
ベースモデル	VGG16
optimizer	rmsprop
学習率	0.001
損失関数	categorical_crossentropy
train:validation	2:1
batch_size	32
epochs	30

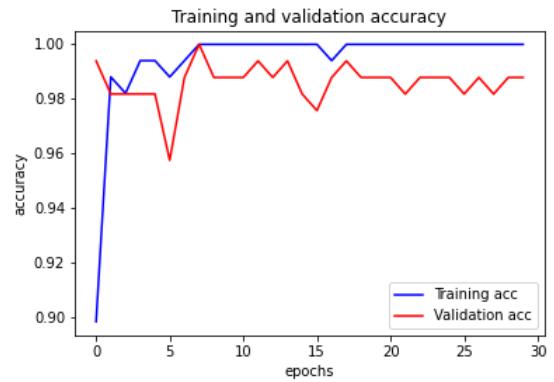


図 1: accuracy の推移

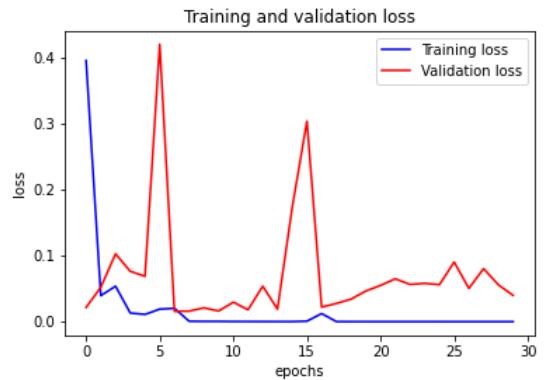


図 2: loss の推移

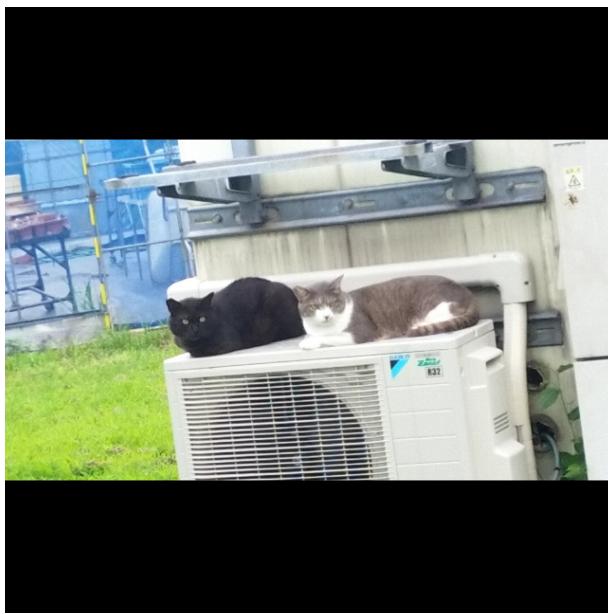


図 3: 猫画像 1

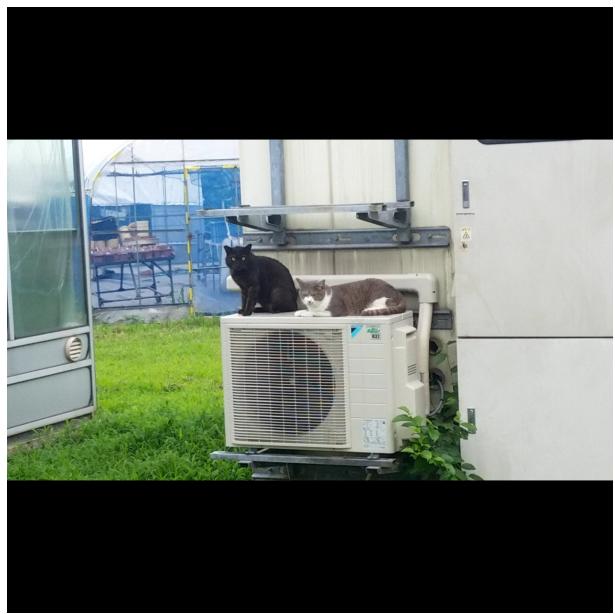


図 5: 猫画像 2



図 4: 猫画像 1 のトリミング



図 6: 猫画像 2 のトリミング

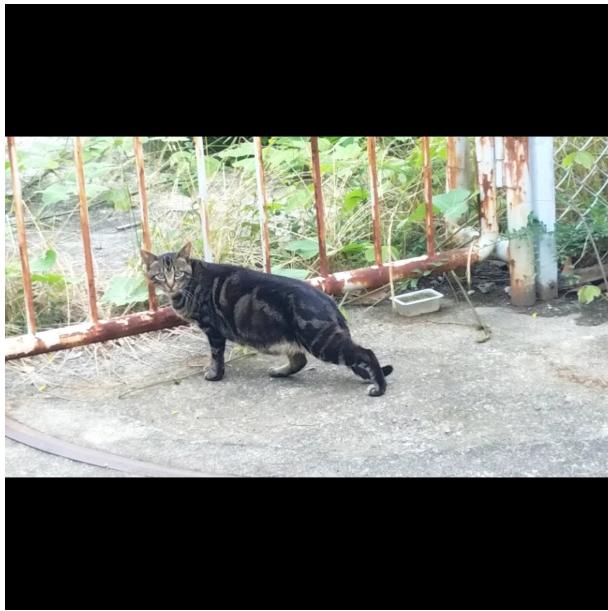


図 7: 猫画像 3



図 9: 猫画像 4



図 8: 猫画像 3 のトリミング

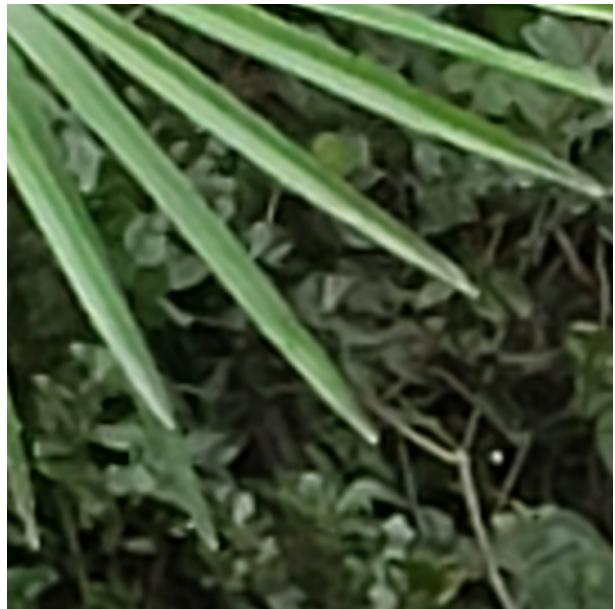


図 10: 猫画像 4 のトリミング