

卒業論文概要書

Summary of Bachelor's Thesis

学科名 Department	情報理工	氏名 Name	渡部宏樹	指導員 Advisor	渡辺 裕 印 Seal
研究指導 Research guidance	オーディオビジュアル情報処理研究	学籍番号 Student ID number	1W120612-1 CD		
研究題目 Title	スタイル行列と畳み込みニューラルネットワークを用いた猫顔画像識別 Cat Face Recognition using Style Representation and Convolutional Neural Network				

1. まえがき

近年、カメラの普及率は高まっており、カメラ所有者の約2割がペットを主な被写体としている[1][2]。また、SNSの流行に伴いインターネット上のペット画像の量は爆発的に増えている。そのため、動物を認識する技術の重要性は高まっているが、動物画像の認識は人の画像に比べて困難であり、開発途上であるのが現状である。猫の画像を対象とする研究は、検出に関する研究は成果を上げている[3][4]が、認識に関する研究はあまり進められていない。そこで、本論文では猫の顔画像を対象とした品種識別精度の向上を目指した。

2. 動物の顔画像による品種識別

本論文では、スタイル行列を用いた事前分類法と、畳み込みニューラルネットワークによる分類を組み合わせた手法を提案した。画像をまずスタイル行列に基づいて分類し、その分類したクラスターに適した畳み込みニューラルネットワークで識別を行う手法である。スタイル行列とは、畳み込みニューラルネットワークのチャンネル間の特徴マップの相関を計算したもので、色合いやテクスチャ等の画風を表す特徴量である[5]。これは、猫の顔画像においては毛並みを表すと考えられる。

3. 実験結果評価

実験では、まず提案手法で用いたネットワークがスタイル行列で分類したクラスターに適した学習を行えているかを調査した。その結果、適切な学習が行えていることがわかった。次に、提案手法と従来手法の平均識別率の比較を行った。その結果、品種によっては精度の向上が見られたものの、全品種の平均識別率を比べると、提案手法が79.2%、従来手法が80.7%と、従来手法の方が高い性能を示した。

4. まとめ

本研究では、猫の顔画像を対象として、スタイル行列を用いた事前分類法および畳み込みニューラルネットワークによる分類を組み合わせて品種識別を行う手法を提案すると共に、実験により手法の評価を行った。評価実験を行ったところ、従来手法の方が高い性能を示した。今後、畳み込みニューラルネットワークの構成の改良やパラメータ調整により、提案手法の精度を向上させることができると考える。

参考文献

[1]内閣府, 主要耐久消費財等の普及率(一般世帯)(平成27年(2015年)3月現在),

<http://www.esri.cao.go.jp/jp/stat/shouhi/2015/201503fukyuritsu.xls>

[2]リサーチバンク, 写真に関する調査, http://research.lifemedia.jp/2014/07/140730_photo.html

[3]草野孝幸, 出口大輔, 井手一郎, 村瀬洋, 猫パーツの抽出とその組み合わせによる猫の顔検出の高精度化, DIA2010 動的画像処理実利用化ワークショップ講演概要, 2010

[4]小坂谷達夫, 伊藤聡, 窪田進, 2種類の異なる特徴を用いた特定物体検出, 東芝レビュー Vol.65 No.7, pp.33-36, 2010

[5] Leon A.Gatys, Alexander S.Ecker, Matthias Bethge, "A Neural Algorithm of Artistic Style", 2015

2015 年度 卒業論文

スタイル行列と畳み込みニューラルネットワークを用いた

猫顔画像識別

Cat Face Recognition using Style Representation and
Convolutional Neural Network

指導教員 渡辺 裕 教授

早稲田大学 基幹理工学部

情報理工学科

1W120612-1

渡部宏樹

目次

第1章	序論	1
1.1	研究の背景	1
1.2	本研究の目的	1
1.3	関連研究	1
1.4	論文の構成	1
第2章	ニューラルネットワークによる画像認識	3
2.1	はじめに	3
2.2	ディープラーニング	3
2.2.1	ニューラルネットワーク	3
2.2.1.1	ユニット	3
2.2.1.2	順伝播型ニューラルネットワーク	4
2.2.2	畳み込みニューラルネットワーク	5
2.2.2.1	畳み込みニューラルネットワークの基本構造	5
2.2.2.2	畳み込み層	6
2.2.2.2.1	畳み込み	6
2.2.2.2.2	パディング	6
2.2.2.2.3	ストライド	7
2.2.2.2.4	畳み込み層	7
2.2.2.3	プーリング層	8
2.2.2.4	全結合層	9
2.2.2.5	ソフトマックス層	9
2.2.2.6	ドロップアウト	10
2.2.2.7	ReLU(Rectified Linear Unit)	10
2.2.2.8	Augmentation	10
2.3	スタイル行列	10
2.4	K-means 法	11
2.5	むすび	11
第3章	動物の顔画像による品種識別	12
3.1	はじめに	12
3.2	提案手法の概要	12
3.3	学習処理	13
3.4	識別処理	13
3.5	むすび	14
第4章	提案手法の評価実験と結果, 考察	15

4.1	はじめに-----	15
4.2	実験の概要-----	15
4.2.1	実験 1-----	15
4.2.2	実験 2-----	15
4.3	実験の条件-----	15
4.3.1	実験環境-----	15
4.3.2	畳み込みニューラルネットワークの構成-----	16
4.3.3	K-means 法のクラスター数 K -----	17
4.3.4	スタイル行列クラスタリングの結果-----	18
4.3.5	画像データセット-----	18
4.4	実験 1 の結果-----	22
4.5	実験 1 の考察-----	23
4.6	実験 2 の結果-----	24
4.7	実験 2 の考察-----	24
第5章	結論と今後の課題	26
5.1	結論-----	26
5.2	今後の課題-----	26
	謝辞	27
	参考文献	28
	図一覧	29
	表一覧	30

第1章 序論

1.1 研究の背景

近年、カメラの普及率は大変高くなっている。内閣府の調査によると、日本では約9割の人が携帯電話を、約7割の人がデジタルカメラを所持している[1]。また、カメラを所持している人のうちの約2割はペットを主な被写体としている[2]。さらに、近年はペット監視カメラも登場しており、我々の身の回りでペットを撮影する機会が増加している。これらのことから、動物を認識する技術の重要性は高まってきている。

従来、人や標識といったものの検出・認識技術の研究は進められているが、代表的な愛玩動物である犬や猫といった特定動物の認識技術の研究はあまり行われていない。犬や猫等動物は、品種によって顔のパーツの大きさや形が異なり、人に比べて検出・認識が困難であり、開発途上であるのが現状である。本研究では、猫の顔画像を対象として、スタイル行列を用いた事前分類法および畳み込みニューラルネットワークによる分類を組み合わせて品種識別を行う手法を提案すると共に、実験により手法の評価を行う。

1.2 本研究の目的

動物の顔画像からの自動品種識別が可能になれば、インターネット上の動物の顔画像の自動識別によるタグ付けが可能となる。そうすれば、データとしての検索が容易になり、画像へのアクセス機能が大幅に向上する。

そこで本研究では、愛玩動物の一種である猫の顔画像を対象として、動物の顔画像による品種識別性能の向上を目的とする。

1.3 関連研究

猫等の動物の顔には多種多様な形状、色、模様があり、人の検出・認識手法をそのまま適用しても、人に適用する場合と比べて精度が下がってしまう。そのため、猫等の動物を対象とした研究が進められている。猫の顔検出手法の研究としては、猫の顔を特徴付けるパーツを検出し、その検出結果を基に猫の顔の検出を行う手法[3]や、2種類の異なる画像特徴量を用いることで猫の顔を検出する手法[4]等高い成果を上げている研究がある。一方で、猫の品種を識別する手法の研究はあまり進められていないのが現状である。

1.4 本論文の構成

本論文の構成は以下のとおりである。

第1章は本章であり、本研究の背景や目的について述べる。

第2章では、本論文で用いる主要な認識技術であるニューラルネットワークや特徴量についての説明や用語の定義について述べる。

第3章では、本論文で提案する手法について述べる。

第4章では、実験概要、結果および考察について述べる。

第 5 章では，本論文の結論と今後の課題について述べる．

第2章 ニューラルネットワークによる画像認識

2.1 はじめに

本章では、本論文で用いる画像認識手法や特徴量についての説明や用語の定義を行う。

2.2 ディープラーニング[5]

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。従来は人が設計した特徴量を用いて機械学習を行っていたが、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ている。近年画像認識や音声認識、自然言語処理といった分野でディープラーニングを用いた手法が高い性能を発揮している。

2.2.1 ニューラルネットワーク

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[6]。まず、ニューラルネットワークの構成要素について説明し、その後代表的なニューラルネットワークの一つである順伝播型ニューラルネットワークについて説明する。

2.2.1.1 ユニット

ニューラルネットワークはユニットと呼ばれるもので構成されている。ユニットの基本構造を図 2.1 に示す。また、総入力 u および出力 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がよく用いられる。

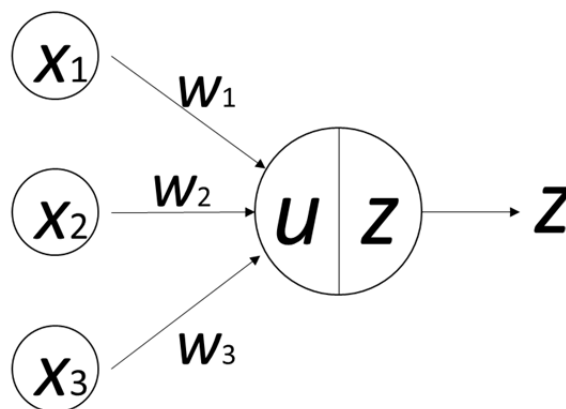


図 2.1 ユニットの基本構造

$$u = x_1 w_1 + x_2 w_2 + x_3 w_3 \quad (2.1)$$

$$z = f(u) \quad (2.2)$$

2.2.1.2 順伝播型ニューラルネットワーク

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。

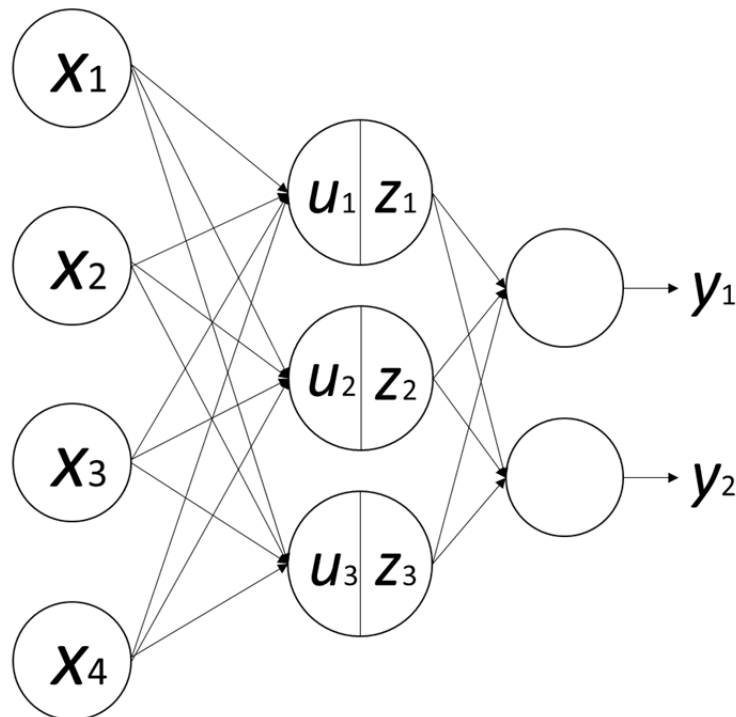


図 2.2 順伝播型ニューラルネットワークの基本構造

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を与えられており、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 および出力 z_j は次のように計算できる。

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

$$z_j = f(u_j) \quad (2.4)$$

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 \mathbf{x} と出力 \mathbf{y} を用いて関数 $\mathbf{y} = \mathbf{y}(\mathbf{x})$ 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。

2.2.2 畳み込みニューラルネットワーク

2.2.2.1 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)を用いたりもする。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。

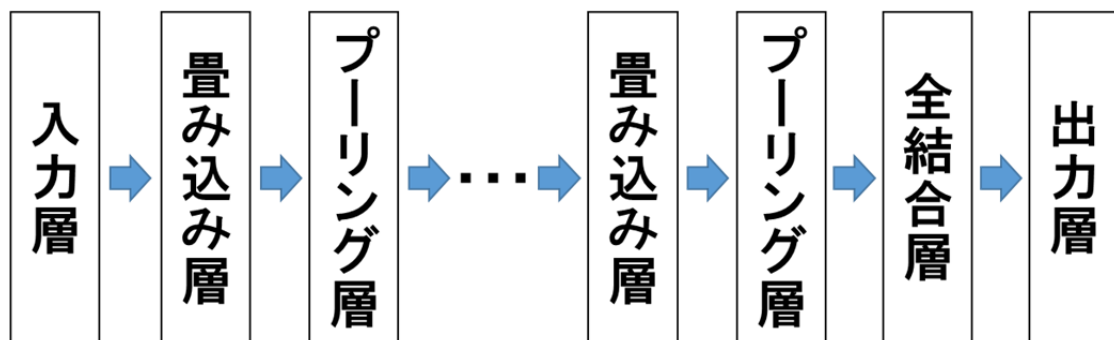


図 2.3 畳み込みニューラルネットワークの基本構造

2.2.2.2 畳み込み層

2.2.2.2.1 畳み込み

畳み込みを施す入力画像として、濃淡値を各画素に格納した画像を考える。画像の縦横サイズを $W \times W$ 、画素 (i, j) ($i = 0, 1, \dots, W - 1, j = 0, 1, \dots, W - 1$) の画素値を x_{ij} とする。また、フィルタと呼ばれるサイズの小さい画像を考える。フィルタのサイズを $H \times H$ とし、画素 (i, j) ($i = 0, 1, \dots, H - 1, j = 0, 1, \dots, H - 1$) の画素値を h_{ij} とする。画像の畳み込みは、次の計算式で表される画像とフィルタ間の積和計算である。畳み込みの例を図 2.4 に示す。

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p, j+q} h_{pq} \quad (2.5)$$

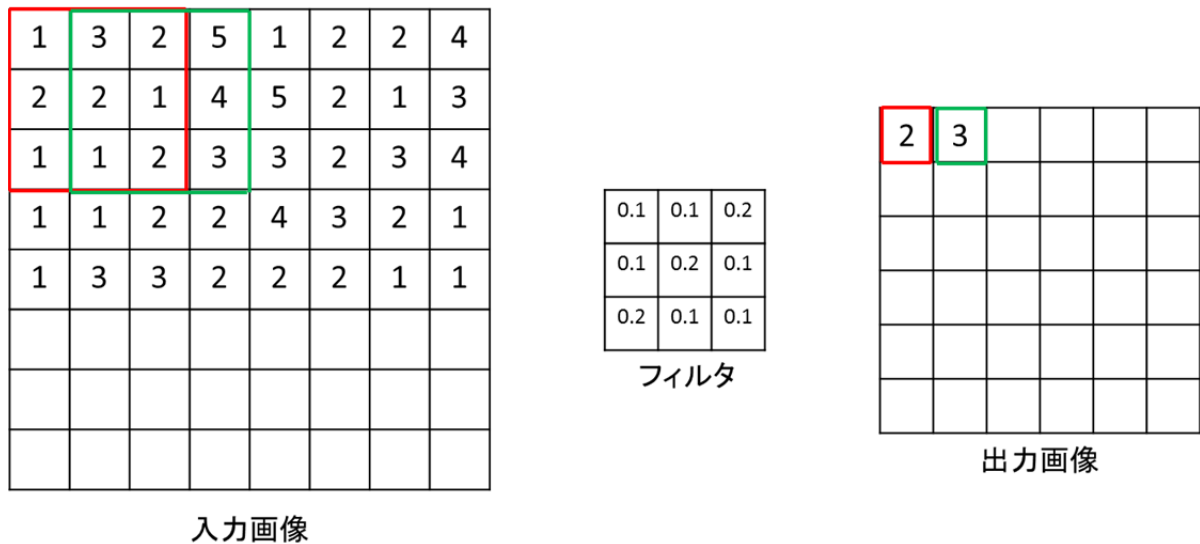


図 2.4 畳み込みの実行例

つまり、畳み込みとは画像にフィルタを重ね合わせたとき、重なり合う画素同士の積を求め、フィルタ全体でその積の和を求める計算である。これを画像全体に対して行う。

2.2.2.2.2 パディング

畳み込みは画像全体に行われるが、画像からフィルタがはみ出すような範囲には適用することができないため、畳み込み結果の画像のサイズは入力画像よりも小さなものになる。また、そのサイズは以下の式で表される。

$$(W - 2[H/2]) \times (W - 2[H/2]) \quad (2.6)$$

これを解決するために入力画像の外側に幅 $[H/2]$ のふちをつけることで、畳み込み結果の画像サイズを入力画像のサイズと同じにすることができる。この手法をパディングという。このふちの画素値は未定であるので、なんらかの方法で決める必要がある。畳み込みニュー

ーラルネットワークではふちの画素値を 0 にセットするゼロパディングが広く用いられている。

2.2.2.2.3 ストライド

画像上を、1画素ずつフィルタを動かすのではなく、複数画素ずつ動かして畳み込みの計算を行っていく場合がある。このときのフィルタを動かす画素の間隔をストライドという。ストライドを s としたとき、畳み込み結果の出力画像の画素値は以下の式で表される。

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{si+p, sj+q} h_{pq} \quad (2.7)$$

また、そのときの出力サイズは以下の式で表される。

$$(\lfloor (W-1)/s \rfloor + 1) \times (\lfloor (W-1)/s \rfloor + 1) \quad (2.8)$$

大きな画像に畳み込みを行う場合、出力ユニット数が大きくなりすぎるのを防ぐために 2 以上のストライドを用いる場合がある。ただし、ストライドを大きくしていくと、その分画像の特徴を取りこぼす可能性があるので注意が必要である。

2.2.2.2.4 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ン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。ここで多チャン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 チャンネルの画像である。各フィルタは入力画像と同じ数のチャンネル数を持つ。3 チャンネルの画像に 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 2.5 に示す。

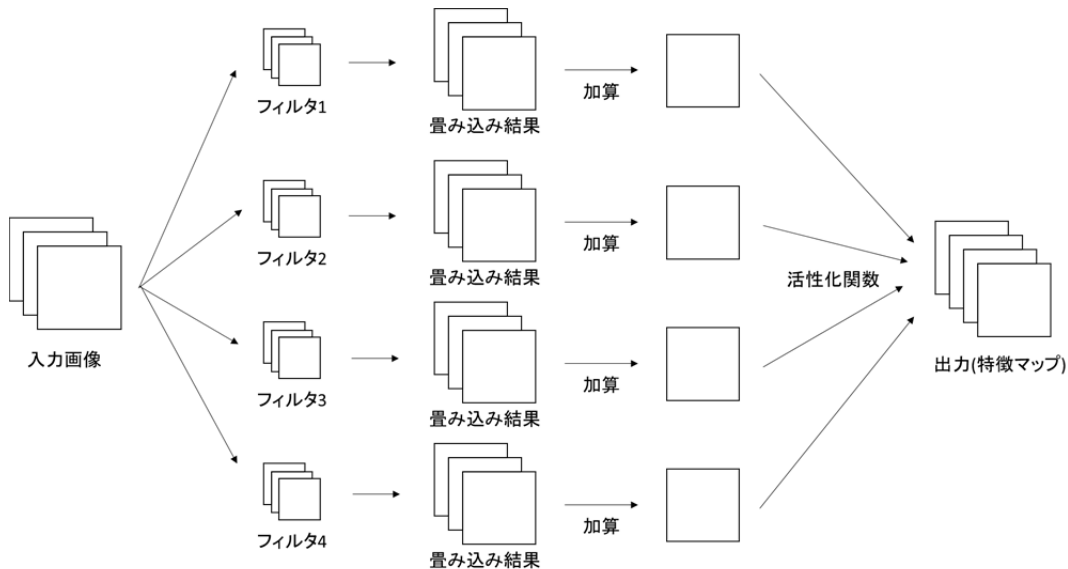


図 2.5 畳み込み層の概要(入力画像 3 チャンネル, フィルタ 4 個)

各チャンネルについて入力画像とフィルタの畳み込みを行うと、チャンネル数の同じ画像が作られる。その画像の全チャンネルの各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チャン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ン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 チャンネルの画像 $z_{ijk}^{(l-1)}$ ($k = 0, 1, \dots, K-1$)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} および畳み込み層の出力 $z_{ijm}^{(l)}$ は次のように計算できる。

$$u_{ijm} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} z_{i+p, j+q, k}^{(l-1)} h_{pqkm} + b_{ijm} \quad (2.9)$$

$$z_{ijm}^{(l)} = f(u_{ijm}) \quad (2.10)$$

2.2.2.3 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[7]。プーリングの例を図2.6に示す。

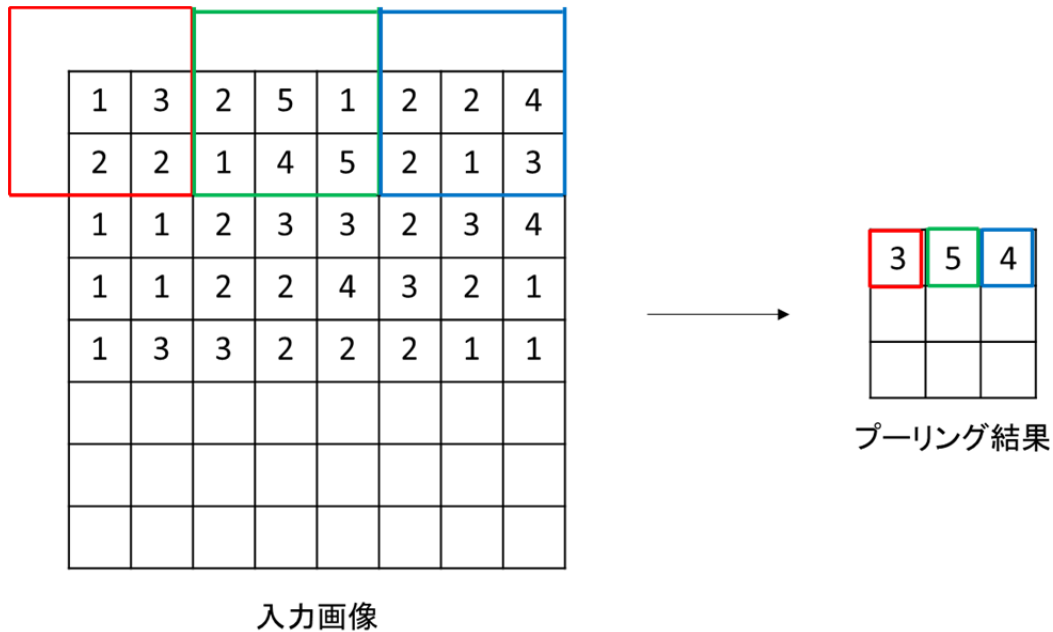


図 2.6 プーリング実行例(最大プーリング)

図 2.6 では、 3×3 のプーリング領域をストライド 3 で動かして、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。

2.2.2.4 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 2.7 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。

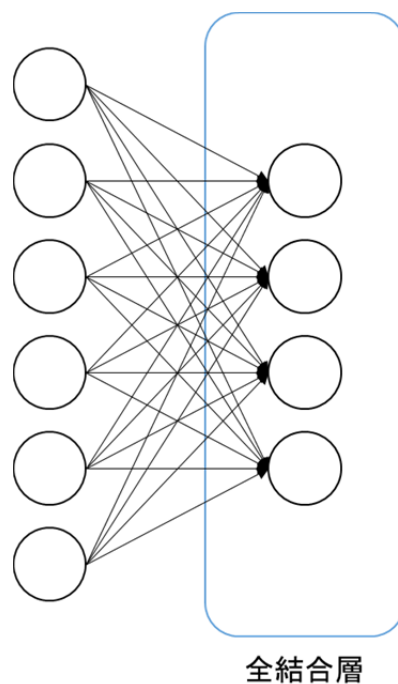


図 2.7 全結合層の概要図

2.2.2.5 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[8].

$$y_i = z_i = \frac{\exp(u_i)}{\sum_{k=1}^K \exp(u_k)} \quad (2.11)$$

このとき y_i は入力画像がクラス i に属する確率を表している。

2.2.2.6 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率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と同じ効果が得られると考えられている。

2.2.2.7 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[8]。

$$f(x) = \max(0, x) \quad (2.12)$$

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。

2.2.2.8 Augmentation

Augmentationとは、画像に加工や変形といった処理を施すことで、畳み込みニューラルネットワークで学習に用いる画像の枚数を増やす手法である。よく用いられる手法としては、画像の反転や回転、並進移動や輝度変化等がある。畳み込みニューラルネットワークは一般的に学習枚数が増えるほど精度が向上するため、このような手法を用いて大量の学習データを作成する。

2.3 スタイル行列[9]

スタイル行列(Style Representation)は、色合いやテクスチャ等の画風を表す特徴量である。畳み込みニューラルネットワークの任意層で得た各チャンネル間の特徴マップの相関を計算したもので、以下の式で求めることができる。チャンネル間の特徴マップの相関を計算していることから、入力層で計算したものは色情報の相関を、畳み込み層等の中間層で計算したものはある特徴間のある関係を表していると考えられる。またこの特徴量は、猫の顔画像においては猫の毛並みを表すと考えられる。

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (2.13)$$

2.4 K-means 法[10]

K-means 法は代表的な非階層クラスタリング手法の一つで、あるデータ集団を K 個のクラスタに分類する手法である。アルゴリズムは次の通りである。

1. データ集団をランダムに K 個のクラスタに分割する。
2. クラスタの中心を計算する。
3. 各データと各クラスタの中心との距離を求め、各データを最も近いクラスタに割り当てる。
4. 2, 3 の処理を変化がなくなるまで繰り返す。

K-means 法は単純で計算量が小さいため、広く用いられている。

2.5 むすび

本章では、本論文で用いる画像認識手法や特徴量について述べた。まず、ニューラルネットワークにおける基本的な概念や用語について説明した。次に、畳み込みネットワークの基本構造や、各層の役割について説明した。その次に、スタイル行列という特徴量について説明した。最後に K-means 法について説明した。

第 3 章では、本章で説明した手法や特徴量を用いた提案手法についての説明を行う。

第3章 動物の顔画像による品種識別

3.1 はじめに

本章では、動物の顔画像による品種識別を目的として、スタイル行列を用いた事前分類法および畳み込みニューラルネットワークによる分類を組み合わせた手法の提案し、処理の詳細について述べる。

3.2 提案手法の概要

提案手法は、スタイル行列を用いた事前分類法と、畳み込みニューラルネットワークによる分類を組み合わせたものである。事前分類で猫の毛並み情報に基づいた大まかな分類を行い、その分類したクラスに適した畳み込みニューラルネットワークでさらに詳細な品種の識別を行うことを目標としている。この手法の概要を図 3.1 に示す。提案手法では、まず入力画像のスタイル行列を取得する。次にそのスタイル行列の属するクラスターの識別を行い、そのクラスターに適した畳み込みニューラルネットワークに画像を渡す。その後、畳み込みニューラルネットワークで分類を行い、結果を出力する。学習と識別の処理について順に述べる。

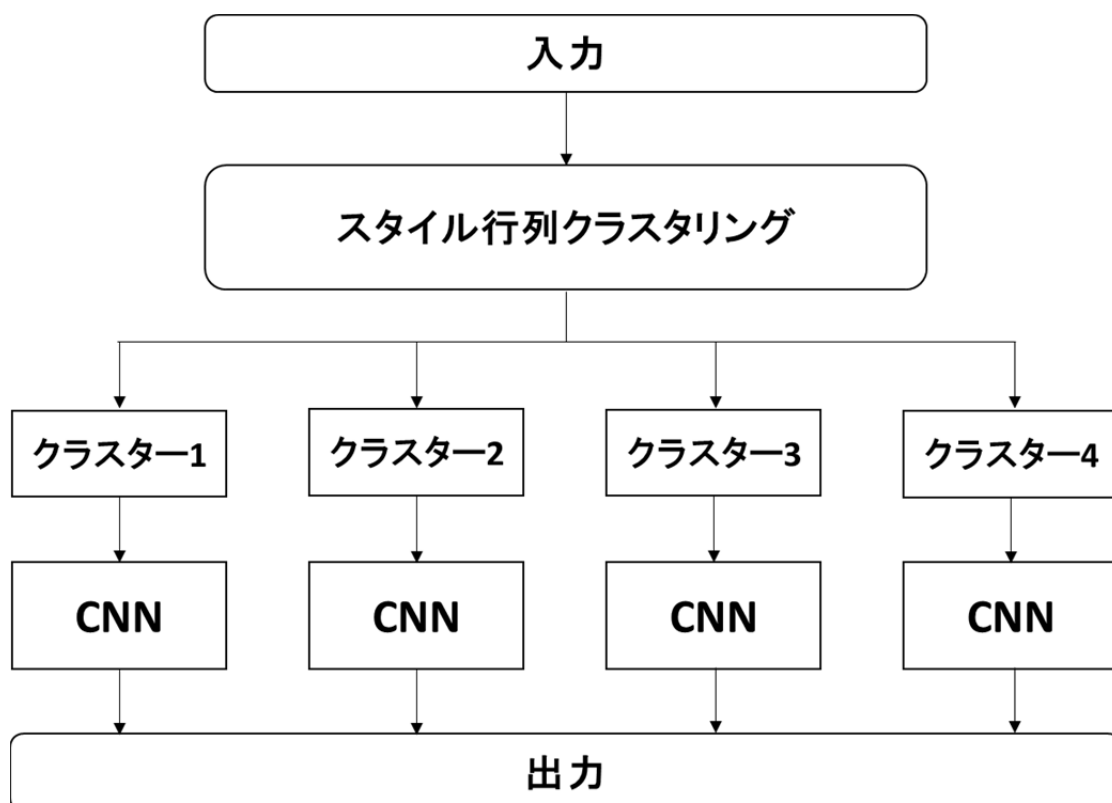


図 3.1 提案手法の概要図

3.3 学習処理

提案手法の学習は以下の処理により実行する。

1. スタイル行列を取得するための畳み込みニューラルネットワークを用意する。
2. 用意した畳み込みニューラルネットワークを用いて、学習画像から複数のスタイル行列を求める。
3. 求めたスタイル行列を結合し、ベクトル化する。
4. それらを **K-means** 法を用いてクラスタリングする。
5. クラスターごとに **Augmentation** を行い、学習セットを作る。
6. それぞれのクラスターに対応した畳み込みニューラルネットワークを学習させる。

学習の処理について以下に詳しく述べる。まず準備として、スタイル行列を計算するための学習済み畳み込みニューラルネットワークを用意する。このネットワークは後で扱うネットワークと異なる構造のものを用意する。異なる構造の複数のネットワークを用いることで、より多くの情報を用いた識別が行えるためである。次に、学習画像からスタイル行列をいくつかの層から取得し、それらを結合しベクトル化する。それを **K-means** 法を用いてクラスタリングする。この時、分類したクラスターの中心ベクトルを保持しておく。そして、分類したクラスターごとに学習データセットを作る。各クラスターの学習データセットは全ての学習画像と、そのクラスターに属する学習画像を **Augmentation** したものを組み合わせたものとする。そのクラスターに属する画像のみを **Augmentation** したものを学習することで、そのクラスターに適した畳み込みニューラルネットワークの学習が行えると考える。最後に、学習データセットを用いて各クラスターに対応した畳み込みニューラルネットワークを学習させる。

3.4 識別処理

提案手法の識別は以下の処理により実行する。

1. 識別したい画像のスタイル行列を複数取得し、結合し、ベクトル化する。
2. ベクトル化したスタイル行列を基に画像の所属するクラスターを求める。
3. 所属するクラスターに対応した畳み込みニューラルネットワークで品種識別を行う。

識別の処理について以下に詳しく述べる。まず、スタイル行列を求めるための学習済み畳み込みニューラルネットワークを用いてスタイル行列を計算する。このとき、いくつかの層から取得したものを組み合わせて学習時と同じ形状のスタイル行列のベクトルを計算する。次に、学習時に保持しておいた各クラスターの中心ベクトルと、先ほど求めたスタイル行列のベクトルとの距離を求め、その距離が最短となるクラスターをその画像の属するクラスターとする。この過程をスタイル行列クラスタリングと呼ぶことにする。最後に、その属するクラスターに対応した畳み込みニューラルネットワークに画像を渡して品種の

識別を行う。

3.5 むすび

本章では，本論文で提案するスタイル行列を用いた事前分類法と，畳み込みニューラルネットワークによる分類を組み合わせた識別手法について述べた。

第4章では，本章で説明した提案手法の評価実験を行う。

第4章 提案手法の評価実験と結果, 考察

4.1 はじめに

本章では, 猫の顔画像を用いて第3章で説明した提案手法の評価実験を行う.

4.2 実験の概要

提案手法の評価実験として, 二つの実験を行う. 一つ目の実験では, クラスタに対応した畳み込みニューラルネットワークが, クラスタに適した学習を行えているかを調べるための実験を行う. 二つ目の実験では, 提案手法が従来手法に比べて識別精度が向上しているかを調べるための実験を行う.

4.2.1 実験1

実験1では, クラスタに対応した畳み込みニューラルネットワークが, クラスタに適した学習を行えているかを調べるために, 提案手法の通りに猫顔画像の識別を行った場合と, クラスタに対応した畳み込みニューラルネットワークの内からランダムに一つを選んで識別を行った場合の識別率の比較を行う.

4.2.2 実験2

実験2では, 提案手法が従来手法に比べて識別精度が向上しているかを調べるために, 提案手法で猫顔画像の識別を行った場合と, 従来手法で畳み込みニューラルネットワークを用いて識別を行った場合の識別率の比較を行った. ここで従来手法とは, 畳み込みニューラルネットワークに, すべての学習画像に **Augmentation** を施したデータセットを学習させ, それを用いて識別を行う手法である.

4.3 実験の条件

実験1および2を行った環境や条件について以下に示す.

4.3.1 実験環境

実験に使用した環境を以下の表4.1に示す. また, ディープラーニングのフレームワーク”Chainer”を使用した[11].

表 4.1 実験環境

OS	Ubuntu 14.04
CPU	Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz
GPU	GeForce GTX 980

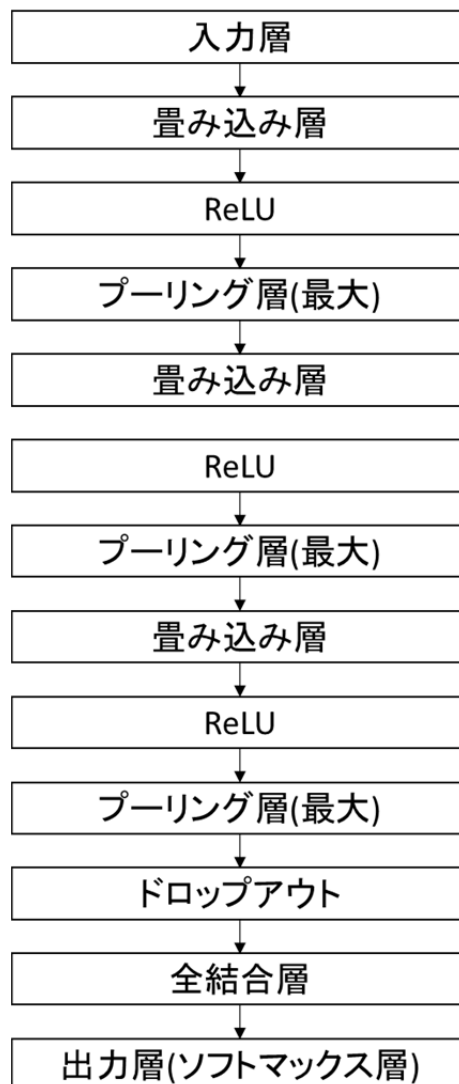


図 4.2 ネットワーク 2 の構成

4.3.3 K-means 法のクラスター数 K

クラスター数 K を変えながら猫顔画像の学習セットのスタイル行列の分類を行ったところ、 K を 5 に設定した場合、所属する画像がほとんど存在しないクラスターが作られた。したがって、本論文ではクラスター数 K を 4 として実験を行う。

4.3.4 スタイル行列クラスタリングの結果

図 4.1 中の①~④の位置でのスタイル行列を取得し、それらを結合し、ベクトル化したものを本実験では特徴量として使用し、クラスタリングを行った。学習画像のスタイル行列ベクトルを K-means 法でクラスタリングした結果を表 4.2 に示す。

表 4.2 スタイル行列ベクトルのクラスタリング結果

品種	クラスター0	クラスター1	クラスター2	クラスター3
Abyssinian	14	20	21	42
Bengal	6	22	17	55
Birman	65	0	33	2
Bombay	73	0	26	1
British Shorthair	9	0	84	5
Egyptian Mau	1	1	88	7
Maine Coon	5	10	54	31
Persian	11	10	64	15
Ragdoll	10	0	82	7
Russian Blue	10	0	87	3
Siamese	60	0	36	4
Sphynx	14	1	61	24
合計	278	64	653	196

4.3.5 画像データセット

学習および評価に用いる猫の顔画像は、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"内の猫画像から顔部分を切り出したものを使用する[12]。猫画像は 12 クラス, 1 クラスにつき約 200 枚の画像が存在する。1 クラスにつき 100 枚を評価に, 残りを学習に使用する。使用する画像の内訳を表 4.3 に示す。また, 各クラスの猫顔画像の例を図 4.3~4.14 にそれぞれ示す。

また, 学習時には **Augmentation** を行って画像の枚数を増やしている。今回の実験では **Augmentation** として画像の左右反転, 輝度変化(+20%, -20%), 左右反転+輝度変化(+20%, -20%)を画像に施した。提案手法のクラスターに対応した畳み込みニューラルネットワークでは, 対応するクラスターに属する画像のみに **Augmentation** を施した学習画像セットを, 従来手法の畳み込みニューラルネットワークでは, 全ての画像に **Augmentation** を施した学習画像セットをそれぞれ使用している。それぞれの学習に用いた画像数を表 4.4 に示す。

表 4.3 ”The Oxford-IIIT-Pet dataset”の内の学習および評価に用いる猫顔画像の内訳

品種	学習用	評価用	合計
Abyssinian	97	100	197
Bengal	100	100	200
Birman	100	100	200
Bombay	100	100	200
British Shorthair	98	100	198
Egyptian Mau	97	100	197
Maine Coon	100	100	200
Persian	100	100	200
Ragdoll	99	100	199
Russian Blue	100	100	200
Siamese	100	100	200
Sphynx	100	100	200
合計	1191	1200	2391

表 4.4 学習に用いた画像数

CNN	学習画像数
クラスター0に対応する CNN	2581
クラスター1に対応する CNN	1511
クラスター2に対応する CNN	4456
クラスター3に対応する CNN	2171
従来手法で用いる CNN	7146



图 4.3 Abyssinian[12]



图 4.6 Bombay[12]



图 4.4 Bengal[12]



图 4.7 British Shorthair[12]

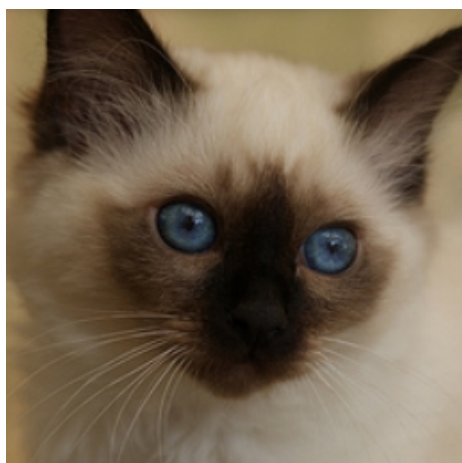


图 4.5 Birman[12]



图 4.8 Egyptian Mau[12]



图 4.9 Maine Coon[12]



图 4.12 Russian Blue[12]



图 4.10 Persian[12]

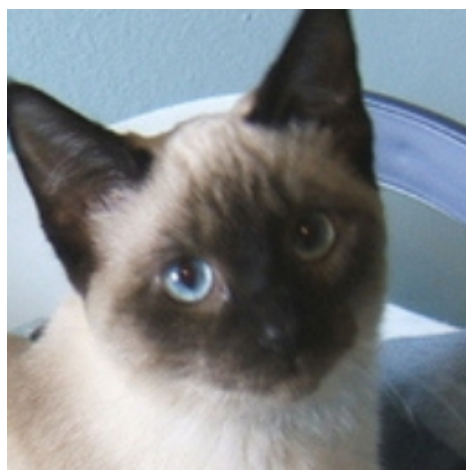


图 4.13 Siamese[12]

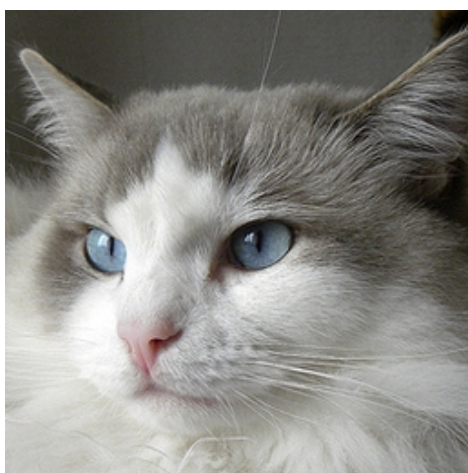


图 4.11 Ragdoll[12]



图 4.14 Sphynx[12]

4.4 実験1の結果

提案手法の識別率とクラスターに対応した畳み込みニューラルネットワークの内からランダムに一つを選んで識別を行った場合の識別率を表 4.5 に示す. また, クラスターに対応した各畳み込みニューラルネットワークの識別率を表 4.6 に示す. なお, ランダムに選択して識別した場合の結果は, 識別実験を 5 回行い, その平均を求めたものである.

ここで, 識別率は以下の式より算出する.

$$\text{識別率} = \frac{\text{正解画像数}}{\text{全画像数}} \quad (4.1)$$

平均識別率を比べると, ランダムに畳み込みニューラルネットワークを選ぶよりも, スタイル行列に合ったネットワークを選ぶ方が識別率は高いことがわかった.

表 4.5 対案手法とランダム識別の識別率

品種	提案手法	ランダム
Abyssinian	80%	73.6%
Bengal	73%	68.2%
Birman	84%	87.2%
Bombay	92%	90.2%
British Shorthair	82%	80.0%
Egyptian Mau	76%	78.6%
Maine Coon	73%	69.2%
Persian	76%	78.8%
Ragdoll	67%	68.4%
Russian Blue	92%	86.8%
Siamese	70%	67.8%
Sphynx	85%	81.6%
平均	79.2%	77.5%

表 4.6 クラスターに対応した各畳み込みニューラルネットワークの識別率

品種	クラスター0	クラスター1	クラスター2	クラスター3
Abyssinian	70%	71%	77%	79%
Bengal	69%	68%	69%	71%
Birman	84%	90%	85%	87%
Bombay	92%	89%	88%	93%
British Shorthair	78%	79%	82%	82%
Egyptian Mau	75%	81%	75%	80%
Maine Coon	70%	66%	75%	73%
Persian	78%	77%	81%	80%
Ragdoll	68%	68%	70%	63%
Russian Blue	86%	86%	92%	83%
Siamese	69%	72%	69%	68%
Sphynx	83%	80%	83%	83%
平均	76.8%	77.3%	78.8%	78.5%

4.5 実験1の考察

提案手法の識別率とクラスターに対応した畳み込みニューラルネットワークの中からランダムに一つを選んで識別を行った場合の識別率を比較したところ、提案手法の平均識別率が 79.2%、ランダムにネットワークを選んだ場合の平均識別率が 77.5%と、提案手法の方が高い識別率を示した。

また、ランダムにネットワークを選択した場合の平均識別率は、各クラスターに対応するネットワークの平均識別率の平均値である 77.85%に近い値を示したが、提案手法は各クラスターに対応するどのネットワークの平均識別率よりも高い平均識別率を示した。したがって、各クラスターに対応するネットワークは、そのクラスターに適した学習を行えていると考えられる。

各クラスの識別率を比べてみると、”Abyssinian”, ”Bengal”, ”Bombay”, ”British Shorthair”, ”Maine Coon”, ”Russian Blue”, ”Siamese”, ”Sphynx”と半数以上のクラスについて提案手法の方が高い識別率を示している。特に”Abyssinian”, ”Bengal”, ”Maine Coon”, ”Russian Blue”, ”Sphynx”といったクラスで、提案手法は高い性能を示した。

4.6 実験2の結果

提案手法の識別率と従来手法の識別率を表 4.7 に示す。

平均識別率を比べると，提案手法の平均識別率が 79.2%，従来手法の平均識別率が 80.7%と従来手法の方が提案手法よりも高い識別率を示した。

表 4.7 提案手法と従来手法の識別率

品種	提案手法	従来手法
Abyssinian	80%	75%
Bengal	73%	74%
Birman	84%	88%
Bombay	92%	94%
British Shorthair	82%	86%
Egyptian Mau	76%	80%
Maine Coon	73%	72%
Persian	76%	82%
Ragdoll	67%	75%
Russian Blue	92%	88%
Siamese	70%	72%
Sphynx	85%	82%
平均	79.2%	80.7%

4.7 実験2の考察

提案手法と従来手法の平均識別率を比較したところ，従来手法の方が高い識別率を示した。実験1でも高い性能を示した”Abyssinian”，”Maine Coon”，”Russian Blue”，”Sphynx”といったいくつかのクラスでは提案手法の方が高い識別率を示しているが，全体的に考えると従来手法の方が高い識別率を示している。

一般的に，畳み込みニューラルネットワークは学習に用いる画像数が多いほど高い性能を発揮すると言われている。提案手法で用いた各クラスターに対応するネットワークは，それぞれの対応したクラスターに適した学習をさせるために，学習枚数が従来手法よりも少ないものとなっている。そのため，従来手法のものよりも個々のネットワークの識別性能は落ちてしまう。これは表 4.6 と表 4.7 の従来手法の結果を比べると明らかである。今回の提案手法では，対応するクラスターに属する画像に適した学習を行うことができたことが実験1から確認できたが，その適する画像に対する精度の向上が，学習枚数を減らしたことによる全体的な精度の低下を下回ってしまったため，提案手法は従来手法よりも低い識別率を示したと考えられる。

今回の提案手法で用いた、クラスターに属する画像のみを **Augmentation** するという方法ではなく、別の方法でそのクラスターに適した学習を行うことが出来れば、従来手法の識別率を超える手法ができる可能性があると考えます。その具体的な方法としては、そのクラスターに適したネットワーク構造の畳み込みニューラルネットワークを用いる、学習時の学習係数等パラメータを変えるとといった方法が考えられる。しかし、そのようなネットワークの構成や学習係数等のパラメータについての理論的な理解はまだ得られていないのが現状であり、最適なネットワークの構成やパラメータは手探りで調べていく必要がある。

また、今回の実験で用いたデータセットは各クラスの画像が 200 枚と少なかったため、学習用に 100 枚のデータしか使うことが出来なかった。このように今回の実験では学習画像が少なかったため、従来手法と提案手法の各クラスターに対応するネットワークとの学習画像数の差が性能に大きく影響を与えた可能性が考えられる。そのため、より様々な方法で **Augmentation** を施したり、他の画像数の多いデータセットに対して提案手法を適用したりすることで、提案手法の性能を再評価する必要があると考えます。

第5章 結論と今後の課題

5.1 結論

本研究では、猫の顔画像を対象として、スタイル行列を用いた事前分類法および畳み込みニューラルネットワークによる分類を組み合わせることで品種識別を行う手法を提案すると共に、実験により手法の評価を行った。提案手法の識別率と、クラスターに対応した畳み込みニューラルネットワークの内からランダムに一つを選んで識別を行った場合の識別率を比較したところ、提案手法の平均識別率が 79.2%、ランダムにネットワークを選んだ場合の平均識別率が 77.5%と、提案手法の方が高い識別率を示した。そのため、提案手法の各クラスターに対応したネットワークは、クラスターに属する画像に適した学習が行えていることがわかった。また、従来手法と提案手法の識別率の比較を行ったところ、提案手法の平均識別率が 79.2%、従来手法の平均識別率が 80.7%と従来手法の方が提案手法よりも高い識別率を示した。

今後、提案手法の各クラスターに適した学習を行う方法に改良を加えることで、より識別率の高い手法とすることができると考える。

5.2 今後の課題

今後の課題としては、以下の二つが挙げられる。

- ・各クラスターに適した学習方法の改良

本研究では、各クラスターに適したネットワークの学習を行うことができたが、その適応による識別率の向上は不十分だった。そのため、各クラスターに適した学習方法の再検討が必要である。具体的には畳み込みニューラルネットワークの構成の検討や学習係数等のパラメータの調整といったことが必要である。

- ・学習画像が十分な場合の評価実験

本研究では、評価実験の対象として”The Oxford-IIIT-Pet dataset”内の猫の顔画像を使用した。学習に用いた画像数が少なかったことで提案手法と従来手法の性能差が生まれた可能性が考えられる。そのため、**Augmentation**の方法を増やすことで学習画像数を増やす、より画像数の多いデータセットを用いる等学習画像数を増やして評価実験を再度行う必要がある。

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。

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。

最後に、私をここまで育ててくださった家族に感謝いたします。

参考文献

- [1] 内閣府, 主要耐久消費財等の普及率 (一般世帯) (平成 27 年 (2015 年) 3 月現在),
<http://www.esri.cao.go.jp/jp/stat/shouhi/2015/201503fukyuritsu.xls>
- [2] リサーチバンク, 写真に関する調査,
http://research.lifemedia.jp/2014/07/140730_photo.html
- [3] 草野孝幸, 出口大輔, 井手一郎, 村瀬洋, 猫パーツの抽出とその組み合わせによる猫の顔検出の高精度化, DIA2010 動的画像処理実利用化ワークショップ講演概要, 2010.
- [4] 小坂谷達夫, 伊藤聡, 窪田進, 2 種類の異なる特徴を用いた特定物体検出, 東芝レビュー Vol.65 No.7, pp.33-36, 2010.
- [5] 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.
- [6] 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/>
- [7] 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>
- [8] 石井智大, 望月義彦, 小山田雄仁, 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.
- [9] Leon A.Gatys, Alexander S.Ecker, Matthias Bethge, “A Neural Algorithm of Artistic Style”, 2015.
- [10] J.MacQUEEN, “Some Methods for classification and Analysis of Multivariate Observations”, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp.281-297, 1967.
- [11] Chainer, <http://chainer.org/>
- [12] The Oxford-IIIT-Pet dataset, <http://www.robots.ox.ac.uk/~vgg/data/pets/>

図一覧

図 2.1	ユニットの基本構造-----	3
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	畳み込みの実行例-----	6
図 2.5	畳み込み層の概要(入力画像 3 チャンネル, フィルタ 4 個)-----	7
図 2.6	プーリング実行例(最大プーリング)-----	8
図 2.7	全結合層の概要図-----	9
図 3.1	提案手法の概要図-----	12
図 4.1	ネットワーク 1 の構成-----	16
図 4.2	ネットワーク 2 の構成-----	17
図 4.3	Abyssinian[12]-----	20
図 4.4	Bengal[12]-----	20
図 4.5	Birman[12]-----	20
図 4.6	Bombay[12]-----	20
図 4.7	British Shorthair[12]-----	20
図 4.8	Egyptian Mau[12]-----	20
図 4.9	Maine Coon[12]-----	21
図 4.10	Persian[12]-----	21
図 4.11	Ragdoll[12]-----	21
図 4.12	Russian Blue[12]-----	21
図 4.13	Siamese[12]-----	21
図 4.14	Sphynx[12]-----	21

表一覧

表 4.1	実験環境-----	15
表 4.2	スタイル行列ベクトルのクラスタリング結果-----	18
表 4.3	”The Oxford-IIIT-Pet dataset”の内の学習および評価に用いる猫顔画像の内訳---	19
表 4.4	学習に用いた画像数-----	19
表 4.5	対案手法とランダム識別の識別率-----	22
表 4.6	クラスターに対応した各量み込みニューラルネットワークの識別率-----	23
表 4.7	提案手法と従来手法の識別率-----	24