

平成 30 年度修士論文

Collaborative Mean Attraction 法による  
一般画像分類

宮崎大学大学院 工学研究科 工学専攻  
機械・情報系コース 情報システム工学分野

学籍番号 T1703009

荻原弘樹

指導教員 椋木雅之教授

平成 3 1 年 1 月 2 8 日

## 修士論文概要

本研究では、これまでカメラ間人物照合に適用されてきた Collaborative Mean Attraction (CMA)法が、一般画像分類にも有効であるか検証する。少数学習用データで分類を行う類似の問題の1つとしてカメラ間人物照合がある。この問題に対して CMA 法が提案された。CMA 法は、全学習用データを用いて分類を行うため、学習用データ枚数が少ない場合においても正解率が高いという利点がある。この特性を利用して CMA 法を少数学習用データ下での一般画像分類に適用し、一般画像分類において有効であることを示す。また学習用データ枚数が増えた際にも認識率の向上を図るため、CMA 法にスパース最適化を導入した Sparse Collaborative Mean Attraction (SCMA)法を提案する。

これらの手法と従来手法である Support Vector Machine (SVM) や、Deep Learning の手法 (DL) とで一般画像分類の実験を行った。また CMA 法の基となった従来手法 CRC、SRC、NRC 法とも一般画像分類において比較を行い、CMA 法、SCMA 法を評価した。CMA 法、SCMA 法は SVM や DL 手法に比べて、学習用データ枚数が少ない場合において高い識別率を得た。SCMA 法は学習用データ枚数が多い時でも DL の手法と同程度以上の認識率を得た。CRC、SRC、NRC 法との比較実験では CMA、SCMA 法は同程度の認識率を得た。

## 目次

第1章 序論 .....	3
第2章 一般画像分類.....	6
2.1 一般画像分類の定義.....	6
2.2 公開データセット .....	8
2.3 特徴量の抽出.....	11
2.4 従来の画像分類の識別器 .....	14
2.5 疎・協働表現を使った識別器 .....	16
2.6 従来手法の問題点 .....	18
第3章 Collaborative Mean Attraction 法 .....	19
3.1 CMA 法 .....	19
3.1.1 最適化段階 .....	21
3.1.2 分類段階 .....	22
3.2 スパース最適化を用いた CMA 法(SCMA 法).....	23
3.2.1 スパース最適化段階 .....	23
3.2.2 分類段階 .....	25
第4章 評価実験.....	26
4.1 従来の識別器との比較実験 .....	26
4.2 従来の疎・協働表現手法との比較実験 .....	29
4.3 複数テストデータの組での実験.....	35
第5章 結論 .....	40
謝辞.....	41
参考文献 .....	42

# 第1章 序論

近年、一般物体認識[1]の研究が盛んに行われ、進展してきた。一般物体認識とは、制約のない実世界で撮られた画像から、画像内の物体の種類等を認識する課題である。この一般物体認識は認識対象となる被写体の種類の多さや、同じ種類の物体での形状や色彩の違いから、画像認識の分野の中でも、困難とされている課題である。一般物体認識の中で、画像から物体を検出する物体検出や、ピクセル単位で画像内の物体の領域を抽出するセマンティックセグメンテーションなど、様々な研究が行われている。本論文では一般物体認識の中でも、画像全体を一つのカテゴリに分類する一般画像分類を対象とする。

一般画像分類の研究では、機械学習の手法が注目されている。機械学習では、学習用データを与えることで、物体の種類の違いや、同じ種類の物体の外観の違いに対処している。特に近年では深層学習の手法が大きな成果をあげている。深層学習では、モデルを学習させるために多くの学習用データを与え、高性能な識別器を得ることができる。しかし、学習用データを多く得られない場合、深層学習の手法では十分にモデルを学習することができないと考えられる。この場合、比較的少数の学習用データから分類するカテゴリを学習し、高性能に分類が行える識別器が必要である。

この問題に対して本研究では Collaborative Mean Attraction (CMA) 法[2]を適用する。CMA 法は、比較的少数の学習用データしか利用できない、カメラ間人物照合の問題に対して提案された手法である。カメラ間人物照合とは、カメラ A で撮影された人物が、別の地点に設置されているカメラ B で撮影された人物像の中のどれと同一人物であるかを判定するものである。カメラ B で撮影された人物像を学習用データのカテゴリとすると、人物がカメラ B の前を通過する間の画像しか得ることができないため、各カテゴリ内で大量の学習用データを得ることはできない。またカメラ B で撮影された人物数がカテゴリ数となるため、数十から数百カテゴリの分類問題となる。

一般的な識別器では、未知のテストデータと既知の各カテゴリの類似度や相違度、距離などを用いて、カテゴリ毎にテストデータのカテゴリらしさを評価し分類する。一方、CMA 法では、初めに既知の全ての学習用データを用いて、未知のテストデータの近似を行う。次に、この近似で最も貢献したカテゴリに、テストデータを分類する。CMA 法は一般的な識別器と異なり、既知の全ての学習用データを用いてテストデータを表現することで、学習用データが比較的少数となるカメラ間人物照合問題で高い性

能を示している。

CMA 法が持つ、「既知の全ての学習用データを用いてテストデータを表現する」というアイデアの基となったのは、**Collaborative Representation Classifier (CRC)** 法 [3] と **Sparse Representation Classifier (SRC)** 法 [4] である。本稿ではこれらの「既知の全ての学習用データを用いてテストデータを表現する」手法を**協働表現手法**と総称する。協働表現手法は、テストデータを表現する際の制約項の違いにより、様々なバリエーションが存在する。初めに提案された SRC 法では、重要な数少ない要素でテストデータを表現するため、L1 ノルムの制約項が使われた。SRC 法は協働表現手法ではあるが、L1 ノルムの制約により、少ない要素での表現をしているため、**疎表現手法**であるともいえる。CRC 法では L2 ノルムを制約項に使うことで、SRC 法の計算速度を向上させながら、SRC 法と同程度以上の識別性能を達成している。Nonnegative representation classifier (NRC) 法[5]では、テストデータを要素の非負の線形和で表すという制約項を使い性能向上を図った。非負であるという制約の結果、より少ない要素での表現になっており、NRC 法も疎表現手法であるといえる。CMA 法ではテストデータを表現する際に、学習用データの平均値から離れないようにする制約項が使われている。これにより、学習用データから離れすぎた近似を行うことがなくなり、性能の向上が期待できる。これらの疎・協働表現手法は、すべての学習用データを用いることで、各カテゴリの学習用データが比較的少数でも、学習用データをより有効に活用して高い識別性能が達成できると考えられる。

本論文では、これまでカメラ間人物照合に適用されてきた CMA 法が、一般画像分類にも有効であるか検証する。カメラ間人物照合では、カテゴリ内の多様性は見えの変動であるのに対して、一般画像分類ではカテゴリに含まれる対象物体自体の多様性が大きいという違いがあり、問題としては大きく異なるものである。そのため、一般画像分類でも CMA 法が有効であるとは限らない。

さらに CMA 法に対してスパース最適化を導入した **Sparse Collaborative Mean Attraction (SCMA)** 法を提案する。CMA 法は協働表現手法であり、すべての学習用データを同様に利用する傾向がある。そのため学習用データが増えた場合、CMA 法は過度に学習用データに適応する可能性がある。疎表現を取り入れた SCMA 法によりこの問題の解消を目指す。公開データセットを用いて、従来の機械学習の手法との比較を行い、少数学習用データ枚数下での CMA 法と SCMA 法の有効性を示す。また疎・協働表現手法を使った従来手法とも比較実験を行い考察する。

本稿の構成は以下の通りである。2章では一般画像分類の定義、画像分類の公開データセット、画像からの特徴量の抽出、従来の機械学習の画像分類手法について述べた後、疎・協働表現の従来手法について示し、本研究での CMA 法の位置づけをする。3章では CMA、SCMA 法について詳細を述べる。4章では SCMA 法の有効性を示すために行った一般画像分類の実験について述べる。5章では本研究の結論と今後の課題について述べる。

## 第2章 一般画像分類

### 2.1 一般画像分類の定義

本稿では、制約のない実世界で撮られた物体の画像に対して行う画像分類を、一般画像分類と定義する。一般画像分類は、一般物体認識の一課題と位置付けられる。

一般物体認識[1]とは、制約のない実世界で撮られた物体の種類を認識する課題である。制約のない実世界で撮られた物体には、その物体以外の物や背景が混じっていることが多く、図 2-1 の様に同じ種類でも画像から得られる特徴量が変わってくる。例として空き缶のカテゴリがあったとすると、赤色のものや、細長いもの、つぶれたものなど、物体そのものの外観の変化が極めて大きくカテゴリ内の変動が高い。これらの理由により一般物体認識は画像認識分野の中でも困難とされている課題である。一般物体認識の中で、画像から物体を検出する物体検出[6]や、ピクセル単位での画像内の物体の領域を抽出するセマンティックセグメンテーション[7]など、様々な研究が行われている。本論文では一般物体認識の中でも、画像全体を一つのカテゴリに分類する一般画像分類を対象とする。



図 2-1 Caltech256 の baseball glove の画像

一般画像分類の手順は通常以下の通りになっている。

- I. 画像から特徴量を抽出
- II. 特徴量を用いて識別器により画像内の主要な1つの物体の種類を識別

まず、扱うデータセットの画像から特徴量を抽出する。画像の特徴量とは、その画像らしさを表すデータである。画像の特徴量を抽出する研究も行われており、画像全体から特徴量を取ったものや、画像内の局所的な部分の特徴量を取ったものなどがある。近年では深層学習を用いて画像から特徴抽出を行い画像分類する研究も行われている[8]。特徴量を抽出する手段については2.4節で述べる。

識別器についても多くの研究が行われている。標準的な手法として **Support Vector Machine (SVM)** [9]などが利用される。本研究では **SVM** を比較手法として扱う。**SVM** については2.5節で述べる。

近年では深層学習 (**Deep Learning**) の手法が大きな成果をあげている。深層学習では、モデルを学習させるために多くの学習用データを与え、高性能な識別器を得ることができる。また、大量の一般画像を用いて学習済みのモデルに対して、任意の画像を入力した際に得られる中間層の出力を、その入力画像の特徴量として利用することもできる。本論文では識別器についての性能評価を行う。



## 2.2 公開データセット

一般に、機械学習の手法で学習、評価を行うために公開データセットを用いる。公開データセットは、世界中のだれでも扱えるデータセットである。公開データセットがあることにより、異なる機械学習の手法を評価する事ができるようになった。また大規模な公開データセットが提供されるようになり、機械学習において十分な学習データを与えることができるようになった。企業、大学や国の機関などが、各分野についての公開データセットを提供している。公開データセットのデータは画像や動画、数値データなどがある。

一般画像分類において有名な大規模公開データセットには ImageNet[10]、ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)データセット[11]などがある。本稿では一般画像分類の公開データセットの中の Caltech101[12]、Caltech256[13]の 2 種類を扱う。Caltech データセットはカリフォルニア工科大学で作成された公開データセットである。

Caltech101 は 101 カテゴリの物体の画像とその他の背景画像からなり、総画像数は 9,144 枚である。各カテゴリ中の最少画像数は 31 枚となっている。それに対して Caltech256 データセットのカテゴリ数は 257 カテゴリで、総画像数は 30,607 枚である。またカテゴリ中の最少画像数は 80 枚となっている。表 2-1 に Caltech101 と Caltech256 の画像数内訳を示す。

Caltech101 は Caltech256 に比べ比較的分類が容易で小規模で扱いやすいものになっている。一方 Caltech256 は、比較的分類が困難で大規模なデータセットとなっている。Caltech256 の画像の例を図 2-2 に示す。

表 2-1 Caltech101 と Caltech256 の画像数内訳

	Caltech101	Caltech256
画像総数	9,144 枚	30,607 枚
カテゴリ数	101 カテゴリ	257 カテゴリ
1 カテゴリの最少画像数	31 枚	80 枚
使用カテゴリ	100 カテゴリ	257 カテゴリ

<b>ak47</b>				
<b>american-flag</b>				
<b>backpack</b>				
<b>baseball-bat</b>				
<b>baseball-glove</b>				
<b>basketball-hoop</b>				
<b>bat</b>				
<b>bathtub</b>				

図 2-2 Caltech256 の画像の例 (1)

<b>bear</b>				
<b>beer-mug</b>				
<b>billiards</b>				
<b>binoculars</b>				
<b>birdbath</b>				
<b>blimp</b>				
<b>bonsai-101</b>				
<b>boom-box</b>				

図 2-2 Caltech256 の画像の例 (2)

## 2.3 特徴量の抽出

画像分類において、画像からの特徴量の抽出は非常に重要な過程である。画像の特徴量とは、その画像らしさを表すデータである。特徴量には様々な種類がある。単純なものは画像全体の色やテクスチャをヒストグラムとして扱う手法がある。また、画像の局所的な部分のみを特徴としてとらえる局所特徴量には、SIFT[14]、HOG[15]などがある。近年、深層学習の手法である Convolutional Neural Network (CNN) が画像認識分野で成果をあげている[16]。CNN とは畳み込み層を導入したニューラルネットワークである。一般に CNN は画像分類器として扱うが、大規模データセットで学習した学習済みモデルを用いて特徴抽出器として扱うこともできる[8]。CNN の概要図を図 2-3 に示す。CNN を特徴抽出器として扱う際には一般的に、大規模データセットで学習した CNN のモデルに入力として画像を与え、全結合層の出力をその画像の特徴量として扱う。入力された画像は、畳み込み層で局所的な特徴を特徴マップとして抽出し、全結合層でそれらをまとめ上げ、識別層でカテゴリに分類する。全結合層で得られる出力は画像の特徴をまとめ上げたものであるため、画像の特徴量として扱うことができる。

ILSVRC のデータで学習した CNN モデルから得られる画像の特徴量は、物体認識や、詳細画像カテゴリ識別などの様々な分野で利用されている[17]。

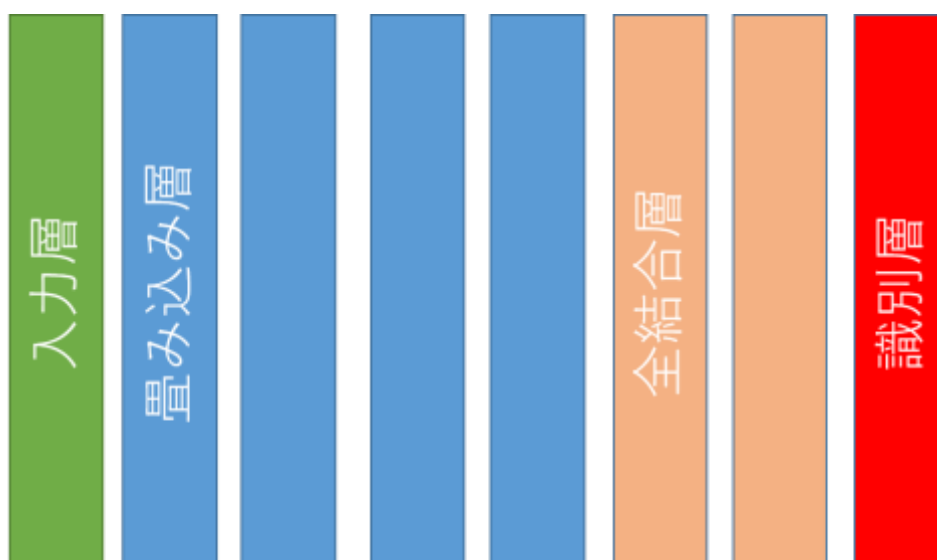


図 2-3 CNN の概要図

本稿では、特徴量の抽出に 2 つの学習済みモデルを使用する。1 つ目のモデルは、AlexNet [16] と呼ばれる構造をもった CNN に ILSVRC2012 の 1000 カテゴリ、120 万枚画像のデータを学習させたもので、学習済みのモデル[18]が公開されている。AlexNet の構造は図 2-4 のようになっている。図中の conv、norm、pool、fc がそれぞれ、畳み込み層、正規化層、プーリング層、全結合層を示している。このネットワークは畳み込み層 5 層、プーリング層 3 層、正規化層 2 層、全結合層 3 層からなっている。畳み込み層では入力されたデータにフィルタをかけて特徴マップを生成する。このフィルタをカーネルといい、カーネルの大きさをカーネルサイズと呼ぶ。特徴マップを生成する際、カーネルをずらす幅のことをストライドという。カーネルサイズやストライドは各畳み込み層によって大きさが違う。このモデルのプーリング層はマックスプーリングを行っている。マックスプーリングでは、与えられたデータのカーネル範囲内の最大値を選択していき、特徴マップを生成する。畳み込み層では、画像の局所的な特徴を抽出する。プーリング層は特徴をまとめて整理する働きをする。これらを組合すことにより画像の抽象化を行うことができる。このモデルに画像を与え、fc7 層の出力をその画像の特徴量とする。特徴量の次元数は 4096 次元である。

2 つ目は VGG19[19] と呼ばれる構造をもった CNN に ILSVRC の 100 万画像のデータを学習させたもので、公開されている学習済みモデル[20]を使用する。VGG19 のモデルは図 2-5 のようになっている。このネットワークは畳み込み層 16 層、プーリング層 5 層、全結合層 3 層からなっている。このモデルに画像を与え、fc18 層の出力を特徴量として扱う。特徴量の次元数は 4096 次元である。

2 つのモデルでは、VGG19 モデルの性能が高く、得られる特徴量の質もよくなる。性能の異なる特徴量を用いた分類実験を行い、特徴量により性能が変化するかを調べるために、この 2 つのモデルを選び、特徴量を抽出した。

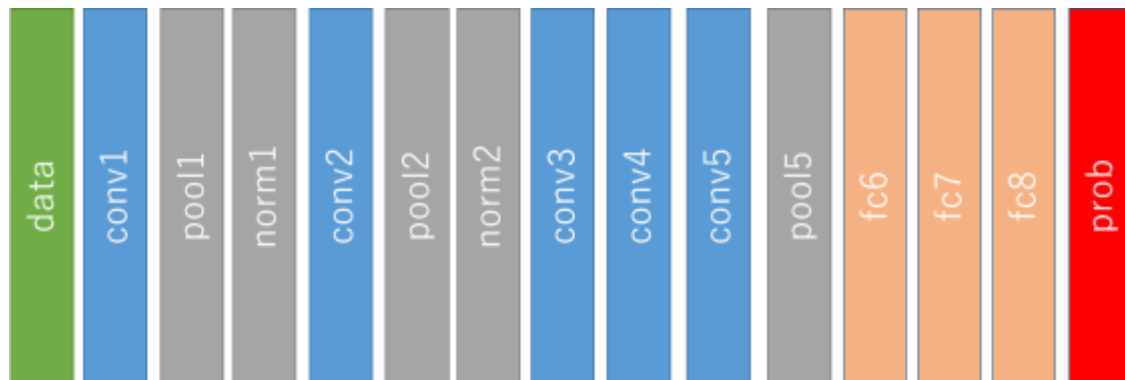


図 2-4 AlexNet の構造

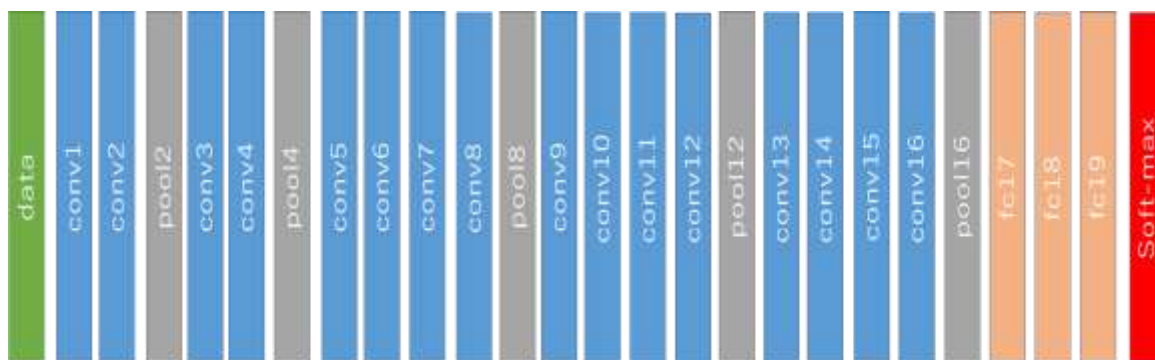


図 2-5 VGG19 の構造

## 2.4 従来の画像分類の識別器

従来の画像分類の識別器として本稿では Support Vector Machine (SVM) [9]と、深層学習 (Deep Learning) を使った手法を扱う。

SVM は学習用データから2クラスを分類する識別境界を求める手法である。2クラス間のデータの-marginが最大となるように境界線を引くことによって、汎化能力を高めている(図 2-6)。

また、SVM ではカーネルトリックと呼ばれる方法を適用することで性能向上を図っている。カーネルトリックとは与えられた学習用データを非線形に変換して、その非線形変換した先の空間で識別境界を求めるものである。非線形空間に変換することで、元の空間で複雑な形状の識別境界を求めることができ、識別性能が向上する。非線形空間への変換は膨大な計算量が必要になる場合が多い。しかし、カーネルトリックでは、実際に学習用データを非線形に変換するのではなく、カーネルと呼ばれる関数の計算のみで、非線形空間内での識別境界を求めることができる。

SVM を多クラス分類に適用するには種々の手法があるが、本研究では全ての2クラスの組み合わせに対して SVM を構築してその結果の多数決で分類を行う「1対1分類」法を用いる。

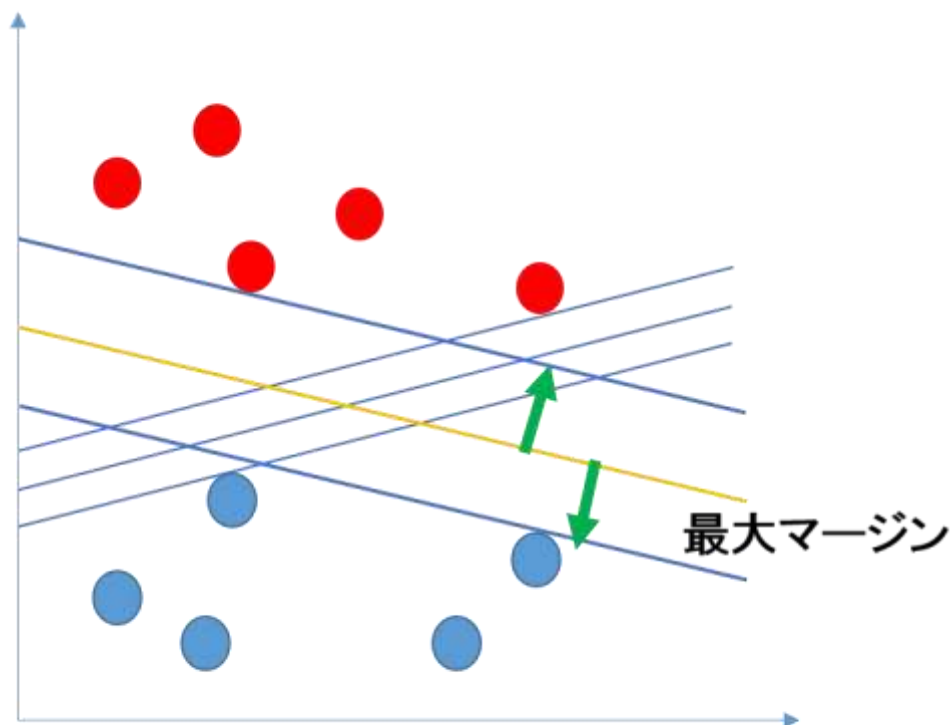


図 2-6 SVM のマージン最大化

深層学習を使った手法では、事前に大量の画像を用いて学習したモデルに対して、ファインチューニングを適用し画像分類を行う方法を扱う。この手法を以下 **DL** 手法と呼ぶ。ファインチューニングとは学習済みのモデルに追加で新たなデータで学習させて、別の分野に適応させることである。ファインチューニングを行うには、学習済みモデルの識別層を対象となるデータセットのクラス数に変更する。学習済みモデルのパラメータを初期値として扱い、学習させる。これにより、過学習を防いで、良い学習結果を得ることができる。ImageNet や ILSVRC のデータセットを学習させた学習済みモデルが配布されている。本稿で用いる DL 手法では学習済みモデルとして 2.3 節で示した AlexNet の学習済みモデル[18]を用いる。

このモデルに本稿で画像分類に利用する学習用データを与え、ファインチューニングを行う。得られたモデルにテストデータを与え、画像分類を行う。



## 2.5 疎・協働表現を使った識別器

本節では疎・協働表現を使った画像分類の識別器（CRC、SRC、NRC法）について説明する。

まず、顔画像分類の問題に対して SRC 法[4]が提案された。次に、SRC法に対して計算速度の向上を図った CRC 法[3]が提案された。この手法は SRC 法と同程度以上の識別性能を有するとされている。CRC 法については、カーネルトリックを導入した手法や、テストデータに近い学習用データのみを利用する局所性を導入した手法など、多くの研究が現在も行われている[21]。その後、認識率を高めるために非負制約をかけた NRC 法が提案された[5]。これらの手法はパターン認識の分野の中でも顔認識、手書き文字認識、物体認識、行動認識などに適用されている。

これらの手法は、「既知の全ての学習用データを用いてテストデータを表現する」という基本的なアプローチは似ているが、その際の制約項に違いがある。まず全学習用データの線形和でテストデータを表現する。この時テストデータと表現されたデータとの差が小さくなるように線形和の係数ベクトルを更新していく。ここで制約項として SRC 法では係数ベクトルの L1 ノルム最小化が、CRC 法では係数ベクトルの L2 ノルム最小化が課せられる。NRC 法では、係数ベクトルが非負であるという制約が課せられる。本稿ではこの係数ベクトルを求める段階を最適化段階と呼ぶ。最適化段階でテストデータを全学習用データで表現する手法を**協働表現手法**と呼ぶ。CRC、SRC、NRC 法はすべて協働表現の手法であるが、SRC、NRC 法では全学習用データの中で有用な要素のみを用いて表現しているため、**疎表現手法**と呼ぶ。

その後、テストデータと各カテゴリのみの学習用データでの線形和との差を測り、最小となったカテゴリにテストデータを分類する。本稿ではテストデータを分類する段階を分類段階と呼ぶ。SRC、CRC、NRC 法では分類段階は同じである。

既知のカテゴリ  $i \in \{1, \dots, n\}$  ( $n$  はカテゴリの種類数) の学習用データの特徴量ベクトルを並べた行列を  $\mathbf{X}_i \in \mathbf{R}^{m \times N_x^i}$  ( $N_x^i$  はカテゴリ  $i$  の学習用データ数)、全学習用データの特徴量ベクトルを並べた行列を  $\mathbf{X} \in \mathbf{R}^{m \times N_x}$  ( $N_x = \sum_i N_x^i$  は全学習用データ数) とする。テストデータの特徴量ベクトルを  $\mathbf{y} \in \mathbf{R}^m$  とする。この時、最適化段階では以下の式を最小化する係数ベクトル  $\boldsymbol{\beta}$  を求める。上から SRC、CRC、NRC 法の最適化段階での式である。

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \quad (1)$$

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \quad (2)$$

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \quad \text{s.t.} \quad \boldsymbol{\beta} \geq 0 \quad (3)$$

$\lambda$  は正則化のパラメータである。 $\lambda$  の値が大きくなるほど制約が強くなる。

**SRC** 法では、**L1** ノルムを利用したスパース最適化により  $\boldsymbol{\beta}$  の係数ベクトルの多くの値が  $0$  となり、係数ベクトル  $\boldsymbol{\beta}$  はスパースなベクトルになる。スパースなベクトルではテスト画像を表現する際に、有用な学習用データのみを厳選して表現することができる。これにより認識率の向上が図れる。

**CRC** 法では正則化項で **L2** ノルムが使われている。**L1** ノルムの最適化では繰り返しによる収束計算が必要になり、計算時間がかかる。一方 **L2** ノルムを利用すると、閉形式の解をもつため、係数ベクトルが一度の線形計算で求まる。そのため、**SRC** 法に比べて少ない計算量で係数ベクトルが求まる。**SRC** 法と比べて係数ベクトルはスパースにならず、全学習用データでテストデータを表現している。

**NRC** 法では係数ベクトルが非負であるという制約が使われている。**NRC** 法では係数ベクトルはスパースなベクトルとなる。正解らしいクラスの係数は非負の値を示しており、係数を非負にすることは自動的に正解の学習用データの表現力を高め、不正解の学習用データの表現力を抑制することができるといわれている。

次に、分類段階での式を示す。**SRC**、**CRC**、**NRC** 法の分類段階での式は同様である。式(1)~(3)で求められた係数ベクトル  $\boldsymbol{\beta}$  はカテゴリごとの係数を  $\boldsymbol{\beta}_i$  として、 $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_n^T)$  と分解することができる。各カテゴリ  $i$  について以下の式を計算し最もテストデータとの誤差の少ないカテゴリにテスト画像を分類する。

$$\text{label}(\mathbf{y}) = \arg \min_i \{\|\mathbf{y} - \mathbf{X}_i \boldsymbol{\beta}_i\|_2\} \quad (4)$$

## 2.6 従来手法の問題点

一般画像分類において、従来手法の問題点をあげる。まず SVM と DL 手法では学習用データ枚数が少ない場合では、学習するデータが少ないため、対象物体の外観の違いに対処できないと考えられる。また、CRC、SRC、NRC 法においては、最適化段階においてテストデータを表現する際に、学習用データによっては、元のデータとかけ離れた近似を行ってしまうおそれがある。一方 CMA 法では学習用データの平均から離れないような制約がかけられており、元のデータからかけ離れた近似を行うおそれが低い。

## 第3章 Collaborative Mean Attraction 法

### 3.1 CMA 法

CMA 法[2]はカメラ間人物照合の問題に対して提案された手法である。カメラ間人物照合は、カメラ A で撮影された人物が、別の地点に設置されているカメラ B で撮影された人物像の中のどれと同一人物であるかを判定するものである (図 3-1)。カメラ B で撮影された人物像を学習用データのカテゴリとすると、人物がカメラ B の前を通過する間の画像のみしか得ることができないため、各カテゴリ内で大量の学習用データを得ることはできない。またカメラ B で撮影された人物数がカテゴリ数となるため、数十から数百カテゴリの分類問題となる。

カメラ間人物照合の対象は人物像であるのでカテゴリ間の違いは少なく、カテゴリ内の見えの違いは少ない。このような問題でテストデータを全学習用データで近似をする場合、学習用データから離れすぎた近似を行うと識別がうまくいかなくなる。CMA 法ではこの問題点を解決するため、学習用データの中心から離れすぎないように、正則化項で制約をする。これにより学習用データから離れすぎた近似を行うことなく、識別率が向上する。

### カメラ間人物照合

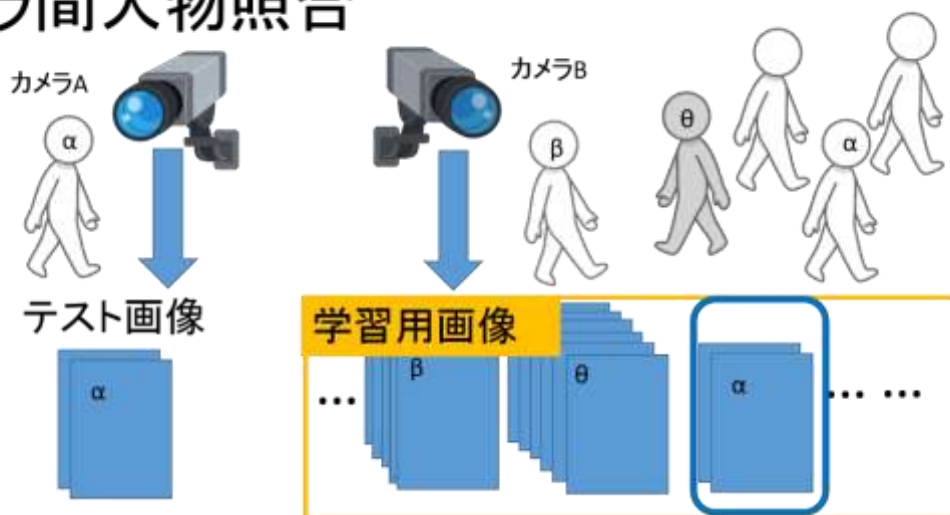


図 3-1 カメラ間人物照合

CMA 法は、未知のテストデータを複数の既知のカテゴリのいずれかに分類する。既知のカテゴリは、そのカテゴリに属する1枚以上の学習用データの特徴量集合として与えられる。テストデータも、テスト画像の特徴量として与えられるが、複数のテスト画像の特徴量を1組として、1組を一括して既知のカテゴリに分類することができる。これは、例えばカメラ間人物照合では、カメラを通過した際に撮影した一連の画像群をテストデータとして既知の人物のいずれかに一括して分類する場面に相当する。CMA 法は、テストデータの代表点を既知のカテゴリ全ての学習用データで近似する最適化段階と、この近似に最も貢献したカテゴリを選択する分類段階の2段階からなる（図 3-2）。

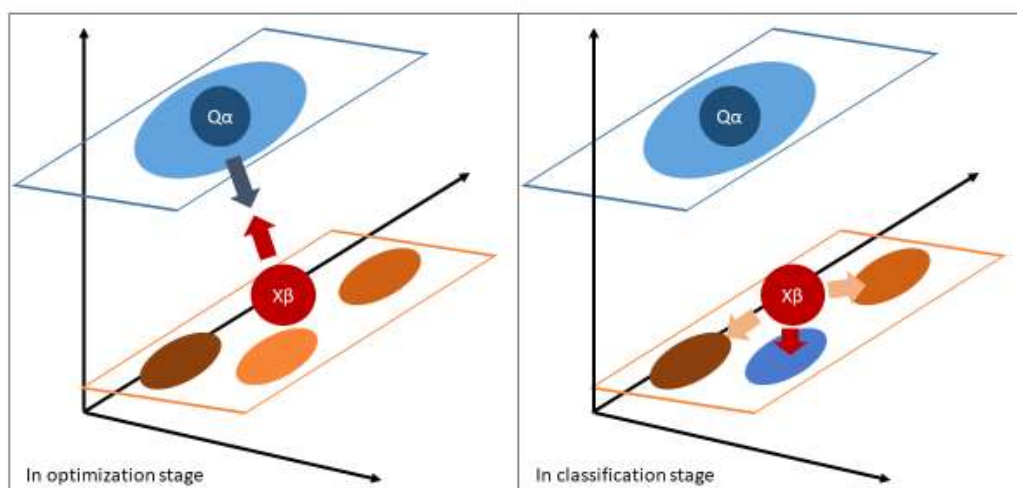


図 3-2 CMA 法の最適化段階と分類段階

### 3.1.1 最適化段階

CMA 法では、テストデータの代表点および、全ての学習用データによる近似点を、それぞれテストデータ、学習用データの線形和により表現する。最適化段階では、この代表点と近似点を構成する線形和の係数ベクトルを求める。この際、代表点と近似点との距離が小さくなるように係数ベクトルを定める。一方で、代表点や近似点がそれぞれのデータの性質を保持するよう、それぞれの平均点に近づけるという制約を課す。テストデータを並べた行列を  $\mathbf{Q} \in \mathbf{R}^{m \times N_q}$  ( $m$  は特徴量ベクトルの次元数、 $N_q$  は1組内のテスト画像の数)、既知のカテゴリ  $i \in \{1, \dots, n\}$  ( $n$  はカテゴリの種類数) の学習用データを並べた行列を  $\mathbf{X}_i \in \mathbf{R}^{m \times N_x^i}$  ( $N_x^i$  はカテゴリ  $i$  の学習用データ数)、全学習用データを並べた行列を  $\mathbf{X} \in \mathbf{R}^{m \times N_x}$  ( $N_x = \sum_i N_x^i$  は全学習用データ数) とする。この時、CMA 法の最適化段階では、以下の式  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  を最小化する係数ベクトル  $\boldsymbol{\alpha} \in \mathbf{R}^{N_q}, \boldsymbol{\beta} \in \mathbf{R}^{N_x}$  を求める。

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \|\mathbf{Q}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda_1 \left\| \boldsymbol{\alpha} - \frac{\mathbf{1}_{N_q}}{N_q} \right\|^2 + \lambda_2 \left\| \boldsymbol{\beta} - \frac{\mathbf{1}_{N_x}}{N_x} \right\|^2 \quad (5)$$

ここで、 $\mathbf{1}_k$  は各要素が 1 の  $k$  次元ベクトル、 $\|\cdot\|$  は、ベクトルの L2 ノルムである。式 (5) の  $\mathbf{Q}\boldsymbol{\alpha}, \mathbf{X}\boldsymbol{\beta}$  がそれぞれ代表点、近似点を表している。第1項は、代表点と近似点を近くするための項である。第2項、第3項は代表点、近似点をそれぞれの平均点に近くするための制約項（正則化項）である。 $\lambda_1, \lambda_2$  は正則化項の重みパラメータである。式 (5) は2次形式であり閉じた解を持つ。具体的には、下記の連立方程式を解けば良い。

$$(\mathbf{Q}^T \mathbf{Q} + \lambda_1 \mathbf{I}) \boldsymbol{\alpha} = \mathbf{Q}^T \mathbf{X} \boldsymbol{\beta} + \lambda_1 \frac{\mathbf{1}_{N_q}}{N_q} \quad (6)$$

$$(\mathbf{X}^T \mathbf{X} + \lambda_2 \mathbf{I}) \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Q} \boldsymbol{\alpha} + \lambda_2 \frac{\mathbf{1}_{N_x}}{N_x} \quad (7)$$

### 3.1.2 分類段階

最適化段階で求めた係数  $\alpha, \beta$  を利用して、近似点を構成する際に最も貢献したカテゴリを求め、そのカテゴリにテストデータを分類する。近似点の係数  $\beta$  の各次元が、各学習用データに対応している。学習用データは、既知のカテゴリのいずれかに含まれている。即ち、カテゴリ  $i$  の学習用データの係数を  $\beta_i$  とし、 $\beta^T = (\beta_1^T, \dots, \beta_n^T)$  と分解できる。この時、各カテゴリ  $i$  について下式を求め、最小となるカテゴリに分類する。

$$d^i = (\|Q\|_* + \|X_i\|_*) \cdot \frac{\|Q\alpha - X_i\beta_i\|^2 \|\beta\|}{\|\beta_i\|} \quad (8)$$

$\|Q\alpha - X_i\beta_i\|^2$  が小さい程、カテゴリ  $i$  のみで代表点の近似がより良く行えていることになる。また、 $\|\beta_i\|/\|\beta\|$  が大きい程（その逆数が小さい程）カテゴリ  $i$  の係数が既知カテゴリ全体の中で大きな役割を果たしていると言える。 $\|\cdot\|_*$  は、行列の核ノルム（行列の特異値の和）である。データのばらつきが大きい程、この値は大きくなる。カテゴリ間のばらつきの違いを考慮するための重み付けに利用している。

## 3.2 スパース最適化を用いた CMA 法(SCMA 法)

CMA 法の最適化段階にて、SRC 法で扱われていたスパース最適化を導入した SCMA 法を提案する。CMA 法では、学習用データの枚数が多くなると一般的な識別器と性能が変わらなくなるという欠点があった。学習用データの枚数が多い場合においても、最適な学習用データのみを用いてテスト画像を表現するために、スパース最適化を CMA 法に導入する。SCMA 法は、テストデータを既知のカテゴリ全ての学習用データで近似するスパース最適化段階と、この近似に最も貢献したカテゴリを選択する分類段階の 2 段階からなる。

### 3.2.1 スパース最適化段階

SCMA 法では、テストデータの近似点を、全学習用データの線形和により表現する。最適化段階では、近似点を構成する線形和の係数ベクトルを求める。この際、近似点との距離が小さくなるように係数ベクトルを定める。一方で、近似点がデータの性質を保持するよう、学習用データの平均点に近づけるという制約を課す。CMA 法では最適化の際 L2 ノルムを用いていたが、SCMA 法では L1 ノルム最適化を用いる。

テストデータを  $\mathbf{y} \in \mathbf{R}^m$  ( $m$  は特徴量ベクトルの次元数)、既知のカテゴリ  $i \in \{1, \dots, n\}$  ( $n$  はカテゴリの種類数) の学習用データを並べた行列を  $\mathbf{X}_i \in \mathbf{R}^{m \times N_x^i}$  ( $N_x^i$  はカテゴリ  $i$  の学習用データ数)、全学習用データを並べた行列を  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) \in \mathbf{R}^{m \times N_x}$  ( $N_x = \sum_i N_x^i$  は全学習用データ数) とする。この時、SCMA 法のスパース最適化段階では、以下の式  $f(\boldsymbol{\beta})$  を最小化する係数ベクトル  $\boldsymbol{\beta} \in \mathbf{R}^{N_x}$  を求める。

$$f(\boldsymbol{\beta}) = \|\mathbf{y} - \{(\mathbf{X} - \bar{\mathbf{X}})\boldsymbol{\beta} + \bar{\mathbf{x}}\}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_1 \quad (9)$$

ここで、 $\bar{\mathbf{x}}$  は、学習用データの平均列ベクトル、 $\bar{\mathbf{X}}$  は  $\bar{\mathbf{x}}$  を  $N_x$  個並べた行列である。式(9)の  $(\mathbf{X} - \bar{\mathbf{X}})\boldsymbol{\beta} + \bar{\mathbf{x}}$  が近似点を表している。第 1 項は、テストデータと近似点を近くし、かつデータの性質を保つため平均点から離れないようにするための項になっている。第 2 項は L1 ノルム正則化項となっている。 $\lambda_2$  は正則化項の係数である。SCMA 法ではスパース最適化を取り入れるため、 $\boldsymbol{\beta}$  に直接平均値に寄せる制約をかけていない。 $\boldsymbol{\beta}$  が 0 になると、近似点が平均点となるように第 1 項で表現されている。この最



適化問題を解くために Alternating Direction Method of Multipliers (ADMM)[22]を用いる。ADMM では以下の式を T 回繰り返し、 $\beta$ を求める。

$$\beta[t + 1] = (X^T X + \mu I)^{-1} (X^T y + \mu(z[t] - u[t])) \quad (10)$$

$$z[t + 1] = \text{Softmax}_{\left(\frac{\lambda}{\mu}\right)}(\beta[t + 1] + u[t]) \quad (11)$$

$$u[t + 1] = u[t] + (\beta[t + 1] - z[t + 1]) \quad (12)$$

式(10)~(12)において、 $\mu$ は最適化のパラメータで、 $z, u$ は補助変数である。Softmax 関数は図 3-3 のような値をとる関数である。

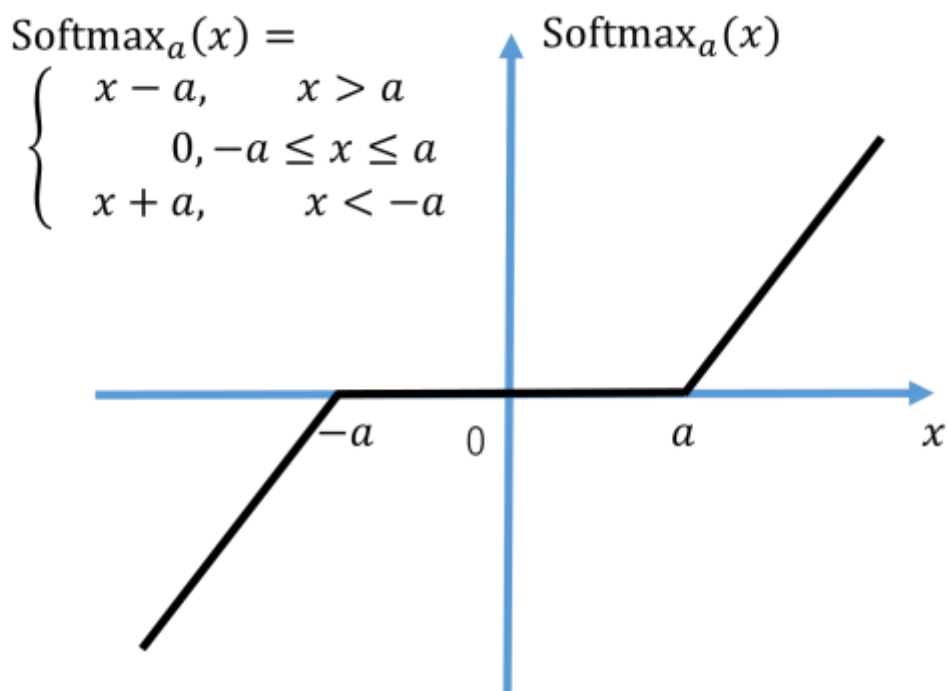


図 3-3 Softmax 関数

### 3.2.2 分類段階

最適化段階で求めた係数  $\boldsymbol{\beta}$  を利用して、近似点を構成する際に最も貢献したカテゴリを求め、そのカテゴリにテストデータを分類する。カテゴリ  $i$  の学習用データの係数を  $\boldsymbol{\beta}_i$  として、 $\boldsymbol{\beta}_i^T = (\beta_i^1, \dots, \beta_i^{N_x})^T$  とする。この時、各カテゴリ  $i$  について下式を求め、最小となるカテゴリに分類する。

$$d^i = - \sum_{j=1}^{N_x} \beta_i^j \quad (13)$$

SCMA 法では  $\boldsymbol{\beta}$  の値がスパースとなっている。  $\boldsymbol{\beta}$  に正値が入っているカテゴリは最適化段階での近似に貢献していたことを示す。CMA 法の分類段階では、 $d^i$  が最小となるカテゴリに分類していた。CMA 法と同様に  $d^i$  が最小となるカテゴリに分類できるようにするため、SCMA 法では  $\beta_i^j$  の総和に  $-1$  をかけた値を  $d^i$  としている。

## 第4章 評価実験

### 4.1 従来の識別器との比較実験

はじめに CMA、SCMA 法が一般画像分類において有効であることを示すために、従来手法である SVM、DL との比較実験を行う。カテゴリあたりの学習用データ枚数  $k$  を 1、2、4、8、16、30 枚と変化させて分類率の変化をみる。扱うデータセットは Caltech256 データセットで、257 カテゴリ全ての画像を使用する。DL と比較するために、同じ学習済みモデルを用いて実験を行う。この実験では AlexNet と呼ばれる構造をもった CNN に ILSVRC2012 の 1000 カテゴリ、130 万画像のデータを学習させたモデル[18]を利用する。本稿ではこのモデルを A モデルと呼ぶ。画像から特徴量を抽出する際、モデルに入力する画像は以下の手順で前処理を行う。

- 1) 縦横比を無視して  $256 \times 256$  に画像をリサイズする
- 2) 画像の中央の  $224 \times 224$  で切り取る

この切り取った画像をモデルに与える(図 4-1)。Caltech256 の画像データを A モデルに入力した際の、結合層(fc7 層)の出力をその画像の特徴量とする。1 枚の画像につき 4096 次元の特徴量が得られる。得られた特徴量は、全データの絶対値の最大値で割ることで、値域が  $[-1,1]$  の範囲に入るようにした。これは、識別器内での計算結果を安定させるためである。A モデルで得られた特徴量を特徴量 A と呼ぶ。

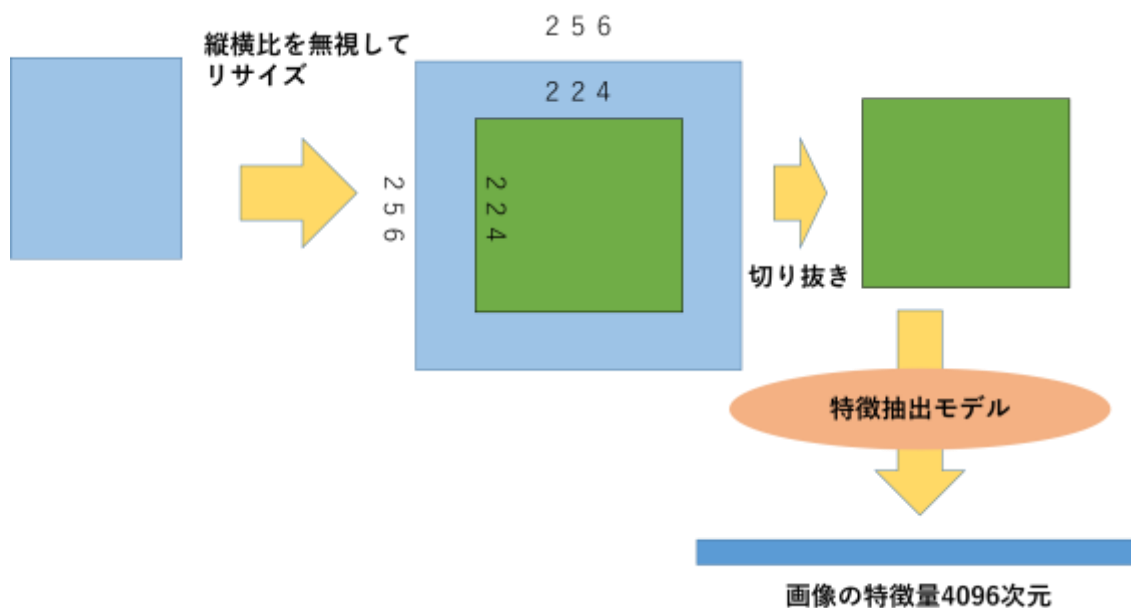


図 4-1 画像の特徴量抽出の前処理

DL では、特徴量抽出に用いた A モデルを元に、ファインチューニングを行った。A モデルの出力層のノード数を Caltech256 のカテゴリ数に合わせて変更したうえで、学習データを学習する。学習は 10000 回まで行う。この時点で残差はほぼ一定となっており、学習が収束していた。

SVM では線形カーネルを用いた。

CMA 法には  $\lambda_1, \lambda_2$  の 2 つのパラメータがある。テストデータ 1 枚を単独で分類する際には  $\lambda_1$  の値は、無視してもよい。CMA 法で種々のパラメータでの予備実験を行い、分類率が良かったパラメータ  $\lambda_2 = 3.0$  を用いた。SCMA 法のスパース最適化段階においてはパラメータを  $\lambda_2 = 1.0$ 、 $\lambda/\mu = 0.01$  とした。

実験に用いるデータの作成方法を図 4-2 に示す。実験では、Caltech256 データセットの各カテゴリからランダムに学習用データ  $k$  枚を抜き出し、これらを学習用データとする。それ以外の残りの画像すべてを評価用データとする。評価用データから一つずつデータを取り出しテストデータとして識別を行う。

各カテゴリについて評価用データに対する識別の正解率を求めた上で、それらを全てのカテゴリについて平均することで全体の正解率を求める。学習用データの選び方による認識率への影響を緩和するために、各  $k$  の値について学習用データの選び方を変えながら生成した 8 通りのパターンについてそれぞれで正解率を求め、その平均を実験結果とする。CMA、SCMA 法及び比較手法の間では、条件が同じになるよう、同じ 8 通りの学習用データを用いる。結果を図 4-3 に示す。

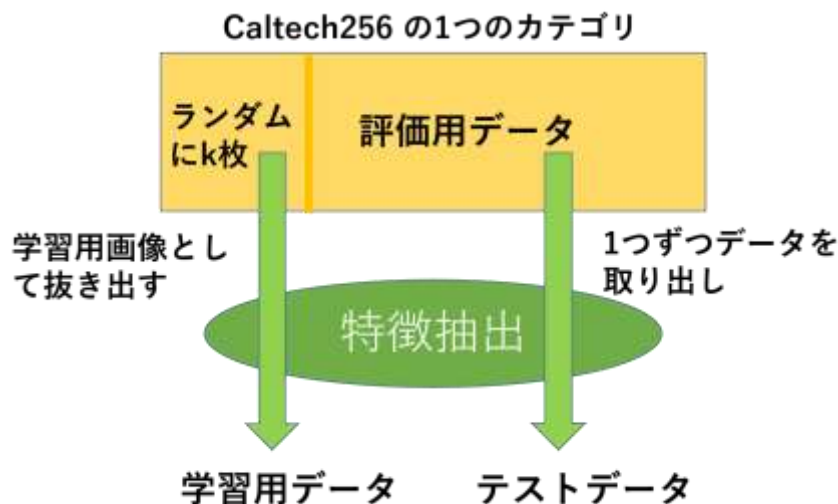


図 4-2 学習用データ、テストデータの作成手順

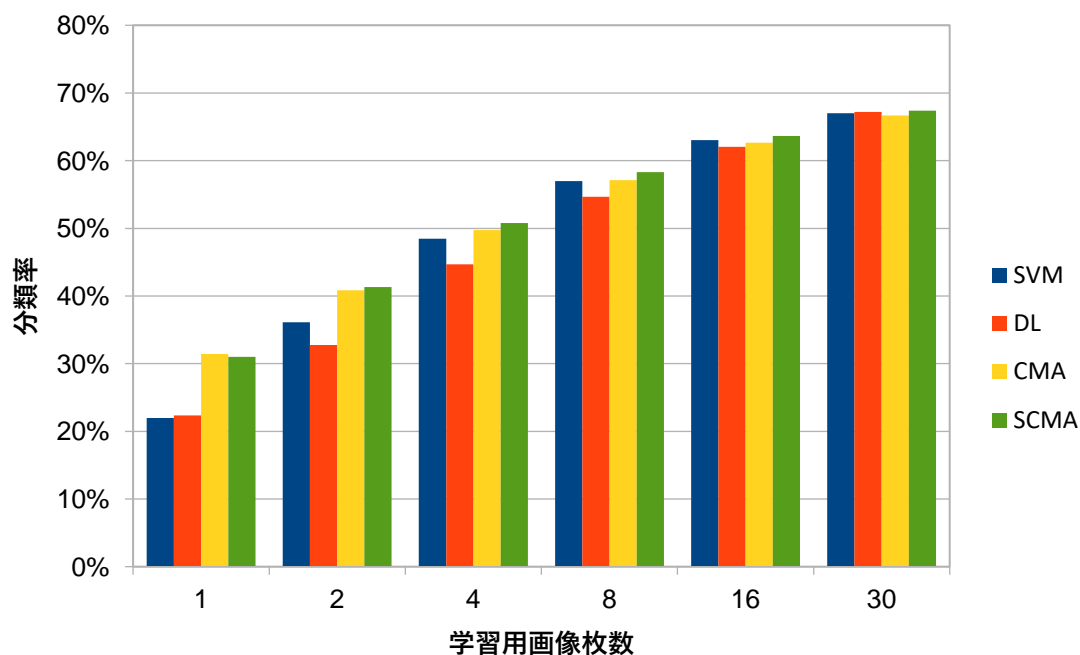


図 4-3 SVM、DL との比較実験結果 (Caltech256 使用)

図 4-3 より、 $k=1$ 、 $2$ 、 $4$ 、 $8$  では CMA 法が他の比較手法より高い正解率を示した。特に、 $k=1$  では、SVM、DL よりそれぞれ 9.4、9.1 ポイント、 $k=2$  ではそれぞれ 4.7、8.1 ポイント高い正解率を示しており、学習用データの数が少ない範囲で相対的により高い性能を示している。また SCMA 法において  $k=1$  の時は CMA 法と同程度であるが、 $k=2$ 、 $4$ 、 $8$ 、 $16$ 、 $30$  では CMA 法よりも高い正解率を示している。CMA 法の欠点である学習用データ枚数の増加による正解率の低下は、SCMA 法では低減できていることが分かる。また、SVM は、いずれの  $k$  でも SCMA 法を上回ることはなく、この実験では SCMA 法の方が、分類率が高かった。DL は学習用データの数が多き  $k=30$  では他の手法よりも高い正解率を示しているが、その差は大きくはない。

## 4.2 従来の疎・協働表現手法との比較実験

次に CMA、SCMA 法と疎・協働表現の従来手法である CRC、SRC、NRC 法との比較実験を行う。各カテゴリ当たりの学習用データ枚数  $k$  を 1、2、4、8、16、30 枚と変化させて分類率の変化をみる。この比較実験は Caltech101、Caltech256 の 2 つのデータセットで行う。

まず Caltech101 データセットでの実験について示す。Caltech101 を用いた実験では重複のあるカテゴリ”face”と背景画像を除外した 100 カテゴリの画像を使用する。特徴量の抽出では 4.1 節の A モデルに加え、VGG19 と呼ばれる構造を持った CNN に ILSVRC の 100 万画像のデータを学習させた、公開されている学習済みモデル[19]を使用する。本稿ではこのモデルのことを B モデルと呼ぶ。B モデルでも 4.1 節と同様に、画像から特徴量を抽出する際は、同様の前処理を行う。前処理を行った画像を B モデルに入力した際の、全結合層(fc18)の出力をその画像の特徴量とする。1 枚の画像につき 4096 次元の特徴量が得られる。得られた特徴量は、長さが 1 のベクトルになるように正規化している。B モデルで得られた特徴量を特徴量 B と呼ぶ。

4.1 節の実験と同様に学習用データと評価用データを作る。即ち、Caltech101 データセットの各カテゴリからランダムに学習用データを  $k$  枚抜き出し、これらを学習用データとする。それ以外の残りの画像すべてを評価用データとする。評価用データから一つずつデータを取り出し、テストデータとして識別を行う。

評価方法も 4.1 節と同様とする。即ち、各カテゴリについて評価用データに対する識別の正解率を求めた上で、それらを全てのカテゴリについて平均することで全体の正解率を求める。学習用データの選び方による認識率への影響を緩和するために各  $k$  の値について学習用データの選び方を変えながら生成した 8 通りのパターンについてそれぞれで正解率を求め、その平均を実験結果とする。CMA、SCMA 法及び比較手法の間では、条件が同じになるよう、同じ 8 通りの学習用データを用いる。

Caltech256 データセットでも 4.1 節の実験と同様にデータセットを扱う。特徴量の抽出は B モデルを使用する。特徴量の抽出については 4.1 節と同様とする。即ち、画像から特徴量を抽出する際は、前処理を行う。前処理を行った画像を B モデルに入力した際の、結合層(fc18)の出力をその画像の特徴量とする。得られた特徴量は、長さが 1 のベクトルになるように正規化している。

4.1 節の実験と同様に学習用データと評価用データを作る。即ち、Caltech256 データセットの各カテゴリからランダムに学習用データ k 枚を抜き出し、これらを学習用データとする。それ以外の残りの画像すべてを評価用データとする。評価用データから一つずつデータを取り出しテストデータとして識別を行う。

評価方法も 4.1 節と同様とする。即ち、各カテゴリについて評価用データに対する識別の正解率を求めた上で、それらを全てのカテゴリについて平均することで全体の正解率を求める。学習用データの選び方による認識率への影響を緩和するために各 k の値について学習用データの選び方を変えながら生成した 8 通りのパターンについてそれぞれで正解率を求め、その平均を実験結果とする。CMA、SCMA 法及び比較手法の間では、条件が同じになるよう、同じ 8 通りの学習用データを用いる。

Caltech101 での実験結果を図 4-4、図 4-5 に、Caltech256 での実験結果を図 4-6 に示す。

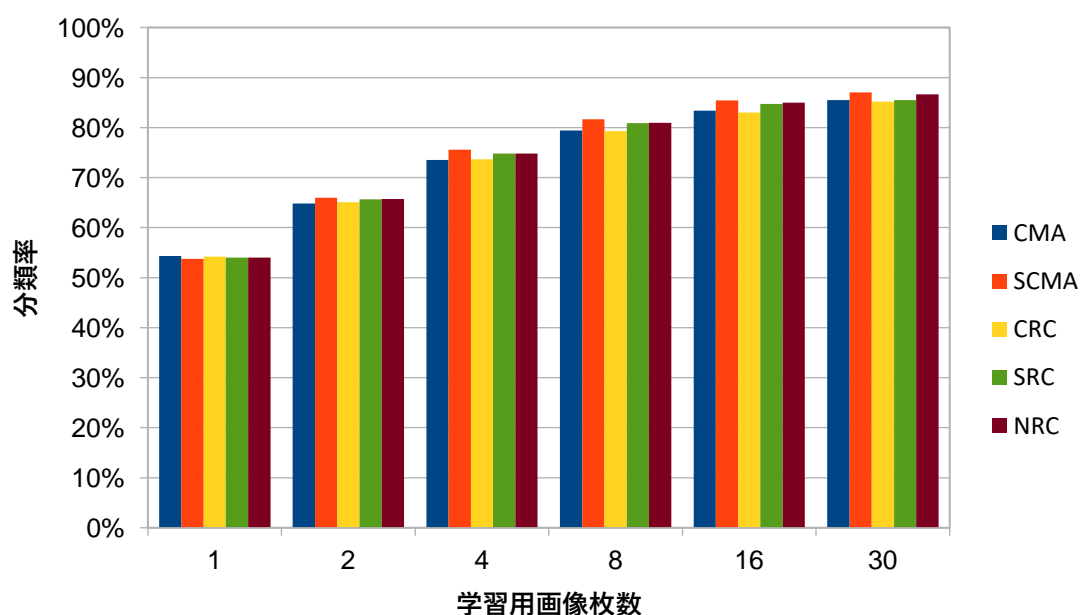


図 4-4 特徴量 A を用いた画像分類実験結果 (Caltech101 使用)

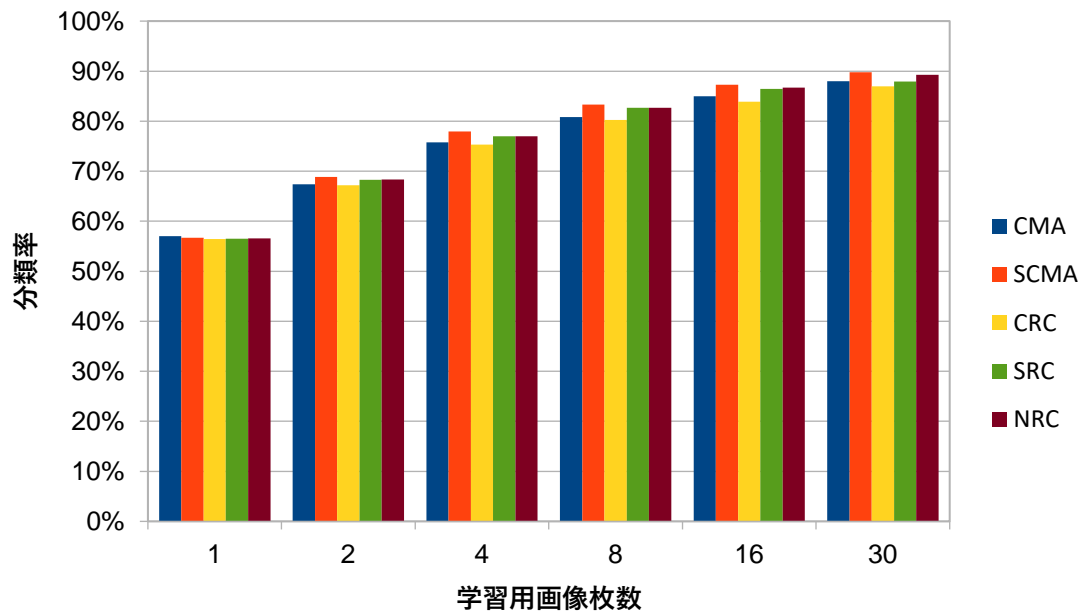


図 4-5 特徴量 B を用いた画像分類実験結果 (Caltech101 使用)

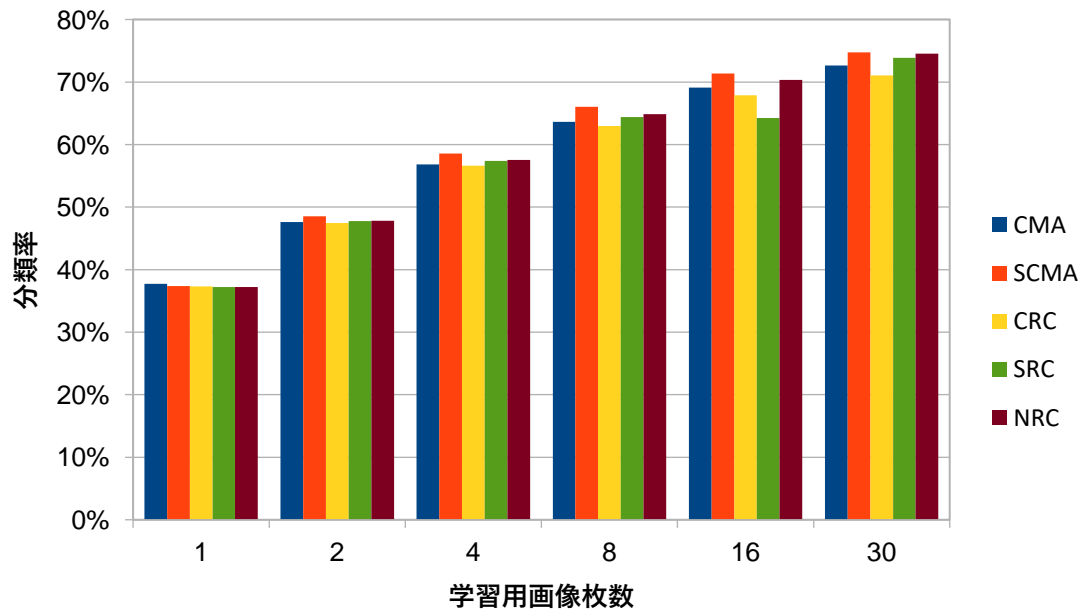


図 4-6 特徴量 B を用いた画像分類実験結果 (Caltech256 使用)



図 4-4、図 4-5 より CMA、SCMA 法と従来手法である CRC、SRC、NRC 法は同程度の分類率を持つことが分かる。特徴量 A、B での実験結果は同傾向であるが、特徴量 B での画像分類実験の分類率のほうが約 2 ポイント高かった。また、CMA、SCMA 法を比べると、学習用データ枚数が増えた  $k=2, 4, 8, 16, 30$  において、約 2 ポイント高くなっていることが分かる。このデータセットにおいても、CMA 法に比べて学習用データ枚数が増加した際正解率がより高くなるという SCMA 法の特徴が見える。

従来手法との比較をすると、 $k=2, 4, 8, 16, 30$  において疎表現手法である SRC、NRC、SCMA 法は、協働表現手法である CRC、CMA 法に対して認識率が高いことが分かる。疎表現手法は協働表現手法に比べて、学習用データ枚数が増加した際正解率がより高くなることが分かる。

図 4-5、図 4-6 より、全手法においてデータセットを変更しても、分類率の傾向はあまり変わらないことが分かる。図 4-5、図 4-6 においても、疎表現手法のほうが協働表現手法よりも比較的認識率が高いことが分かる。だが CMA、SCMA 法と従来手法は同程度の分類率であることが分かる。

NRC の文献[5]では、CRC、SRC、NRC 法で画像分類の比較実験を行っており、そこでは NRC 法が比較手法よりも特に正解率が高いと記されていた。NRC の文献でのデンスサンプリングを用いた特徴量と同様の手順で抽出した特徴量を使用し、Caltech256 データセットで実験を行い、CMA、SCMA 法と従来手法は同程度の分類率であるかを調べた。

Caltech256 データセットでは、全 257 カテゴリの画像を使用する。この実験では特徴抽出に B モデルを使用する。デンスサンプリングは以下の手順で行う(図 4-7)。

- 1) 縦横比を変えずに画像の短辺が 256 になるようにリサイズする
- 2) この画像に対して  $224 \times 224$  のサイズでストライドが 15 で画像を切り出していく
- 3) それらをフリップ (左右反転) したのも切り出していく
- 4) 切り出されたこれらの画像を 1 つずつ B モデルに与え、得られた特徴量の平均をとる
- 5) この処理を画像の短辺が 384 になるようにリサイズしたものに對しても行う
- 6) 得られた 2 つの値を平均して画像の特徴量とする。

ここで得られた特徴量を、長さが 1 のベクトルになるように正規化している。デンスサンプリングを用いて B モデルで抽出した特徴量を特徴量 B2 と呼ぶ。

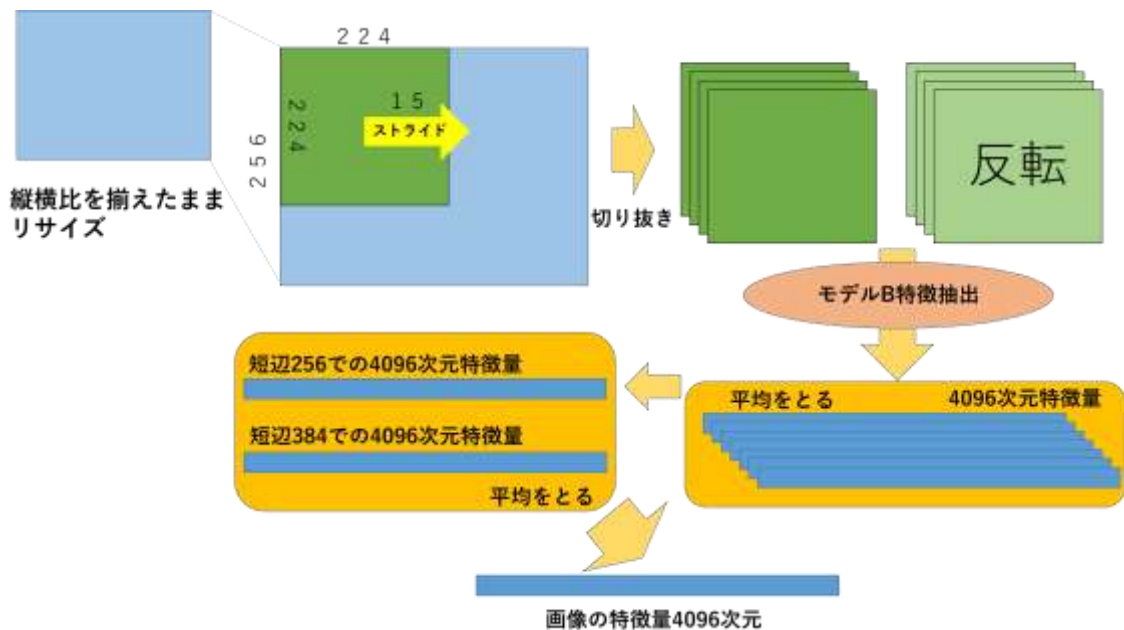


図 4-7 デンスサンプリングを用いた画像の特徴量抽出の手順

4.1 節の実験と同様に学習用データと評価用データを作る。即ち、Caltech256 データセットの各カテゴリからランダムに学習用データ  $k$  枚を抜き出し、これらを学習用データとする。それ以外の残りの画像すべてを評価用データとする。評価用データから一つずつデータを取り出しテストデータとして識別を行う。

評価方法も 4.1 節と同様とする。即ち、各カテゴリについて評価用データに対する識別の正解率を求めた上で、それらを全てのカテゴリについて平均することで全体の正解率を求める。学習用データの選び方による認識率への影響を緩和するために各  $k$  の値について学習用データの選び方を変えながら生成した 8 通りのパターンについてそれぞれで正解率を求め、その平均を実験結果とする。CMA、SCMA 法及び比較手法の間では、条件が同じになるよう、同じ 8 通りの学習用データを用いる。実験結果を図 4-8 に示す。

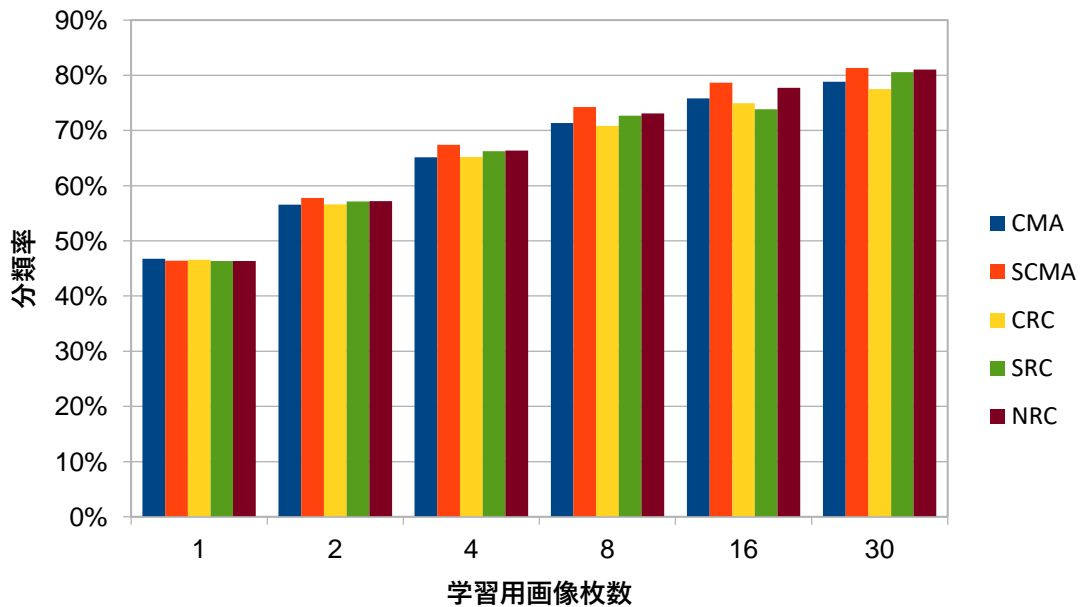


図 4-8 特徴量 B2 を用いた画像分類実験結果 (Caltech256 使用)

実験結果では CMA、SCMA 法と従来手法である CRC、SRC、NRC 法は同程度の分類率を持つことが分かる。特徴量 B、B2 での実験結果は同傾向であるが、特徴量 B2 での画像分類実験の分類率のほうが  $k=1, 2$  では約 9 ポイント、 $k=4, 8$  では約 8 ポイント、 $k=16, 30$  では約 7 ポイント高かった。デンスサンプリングの効果が出ていることが分かる。また、CMA、SCMA 法を比べるとここでも学習用データ枚数が増えた  $k=2, 4, 8, 16, 30$  において、SCMA 法が約 2 ポイント高くなっていることが分かる。また学習用データ枚数が増えるごとに、この 2 手法の分類率の差が出ていることが分かった。このデータセットにおいても、SCMA 法の特徴が見える。従来手法の NRC 法は SRC、SCMA 法と同程度の分類率を持つ。疎表現手法の SRC、NRC、SCMA 法は同程度の分類率であることが分かる。また協働表現手法の CRC、CMA 法に対して疎表現手法での分類率は高いことが分かる。

### 4.3 複数テストデータの組での実験

CMA 法の特長の一つは、複数のテストデータを1組として与え、1組を一括して識別できる点が挙げられる。本節では、テストデータが複数の画像からなる場合についての実験を行う。テストデータの枚数  $j$  と学習用データの枚数  $k$  を変化させた時、識別結果がどのように変化するか調査する。

この実験では Caltech256 データセットを用いた。4.1 節と同様にして Caltech256 の全カテゴリを実験に用いた。特徴量も 4.1 節と同様にモデル A を用いたものとなっている。

複数テストデータの実験における学習用データとテストデータの作成手順を図 4-9 に示す。各カテゴリ  $k$  個の学習用データを作成し、残りを評価用データとする。評価用データからランダムに  $j$  枚ずつデータを取り出し、それらを1組のテストデータとして識別を行う。これを評価用データがなくなるまで (残りが  $j$  枚未満となるまで) 繰り返す。テストデータを1つも取り出せないカテゴリについては、評価対象から除く。テストデータの枚数は  $j = 2, 4, 8, 16, 30$  とする。テストデータの組み合わせによる変動の影響を緩和するために、同じ  $j$  について3回の試行を行い、その平均を求める。さらに、4.1 節と同様に同じ  $k$  について8通りの試行を行いその平均を実験結果とする。学習用データの枚数は  $k = 1, 2, 4, 8, 16, 30$  とする。

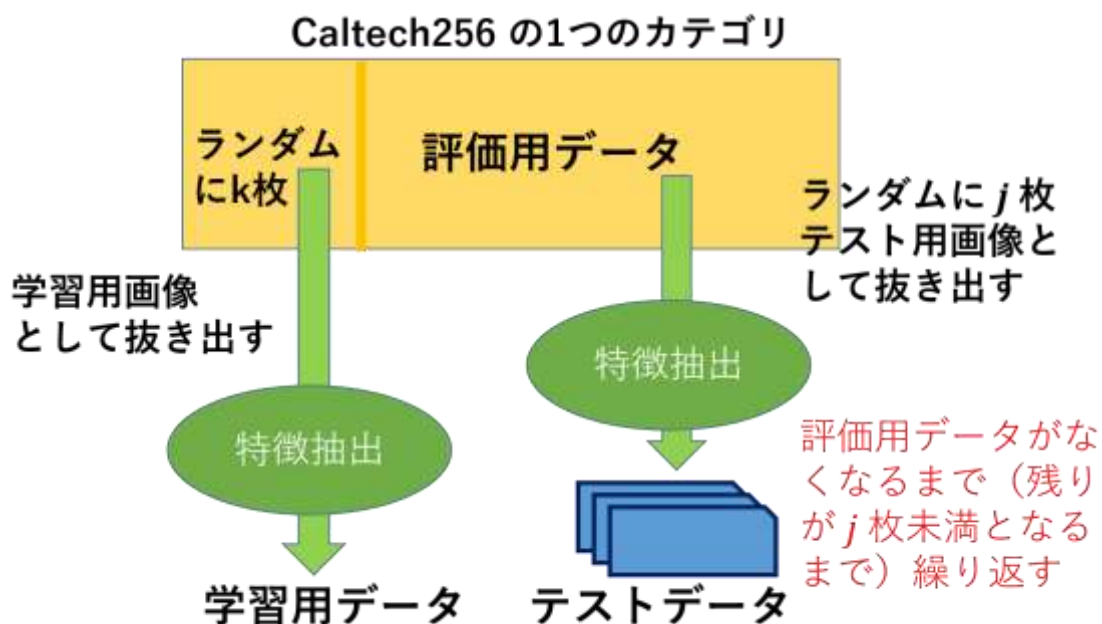


図 4-9 複数テストデータの実験における学習用データとテストデータの作成手順

組で与えたテストデータについて、それらの線形結合で代表点を求め、分類した本来の CMA 法の結果を set とする。比較のために、 $j$  枚のテストデータそれぞれについて、CMA 法で単独テストデータとして式(8) の  $d^i$  を求め、 $j$  枚全ての中で最小となった  $d^i$  のカテゴリに分類した結果を group とする。

実験結果を図 4-10~図 4-15 に示す。複数テストデータを組で与えることにより、テストデータを単独で与えた時(図 4-3)よりも全体に正解率が大きく向上している。特に、set の結果において、学習用データ数  $k=16$  以上、テストデータ枚数  $j=8$  以上では、95 %以上の分類率となっている。

set と group で比較を行うと、group の手法が set の手法よりも分類率が低いことが分かる。set の手法では、組で与えたテストデータについて、それらの線形結合で代表点を求めている。このことが、テストデータをそれぞれ評価した結果の中から最大値を選ぶ方法よりも有効に働いていると言える。

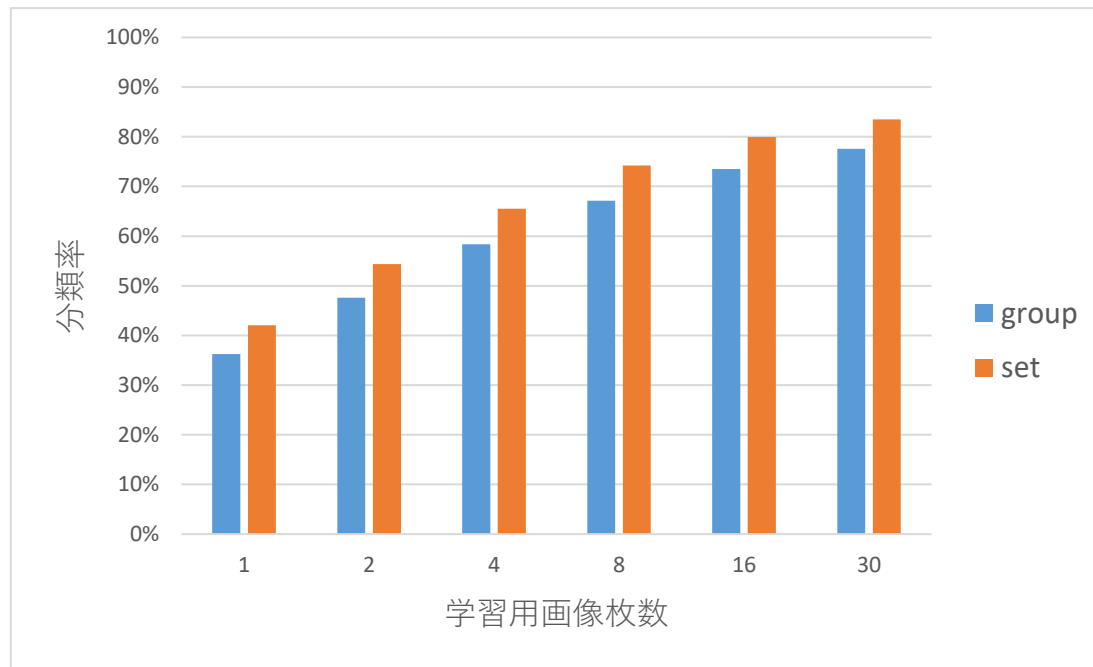


図 4-10 複数テストデータでの画像分類実験結果 ( $j=2$ )

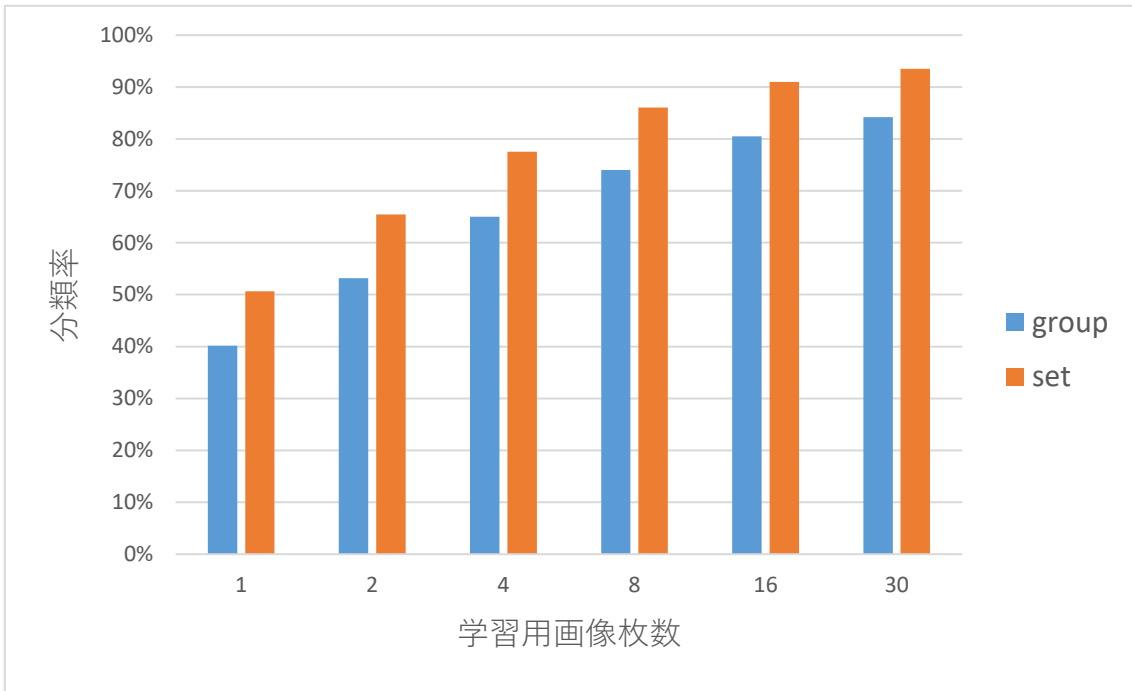


図 4-11 複数テストデータでの画像分類実験結果 ( $j = 4$ )

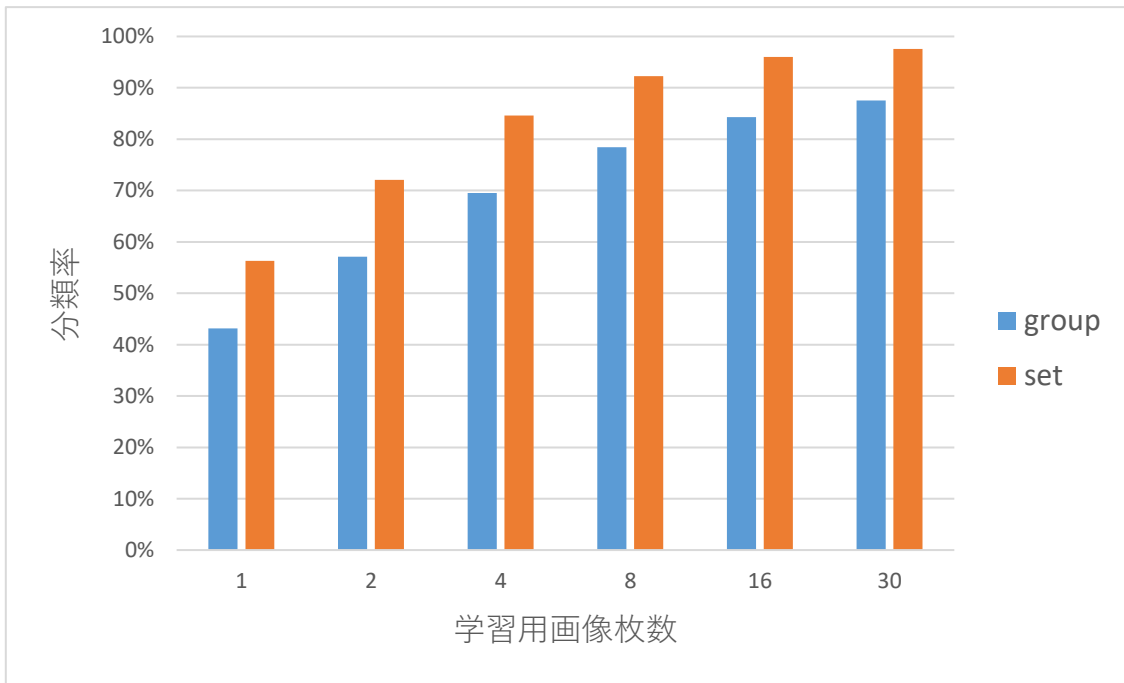


図 4-12 複数テストデータでの画像分類実験結果 ( $j = 8$ )

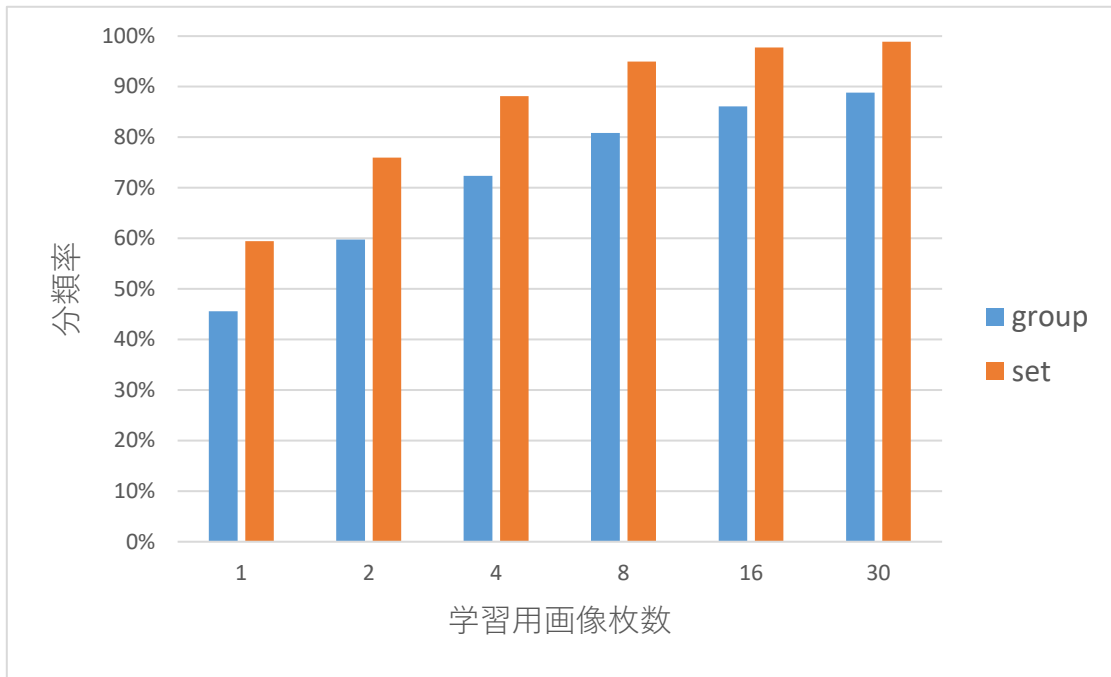


図 4-13 複数テストデータでの画像分類実験結果 ( $j = 16$ )

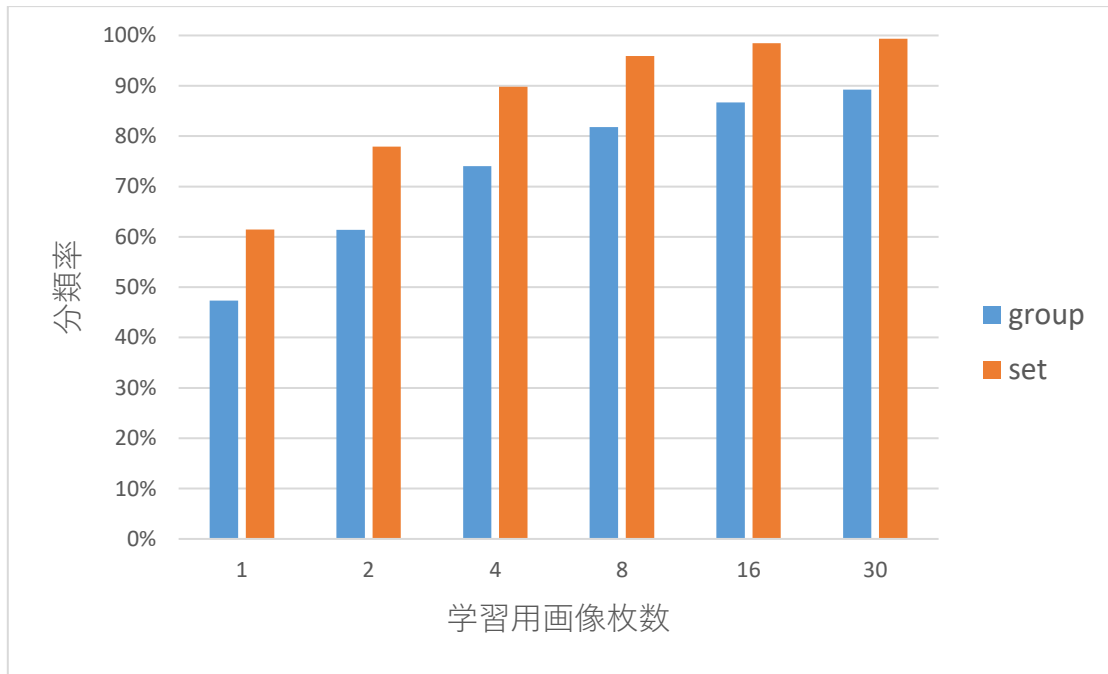


図 4-14 複数テストデータでの画像分類実験結果 ( $j = 30$ )

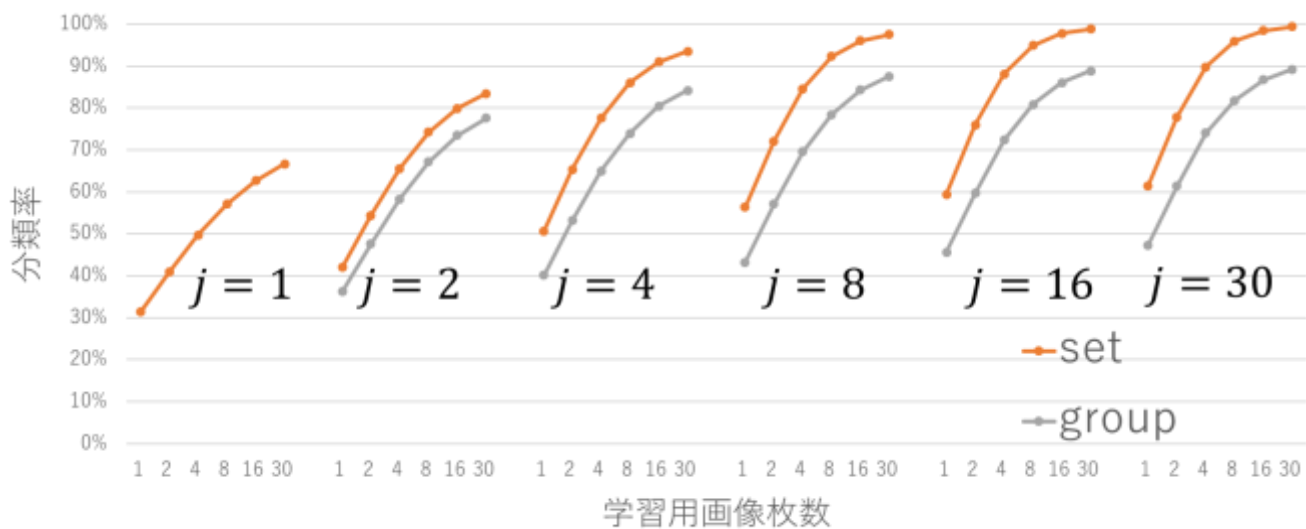


図 4-15 複数テストデータでの画像分類実験結果 ( $j = 1, 2, 4, 8, 16, 30$ )



## 第5章 結論

本研究では、これまでカメラ間人物照合に適用されてきた CMA 法が、一般画像分類にも有効であるか検証した。さらに、CMA 法に対してスパース最適化を導入した SCMA 法を提案した。一般画像分類において従来の機械学習手法との比較を行い、少数学習用データ枚数下での CMA、SCMA 法の有効性を示した。

CMA 法は学習用データを全て用いてテストデータを表現する協働表現手法の一種であり、少数学習用データ枚数下でも比較的高い認識率を持つ。しかし CMA 法は協働表現手法であり、すべての学習用データを同様に利用する傾向がある。そのため、学習用データが増えた場合、CMA 法は過度に学習用データに適応する恐れがある。この問題を解消するためにスパース最適化を導入した SCMA 法を提案した。

CMA、SCMA 法は従来の識別器である SVM や DL に対して、高い認識率を持ち、学習用データ枚数が少ない範囲で相対的に高い性能を示した。SCMA 法は CMA 法に比べて学習用データ枚数が増加した場合、正解率がより向上できていた。

Caltech データセットではクラス数は多いが、同じクラスの物体の向きが同じであったり、画像内に対象となる物体が中心にあるデータが多い。Pascale Visual Object Classes (VOC) データセットは、クラス数が少ないが、Caltech データセットよりも一般的な状況での画像が扱われている。このようにデータセットごとに性質の違いがあるため、他のデータセットにおいても、CMA、SCMA 法で画像分類の実験を行うことを今後の課題とする。

疎・協働表現手法の分類段階は、本論文で 3 パターンあげている。これらは最適化段階の結果と任意に組み合わせることができる。最適化段階と分類段階の組み合わせの違いで分類率がどう変化するかを調査することも今後の課題とする。

## 謝辞

本研究を進めるにあたり、ご指導いただいた修士論文指導教員の椋木雅之教授に感謝いたします。指導教員である椋木教授には、本研究や本論文の作成に関して様々のご指導を頂きました。学部4年ころから実験を行うコツや、研究の進捗についての相談に乗っていただき、研究を進めることができました。その結果多くの学会に参加することができ、人生初の海外に行くこともできました。この3年間とても貴重な体験ができました。3年間ありがとうございました。

廿日出教授、井上助教には、本修士論文の副査を務めていただきました。お忙しい中、貴重なお時間を頂き、ありがとうございます。

椋木研究室の皆様には、日々研究を進めるにあたって、様々な助言をいただきました。皆さんのおかげで充実した研究生生活を送ることができました。皆さんありがとうございました、またどこかで会いましょう。

## 参考文献

- [1] 柳井啓司, ”一般物体認識の現状と今後”, 情報処理学会論文誌コンピュータビジョン・イメージメディア, vol.48, no.SIG16(CVIM19), pp.1-24, 2007.
- [2] Y.Wu, M.Mukunoki, M.Minoh, ”Collaborative Mean Attraction for Set Based Recognition”, MIRU, 2014.
- [3] L.Zhang, M.Yang, X.Feng, ”Sparse Representation or collaborative representation: which helps face recognition?”, ICCV, pp.471-478, 2011.
- [4] J.Wright, A.Y. Yang, A. Ganesh, ”Robust face recognition via sparse representation”, Transactions on Pattern Analysis and Machine Intelligence, vol.31, issue2, pp.210 – 227, 2009
- [5] J.Xu, W.An, L.Zhang, D.Zhang, ”Sparse, Collaborative, or Nonnegative Representation: Which Helps Pattern Classification?”, arXiv:1806.04329v2 [cs.CV], 2018
- [6] W.Liu, D.Anguelov, D.Erhan, C.Szegedy, S.Reed, C.Fu, A.C. Berg, ”SSD: Single Shot MultiBox Detector”, arXiv:1512.02325v5 [cs.CV], 2016
- [7] M.Thoma, ”A Survey of Semantic Segmentation”, arXiv:1602.06541v2 [cs.CV], 2016
- [8] S.Razavian, A.Azizpour, H.Sullivan, ”CNN features off-the-shelf: an astounding baseline for recognition”, arXiv:1403.6382 [cs.CV], 2014
- [9] 小野田崇, 人工知能学会: ”知の科学 サポートベクターマシン”, オーム社, 4章, 2007
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. FeiFei, ”ImageNet : A large-scale hierarchical image database.”, CVPR, pp. 2–9, 2009
- [11] ILSVRC <http://www.image-net.org/challenges/LSVRC/>
- [12] Caltech101  
[http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)
- [13] Caltech256  
[http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/)
- [14] D. Lowe, ”Object recognition from local scale invariant features.” ICCV, vol.2, pp.1150–1157, 1999
- [15] N. Dalal and B. Triggs, ”Histograms of Oriented Gradients for

- Human Detection.” CVPR, 2005
- [16] A. Krizhevsky, I. Sutskever, G.E. Hinton, ”ImageNet Classification with Deep Convolutional Neural Networks”, NIPS, pp. 1097-1105, 2012
- [17] L.Zheng, Y.Zhao, S.Wang, J.Wang, Q.Tian, ”Good Practice in CNN Feature Transfer”, arXiv:1604.00133 [cs.CV],2016
- [18] AlexNet 学習済みモデル  
[github.com/BVLC/caffe/tree/master/models/bvlc\\_reference\\_caffenet /](https://github.com/BVLC/caffe/tree/master/models/bvlc_reference_caffenet/)
- [19] K.Simonyan, ”Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv:1409.1556v6 [cs.CV] , 2015
- [20] VGG19 学習済みモデル  
<https://gist.github.com/ksimonyan/3785162f95cd2d5fee77>
- [21] B.Wang, W.Li, N.Poh, Q.Liao, ”Kernel collaborative representation-based classifier for face recognition”, International Conference on Acoustics, Speech and Signal Processing, pp. 2877-2881, 2013
- [22] ADMM  
[http://web.stanford.edu/~boyd/papers/pdf/admm\\_slides.pdf](http://web.stanford.edu/~boyd/papers/pdf/admm_slides.pdf)