畳み込みニューラルネットワークによるパグの個体認識

谷聖一 研究室　佐藤光貴
Koki Sato

概要

犬種がパグである犬が写った画像を入力すると自分が飼っている犬かどうかを分類する方法を機械学習手法を用いて獲得した．用いた手法は，学習済み MobileNet v-1，学習済み VGG16, 未学習の VGG16 の三種類である．

1　はじめに

報告者は犬 を飼っていおり，その犬種はパグである．震災などが起こった際に飼っている犬が迷子になった場合に，自分もしくは他者が自分の犬を探し，その犬が自分の犬であるかを判断する必要がある．しかし，他者は見分けがつかず判断することは難しいと思われる．そこで画像を入力し，その犬が自分の飼っている犬かどうかを判定できる判定法を獲得することにした．どのように入力された画像が自分の飼っている犬であるかどうかを判定すれば良いだろうか.[1] によれば,「Google Cloud Platform」が提供するサービスの一つである「Google Cloud Vision API」は, 学習モデルを利用し, 画像内のさまざまな物体を検出することができる. そして, 画像を数千ものカテゴリ (「ヨット」,「エッフェル塔」など) に高速で分類することもできる. そこで，本演習では, 学習モデルを利用し, 撮影された動物が自分の飼っているパグであるかどうかを判定することにした. そのためには, 判定に用いる学習モデルを獲得する必要がある. 学習モデルの獲得手法として, 深層学習が有用な事例が多く報告されている. 例えば,2012 年に開催された一般画像認識のコンペティション「ILSVRC」において「Su-perVision team」は突出した成績をおさめた ([2]) が,「SuperVision team」は, 深層学習を用いて学習モデルを獲得している ([3]). そこで, 本演習では, 画像に写ってる動物が自分の飼っているパグかどうかの判定に, 深層学習を用いて学習モデルを獲得することにした.

2　準備

この節では，パグ，機械学習・ニューラルネットワーク・深層学習・CNN について説明する．

2.1　パグ

哺乳綱食肉目イヌ科の動物．家畜犬の一種で，中国原産の小型愛玩犬．鼻梁が短く，ひたいにはしわが深く刻まれ，目は大きく丸い．顔全体の印象は握り拳を思わせるが，パグという名も，ラテン語の握り拳に由来する．小柄ながら，体格にひ弱なところはなく，筋腱がよく発達した四角い体つきのイヌである．ほえ声も小型犬にしては太くて低く，喧騒でない．耳が垂れ耳で，尾は付け根が高く巻尾である．体重 6.5〜8.5 キログラム ([4]).

2.2　機械学習

本演習では機械学習 (Machine Learning) とは，人工知能の研究テーマの一つで「明示的にプログラムしなくても学習する能力をコンピュータに与えることである.」([5]) とする．また [6] によれば，明示的に表現することが困難なものを、具体的な事例などから計算機に帰納的に学習させることとされている．機械学習は画像認識や文字認識, 音声認識など様々な場面に使用される.

2.3　ニューラルネットワーク

ニューラルネットワークとは，脳神経系の情報処理機構を模倣した数理モデルであり，相互結合型や階層型などがある ([6]). 階層型ニューラルネットワークは入力層, 中間層 (隠れ層), 出力層の 3 層から構成される。深層学習に用いられる多層ニューラルネットワークは，中間層を複数層重ねた構造になっている．多層ニューラルネットワークは，多くの特徴量を抽出し，精度の高い分析をすることができる.

2.4　深層学習

深層学習とは，人工知能の学習手法の一つである.「深層というのは, 学習を行うニューラルネットワーク (ないしはそれに相当するもの) において層が深い、すなわち、何段にも層が積み重なっているということであり，深層学習とは多段の層が積み重なっているということである」([6]). 人間の神経回路に近いパーセプトロンの接続を行うことができ、入力された情報から多くの特徴を認識できるようになった．その為、画像認識や文字認識, 音声認識などで高い制度を誇っているとされている.

2.5 CNN

CNN とは Convolutional Neural Networks の略で畳み込みニューラルネットワークともいう.「人間の視覚神経系の構造にヒントを得たニューラルネットワーク 」([7]) である．深層学習における応用的な手法の一つで，局所領域の畳み込みとプーリングを繰り返すことで局所的な相関パターンの抽出や局所的な位置依存性をなくし，より精度の高い分析を可能としている．CNN は画像認識や音声認識などを至るところで用いられる．特に画像認識では移動普遍性を持たせることができる為、優れた実績を納めている．CNN は入力層，畳み込み層，プーリング層，全結合層、出力層から構成されている．以下の役割について述べる．

畳み込み層

畳み込み層は画像の特徴を検出する層である．入力のあらゆる位置で特徴の比較と一致点の検出を試み，画像全体にわたって特徴の一致件数を計算し特徴マップを作成する．

プーリング層

プーリング層は畳み込み層から出力された特徴マップを縮小することができる．領域内から新たな値を求める方法として，最大値を取り出す,「max pooling」や最小値を取り出す「min pooling」，平均値を取り出す「avaragepooling」などがある．本演習では最大値を取り出す「max pooling」を使用した．

全結合層

全結合層とはプーリング層からの出力を入力として，出力層に渡すユニットの値を決定する層である．

2.6 転移学習

転移学習 (Transfer Learning) は，ある異なるタスクに 対して学習された DCNN モデルから，目的とするタスク において利用するモデルへ重みを転移し，目的に対応する 学習データを用いて再学習を行うという手法である ([8])． 転移学習を行うことにより，データセットが小規模である 場合においても，汎化能力の高いモデルが得られる場合が 多いことが報告されている ([9])．本演習では，2014 年に開催された一般画像認識のコンペティション「ImageNet Large Scale Visual Recognition Challenge 2014(ILSVRC 2014)」において第二位を獲得し た CNN モデルである，VGG-16 ([10]) と同じく CNN モデルである MobileNet([11]) を用いて転移学習を 行った．VGG-16 は 100 万枚を超える自然画像によって学 習されており，13 層の畳み込みおよびプーリング層から なる特徴抽出部を介して，入力画像から 4096 次元の特徴 量を抽出する. 学習済み VGG-16 と同一構造を持つ CNN を構築し，自分で用意したデータセットを用いて，転移学習無しのモデル学習を 行った. この学習済み MobileNet, 学習済み VGG16，転移学習なしの VGG16 の 3 種類で精度を比較する．

3 演習内容

本演習では，パグが写った画像を入力すると自分が飼っている犬かどうかを分類する方法を以下の三つの CNN を用いて学習した．

- 学習済みモデルの MobileNetv-1 から転移学習
- 学習済みモデルの VGG16 から転移学習
- 未学習の VGG16 で学習

演習の手順を以下に示す．

1. データの収集
2. データの整形
3. データの前処理
4. モデルの構築
5. 学習モデルの評価

以下に、各手順の詳細を示す．本演習に使用したネットワークは以下のようにして取得．

表 1: 使用したネットワーク

事前学習済み MobileNet v-1	「ImageNet」で事前学習されたネットワーク（Keras から利用）
事前学習済み VGG16	「ImageNet」で事前学習されたネットワーク（Keras から利用）
未学習 VGG16	自分で実装

未学習 VGG16 のネットワーク構成を [10] を参考に Keras を用いて構築した. 未学習の VGG16 のネットワーク構成を表 2 に示す.

表 2: 未学習 VGG16 のネットワーク

操作	カーネル	ストライド	チャンネル数	活性化関数
入力: 224 × 224	-	-	3	-
Conv2D	3 × 3	1	64	ReLU
Conv2D	3 × 3	1	64	ReLU
MaxPooling	2 × 2	2	64	-
Conv2D	3 × 3	1	128	ReLU
Conv2D	3 × 3	1	128	ReLU
MaxPooling	2 × 2	2	128	-
Conv2D	3 × 3	1	256	ReLU
Conv2D	3 × 3	1	256	ReLU
Conv2D	3 × 3	1	256	ReLU
MaxPooling	2 × 2	2	256	-
Conv2D	3 × 3	1	512	ReLU
Conv2D	3 × 3	1	512	ReLU
Conv2D	3 × 3	1	512	ReLU
MaxPooling	2 × 2	2	512	-
Conv2D	3 × 3	1	512	ReLU
Conv2D	3 × 3	1	512	ReLU
Conv2D	3 × 3	1	512	ReLU
MaxPooling	2 × 2	2	512	-
Flatten	-	-	25088	-
Dense	-	-	4096	ReLU
Dense	-	-	4096	ReLU
Dense	-	-	2	softmax

3.1 データ収集

パグ の画像をエキサイト画像検索から 2250 枚スクレイピングにより収集し，自分の犬の画像は 1131 枚撮影し，収集した．本演習では「github」上のコードを利用させてもらった ([12]).

3.2 データの整形

収集してきた各画像において，OpenCV を用いて正面を向いたぱぐの顔をトリミングした．本演習ではパグの顔を検出するモデルは OpenCV のデフォルトにないので「github」上のコードを利用しパグ の顔検出モデルを構築，トリミングを行った ([13]).

表 3: 演習環境

	自分の犬以外のパグ	自分の犬
収集枚数	2250	1131
正面を向いた顔が写っていた枚数	685	499

3.3 データの前処理

「Pillow」([14]) を用いて画像データのカラーモードを「RGB」に変換し，画像サイズを「224 × 224」に統一した．

3.4 モデルの構築

本演習においてモデル構築は「Keras」([15]) を用いて行う．

使用したライブラリを表 4 に示す．

表 4: 演習環境

ライブラリ	概要
Keras	深層学習ライブラリ
TensorFlow	計算ライブラリ
OpenCV	画像解析ライブラリ

表 5: データセット

	自分の犬以外のパグ	自分の犬
訓練データ	565	379
テストデータ	60	120

学習済み MobileNet

- 入力層
  ILSVR の入力サイズが「224 × 224」なので「224 × 224」に統一．
- 中間層
  10 層までの重みを固定し，11 層以降の重みを更新
- 出力層
  出力層はネットワークの最後に配置され出力値は精度を計算する際に使用される．クラス分類では，ソフトマックス関数 (softmax function) を用いた計算が比較的に有名で，出力値の合計値が 1 となるように変換できるので出力を確率とみなすことができる．尤度の大きいクラスをクラス分類とした．出力層の次元は 2 クラス分類するので 2 次元となる．

学習済み VGG16

- 入力層
  ILSVR の入力サイズが「224 × 224」なので「224 × 224」に統一．
- 中間層
  4 層までの重みを固定し，5 層以降の重みを更新
- 出力層
  出力層はネットワークの最後に配置され出力値は

精度を計算する際に使用される．クラス分類では，ソフトマックス関数 (softmax function) を用いた計算が比較的に有名で，出力値の合計値が 1 となるように変換できるので出力を確率とみなすことができる．尤度の大きいクラスをクラス分類とした．出力層の次元は 2 クラス分類するので 2 次元となる．

未学習 VGG16

- 入力層
  ILSVR の入力サイズが「224 × 224」なので「224 × 224」に統一．
- 中間層
  全層の重みを更新
- 出力層
  出力層はネットワークの最後に配置され出力値は精度を計算する際に使用される．クラス分類では，ソフトマックス関数 (softmax function) を用いた計算が比較的に有名で，出力値の合計値が 1 となるように変換できるので出力を確率とみなすことができる．尤度の大きいクラスをクラス分類とした．出力層の次元は 2 クラス分類するので 2 次元となる．

以上の 3 種類に対して，表 5 のデータセットを学習させ，学習モデルを構築した．

学習回数

本演習の学習方法はミニバッチ学習法を用いてバッチサイズを 32, エポック数を 100 とした．

3.5 学習モデルの評価

学習モデルの評価はホールドアウト法を用いて行う．ホールドアウト法はデータ全体を訓練データとテストデータに分割し，モデルの精度を確かめる手法である．本演習における学習モデルの評価を表 6 に示す．

表 6: 学習モデルの評価

	MobileNet (学習済み)		VGG16 (学習済み)		VGG16 (未学習)	
正解とする識別度のボーダー	50%	90%	50%	90%	50%	90%
正解率	0.942	0.500	0.983	0.950	0.983	0.853
適合率	0.964	0.625	0.983	0.950	0.983	0.939
再現率	0.917	0.417	0.983	0.950	0.833	0.754
F-値	0.940	0.500	0.983	0.950	0.893	0.836

3.6 考察

本演習ではこの三種類の中では，正解率，適合率，再現率，F-値　のすべてにおいて学習済み VGG16 が良い結果となった．

3.7 今後の課題

今後の課題は 3 点ある．1 点目は，アプリケーション化である．アプリケーション化を行うことで誰にでも使ってもらえるようにすることが目的である．2 点目は汎用化である．本演習では今の自分の犬かどうかの判定しかできないので，他の犬を入れたときにも分類・判定できるようにしたいと考える．最後に 3 点目は，他の手法での学習である．本演習では，MobileNet，VGG16 を使用したが別の手法でも精度を出し，比較し，より良い手法を用いたい．

参考文献

[1] Google LLC． Cloud Vision
https://cloud.google.com/vision/?hl=ja (参照:2020-02-10)

[2] Olga Russakovsky*，Jia Deng*，Hao Su，Jonathan Krause，Sanjeev Satheesh，Sean Ma，Zhiheng Huang，Andrej Karpathy，Aditya Khosla，Michael Bernstein，Alexander C. Berg and Li Fei-Fei
ImageNet Large Scale Visual Recognition Challenge
International Journal of Computer Vision，December (2015)，Volume 115，pp 2112013252

[3] Alex Krizhevsky，Ilya Sutskever and Geoffrey E. Hinton
Imagenet classification with deep convolutional neural networks
In Proc． NIPS，(2012)

[4] 日本大百科全書（ニッポニカ）小学館

[5] Arthur Samuel Machine Learning
http://holehouse.org/mlclass/$01_02_{I}ntroduction_{r}egression_{a}nalysis_{a}nd_{g}r.html$/(参照 : 2019 − 02 − 06)

[6] 麻生 英樹，安田 宗樹，前田 新一，岡野原 大輔，岡谷 貴之，久保 陽太郎，ボレガラ ダヌシカ．人工知能学会監修.「深層学習」第 4 版．近代科学社．(2016)

[7] 小高知宏「基礎から学ぶ人工知能の教科書」オーム社 (2019)

[8] Pan, S. and Yang, Q. : A survey on transfer learning, IEEE Trans Knowl Data Eng 22, 1345-1359, 2010

[9] 上野洋典, 東耕平, 近藤正章:画像認識における効率的な転移学習のための学習モデル選択手法の検討（2017）

[10] K. Simonyan, et al. : Very deep convolutional networks for large-scale image recognition, arXiv technical report （https://arxiv.org/pdf/1409.1556.pdf: 2016.12.30 accessed）. 2014.

[11] A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" arXiv, 2017

[12] github excite-image-scraping
https://github.com/Doarakko/scraping-challenge/tree/master/excite-image-scraping/(参照:2020-02-10)

[13] github TraningAssistant
https://github.com/shkh/TrainingAssistant

[14] Pillow
http://pillow.readthedocs.io/en/latest//(参照:2020-02-10)

[15] Keras
https://keras.io/ja/(参照:2020-02-10)