2017 年度 修士論文

DNN を用いたペット顔検出・品種識別

Pet Face Detection and Breed Identification using DNN

提出日： 2018 年 1 月 30 日
指導： 渡辺 裕 教授

早稲田大学大学院 基幹理工学研究科
情報理工・情報通信専攻
学籍番号： 5116F094-5

渡部 宏樹

目次

第 1 章　序論 1
1.1　研究背景 1
1.2　研究目的 1
1.3　本論文の構成 2
第 2 章　関連用語 3
2.1　ディープラーニング [9] 3
2.1.1　ニューラルネットワーク [12] 3
2.2　畳み込みニューラルネットワーク 5
2.2.1　畳み込みニューラルネットワークの基本構造 5
2.2.2　畳み込み層 6
2.2.3　プーリング層 7
2.2.4　全結合層 7
2.2.5　ソフトマックス層 8
2.2.6　ドロップアウト 8
2.2.7　ReLU (Rectified Linear Unit) 9
2.2.8　データオーギュメンテーション 9
2.3　DCGAN 9
2.3.1　GAN 9
2.3.2　DCGAN 10
2.4　YOLO 12
第 3 章　提案手法 13
3.1　提案手法の概要 13
3.2　CNN と YOLO の二段階手法 14
3.3　DCGAN を用いたデータオーギュメンテーション 14
3.3.1　DCGAN を用いたデータオーギュメンテーションの概要 14
3.3.2　画像の品質評価値 16
3.4　提案手法 3 17
第 4 章　評価実験と考察 18
4.1　実験の概要 18
4.1.1　データセット 18
4.1.2　評価指標 20
4.2　実験 1：提案手法 1 の評価実験 22
4.2.1　実験概要 22

4.2.2 実験結果 22
4.2.3 考察 23
4.3 DCGAN により生成された画像 24
4.4 実験 2：提案手法 2 の評価実験 1 (生成画像付加の有無) 26
4.4.1 実験概要 26
4.4.2 実験結果 26
4.4.3 考察 27
4.5 実験 3：提案手法 2 の評価実験 2 (生成画像品質の与える影響) 28
4.5.1 実験概要 28
4.5.2 実験結果 28
4.5.3 考察 29
4.6 実験 4：提案手法 2 の評価実験 3 (品質評価指標) 32
4.6.1 実験概要 32
4.6.2 実験結果 32
4.6.3 考察 33
4.7 実験 5：提案手法 2 の評価実験 4 (追加画像枚数) 34
4.7.1 実験概要 34
4.7.2 実験結果 34
4.7.3 考察 36
4.8 実験 6：提案手法 3 の評価実験 37
4.8.1 実験概要 37
4.8.2 実験結果 37
4.8.3 考察 38
第 5 章 結論 39
謝辞 40
参考文献 41
図一覧 43
表一覧 44
研究業績 45

第1章
序論

1.1 研究背景

内閣府の調査によると，デジタルカメラの普及率は 58.9%，スマートフォンの普及率は 60.3%，特に 29 歳以下の若者のスマートフォン普及率は 93.9%とカメラが人々にとって身近なものとなっている [1]. また，コミュニケーションツールや情報発信ツールとして SNS (Social Networking Service)が台頭している．総務省の調査によると，SNS の利用者の割合は年々増加しており，10 代～30 代の約 7 割が SNS を利用している [2]．このようなカメラの普及と SNS の利用増加により，インターネット上には多くの写真がアップロードされるようになっている. それらの画像データを活用するために，画像中から特定の物体を切り出したり，画像にタグ情報を付加したりする必要が生じることがある.

代表的な被写体の一つにペットがある．スマートフォンカメラの利用者を対象にした調査によると，約 2 割の人がペットを主な被写体としている [3]．ペットの画像を活用する際には，ペットの顔領域の検出や品種タグ付けのための品種識別が必要になると考えられる．猫・犬の顔検出を行う研究として，Deformable Part Model を用いた研究が行われているが，犬の検出率は 91.8%と高いものの，猫の検出率が 74.5%と低い検出率であり，猫の検出率の改善が課題となっている [4].

コンピュータビジョンの分野において，ディープラーニングを用いた手法が高い成果をあげている．物体識別の分野では畳み込みニューラルネットワーク(CNN) [5]を用いた手法が, 物体検出の分野では You Only Look Once (YOLO) [6]や Single Shot MultiBox Detector (SSD) [7]といった CNN を応用したアルゴリズムが活躍している．このようなディープラーニングを用いた手法は猫や犬といったペットの顔検出や品種識別にも有効であると考えられる.

1.2 研究目的

物体検出や物体識別の性能を向上させることで, 正しい物体を映しており, 正しいタグが

付いた利用価値の高いデータを増やすことができる．本研究では，代表的なペットである猫・犬画像に焦点を当て，猫・犬の顔検出及び品種識別の性能向上を目指す．手法としては，ディープラーニング手法を検出・識別に適用する．顔検出には YOLO を使用する．YOLO は物体の検出と識別を同時に行うため，動物の顔を検出すると同時にその品種を求めることができる．しかし，本研究では顔を検出した後品種までは識別せず，猫か犬かの識別のみを行い，品種の識別はそれぞれの動物に対応した CNN の品種識別機で行うという 2 段階の手法を提案する．また，品種識別用の CNN について，類似画像生成アルゴリズムである DCGAN [8]を用いたデータオーギュメンテーションを行うことで更なる識別精度の向上を図る．

1.3 本論文の構成

本論文の構成は以下の通りである．

第1章　本章であり，本研究の背景や目的について述べている．

第2章　本論文で用いるディープラーニング手法関連の用語について述べる．

第3章　本論文で提案する手法について述べる．

第4章　提案手法の評価実験の概要，結果及び考察について述べる．

第5章　本論文のまとめを行う．

第2章
関連用語

本章では，本研究に関連するディープラーニング技術についての説明を行う．

2.1 ディープラーニング [9]

ディープラーニングは多層ニューラルネットワークを用いた機械学習手法の総称である．従来の機械学習では，人が設計した特徴量を使用していた．一方，ディープラーニングは機械が自ら特徴を設計，抽出し，学習を行う．その結果，ディープラーニングを用いた手法は従来手法よりも高い識別率を示し，画像認識や音声認識，自然言語処理といった分野で用いられるようになった．近年では，識別に限らず画像生成 [8]や画像のカラー化 [10]，囲碁のAI [11]などさまざまな分野で活躍している．

2.1.1 ニューラルネットワーク [12]

ニューラルネットワークとは，人間の脳神経をモデルにした情報処理システムである．ニューラルネットワークはユニットで構成されている．ユニットの基本構造を図 2.1 に示す．また，総入力u及び出力zの計算式を以下に示す．ユニットは複数の入力を受け取り，一つの出力を計算する．ユニットの総入力uは，各入力x_iにそれぞれ異なる重みw_iを掛け合わせたものの総和を求め，それにバイアス値bを足し合わせたものである．そして，総入力を活性化関数と呼ばれる関数fに適用した結果を出力とする．活性化関数としてはシグモイド関数やロジスティック関数などが用いられる．

$$u = x_1 w_1 + x_2 w_2 + x_3 w_3 \tag{2.1}$$

$$z = f(u) \tag{2.2}$$

代表的なニューラルネットワークの一つに順伝播型ニューラルネットワークがある．順伝播型ニューラルネットワークは層状に並べたユニットが隣接層間でのみ結合した構造を持ち，情報が入力側から出力側へと一方向にのみ伝播するニューラルネットワークである．その基本構造を図 2.2 に示す．

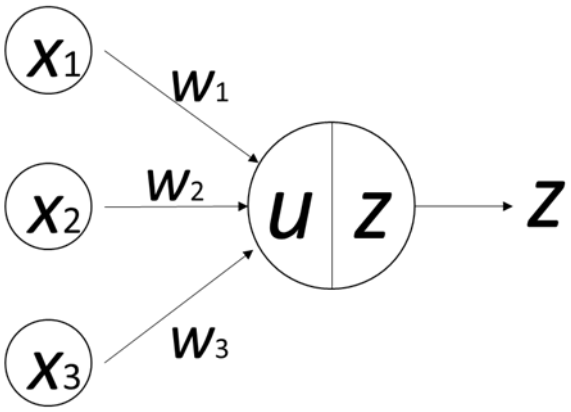


図 2.1　ユニットの基本構造

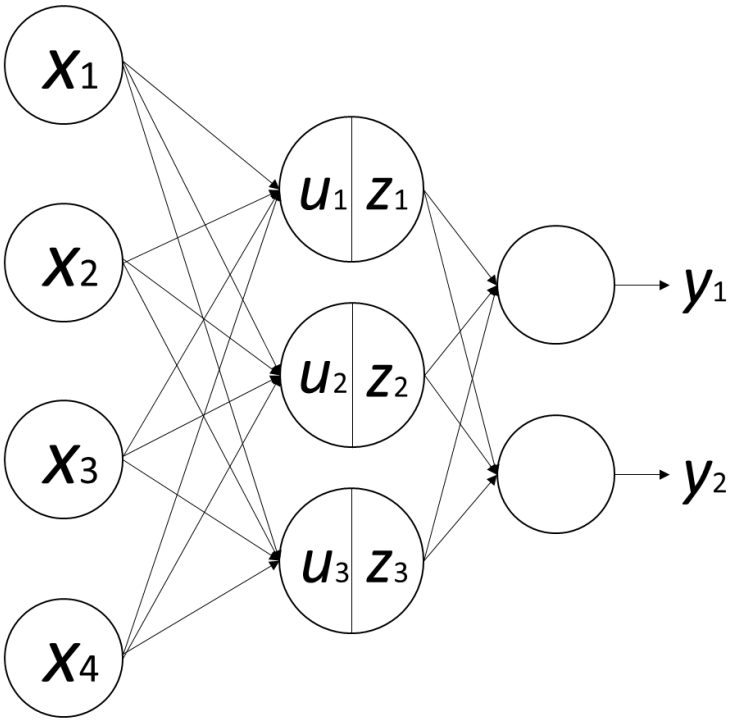


図 2.2　順伝播型ニューラルネットワークの基本構造

このような構造のネットワークにおいて，左の層を入力層，真中の層を隠れ層，右の層を出力層と呼ぶ．隠れ層の 3 つのユニットは入力層の四つのユニットからx_1, x_2, x_3, x_4を入力として受け取る．ユニット間の個々の結合にはそれぞれ異なる重みが与えられており，それらを用いてそれぞれの隠れ層のユニットで総入力が計算される．それらに活性化関数が適用されたものが隠れ層の出力となる．入力層のユニットを$i = 1,2, \dots, I$，隠れ層のユニットを$j = 1,2, \dots, J$，入力層のユニットiから出力層のユニットjへの重みをw_{ji}，出力層の各ユニ

ットのバイアスをb_j，活性化関数をfとすると，隠れ層の総入力u_j及び出力z_jは次のように計算できる．

$$u_j = \sum_{i=1}^{I} w_{ji} x_i + b_j \tag{2.3}$$

$$z_j = f(u_j) \tag{2.4}$$

順伝播型ニューラルネットワークでは上記の計算を入力層から出力層へ向けて順に行っていくことで情報を一方向へと伝達する．この関係は入力$\boldsymbol{x}$と出力$\boldsymbol{y}$を用いて関数$\boldsymbol{y} = y(\boldsymbol{x})$と表現することができる．この関数は各層間の結合重みとユニットのバイアスによって決定される．よって，これらの値を変えることで順伝播型ニューラルネットワークは様々な関数を表現することができる．

2.2 畳み込みニューラルネットワーク

2.2.1 畳み込みニューラルネットワークの基本構造

畳み込みニューラルネットワーク（Convolutional Neural Network)は画像認識分野においてよく用いられるディープラーニングの一種である．CNN と呼ばれることもある．畳み込みニューラルネットワークの基本構造を図 2.3 に示す. 畳み込みニューラルネットワークでは，畳み込み層とプーリング層と呼ばれる二つの層が交互に複数回繰り返される．畳み込み層が複数回続いた後，プーリング層が 1 層続く場合もある．また，認識の精度を向上させるため，プーリング層の後にドロップアウトを行ったり，ユニットに ReLU（Rectified Linear Unit を用いたりする．畳み込み層とプーリング層の繰り返しが終わると，全結合層と呼ばれる層でユニットの全結合が行われる．そして，クラス分類が目的の場合は最後に出力層としてソフトマックス層が用いられる．

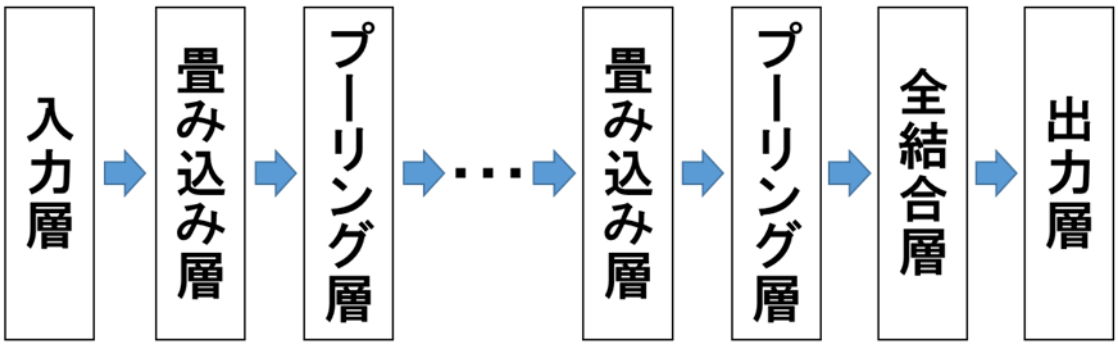


図 2.3　畳み込みニューラルネットワークの基本構造

2.2.2　畳み込み層

畳み込み層は，畳み込み演算を行う単層ネットワークである．一般的に，畳み込み層は多チャネルの入力画像に対し，複数のフィルタを畳み込む演算を行う．ここで多チャネルの画像とは一つの画素に複数の値を持つ画像である．例えば，RGB 画像は 3 チャネルの画像である．各フィルタは入力画像と同じ数のチャネルを持つ．3 チャネルの画像に 4 種類のフィルタを畳み込む際の畳み込み層の概要を図 2.4 に示す．

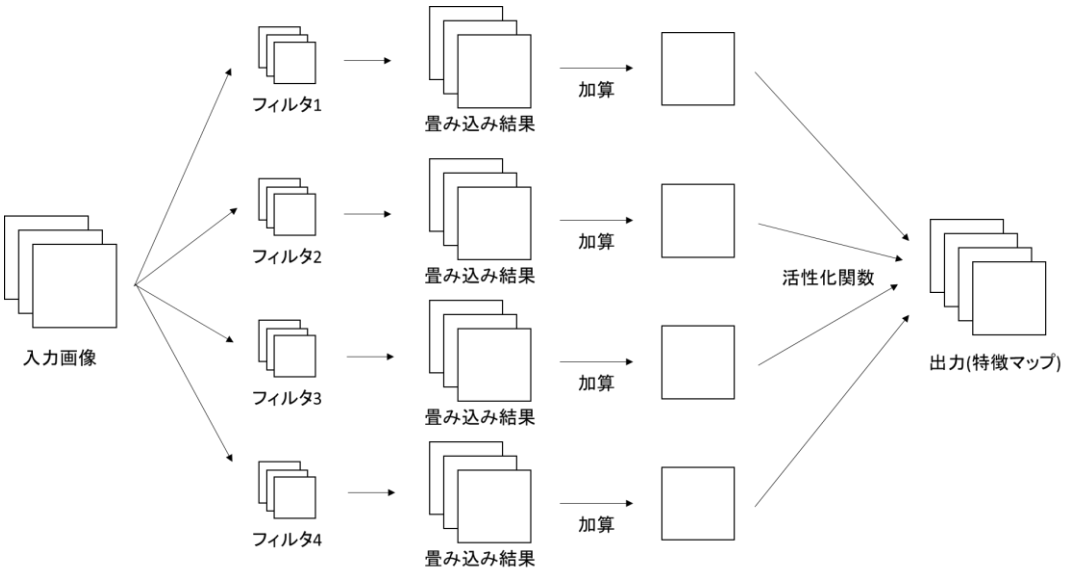


図 2.4　畳み込み層の概要(入力画像 3 チャネル，フィルタ 4 個)

あるフィルタについて，入力画像とフィルタの畳み込みを行うと，入力画像とチャネル数の同じ画像が作られる．その画像の全チャネルの画素値を加算し一つの画像とする．こうして得られた画像に活性化関数を適用する．この処理が各フィルタで行われた後，それらをまとめて畳み込み層の出力とする．このように，1 種類のフィルタから 1 チャネルの画像が計算され，それらをまとめた結果，フィルタの種類数と同じチャネルの画像が出力となる．また，この出力を特徴マップとも呼ぶ．

第l層の畳み込み層において，直前の第$l-1$層から K チャネルの画像$z_{ijk}^{(l-1)}$ $(k = 0,1,\dots,K-1)$を入力として受け取り，M 種類のフィルタ$h_{pqkm}(m = 0,1,\dots,M-1)$を適用することを考える．また，バイアスを$b_{ijm}$とする．このとき，フィルタからの出力$u_{ijm}$及び畳み込み層の出力$z_{ijm}^{(l)}$は次のように計算できる．

$$u_{ijm} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} z_{i+p,j+q,k}^{(l-1)} h_{pqkm} + b_{ijm} \tag{2.5}$$

$$z_{ijm}^{(l)} = f(u_{ijm}) \tag{2.6}$$

2.2.3　プーリング層

プーリング層は，一定領域内の畳み込み結果をまとめることで，一定領域内での平行移動普遍性を獲得した特徴を抽出する層である [13]．プーリングの例を図 2.5 に示す．

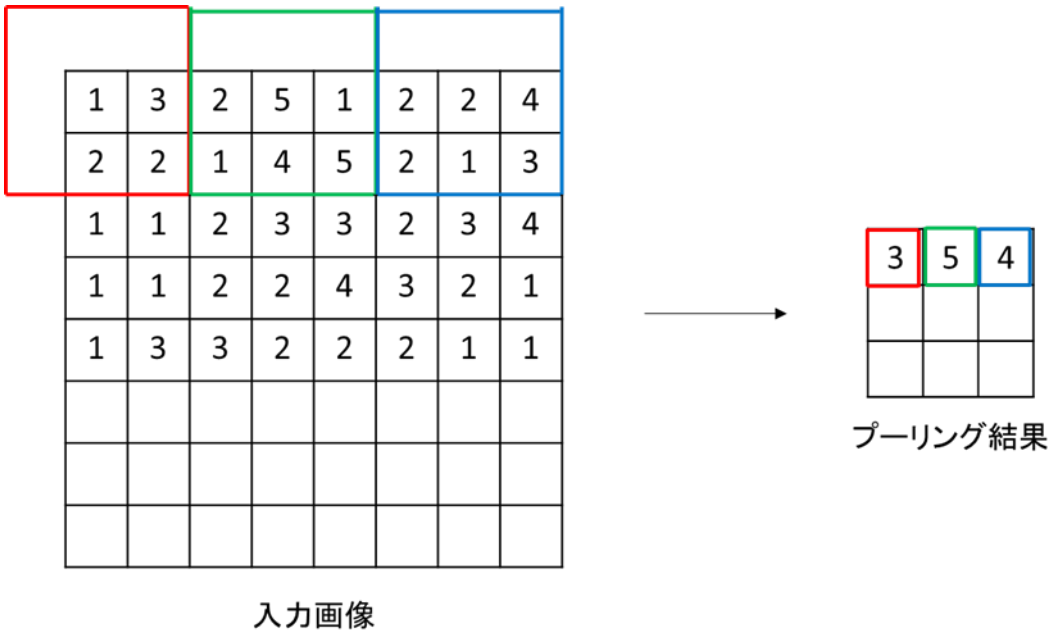


図 2.5　プーリング実行例(最大プーリング)

図 2.5 では，3×3 のプーリング領域を 3 画素ずつ動かして，領域内の最大の画素値を抽出している．このようなプーリングを最大プーリングと呼ぶ．この他にも，領域内の平均値を計算する平均プーリングが代表的なプーリング手法として知られている．

2.2.4　全結合層

全結合層は，隣接するひとつ前の層の全てのユニットが結合する層である．全結合層の概要を図 2.6 に示す．一般的に畳み込みニューラルネットワークの出力層の前の層に全結合層は用いられる．全結合層が何層か続く場合もある．

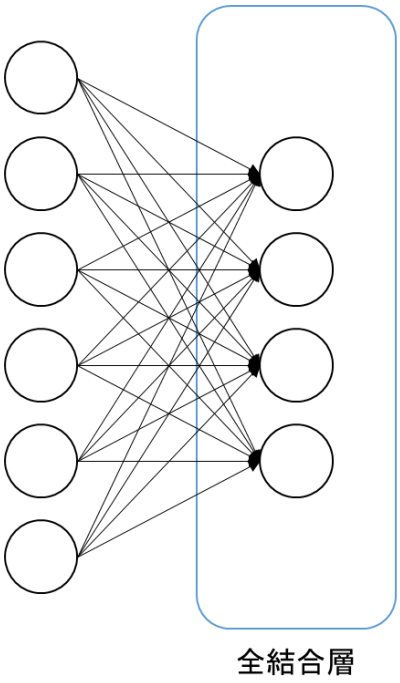


図 2.6　全結合層の概要図

2.2.5　ソフトマックス層

ソフトマックス層は多クラス分類問題を対象としたときに出力層として用いる層で，活性化関数にソフトマックス関数を用いる層である．ソフトマックス層には分類したいクラス数 K と同じ数のユニットを並べ，活性化関数として以下の式に示すソフトマックス関数を用いる [14].

$$y_i = \frac{\exp(u_i)}{\sum_{k=1}^{K} \exp(u_k)} \tag{2.7}$$

このときu_iは最終層のそれぞれのユニットの出力を，y_iは入力画像がクラスiに属する確率を表している．

2.2.6　ドロップアウト

ドロップアウトとは，多層ニューラルネットワークのユニットを確立的に選択して学習する手法である．学習時にユニットを決まった割合pで選択し．それ以外のユニットを無効化する．そして，その選択されたユニットのみのネットワークで最適化を行い，重みを更新する．ユニットの選択は重みの更新のたびに行う．推論時はすべてのユニットを用い

て順伝播計算を行うが，ドロップアウトで無効化の対象とした層のユニットは，出力をp倍する．

ドロップアウトは学習時にネットワークの自由度を下げることで，過適合を避ける狙いがある．また，単一のネットワークを使って，実質的に複数のネットワークの平均を取るのと同じ効果が得られると考えられている．

2.2.7 ReLU (Rectified Linear Unit)

ReLU とは，活性化関数として以下に示す正規化線形関数を用いるユニットである[14].

$$f(x) = \max(0, x) \tag{2.8}$$

このユニットは単純で計算量が小さいため，シグモイド関数やロジスティック関数を使うよりも学習が速く進み，また良い結果が得られることが多いためよく使われている．

2.2.8 データオーギュメンテーション

データオーギュメンテーションとは，CNN の学習に用いる画像の枚数を増やすために施される，画像の加工や変形などの処理の総称である．良く用いられる手法としては，画像の反転や回転，スケール変化，平行移動や輝度変化などがある．データオーギュメンテーションは過学習を防ぎ，CNN の性能を向上させることにつながる．

2.3 DCGAN

2.3.1 GAN

Generative Adversarial Network (GAN) [15]は，学習させた画像と似た特徴を持つ画像を生成するアルゴリズムである．GAN は Generator と Discriminator の二つのネットワークで構成されている．GAN の概念図を図 2.7 に示す．Generator は，一様分布などからサンプルされた乱数を入力として受け取り，これを種として学習した画像に似た特徴を持つ画像を生成する．Discriminator は，入力画像が学習データセット由来か生成器由来かを判別する．GAN では，この Generator と Discriminator の二つを競い合わせるように学習させる．Generator は，Discriminator が学習データセット由来の画像だと判断するような生成画像を作るように学習を進める．Discriminator は入力された画像が学習データセットから来たものか，Generator が生成したものかを正しく判断できるように学習を進める．この

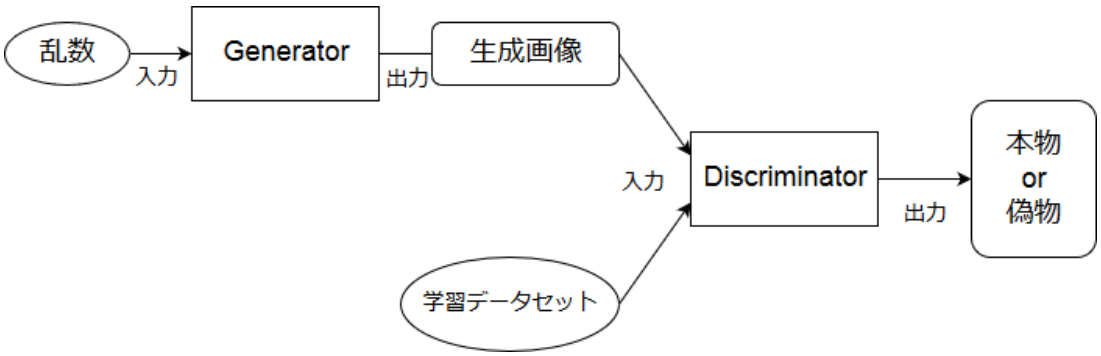


図 2.7　GAN の概念図

Generator と Discriminator の学習を交互に進めていくことで，互いに高め合うように学習が進んでいき，最終的に高性能な Generator を得ることができる．Generator 及び Discriminator の目的関数は以下の式のようになる．

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}[\log(1 - D(\boldsymbol{G}(\boldsymbol{z})))] \tag{2.9}$$

ここで，G は Generator，D は Discriminator，$\boldsymbol{x}$ は学習データ，$\boldsymbol{z}$ は乱数を．Discriminator が正しく画像を判別できるようになると，$D(\boldsymbol{x})$は大きくなり，$D(\boldsymbol{G}(\boldsymbol{z}))$は小さくなる．そのため，$\log D(\boldsymbol{x}), \log(1 - D(\boldsymbol{G}(\boldsymbol{z}))$は大きくなるため，Discriminator はこの目的関数を最大化しようとする．一方，Generator が Discriminator を欺けるようになると$D(\boldsymbol{G}(\boldsymbol{z}))$は大きくなるため，$\log(1 - D(\boldsymbol{G}(\boldsymbol{z}))$は小さくなる．そのため，Generator はこの目的関数を最小化しようとする．このように，Generator と Discriminator は対立的に目的関数を最適化する．

2.3.2　DCGAN

Deep Convolutional Generative Adversarial Network (DCGAN) [8] は，GAN の Generator と Discriminator を CNN で実装したものである．GAN の学習をうまく進めるために，次のような工夫を取り入れている [16].

・プーリング層の廃止

CNN では畳み込み層とプーリング層を交互に繰り返すことが一般的だが，プーリング層を次の畳み込み層に置き換える．Generator では fractional-strided convolution 層に，Discriminator ではストライドする畳み込み層に置き換える．

fractional-strided convolution とは，元となる特徴マップを拡張してから畳み込むことで deconvolution や transposed convolution とも呼ばれる [17]．fractional-strided

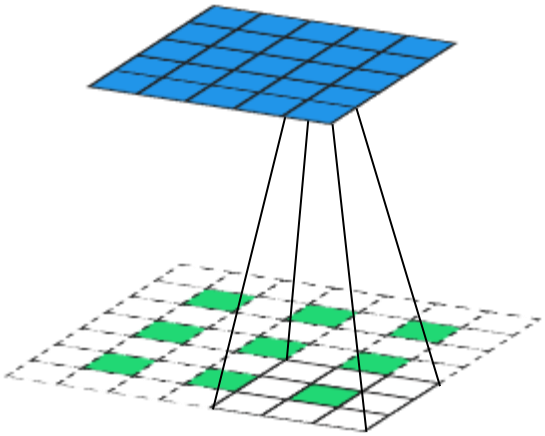


図 2.8　fractional-strided convolution の概要図

convolution の概要図を図 2.8 に示す．緑のマスが元の特徴マップ，青いマスが畳み込み後の特徴マップである．図に示されるように，元の特徴マップを隙間や周囲に 0 を埋めることで拡大した後に畳み込みが行われる．

・全結合層の廃止

CNN では最終層で全結合層を用いることがよくあるが，DCGAN の discriminator では全結合層を global average pooling [18]に置き換える．global average pooling は，一つの特徴マップに一つのクラスを対応させ，特徴マップの平均値をソフトマックス層に入力することでクラスを識別する．global average pooling を使うことで，学習の収束は遅くなるものの，学習の安定性が増すという利点がある．

・バッチ正規化の適用

Generator の出力層以外の層と Discriminator の入力層以外の層にバッチ正規化 [19]を適用する．バッチ正規化とは，バッチごとに入力の平均を 0 に，標準偏差を 1 になるように正規化を行うことである．これにより勾配の消失や爆発を防ぐことができるため，大きな学習係数を使用することができ，学習の収束速度が向上するというメリットがある [20].

・ReLU, Leaky ReLU の適用

Generator の出力層以外の各層に ReLU を，Discriminator の全ての層に Leaky ReLU を適用する．Leaky ReLU は以下の式であらわされる．また，そのグラフを図 2.9 に示す．

$$f(x) = \max(\alpha x, x) \ (\text{Leaky ReLU}) \tag{2.10}$$

ここでαはパラメータであり，DCGAN では$\alpha = 0.2$が使用されている．

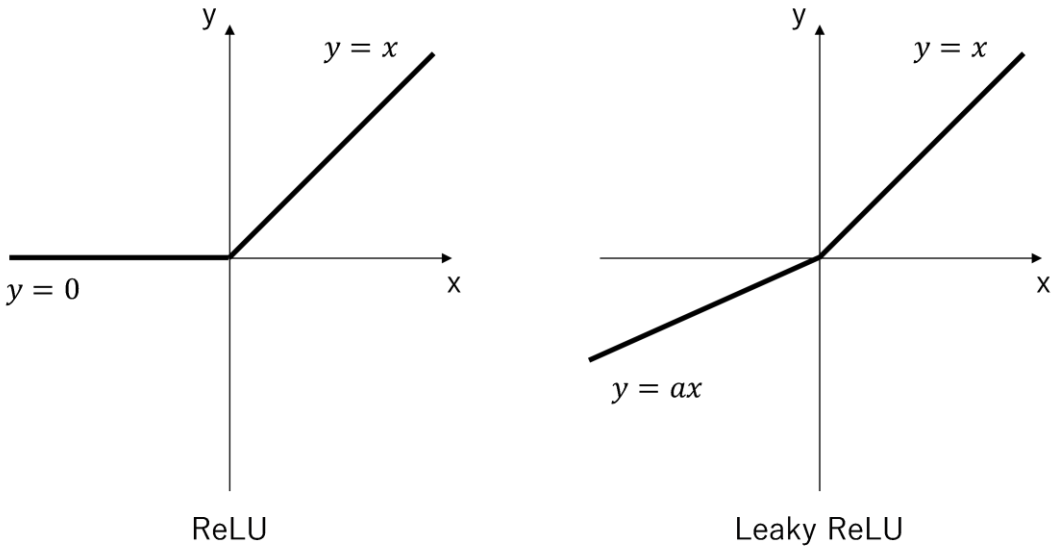


図 2.9　ReLU, Leaky ReLU のグラフ

2.4 YOLO

You Only Look Once (YOLO) [6]は CNN を用いた物体検出アルゴリズムの一つで，物体の検出と識別を同時に行う，高速・高精度なアルゴリズムである．YOLO は次のような流れで検出を行う．

1. 入力画像を S×S 個のセルに分割する．
2. セルごとに B 個の物体候補領域を予測する．
   各候補領域は候補領域内に物体が存在する確率$P(object)$を持つ．
3. セルごとにセル内の物体のクラス確率$P(\text{class}|object)$を予測する．
4. $P(object)$と$P(\text{class}|object)$を掛け合わせ，閾値以上となった候補領域を物体領域として検出する．

$$P(\text{class}) = P(object) \times P(\text{class}|object) \geq \text{閾値} \tag{2.11}$$

第3章
提案手法

本章では，提案手法についての説明を行う．

3.1 提案手法の概要

本研究では，猫・犬の顔検出及び品種識別の性能向上を目指している．提案手法の概要を図 3.1 に示す．

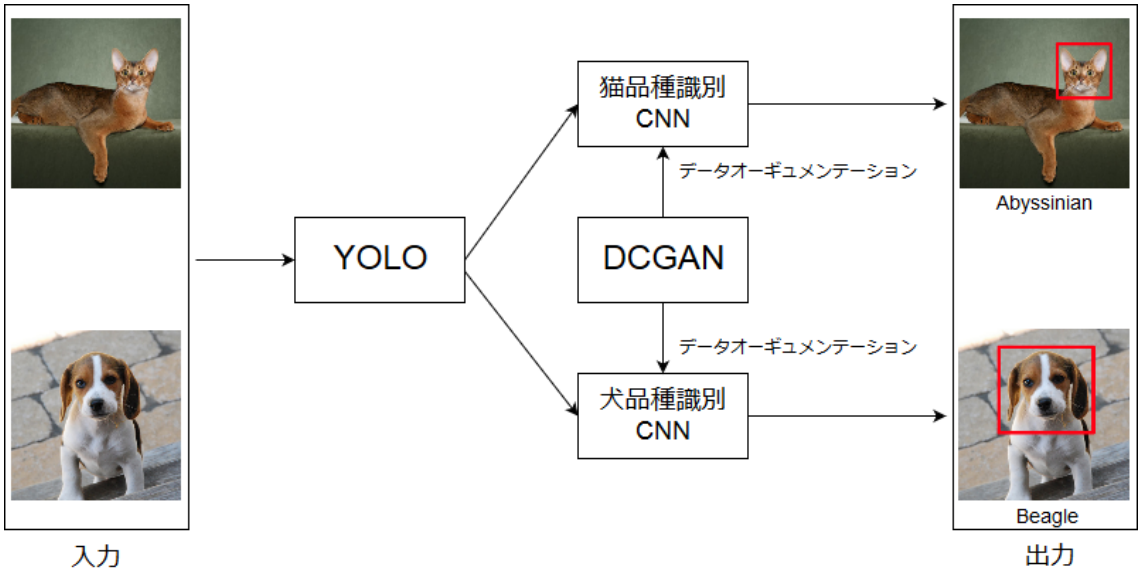


図 3.1　提案手法概要図（”The Oxford-IIIT-Pet Dataset” [21]）

提案手法は二つの手法で構成されている．一つ目は，YOLO で猫・犬の顔領域を検出し，その後検出した顔領域を猫・犬それぞれに対応した品種識別用 CNN で品種識別を行うという二段階の検出手法である．これを提案手法 1 とする．二つ目は，品種識別用 CNN を学習させる際に，DCGAN を用いて学習用データセットから猫・犬の類似画像を生成し，生成した類似画像を CNN の学習データセットに加えるというデータオーギュメンテーションである．これを提案手法 2 とする．そして，提案手法 1 と提案手法 2 を組み合わせたものを提案手法 3 とする．それぞれの提案手法について，次節より説明を行う．

3.2 CNN と YOLO の二段階手法

提案手法では，YOLO と CNN を組み合わせた二段階手法による猫・犬の顔検出及び品種識別を行う．これを提案手法 1 とする．提案手法 1 の概要図を図 3.2 に示す．

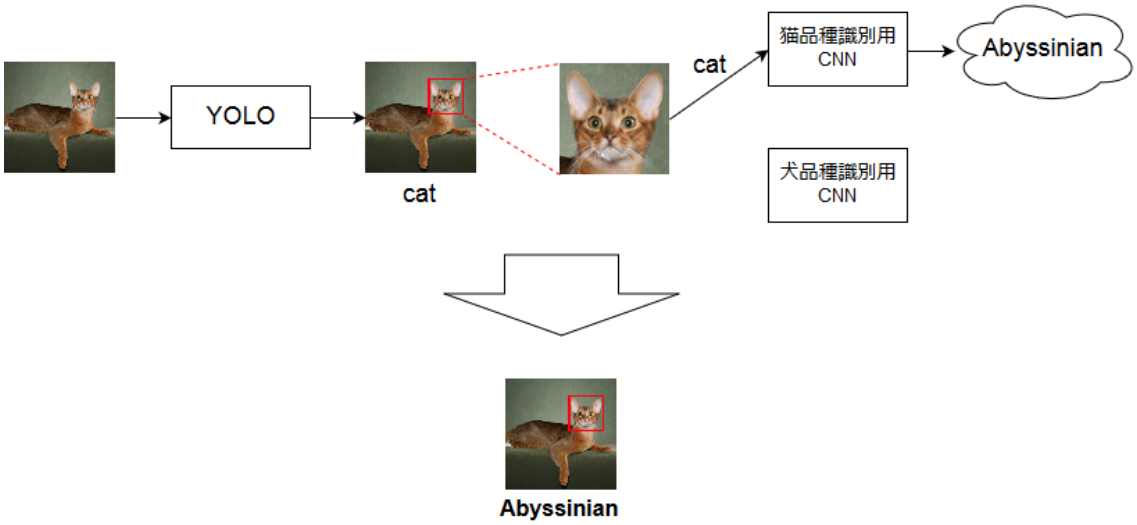


図 3.2　提案手法 1 の概要図（”The Oxford-IIIT-Pet Dataset” [21]）

提案手法 1 では，まず YOLO により猫・犬の顔領域の検出を行う．次に，検出した顔領域を，猫の顔領域ならば猫の品種識別用 CNN に，犬の顔領域ならば犬の品種識別用 CNN に入力として渡す．そして，品種識別用 CNN により品種の特定を行い，YOLO の結果と統合して最終的な検出結果を出力する．

3.3 DCGAN を用いたデータオーギュメンテーション

3.3.1　DCGAN を用いたデータオーギュメンテーションの概要

提案手法 1 の後段の品種識別用 CNN の学習において，DCGAN で生成した類似ペット顔画像を CNN の学習データセットに加えるデータオーギュメンテーションを行う．これを提案手法 2 とする．提案手法 2 の概要図を図 3.3 に示す．提案手法 2 の流れは次のようになっている．

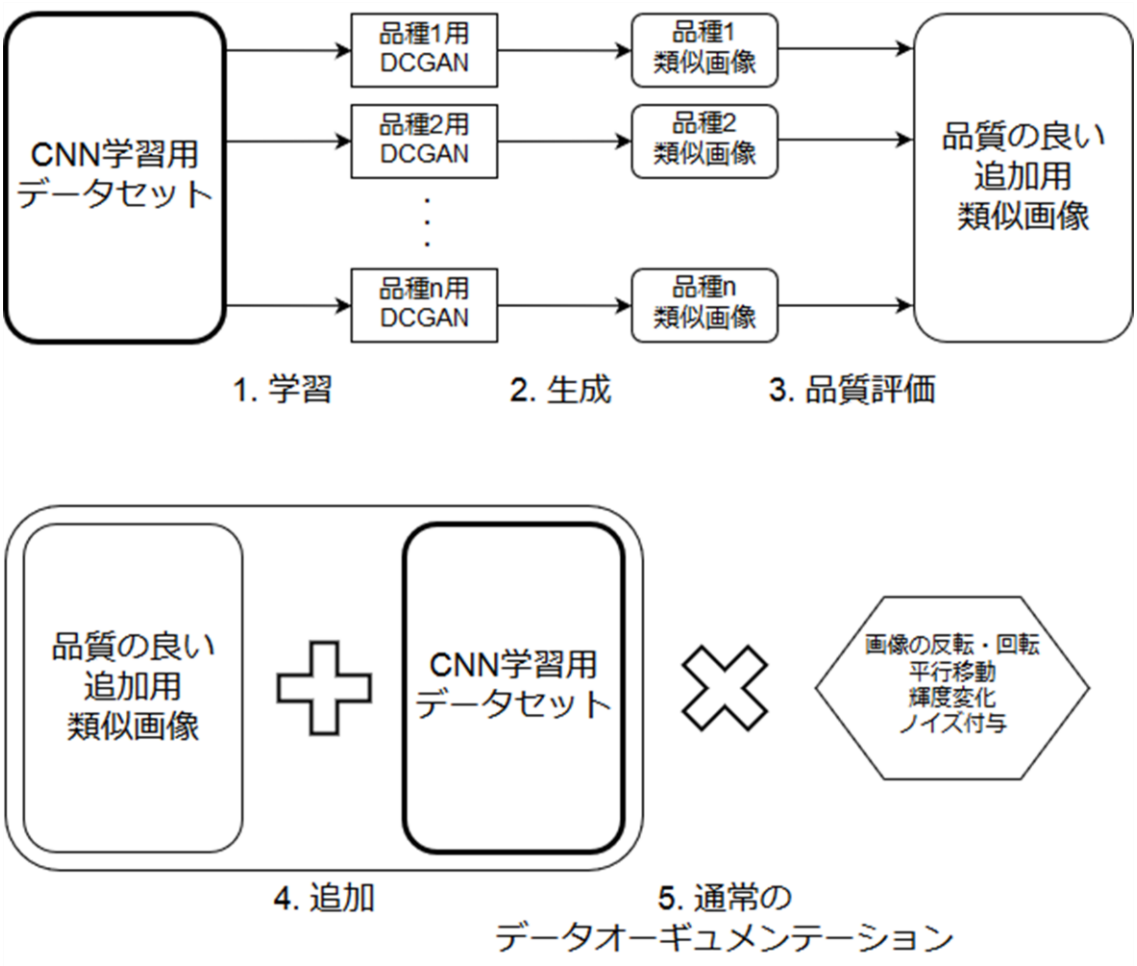


図 3.3　提案手法 2 の概要図

1. CNN の学習用データセットを用いて，品種ごとに DCGAN を学習させる．
2. 学習させた DCGAN で類似画像を生成する．
3. 生成した類似画像の品質評価を行い，品質の良い生成画像を取り出す．
4. 品質の良い生成画像を CNN の学習用データセットに加える．
5. 類似画像を加えた CNN の学習用データセットに通常のデータオーギュメンテーションを施す．

提案手法 2 では，品種ごとに DCGAN の類似画像生成機を用意し，それらに画像を生成させる．DCGAN からは人の目で見て明らかに猫や犬の画像に見えないような，品質の悪い画像も生成される．それらの画像を取り除くために，評価値による品質評価を行い，品質の良いものだけを CNN の学習用データセットに加えることにする．評価値については次項で説明する．そして，類似画像を加えた CNN の学習用データセットに通常のデータオーギュメンテーションを施す．本研究では，通常のデータオーギュメンテーションとして，画像の

反転，回転，平行移動，輝度変化，ガウスノイズ付与を施した．そして，提案手法 2 のデータオーギュメンテーションを施したデータセットで品種識別用の CNN を学習させる．

DCGAN の Generator は Discriminator の CNN が本物かどうか区別できないような画像を生成する．よって，その画像は DCGAN の学習画像に類似した CNN 特徴量を持つと考えられる．そのため，DCGAN の生成画像は CNN の学習画像としても働くと考えられ，CNN のデータオーギュメンテーションに使用できると考えられる．

3.3.2　画像の品質評価値

本研究で用いた DCGAN からは，品質の悪い画像も生成される．そこで評価指標を用いた品質評価を行い，品質の良いものを CNN の学習用データセットに加えることにする．評価指標として，次の三つを検討する．

・評価指標 1：DCGAN の Discriminator の出力

DCGAN の Discriminator は入力された画像が本物か生成された偽物かを判別する．品質の悪い画像は Discriminator が偽物と判断すると考えられるため，Discriminator の真偽識別結果を評価指標の候補の一つとする．

・評価指標 2：事前学習した品種識別用 CNN の出力

提案手法 2 を施す前のデータセットで学習した品種識別用 CNN を用意する．品質の悪い画像は顔の輪郭等が不完全であるため，品種識別用 CNN でうまく識別されないと考えられる．そのため，品種識別用 CNN による品種識別結果を評価指標の候補の一つとする．

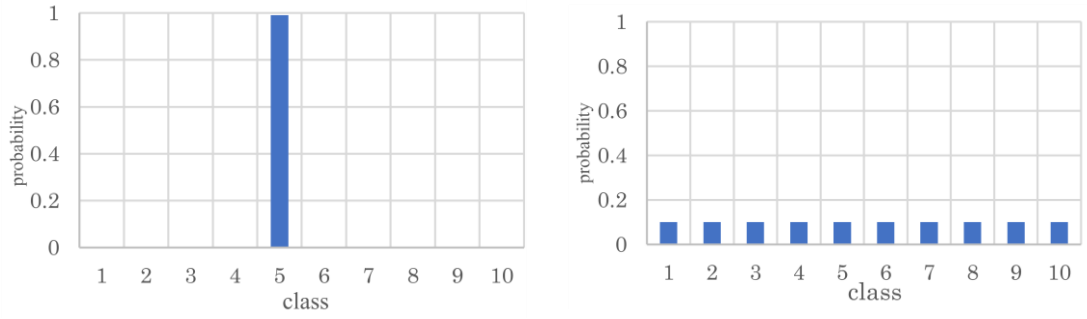
・評価指標 3：事前学習した品種識別用 CNN の出力から求める Inception score

Inception score[22]は GAN の Generator の評価指標の一つである．これは二つの確率分布の KL ダイバージェンスを求めたもので，次式のように表される．

$$\begin{aligned} \mathrm{I} &= \exp\left(\mathbb{E}_x KL(p(y|x)||p(y))\right) \\ &= \exp\left(\mathbb{E}_x \sum_x p(y|x) \log \frac{p(y|x)}{p(y)}\right) \end{aligned} \tag{3.1}$$

$p(y|x)$ は個々の生成画像の CNN による識別結果の確率分布を，$p(y)$ は生成画像全体の識別結果の確率分布の平均を表す．一般的な GAN は様々なクラスの画像を学習に使う．品質の良い画像は CNN で識別が正しくできる画像であると考えられ，理想的な$p(y|x)$の分布は図 3.4(a)のような一つのクラスに突出した分布となる．また，GAN からは様々なクラ

スの画像が生成されるため，理想的な$p(y)$の分布は図 3.4(b)のような全てのクラスに均一な分布となる．KL ダイバージェンスは二つの確率分布の差異を計る尺度であり，差異が大きいほど大きな値を取る．そのため，$p(y|x)$及び$p(y)$が理想的な分布の場合，Inception score は大きな値を取る．Inception score は人の主観的な品質評価と相関があるとされている．



(a)　　(b)

図 3.4　理想的な確率分布の例

(a: 一つのクラスに突出した分布, b: 全てのクラスに均一な分布)

本研究では，特定の品種の画像のみを学習した DCGAN を使用する．そのため，理想的な生成画像全体の識別結果の確率分布の平均$p(y)$は図 3.4(a)のような特定の品種に突出したものとなる．よって理想的な分布の場合，Inception score は小さな値を取る．そこで本研究では，Inception score の逆数を画像の評価指標として用いる．また，個々の画像の品質を評価するために，個々の生成画像の CNN による識別結果の確率分布と理想的な分布の KL ダイバージェンスを使用する．

3.4 提案手法 3

提案手法 3 は，提案手法 1 の後段の品種識別用 CNN の学習に，提案手法 2 の DCGAN を用いたデータオーギュメンテーションを適用したものである．提案手法 1 と 2 を組み合わせることで，さらなる性能の向上を図る．

第4章
評価実験と考察

本章では，提案手法の評価実験について述べる．

4.1 実験の概要

本研究では，提案手法の評価実験として次の 5 つの評価実験を行った．

実験 1： 提案手法 1 の評価実験
実験 2： 提案手法 2 の評価実験 1 (生成画像付加の有無)
実験 3： 提案手法 2 の評価実験 2 (生成画像品質の与える影響)
実験 4： 提案手法 2 の評価実験 3 (品質評価指標)
実験 5： 提案手法 2 の評価実験 4 (追加画像枚数)
実験 6： 提案手法 3 の評価実験

実験 1 で提案手法 1 の評価を，実験 2-5 で提案手法 2 の評価及び最適なパラメータの探索を，そして実験 6 で提案手法 1 と 2 をまとめた提案手法 3 の評価を行った．

4.1.1 データセット

本研究では，オックスフォード大学が公開している動物画像データセット，”The Oxford-IIIT-Pet Dataset” [21]より猫・犬それぞれ 12 品種ずつを使用する．1 クラスに約 200 枚の画像が存在し，そのうち 50 枚をテストに，残りを学習に使用した．使用した品種の一覧を表 4.1 に，画像の例を図 4.1 に示す．また、品種識別用の CNN や DCGAN の学習には、データセット中の動物画像の顔領域を切り出したものを使用した。

表 4.1　使用した品種の一覧

猫	犬
Abyssinian (Ab)	American Bulldog (AB)
Bengal (Ben)	American Pit Bull Terrier (AP)
Birman (Bi)	Basset Hound (BH)
Bombay (Bom)	Beagle (Bea)
British Shorthair (BS)	Boxer (Box)
Egyptian Mau (EM)	Chihuahua (Ch)
Maine Coon (MC)	English Cocker Spaniel (EC)
Persian (Pe)	English Setter (ES)
Ragdoll (Ra)	German Shorthaired (GS)
Russian Blue (RB)	Great Pyrenees (GP)
Siamese (Si)	Havanese (Ha)
Sphynx (Sp)	Japanese Chin (JC)



図 4.1　使用した品種の例（"The Oxford-IIIT-Pet Dataset" [21]）

4.1.2　評価指標

本研究では，提案手法の評価に次の評価指標を用いる．

Accuracy：全体の精度．
Precision：検出した物体のうち，正しく検出できたものの割合．
Recall：正解データのうち，正しく検出できたものの割合．
F 値：Precision と Recall の調和平均．Precision と Recall を統合して評価する指標

各評価指標の式を式(4.1) ~ (4.4)に，式中で用いた用語の表を表 4.2 に示す．

$$\text{accuracy} = \frac{\text{正解数}}{\text{テストデータ数}} \tag{4.1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.3}$$

$$\text{F 値} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.4}$$

表 4.2　スレットスコア

		正解	
		正	負
予測	正	TP (True Positive)	FP (False Negative)
	負	FN (False Positive)	TN (True Negative)

4.2 実験 1：提案手法 1 の評価実験

4.2.1 実験概要

実験 1 では，提案手法 1 である YOLO と CNN を組み合わせた二段階手法の評価実験を行った．YOLO 単体で猫・犬の顔の検出と品種識別を行う一段階の手法（以下，従来手法）と提案手法 1 である二段階の手法の実験をそれぞれ行い，その結果の比較を行った．

4.2.2 実験結果

従来手法の結果を表 4.3 に，提案手法 1 の結果を表 4.4 にそれぞれ示す．

表 4.3 従来手法の実験結果

	cat												
	Ab	Ben	Bi	Bo	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
Precision	0.840	0.700	0.654	0.787	0.729	0.556	0.732	0.745	0.622	0.642	0.705	0.794	0.709
Recall	0.412	0.135	0.680	0.960	0.700	0.800	0.600	0.760	0.560	0.680	0.620	0.540	0.621
F値	0.553	0.226	0.667	0.865	0.714	0.656	0.659	0.752	0.589	0.660	0.660	0.643	0.662

	dog													
	AB	AP	BH	Bea	Bo	Ch	EC	ES	GS	GP	Ha	JC	dog	all
Precision	0.589	0.444	0.818	0.667	0.643	0.842	0.571	0.400	0.700	0.717	0.772	0.974	0.678	0.693
Recall	0.660	0.640	0.180	0.200	0.180	0.640	0.320	0.400	0.560	0.860	0.880	0.760	0.523	0.572
F値	0.622	0.524	0.295	0.308	0.281	0.727	0.410	0.400	0.622	0.782	0.822	0.854	0.591	0.627

表 4.4 提案手法 1 の実験結果

	cat												
	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
Precision	0.786	0.675	0.769	0.845	0.860	0.917	0.841	0.917	0.868	0.889	0.816	0.696	0.823
Recall	0.863	0.519	0.800	0.980	0.860	0.660	0.740	0.880	0.660	0.800	0.800	0.780	0.779
F値	0.823	0.587	0.784	0.908	0.860	0.768	0.787	0.898	0.750	0.842	0.808	0.736	0.800

	dog													
	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog	all
Precision	0.800	0.698	0.646	0.615	0.756	0.702	0.549	0.647	0.795	0.778	0.875	0.930	0.733	0.778
Recall	0.720	0.600	0.620	0.480	0.680	0.660	0.560	0.660	0.700	0.840	0.840	0.800	0.680	0.729
F値	0.758	0.645	0.633	0.539	0.716	0.680	0.554	0.653	0.744	0.808	0.857	0.860	0.705	0.753

4.2.3　考察

実験の結果，従来手法では Precision が 0.693，Recall が 0.572，F 値が 0.627 となった．一方提案手法 1 では，Precision が 0.778，Recall が 0.729，F 値が 0.753 となった．よって，提案手法 1 が従来手法より高い性能を示すことが確認できた．特に Recall の向上が大きいことから，提案手法 1 は従来手法では検出することができなかった動物の顔を検出できるようになったのだと考えられる．

従来手法と提案手法 1 により検出された物体数を表 4.5 に示す．表 4.5 から，提案手法 1 は従来手法よりもペットの顔を 114 個多く検出していることがわかる．この検出数の差が，従来手法と提案手法 1 の性能差の一因であると考えられる．提案手法 1 の YOLO では猫か犬かの 2 クラス識別を行うが，従来手法の YOLO では猫 12 品種，犬 12 品種の計 24 クラス識別を行う．識別するクラス数の多い従来手法の YOLO の方が識別は難しく，識別精度は低くなる．YOLO では，検出候補領域から最終検出領域を決定する際に，物体の存在確率$P(object)$と物体のクラス確率$P(\mathrm{class}|object)$を掛け合わせ，閾値以上となったものを物体領域として決定する．識別するクラス数の多い従来手法の YOLO の場合，このクラス確率$P(\mathrm{class}|object)$の精度が低くなるため，誤って小さな値を取ってしまい，$P(object)$の値が大きくても物体として検出されないという現象が起きてしまう．この結果，従来手法と提案手法 1 の検出数の差が生じたと考えられる．

表 4.5　検出した物体数

	従来手法			提案手法1		
	cat	dog	all	cat	dog	all
検出した物体数	536	480	1016	572	558	1130
正解数	373	314	687	469	408	877
正解テストデータ数	603	600	1203	603	600	1203

4.3 DCGAN により生成された画像

DCGAN により生成された画像の例を図 4.2, 4.3 に示す．生成された画像には，比較的動物の顔の形を保った品質の良い画像と，輪郭の形が崩れていたり，ぼやけていたり，ただの毛並みのテクスチャのようなものになったりしている品質の悪い画像が見られた．これは DCGAN の学習枚数不足が原因だと考えられる．



図 4.2　DCGAN の生成画像例(猫)



図 4.3　DCGAN の生成画像例(犬)

4.4 実験 2：提案手法 2 の評価実験 1 (生成画像付加の有無)

4.4.1 実験概要

実験 2 では，提案手法 2 である DCGAN を用いたデータオーギュメンテーションの評価実験として，DCGAN の生成画像を CNN の学習データセットに加えることが CNN の性能に与える影響の調査を行った．実験としては，生成画像を CNN の学習データセットに加えない場合（以下，従来手法)と，品質をランダムに一品種あたり 100 枚の生成画像を学習データセットに加えた場合（以下，提案手法 2)の CNN の性能の比較を行った．

4.4.2 実験結果

猫の実験結果を表 4.6 に，犬の実験結果を表 4.7 にそれぞれ示す．

表 4.6 実験 2 の実験結果(猫)

	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
従来手法	0.902	0.750	0.840	1.000	0.880	0.820	0.840	0.960	0.720	0.900	0.820	0.780	0.851
提案手法	0.902	0.750	0.840	0.980	0.860	0.800	0.880	0.900	0.780	0.860	0.860	0.860	0.856

表 4.7 実験 2 の実験結果(犬)

	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog
従来手法	0.800	0.620	0.820	0.600	0.900	0.940	0.840	0.660	0.840	0.960	0.840	0.860	0.807
提案手法	0.800	0.660	0.900	0.520	0.900	0.940	0.780	0.720	0.860	0.940	0.900	0.820	0.812

4.4.3　考察

実験の結果，猫の場合，従来手法の accuracy が 0.851，　提案手法 2 の accuracy が 0.856，犬の場合，従来手法の accuracy が 0.807，　提案手法 2 の accuracy が 0.812 と猫・犬どちらの場合も提案手法 2 が高い性能を示した．よって，提案手法 2 の有効性が確認できた．

4.5 実験 3：提案手法 2 の評価実験 2 (生成画像品質の与える影響)

4.5.1 実験概要

実験 3 では，提案手法 2 である DCGAN を用いたデータオーギュメンテーションの評価実験として，DCGAN の生成画像の品質が CNN の性能に与える影響の調査を行った．品質が良いものだけ，品質が悪いものだけ，品質をランダムに一品種あたり 100 枚の生成画像を学習データセットに加えた際の CNN の性能の比較を行った．本実験では，画像品質の評価指標として評価指標 1：DCGAN の Discriminator の出力を使用した．

4.5.2 実験結果

猫の実験結果を表 4.8 に，犬の実験結果を表 4.9 にそれぞれ示す．

表 4.8 実験 3 の実験結果(猫)

	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
従来手法	0.902	0.750	0.840	1.000	0.880	0.820	0.840	0.960	0.720	0.900	0.820	0.780	0.851
提案手法(品質ランダム)	0.902	0.750	0.840	0.980	0.860	0.800	0.880	0.900	0.780	0.860	0.860	0.860	0.856
提案手法(品質良い)	0.922	0.769	0.840	1.000	0.900	0.740	0.920	0.940	0.720	0.880	0.840	0.880	0.862
提案手法(品質悪い)	0.882	0.788	0.860	1.000	0.960	0.720	0.900	0.940	0.700	0.860	0.860	0.820	0.857

表 4.9 実験 3 の実験結果(犬)

	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog
従来手法	0.800	0.620	0.820	0.600	0.900	0.940	0.840	0.660	0.840	0.960	0.840	0.860	0.807
提案手法(品質ランダム)	0.800	0.660	0.900	0.520	0.900	0.940	0.780	0.720	0.860	0.940	0.900	0.820	0.812
提案手法(品質良い)	0.760	0.680	0.900	0.680	0.920	0.940	0.800	0.680	0.840	0.920	0.860	0.860	0.820
提案手法(品質悪い)	0.820	0.740	0.820	0.600	0.920	0.900	0.780	0.640	0.780	0.900	0.900	0.880	0.807

4.5.3　考察

実験の結果，猫の場合，従来手法の accuracy が 0.851，提案手法 2 で品質ランダムに画像を追加した場合の accuracy が 0.856, 品質の良い画像のみを追加した場合の accuracy が 0.862，品質の悪い画像のみを追加した場合の accuracy が 0.857 となった．犬の場合，従来手法の accuracy が 0.807，提案手法 2 で品質ランダムに画像を追加した場合の accuracy が 0.812，品質の良い画像のみを追加した場合の accuracy が 0.820，品質の悪い画像のみを追加した場合の accuracy が 0.807 となった．実験結果から，猫・犬どちらの場合も品質の良い画像のみを追加する方が高い性能を発揮することが確認できた．

評価指標 1 で分類した，品質の良い画像の例を図 4.4 に，品質の悪い画像の例を図 4.5 に示す．評価指標 1 で分類した品質の良い画像には，人の目で見て動物の顔に見える画像が，品質の悪い画像には，動物の顔に見えないような画像が多く入っていることが確認できた．品質の悪い画像は毛の色や模様の情報がある程度含まれているものの，輪郭等の情報も含むような品質の良い画像には特徴の情報量が及ばないため，品質の良い画像のみを追加する方が高い性能を発揮したと考えられる．



図 4.4　品質の良い画像の例



図 4.5　品質の悪い画像の例

4.6 実験 4：提案手法 2 の評価実験 3 (品質評価指標)

4.6.1 実験概要

実験 3 より，品質の良い画像を加える方が高い性能を示すことがわかった．実験 4 では，提案手法 2 である DCGAN を用いたデータオーギュメンテーションの評価実験として，以下の三つの画像品質の評価指標の評価実験を行った．三つの評価指標に基づき選び出された品質の良い画像を，一品種あたり 100 枚ずつ加えた際の CNN の性能の比較を行った．

評価指標 1：DCGAN の Discriminator の出力
評価指標 2：事前学習した品種識別用 CNN の出力
評価指標 3：事前学習した品種識別用 CNN の出力から求める Inception score

4.6.2 実験結果

猫の実験結果を表 4.10 に，犬の実験結果を表 4.11 にそれぞれ示す．

表 4.10 実験 4 の実験結果(猫)

	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
評価指標1	0.922	0.769	0.840	1.000	0.900	0.740	0.920	0.940	0.720	0.880	0.840	0.880	0.862
評価指標2	0.882	0.808	0.800	1.000	0.880	0.780	0.900	0.900	0.680	0.900	0.800	0.860	0.849
評価指標3	0.882	0.769	0.860	0.980	0.900	0.860	0.860	0.960	0.740	0.920	0.820	0.800	0.862

表 4.11 実験 4 の実験結果(犬)

	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog
評価指標1	0.760	0.680	0.900	0.680	0.920	0.940	0.800	0.680	0.840	0.920	0.860	0.860	0.820
評価指標2	0.760	0.660	0.820	0.620	0.880	0.940	0.800	0.720	0.820	0.900	0.840	0.880	0.803
評価指標3	0.740	0.660	0.860	0.640	0.880	0.940	0.860	0.660	0.820	0.940	0.900	0.920	0.818

4.6.3　考察

実験の結果，猫の場合は評価指標 1 を用いた場合 accuracy が 0.862，評価指標 2 を用いた場合 accuracy が 0.849，評価指標 3 を用いた場合 accuracy が 0.862 となった．犬の場合は評価指標 1 を用いた場合 accuracy が 0.820, 評価指標 2 を用いた場合 accuracy が 0.803, 評価指標 3 を用いた場合 accuracy が 0.818 となった．つまり，猫では評価指標 1 及び評価指標 3 を用いた場合が, 犬では評価指標 1 を用いた場合が最も高い性能を示した．よって，ペット画像を対象とした場合は，評価指標 1 である DCGAN の Discriminator の出力を画像品質の評価指標として用いるのが妥当だと考えられる．

4.7 実験 5：提案手法 2 の評価実験 4 (追加画像枚数)

4.7.1 実験概要

実験 5 では，提案手法 2 である DCGAN を用いたデータオーギュメンテーションの評価実験として，加える DCGAN の生成画像の枚数が CNN の性能に与える影響の調査を行った．一品種あたりに加える DCGAN の生成画像の枚数を 0, 50, 100, 200, 300 枚と変えて CNN の学習を行い，それぞれの CNN の性能の比較を行った．加える画像には，評価指標 1 に基づき選び出された品質の良い画像のみを使用した．

4.7.2 実験結果

猫の実験結果を表 4.12 に，犬の実験結果を表 4.1 にそれぞれ示す．また，結果をグラフ化したものを図 4.6, 4.7 に示す．

表 4.12 実験 5 の実験結果(猫)

追加枚数	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
0	0.902	0.750	0.840	1.000	0.880	0.820	0.840	0.960	0.720	0.900	0.820	0.780	0.851
25	0.863	0.808	0.800	1.000	0.940	0.760	0.920	0.960	0.680	0.920	0.800	0.880	0.861
50	0.843	0.846	0.840	1.000	0.900	0.740	0.920	0.920	0.700	0.920	0.820	0.900	0.862
100	0.922	0.769	0.840	1.000	0.900	0.740	0.920	0.940	0.720	0.880	0.840	0.880	0.862
150	0.804	0.827	0.900	1.000	0.940	0.720	0.900	0.940	0.680	0.860	0.800	0.880	0.854
200	0.843	0.750	0.840	1.000	0.900	0.760	0.880	0.900	0.680	0.880	0.800	0.840	0.839
250	0.902	0.769	0.840	1.000	0.880	0.780	0.900	0.900	0.780	0.820	0.820	0.840	0.852
300	0.843	0.788	0.900	1.000	0.900	0.780	0.900	0.920	0.760	0.900	0.780	0.780	0.854

表 4.13 実験 5 の実験結果(犬)

追加枚数	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog
0	0.800	0.620	0.820	0.600	0.900	0.940	0.840	0.660	0.840	0.960	0.840	0.860	0.807
25	0.720	0.680	0.860	0.680	0.900	0.920	0.800	0.640	0.820	0.940	0.880	0.940	0.815
50	0.760	0.680	0.860	0.640	0.920	0.920	0.780	0.700	0.840	0.900	0.880	0.900	0.815
100	0.760	0.680	0.900	0.680	0.920	0.940	0.800	0.680	0.840	0.920	0.860	0.860	0.820
150	0.720	0.660	0.840	0.660	0.900	0.940	0.780	0.700	0.860	0.940	0.840	0.900	0.812
200	0.760	0.700	0.900	0.660	0.920	0.900	0.840	0.680	0.800	0.900	0.840	0.900	0.817
250	0.720	0.640	0.880	0.700	0.920	0.940	0.700	0.720	0.800	0.900	0.860	0.840	0.802
300	0.780	0.660	0.820	0.640	0.880	0.940	0.800	0.720	0.800	0.920	0.860	0.820	0.803

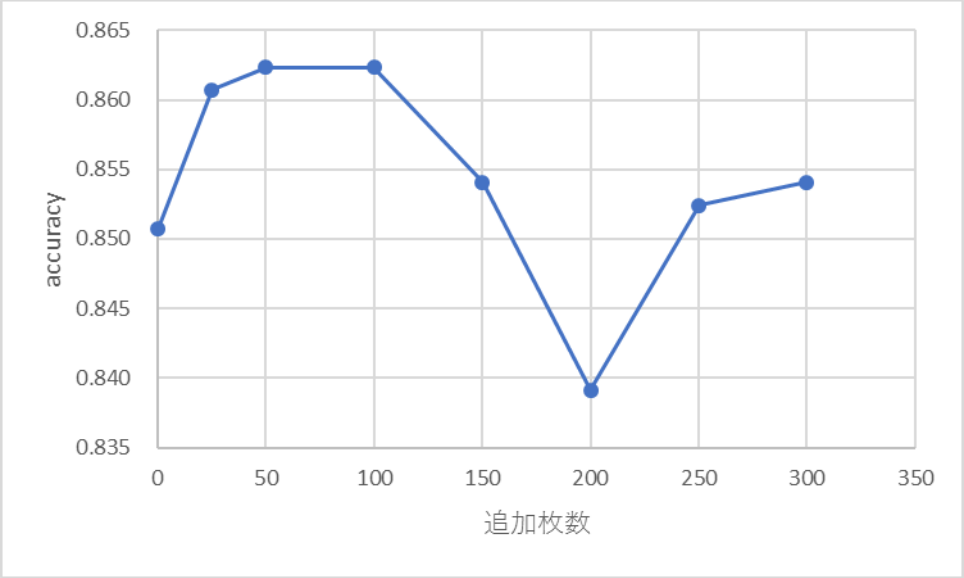


図 4.6　実験 5 の実験結果(猫)のグラフ

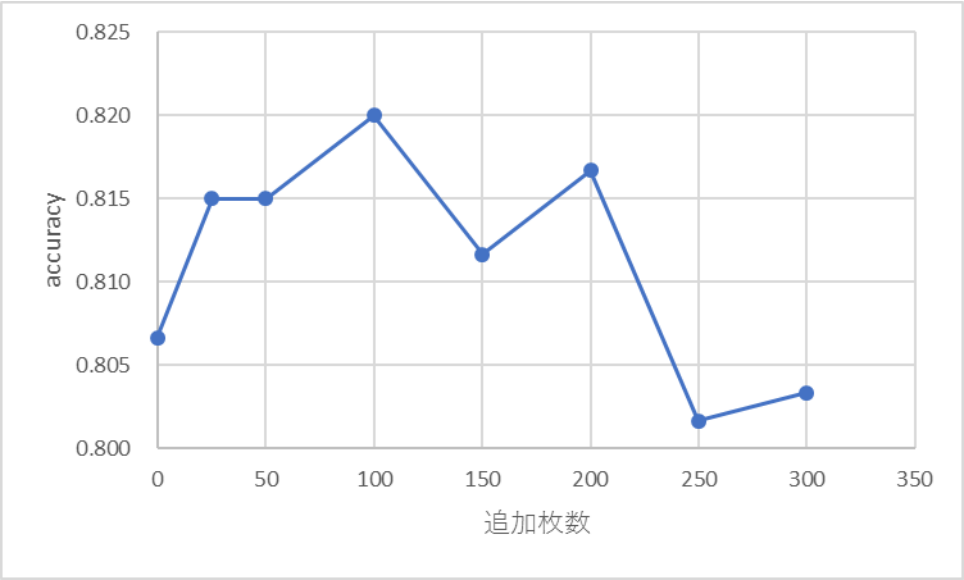


図 4.7　実験 5 の実験結果(犬)のグラフ

4.7.3　考察

実験の結果, 猫の場合は追加枚数が 50, 100 枚の時が最も性能が高く, accuracy が 0.862, 犬の場合は追加枚数が 100 枚の時が最も性能が高く, accuracy が 0.820 となった. よって, 本実験のデータセットにおいては，猫・犬どちらの場合も 1 クラスあたり 100 枚の生成画像を追加するのが最も良いということがわかった.

猫・犬どちらの場合も追加画像枚数を多くすると，性能が下がってしまった．この原因について，次のように考察する．DCGAN により生成される類似画像は，評価指標 1 により品質が良いと判断された画像であっても元々の学習画像に比べると品質が低い．また，評価指標 1 の値が高い順に追加しているため，追加枚数が増えるにつれて，追加する画像の品質は低くなっていく．そのため，追加枚数が多くなると，学習データセットを占める類似画像の割合が多くなり，データセット全体の質が少しずつ悪くなっていく．その結果，少量の追加の場合は CNN の学習を補っていたのが，だんだんと類似画像が足枷となっていき，性能が低くなっていってしまうと考えられる．そのため，提案手法 2 の DCGAN を用いたデータオーギュメンテーションにおいては，追加する画像枚数をしっかりと考慮し，選択する必要があると考えられる．

4.8 実験 6：提案手法 3 の評価実験

4.8.1 実験概要

実験 6 では，提案手法 1 である YOLO と CNN を組み合わせた二段階手法と提案手法 2 である DCGAN を用いたデータオーギュメンテーションを組み合わせた提案手法 3 の評価実験を行った．YOLO 単体で猫・犬の顔の検出と品種識別を行う一段階の手法(以下，従来手法)と提案手法 1，そして提案手法 3 の性能の比較を行った．実験 2~5 の結果を踏まえ，提案手法 2 として画像の品質評価指標 1 に基づき選び出された品質の良い画像を 1 クラスあたり 100 枚追加して実験を行った．

4.8.2 実験結果

提案手法 3 の実験結果を表 1.14 に示す．また，従来手法，提案手法 1，提案手法 3 の結果をまとめたものを表 4.15 に示す．

表 4.14 提案手法 3 の結果

	cat												
	Ab	Ben	Bi	Bom	BS	EM	MC	Pe	Ra	RB	Si	Sp	cat
Precision	0.768	0.730	0.818	0.845	0.885	0.971	0.837	0.932	0.844	0.976	0.843	0.771	0.852
Recall	0.843	0.519	0.900	0.980	0.920	0.680	0.820	0.820	0.760	0.820	0.860	0.740	0.805
F値	0.804	0.607	0.857	0.907	0.902	0.800	0.828	0.872	0.800	0.891	0.851	0.755	0.828

	dog													
	AB	AP	BH	Bea	Box	Ch	EC	ES	GS	GP	Ha	JC	dog	all
Precision	0.786	0.651	0.653	0.649	0.791	0.655	0.605	0.632	0.745	0.804	0.891	0.933	0.733	0.792
Recall	0.660	0.560	0.640	0.480	0.680	0.720	0.520	0.720	0.700	0.820	0.820	0.840	0.680	0.743
F値	0.717	0.602	0.646	0.552	0.731	0.686	0.559	0.673	0.722	0.812	0.854	0.884	0.705	0.767

表 4.15 実験結果のまとめ

	Precision	Recall	F値
従来手法	0.693	0.572	0.627
提案手法1	0.778	0.729	0.753
提案手法3	0.792	0.743	0.767

4.8.3　考察

実験の結果，提案手法 3 では Precision が 0.792，Recall が 0.743，F 値が 0.767 となった．従来手法と提案手法 1，提案手法 3 の結果を比較すると，提案手法 3 が最も高い性能を発揮し，従来手法と比較すると Precision で 0.099，Recall で 0.171，F 値で 0.14 の性能向上が確認できた．よって、提案手法の有効性を確認することができた。

第5章
結論

本研究では，代表的なペットである猫・犬の顔検出及び品種識別を行う手法を提案した．YOLO で猫・犬の顔領域を検出し，その後検出した顔領域を猫・犬それぞれに対応した品種識別用 CNN で品種識別を行うという 2 段階の検出手法と，品種識別用 CNN を学習させる際に，DCGAN を用いて学習用データセットから猫・犬の類似画像を生成し，生成した類似画像を CNN の学習データセットに加えるというデータオーギュメンテーションの二つの手法を提案し，猫・犬の顔検出及び品種識別の性能向上を図った．実験の結果，提案手法は従来手法と比べて Precision で 0.099，Recall で 0.171，F 値で 0.14 の性能向上を達成し，提案手法の有効性を確認することが出来た．また，提案手法 2 の DCGAN を用いたデータオーギュメンテーションにおいて，追加する画像の品質，画像の枚数を考慮し，品質の良い画像を適切な枚数追加することでより性能が向上することがわかった．

謝辞

本研究の実験環境を与えてくださり，研究の方向性等の丁寧かつ熱心なご指導を頂いた渡辺教授に心から感謝いたします．また，日頃から御意見やアドバイスをくださった研究室の皆様に御礼申し上げます．

参考文献

[1] 内閣府, “消費動向調査 主要耐久消費財等の普及・保有状況(総世帯),” 2017 年 3 月.

[2] 総務省, “平成 28 年通信利用動向調査,” 2017.

[3] MMD 研究所, “スマートフォンカメラの利用に関する調査,” https://mmdlabo.jp/investigation/detail_1578.html, 2016.

[4] H. Watanabe, H. Zhang, “Cat and Dog Face Detection Using Deformable Part Model,” PCSJ/IMPS, 2015.

[5] A. Krizhevsky, I. Sutskever, G. F. Hinton, “ImageNet Classification with Deep Convolutional,” Neural Information Processing Systeems (NIPS), 2012.

[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified Real-Time Object Detection,” CVPR, 2016.

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, “SSD: Single Shot MultiBox Detector,” European conference on computer vision, 2016.

[8] A. Radford, L. Metz, S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” ICLR, 2016.

[9] 岡谷貴之, 機械学習プロフェッショナルシリーズ 深層学習, 講談社, 2015.

[10] S. Iizuka, E. Simo-Serra, H. Ishikawa, “Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification,” ACM Transactions on Graphics (TOG), 2016.

[11] D. Silver, "Mastering the game of go without human knowledge," *Nature,* vol. 550, p. 354, 2017.

[12] “村上研究室 コラム/ニューラルネットワーク,” http://ipr20.cs.ehime-u.ac.jp/column/neural/.

[13] 中山英樹, “画像認識分野における deep learning の発展と最新動向,” http://www.nlab.ci.i.u-tokyo.ac.jp/pdf/asj20141215.pdf#search=%27%E7%94%BB%E5%83%8F%E8%AA%8D%E8%AD%98%E5%88%86%E9%87%8E%E3%81%AB%E3%81%8A%E3%81%91%E3%82%8Bdeep+learning%27.

[14] 石井智, 望月義, 小山田雄, 石川博, “Convolutional Neural Network を用いた一般物体認識手法の解析,” 情報処理学会研究報告, 2014.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Networks,” Neural Information Processing Systeems (NIPS), 2014.

[16] S. Yuki, “はじめての GAN,” https://elix-tech.github.io/ja/2017/02/06/gan.html.

[17] “ニューラルネットワークにおける Deconvolution,” https://qiita.com/shngt/items/9c86e69e16ce6d61a0c6.

[18] M. Lin, Q. Chen, S. Yan, “Network In Network,” ICLR, 2014.

[19] S. Ioffe, C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” ICLR, 2015.

[20] Batch Normalization：ニューラルネットワークの学習を加速させる汎用的で強力な手法,” https://deepage.net/deep_learning/2016/10/26/batch_normalization.html.

[21] O. M. Parkhi, A. Vedaldi, A. Zisserman, C. V. Jawahar, “Cats and Dogs,” CVPR, 2012.

[22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, “Improved techniques for training GANs,” Neural Information Processing Systeems (NIPS), 2016.

図一覧

図 2.1　ユニットの基本構造 4
図 2.2　順伝播型ニューラルネットワークの基本構造 4
図 2.3　畳み込みニューラルネットワークの基本構造 5
図 2.4　畳み込み層の概要(入力画像 3 チャネル，フィルタ 4 個) 6
図 2.5　プーリング実行例(最大プーリング) 7
図 2.6　全結合層の概要図 8
図 2.7　GAN の概念図 10
図 2.8　fractional-strided convolution の概要図 11
図 2.9　ReLU, Leaky ReLU のグラフ 12
図 3.1　提案手法概要図（”The Oxford-IIIT-Pet Dataset” [21]) 13
図 3.2　提案手法 1 の概要図（”The Oxford-IIIT-Pet Dataset” [21]) 14
図 3.3　提案手法 2 の概要図 15
図 3.4　理想的な確率分布の例 17
図 4.1　使用した品種の例（”The Oxford-IIIT-Pet Dataset” [21]) 20
図 4.2　DCGAN の生成画像例(猫) 24
図 4.3　DCGAN の生成画像例(犬) 25
図 4.4　品質の良い画像の例 30
図 4.5　品質の悪い画像の例 31
図 4.6　実験 5 の実験結果(猫)のグラフ 35
図 4.7　実験 5 の実験結果(犬)のグラフ 35

表一覧

表 4.1　使用した品種の一覧 19
表 4.2　スレットスコア 21
表 4.3　従来手法の実験結果 22
表 4.4　提案手法 1 の実験結果 22
表 4.5　検出した物体数 23
表 4.6　実験 2 の実験結果(猫) 26
表 4.7　実験 2 の実験結果(犬) 26
表 4.8　実験 3 の実験結果(猫) 28
表 4.9　実験 3 の実験結果(犬) 28
表 4.10　実験 4 の実験結果(猫) 32
表 4.11　実験 4 の実験結果(犬) 32
表 4.12　実験 5 の実験結果(猫) 34
表 4.13　実験 5 の実験結果(犬) 34
表 4.14　提案手法 3 の結果 37
表 4.15　実験結果のまとめ 37

研究業績

渡部，渡辺: “DCGAN を用いたデータオーギュメンテーションによる猫の品種識別について“, 2016 映像情報メディア学会年次大会, No.24D-5, Sep. 2016

渡部，渡辺: “DCGAN による生成画像品質の猫品種識別性能に及ぼす影響について”, PCSJ/IMPS 2016, No. P-4-16, Nov. 2016

H. Watabe and H. Watanabe: “Data Augmentation Using DCGAN for Breed Identification,” International Workshop on Image Electronics and Visual Computing 2017 (IEVC2017), No.4C-3, DaNang, Vietnam, Mar. 2017

渡部，渡辺: “生成画像品質を考慮した DCGAN を用いたデータオーギュメンテーション”, FIT2017(第 16 回情報科学技術フォーラム)，H-038, Sep. 2017

渡部，渡辺 : ”多段ニューラルネットワークを用いた動物の品種識別“，映像情報メディア学会冬季大会，13B-1，Dec. 2017