

平成 28 年度卒業論文

Collaborative Mean Attraction 法による  
画像分類の特性調査

宮崎大学 工学部 情報システム工学科  
荻原 弘樹

指導教員 椋木雅之

## 目次

1. はじめに.....	1
2. CMA 法.....	2
2.1. 最適化段階.....	3
2.2. 分類段階.....	5
3. 画像分類.....	6
3.1. データセット.....	6
3.2. 特徴抽出.....	7
3.3. 比較手法.....	8
4. 実験.....	11
4.1. 画像分類の実験.....	11
4.1.1. 単独テストデータでの実験.....	11
4.1.2. 複数テストデータでの実験.....	14
4.2. パラメータ $\lambda_1, \lambda_2$ の調節実験.....	17
5. 終わりに.....	19

## 1. はじめに

近年、携帯端末の普及により個人が多数の画像を所有するようになってきた。一般にこれらの画像には様々な被写体の画像が含まれている。これらの画像を個人の観点に基づいて自動分類したいという要望が生じている（図 1）。この場合、分類したいカテゴリの訓練用画像は比較的少数であることが望ましい。訓練用画像が比較的少数であるという事は、分類を行う際コストが少ないという点で有用である。このような画像分類においては、比較的少数の訓練用画像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。

少数訓練用画像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 Collaborative Mean Attraction (CMA) 法[1]が提案された。

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 法を一般物体認識の一種である画像分類に適用可能であるか調査する。また CMA 法の特性を解明する為に 2 種類の実験を行う。

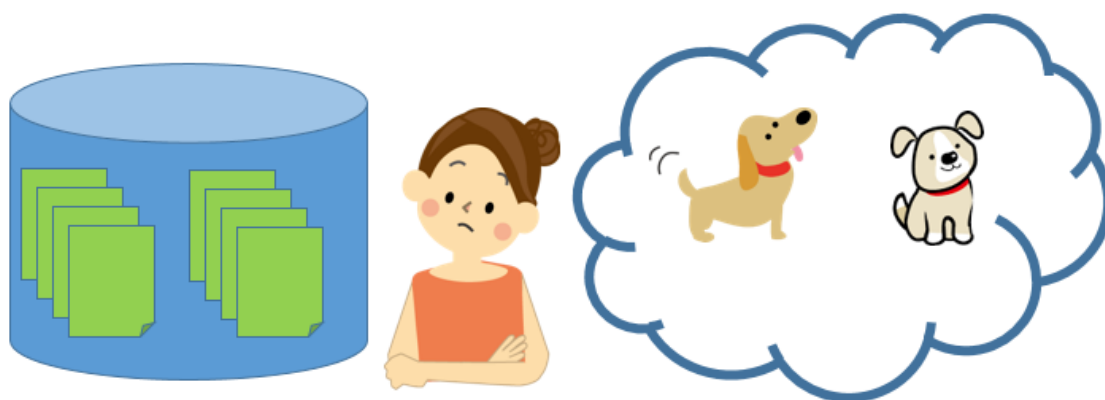


図 1 画像の自動分類の要望例

## 2. CMA 法

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識別手法である。既知のカテゴリは、そのカテゴリに属する 1 枚以上の訓練用画像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、組を一括して既知のカテゴリに分類することができる。CMA 法は、テストデータの代表点を既知のカテゴリ全ての訓練用画像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 2 に CMA 法のイメージ図を示す。

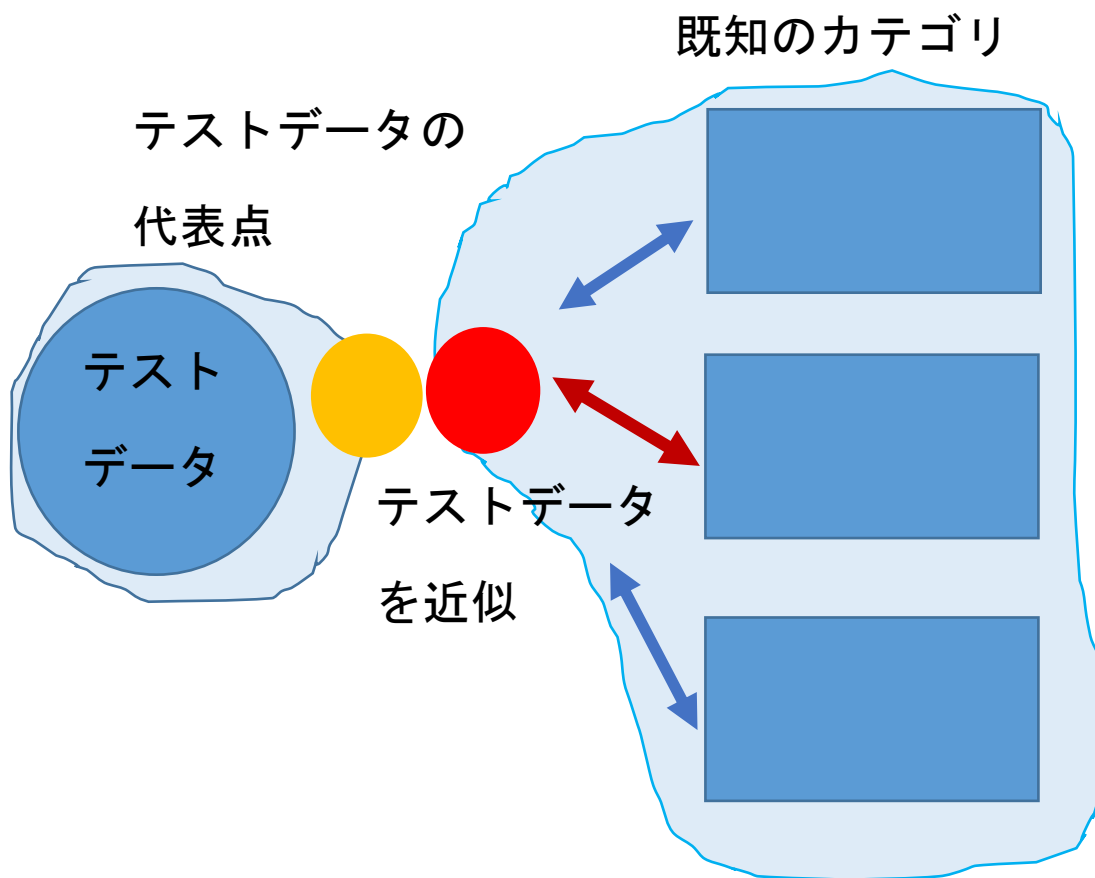


図 2 CMA 法のイメージ図

## 2.1. 最適化段階

CMA 法では，テストデータの代表点および既知のカテゴリ全ての訓練用画像による近似点を，それぞれテストデータの特徴量，訓練用画像の特徴量の線形和により表現する．最適化段階では，代表点と近似点を構成する線形和の係数を求める．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（図 3）．

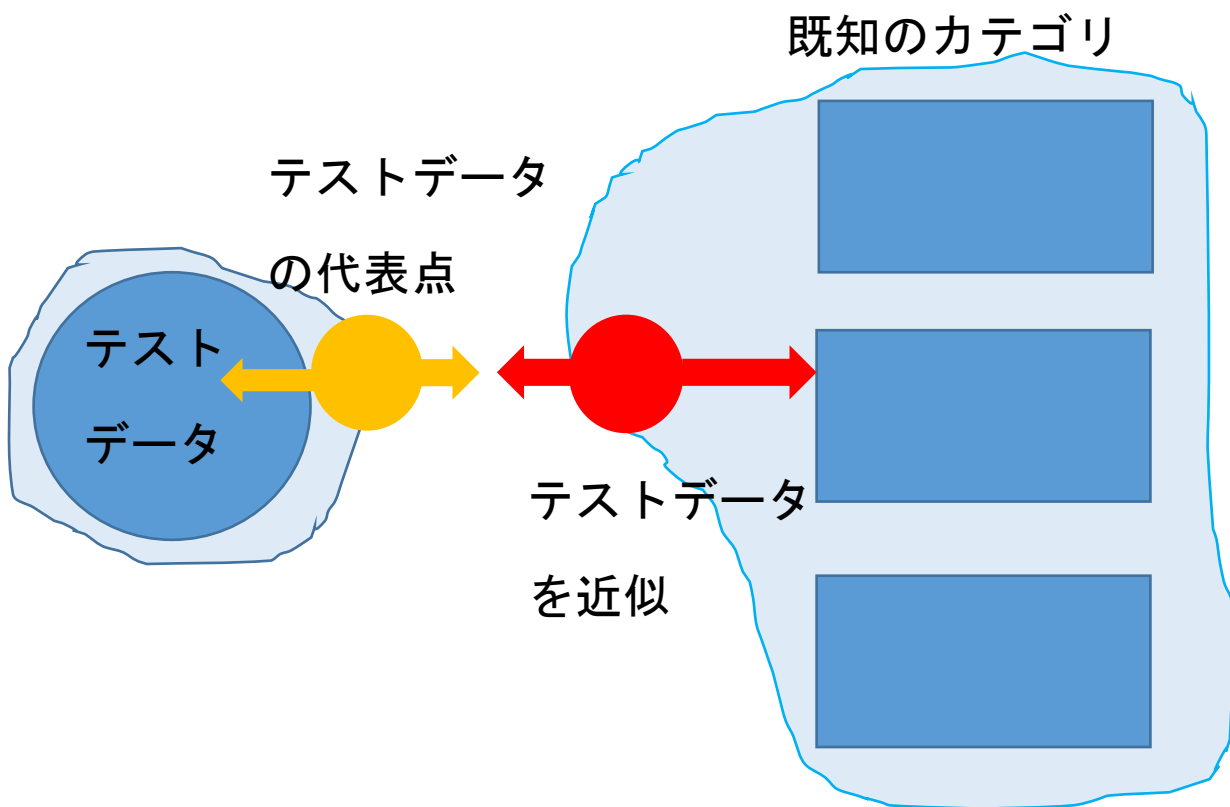


図 3 最適化段階

テスト画像の特徴量ベクトルを並べた行列を  $\mathbf{Q} \in R^{m \times N_q}$  ( $m$  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) の訓練用画像の特徴量ベクトルを並べた行列を  $\mathbf{X}_i \in R^{m \times N_x^i}$  ( $N_x^i$  はカテゴリ  $i$  の訓練用画像数), 全訓練用画像の特徴量ベクトルを並べた行列を  $\mathbf{X} = (\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_n) \in R^{m \times N_x}$  ( $N_x = \sum_i N_x^i$  は全訓練用画像数) 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  $\boldsymbol{\alpha}, \boldsymbol{\beta} \in R^m$  を求める.

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

ここで,  $\mathbf{1}_N$  は各要素が 1 の  $N$  次元ベクトル,  $\|\cdot\|$  は, ベクトルの  $L_2$  ノルムである. 式(1) 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.

## 2.2. 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。カテゴリ  $i$  の訓練用画像の係数を  $\beta_i$  として、 $\beta^T = (\beta_1^T, \dots, \beta_n^T)$  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。

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

$\|Q\alpha - X_i\beta_i\|^2$  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. 画像分類

#### 3.1. データセット

手法の評価には Caltech101 データセット[2]と Caltech256 データセットを用いた (表 1).

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 (図 4). Caltech101 データセットは 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. 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.

Caltech101 を用いた実験では重複のあるカテゴリ “face”, 背景画像のカテゴリ “background” を除外した 100 カテゴリの画像を使用する. Caltech256 を用いた実験では全てのカテゴリの画像を使用する.

表 1 使用したデータセット

	Caltech101	Caltech256
画像総数	9144 枚	30607 枚
カテゴリ数	101 カテゴリ	257 カテゴリ
1 カテゴリの	31 枚	80 枚
最小画像数		
使用カテゴリ	100 カテゴリ	257 カテゴリ





図 4 Caltech101

### 3.2. 特徴抽出

画像分類では、画像から抽出した特徴量に基づいて分類を行う。特徴抽出には、Deep Learning フレームワークである Caffe [3] で ImageNet 画像 [4] を学習したモデル (bvlc\_reference\_caffenet [5]) を用いる。このモデルは、ILSVRC2012 の 1,000 カテゴリ、120 万画像の学習用データを用いて構築されたもので、学習済みのモデルが公開されている。このモデルに画像を与えると結合層の出力として与えられた画像の特徴量が出力される (図 5)。この特徴量の次元数は 4,096 次元となる。実装には、OpenCV 3.1 の Deep Neural Network Module を用いた。

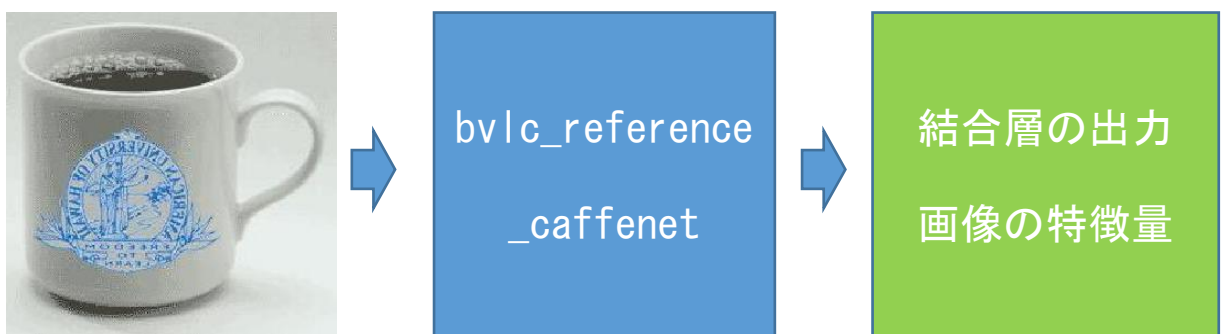


図 5 特徴抽出

### 3.3. 比較手法

CMA 法の特性を考察するために3つの従来手法と比較する。1つ目はサポートベクターマシンを使った方法 (SVM) である。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6)。

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」法を用いた。また、カーネルには線形カーネルを用いた。SVM は現在のパターン認識の標準的な手法である。実装には libsvm[6] を利用した。

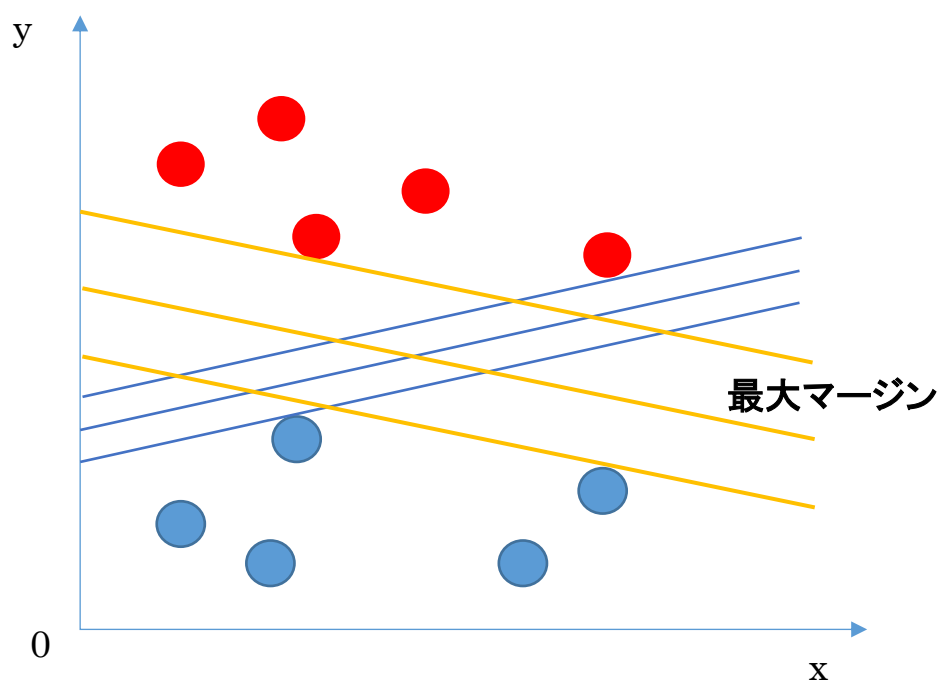


図 6 SVM のマージン最大化

2つ目は、重心間距離（Center Point Distance; CPD）を用いた手法である（図7）。あるカテゴリの訓練用データの重心とテストデータの重心との距離を、そのカテゴリとテストデータとの距離とする。その距離が最小となるカテゴリにテストデータを分類する。CPDは群間の距離を求める比較的単純な手法である。

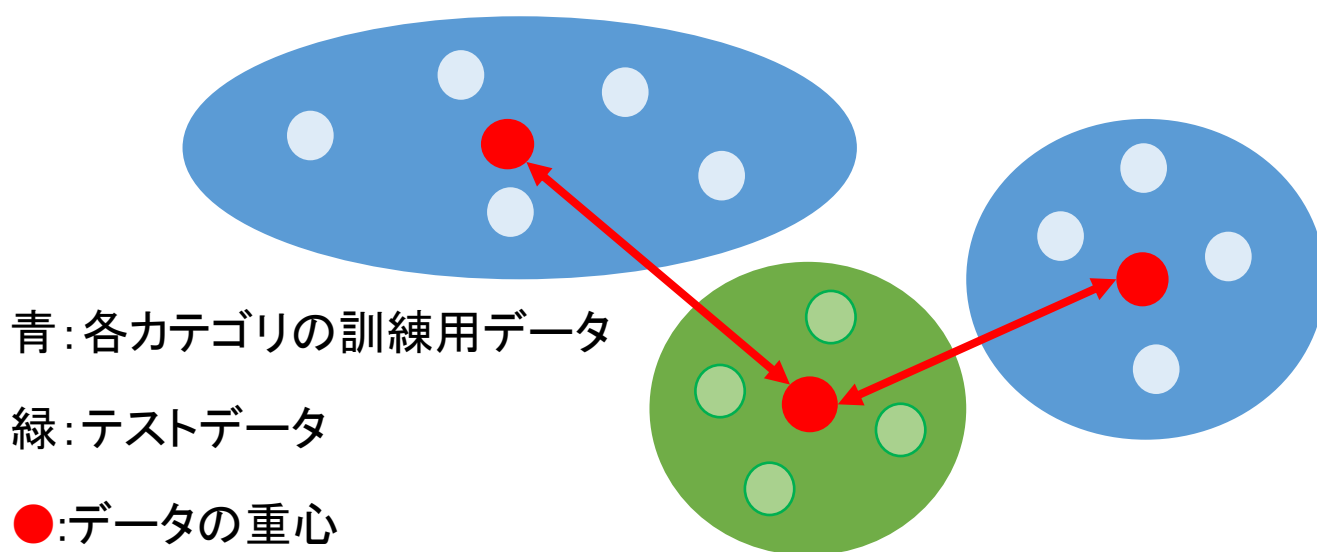


図7 CPDの説明

3つ目は、データ間の最小距離（Minimum Point Distance; MPD）を用いた手法である（図8）。あるカテゴリの訓練用データとテストデータ間で、全てのデータの組み合わせについて2点間の距離を求める。距離の中で最小値をそのカテゴリとテストデータの距離とする。この距離が最小となるカテゴリにテストデータを分類する。MPDは個々のデータ点を元にしており、データの広がりや形状を反映した分類が行える。

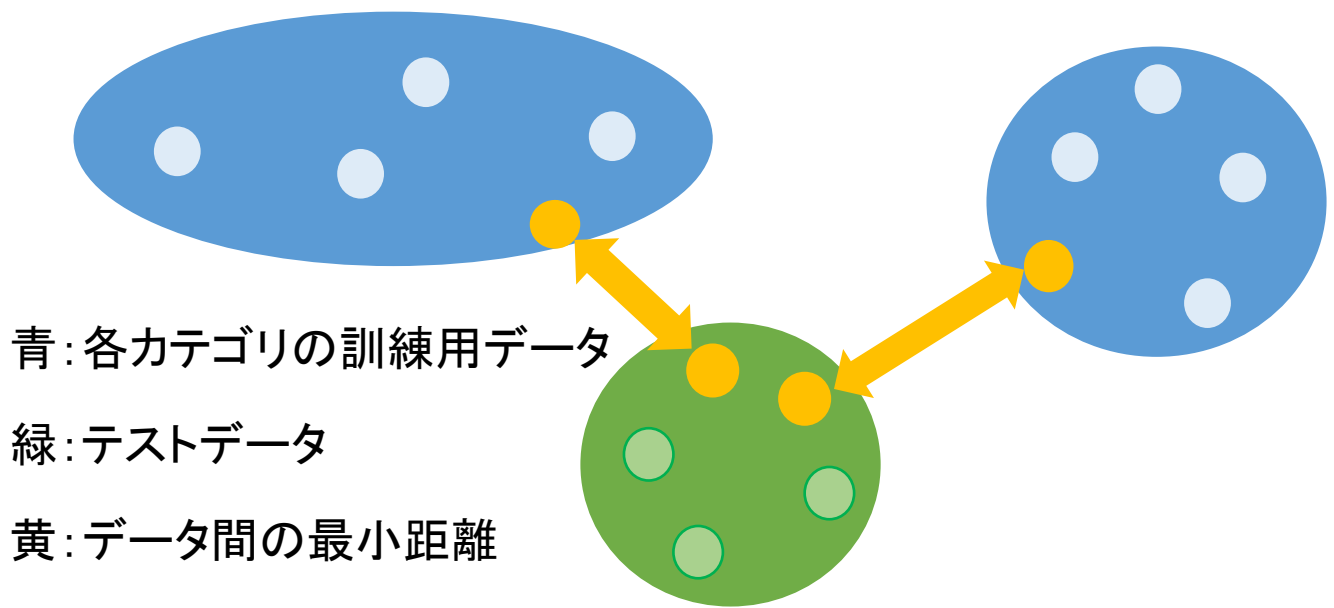


図 8 MPD の説明

## 4. 実験

### 4.1. 画像分類の実験

#### 4.1.1. 単独テストデータでの実験

本節では、テストデータが1枚の画像からなる場合についての実験を行う。訓練用画像の枚数  $k$  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。Caltech101 の 100 カテゴリそれぞれから  $k$  枚の画像を訓練用画像としてランダムに抜き出し、特徴抽出を行い訓練用データとする。残りの画像全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 (図 9)。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。訓練用画像の枚数は  $k = 1, 2, 4, 8, 16, 30$  とする。訓練用画像の選び方の影響を緩和するため、各  $k$  の値について訓練用画像の選び方を変えながら生成した 8 通りの訓練用画像について正解率をもとめ、その平均を実験結果とする。CMA 法および比較手法の間では、条件が同じとなるよう、同じ 8 通りの訓練用画像を用いる。結果を図 10 に示す。

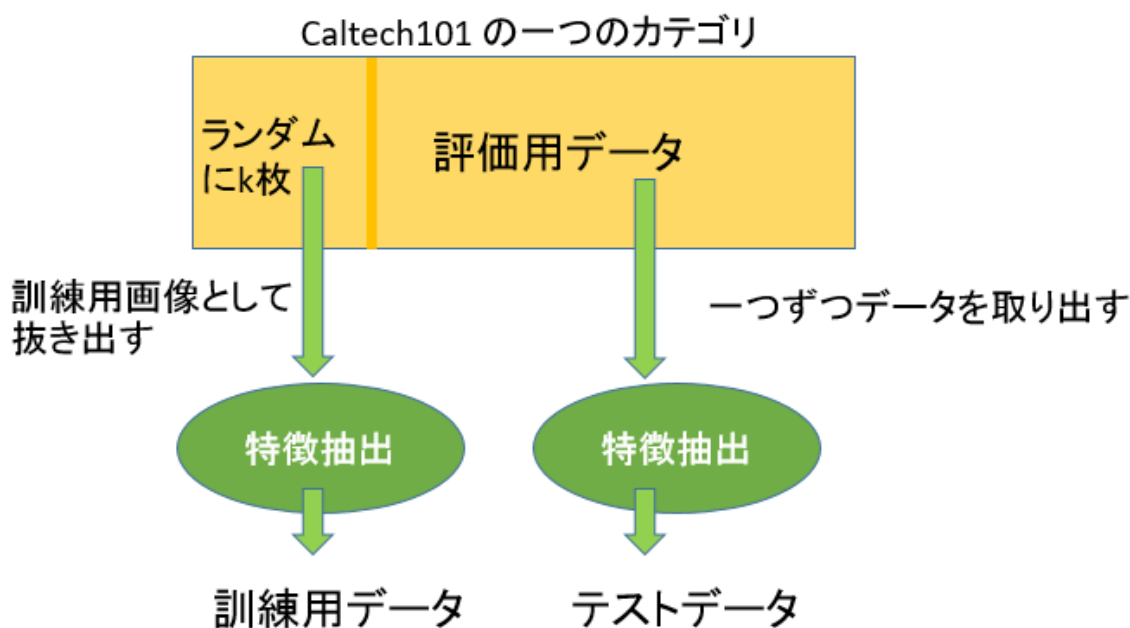


図 9 単独テストデータでのデータセットの扱い方

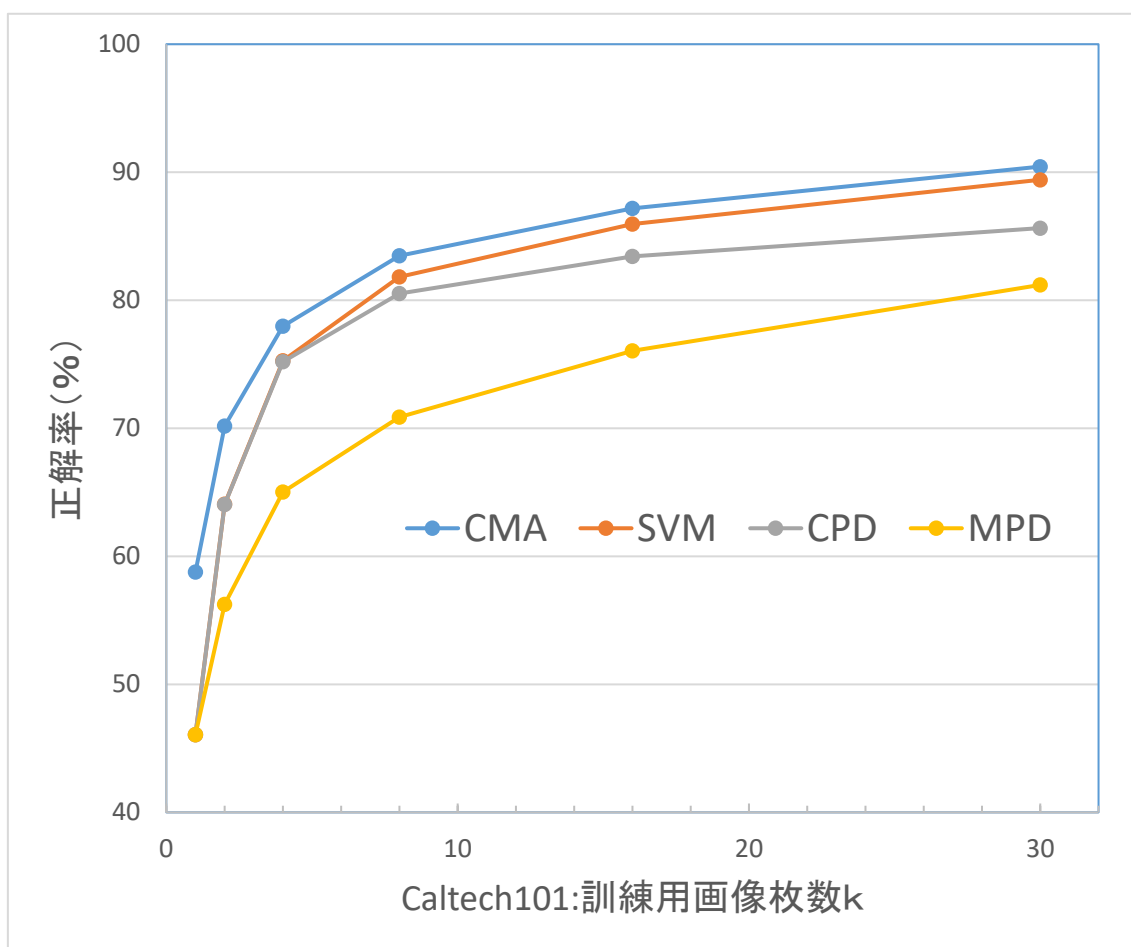


図 10 Caltech101 での単独テストデータの実験結果

実験結果では、いずれの  $k$  においても CMA が他の比較手法より高い正解率を示した。特に、 $k = 1$  では、比較手法より約 12 ポイント高い正解率を示しており、訓練用画像の数が少ない範囲で相対的により高い性能を示すことが分かった。

同様の実験を Caltech256 について行った。正解率の求め方やデータセットの扱い方などは同じである。結果を図 11 に示す。

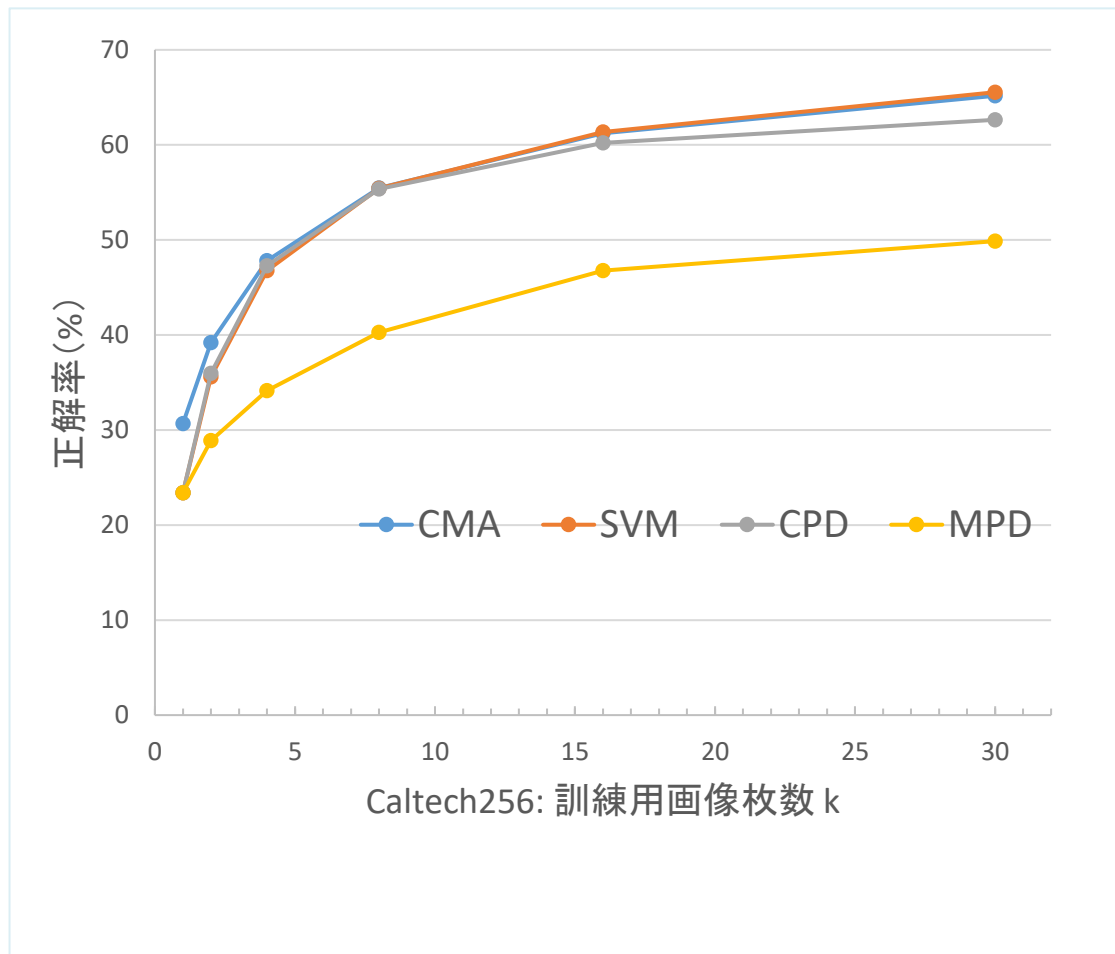


図 11 Caltech256 での単独テストデータの実験結果

Caltech101 の実験結果と同様に、いずれの  $k$  においても CMA が他の比較手法より高い正解率を示した。特に、 $k = 1$  では、比較手法より約 7 ポイント高い正解率を示しており、訓練用画像の数が少ない範囲で相対的により高い性能を示すことが分かった。

#### 4.1.2. 複数テストデータでの実験

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  $l$  と訓練用画像の枚数  $k$  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。SVM については、テストデータが複数の場合は扱えないため、比較対象から除く。前節の単独テストデータでの実験と同様にして訓練用データを作成し、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  $l$  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繰り返す (図 12)。テストデータの枚数は  $l = 2, 4, 8, 16, 30$  とする。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。テストデータの組み合わせによる変動の影響を緩和するために、同じ  $l$  について 3 回の試行を行い、その平均を求める。さらに、前節と同様に同じ  $k$  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。Caltech101 での実験結果を図 13 に示す。

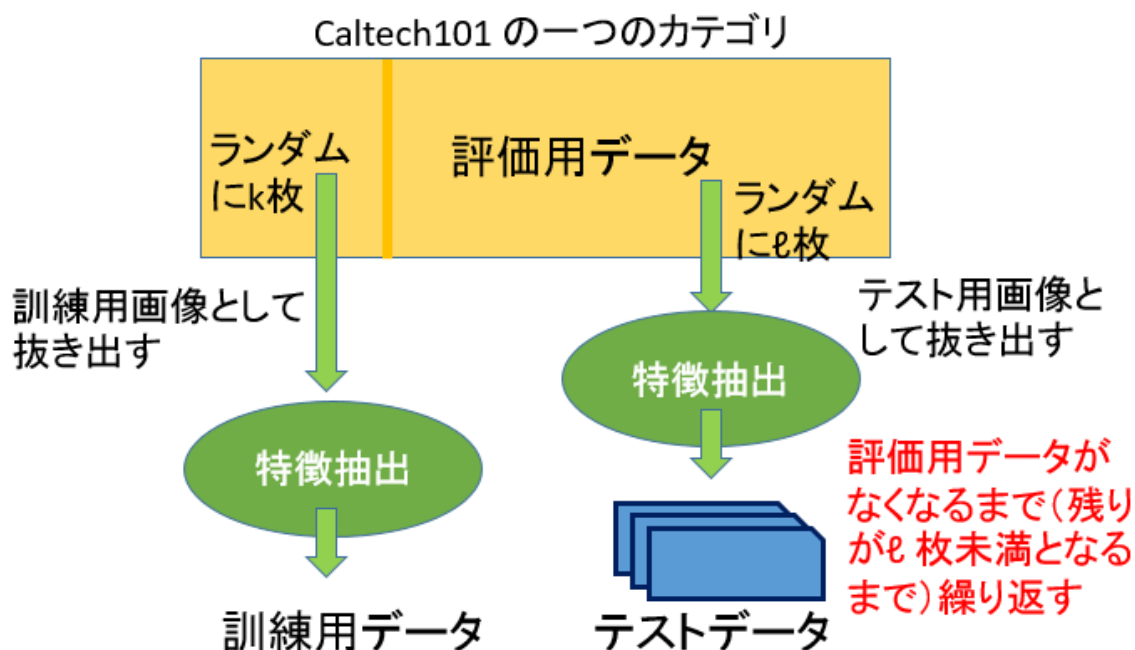


図 12 複数テストデータでのデータセットの扱い方



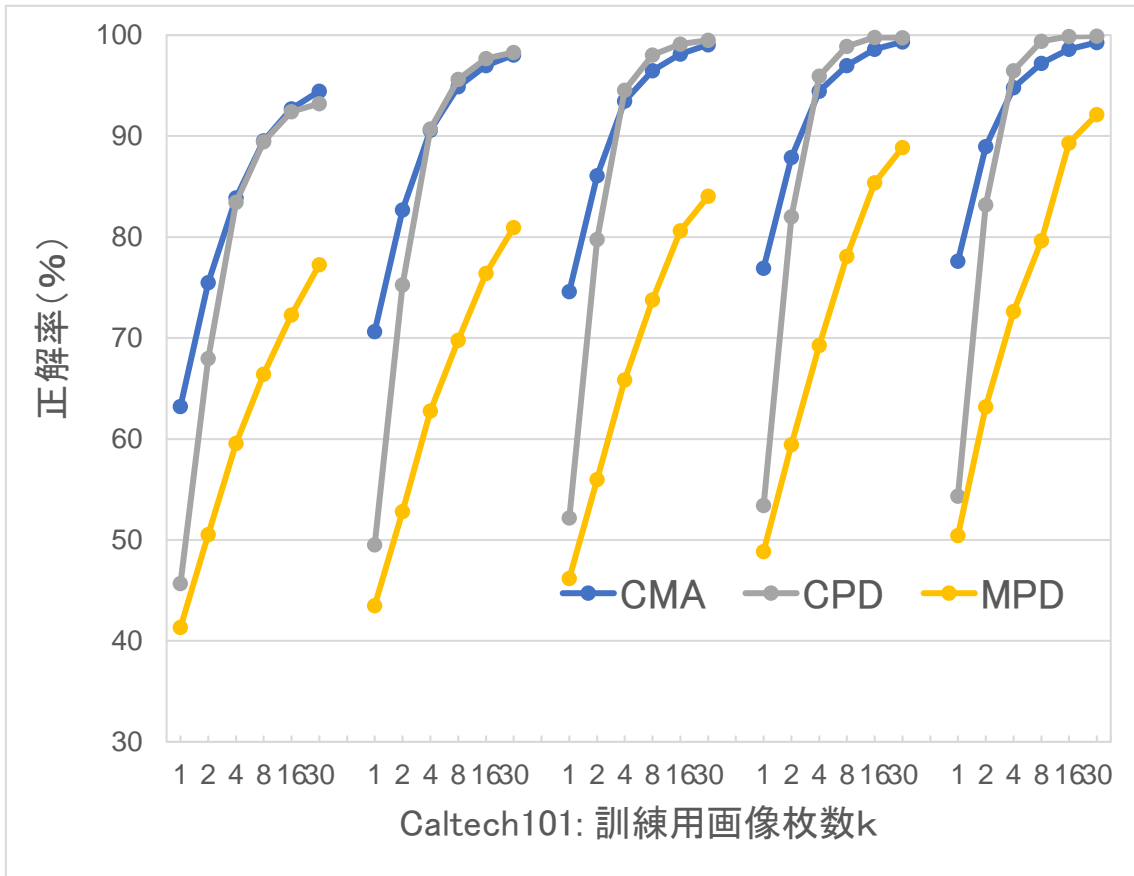


図 13 Caltech101 での複数テストデータの実験結果

複数テストデータの場合でも、訓練用画像の枚数が少ない  $k = 1, 2$  の範囲では CMA 法が比較手法より高い正解率を示している。また、全体として、テストデータの枚数  $l$  を大きくするほど正解率は上がっており、複数テストデータを組にして画像分類を適用できる CMA 法には利点があると考えられる。

同様の実験を Caltech256 について行った。正解率の求め方やデータセットの扱い方などは同じである。結果を図 14 に示す。

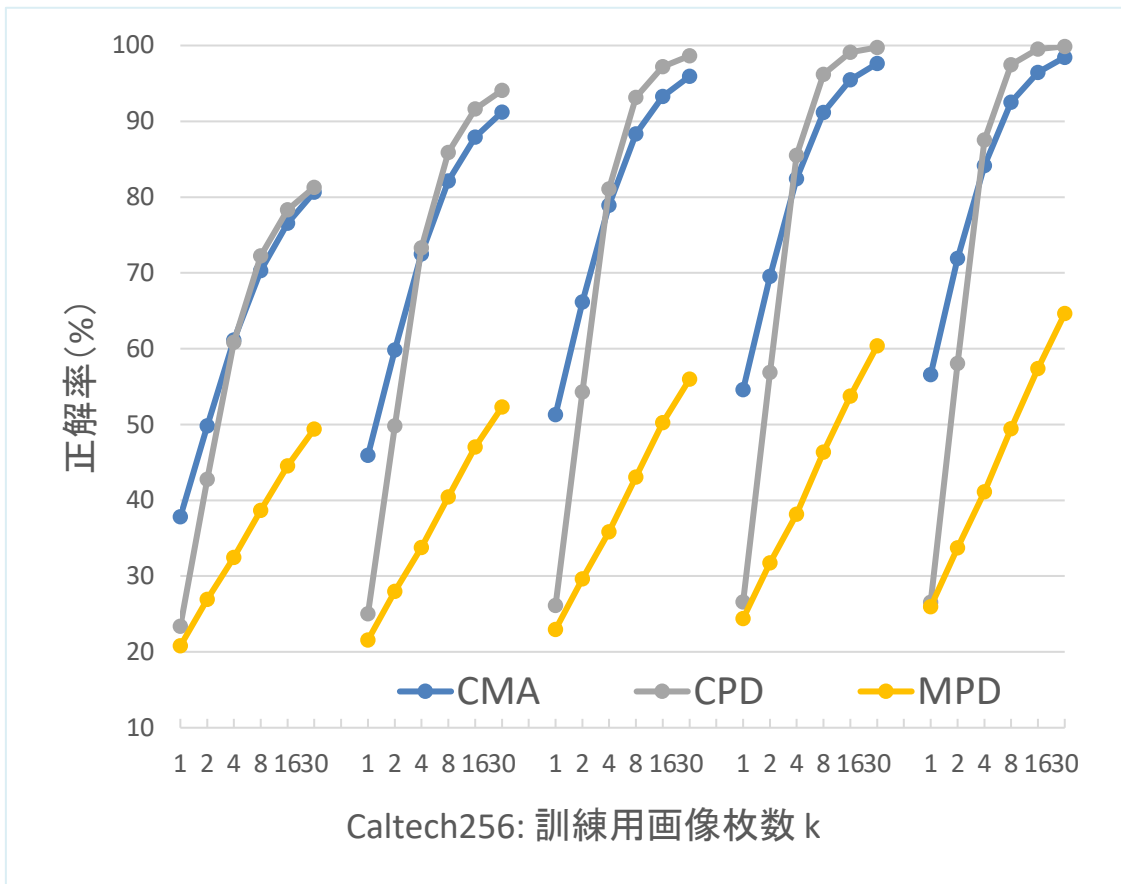


図 14 Caltech256 での複数テストデータの実験結果

Caltech101 の実験結果と同様に、訓練用画像の枚数が少ない  $k = 1, 2$  の範囲では CMA 法が比較手法より高い正解率を示している。また、全体として、テストデータの枚数  $l$  を大きくするほど正解率は上がっており、複数テストデータを組にして画像分類を適用できる CMA 法には利点があると考えられる。

#### 4.2. パラメータ $\lambda_1, \lambda_2$ の調節実験

最適化段階で用いた正則化項の重みのパラメータである $\lambda_1, \lambda_2$ はあらかじめ適切な値を与える必要がある。前節の実験では $\lambda_1 = 16, \lambda_2 = 16$ と設定していた。このパラメータ値 $\lambda_1, \lambda_2$ の影響を評価する為に、 $\lambda_1, \lambda_2$ の値を32,16,8と変化させて正解率がどのように変化するか実験を行った。この実験では Caltech256 データセットを用いた。訓練用画像枚数  $k = 1,30$  とし、テストデータの枚数  $\ell = 30$  とする。実験結果を表 2,3 に示す。

表 2 パラメータの割合を変えた実験結果 (単独テストデータ)

$\lambda_2$	k=1(%)	k=30(%)
32	30.7	64.2
16	31.1	65.2
8	<b>31.4</b>	<b>65.7</b>

表 3 パラメータの割合を変えた実験結果 (複数テストデータ)

$\lambda_1$	$\lambda_2$	k=1(%)	k=30(%)
32	32	53.4	97.7
32	16	55.1	98.7
32	8	<b>55.7</b>	<b>98.9</b>
16	32	54.7	97.9
16	16	55.4	98.6
16	8	<b>56.6</b>	<b>99.2</b>
8	32	54.8	98.1
8	16	55.9	98.5
8	8	<b>56.3</b>	<b>98.9</b>

単独テストデータの実験では、 $\lambda_1$ の値は影響がないため、 $\lambda_1$ の評価は行わない。両方の結果より $\lambda_2$ の値がより低い程正解率が高くなっていることが分かる。これは訓練用データに異なる種類の対象が含まれている為、近似点がデータの平均に寄らない方が良くと解釈できる。

単独テストデータに対して $\lambda_2$ の値を下げて正解率を測った結果、 $\lambda_2 = 3$ で最も高い正解率を得た。 $\lambda_2 = 3$ での実験結果を図 15 に示す。正解率の求め方やデータセットの扱い方は前節での実験と同様である。凡例にある CMAr3 は $\lambda_2 = 3$ に設定した時の CMA 法の正解率を示し、CMAr16 は $\lambda_2 = 16$ に設定した時の CMA 法の正解率を示している。

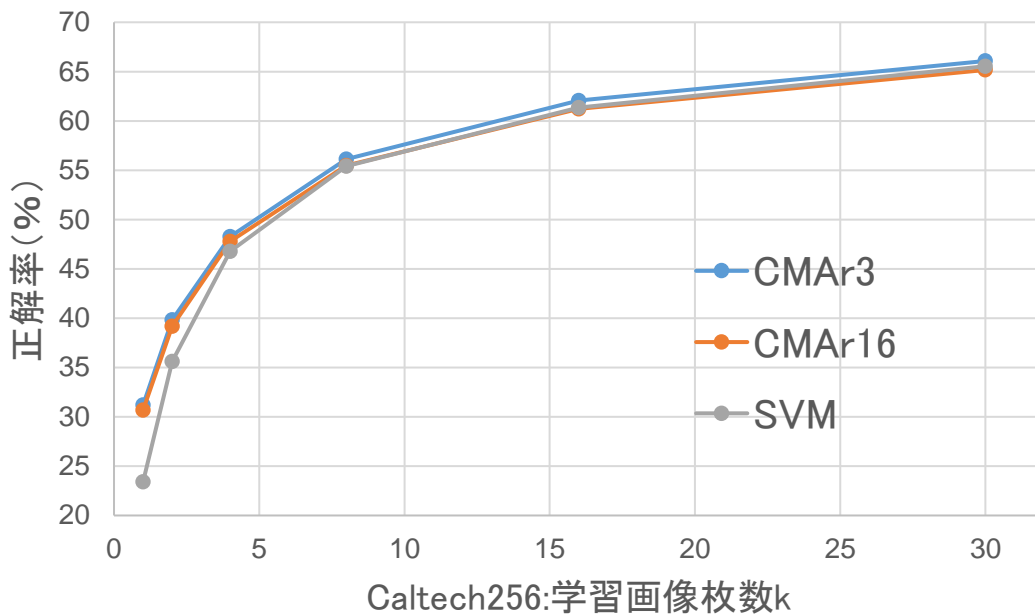


図 15  $\lambda_2$ を変えた実験結果 (単独テストデータ)

$\lambda_2 = 3$ に設定した CMA 法は $\lambda_2 = 16$ に設定した CMA 法よりも正解率が約 1 ポイント高い。また、 $\lambda_2 = 16$ の CMA 法は訓練用画像枚数が多いときには SVM と同等の正解率であったが、 $\lambda_2 = 3$ の CMA 法は訓練用画像枚数が多いときでも SVM より高い正解率を持つ。

正則化項の重みのパラメータは正解率に大きな影響は与えておらず、パラメータの設定に過敏になる必要はない。

## 5. おわりに

CMA 法が一般物体識別の研究の 1 つである画像分類に対して有効であることを示した。特に訓練用画像が少ない状況では比較手法よりも良い正解率を示した。また正則化項の重みのパラメータの影響を評価した結果、 $\lambda_2 = 3$  で設定した時に正解率が約 1 ポイント高くなったが、パラメータの値が正解率に大きな影響を与えていないという事が明らかになった。

今後の課題として、今回は与えられたテストデータをどれかのカテゴリに分類をしていたが、与えられたテストデータが分類されるカテゴリの中に必ずしもあるとは限らない場合についての分類も行えるようにしたい。そのために分類段階で求められる各カテゴリとの距離  $d^i$  の解析を行い、最適な閾値を求め分類を行えるよう改良したい。

## 謝辞

本研究を進めるにあたり，ご指導を頂いた卒業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雅之教授には，実験を行う上でのコツや，論文に関する助言やご指導を沢山いただきました。また，本研究で調査したCMA法を提案されたYang Wuさんを始めとする研究者の皆様に感謝いたします。そして，日常の議論を通じて多くの知識や示唆を頂いた椋木研究室の皆様に感謝します。

## 参考文献

- [1] Y.Wu, M.Mukunoki, M.Minoh: Collaborative Mean Attraction for Set Based Recognition, MIRU2014, OS3-3, 2014-07.
- [2] L.Fei-Fei, R.Fergus, P.Perona: Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, CVPR2004 Workshop on Generative-Model Based Vision, 2004.
- [3] Y.Jia, E.Shelhamer, J.Donahue, et.al.: Caffe: Convolutional Architecture for Fast Feature Embedding, arXiv preprint arXiv:1408.5093, 2014.
- [4] L.Fei-Fei: ImageNet: crowdsourcing, benchmarking and other cool things, CMU VASC Seminar, 2010-03.
- [5][http://dl.caffe.berkeleyvision.org/bvlc\\_reference\\_caffenet.caffemodel](http://dl.caffe.berkeleyvision.org/bvlc_reference_caffenet.caffemodel)
- [6] C.C.Chang, C.J.Lin: LIBSVM - A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.