

# Deep CNN と SVM を用いた物体画像と文字パターンに共通な カテゴリ分類器の一提案

嶋 好博<sup>†</sup>      中島由美<sup>‡</sup>      安田道夫<sup>‡</sup>  
 明星大学理工学部<sup>†</sup>      明星大学情報学部<sup>‡</sup>

## 1. はじめに

画像や文字パターンの分類システムの開発が近年盛んに行われている[1][2][3]。物体画像と文字パターンはそれぞれ別個に独立してカテゴリ分類器が提案されている。一般の周囲環境をコンピュータが理解するためには、物体画像と文字パターンを共通に分類する必要がある。

本研究の目的は物体画像と文字パターンの混在状態を対象にカテゴリを分類する共通分類器を提案することである。公開データセット[4][5]に対し、分類精度を実験的に確認する。

## 2. 物体画像と文字パターンに共通な分類器

### (1) 分類対象

図1に物体画像と文字パターンを示す。物体画像は Caltech-256 オブジェクトデータセット[5]からサンプルを 125 枚以上確保できるカテゴリ 10 種を先頭から順に選択する。カテゴリ内の先頭から評価用 25 枚、学習用残り 100 枚以上を選択する。文字パターンは MNIST 手書き数字データセット[4]から評価用 25 枚、学習用 100 枚を先頭から順に選択する。

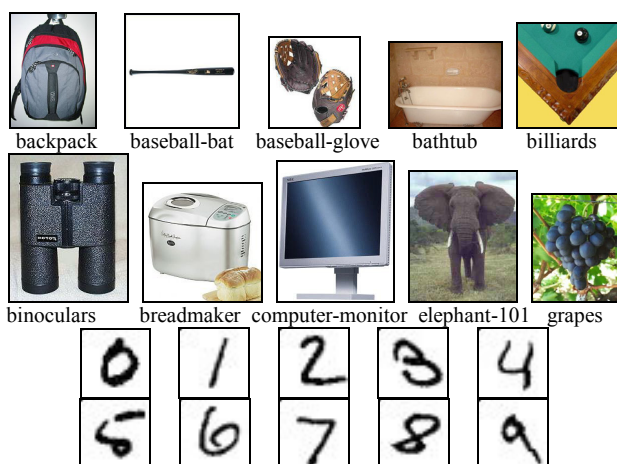


図1 分類評価対象の例(Caltech-256, MNIST[4,5])

物体画像のカテゴリは、backpack, baseball-bat, baseball-glove, bathtub, billiards, binoculars, breadmaker, computer-monitor, elephant-101, grapes の 10 種である[5]。サイズは様々でおおよそ 200x300 画素程度である。また、文字パターンは手書き数字 '0' から '9' の 10 字種の濃淡画像である。サイズは 28x28 画素である。

### (2) 共通分類器の構成

分類器は特徴抽出部と識別部からなる。図2に共通カテゴリ分類器の構成を示す。特徴抽出部の構成は物体画像と文字パターンに共通とし、学習済みの CNN(Convolutional Neural Network) を特徴抽出に用いる。識別部はカテゴリ毎に学習が可能な複数カテゴリ対応の SVM(Support Vector Machines)[7]を用いる。

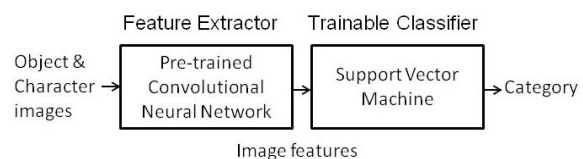


図2 共通カテゴリ分類器の構成

## 3. 共通分類器によるイメージ分類の実験

物体画像と文字パターンの単独での分類精度を求める。次いで、物体画像と文字パターンを混在した状態で、学習を行い分類精度を求める。

### (1) 学習済み CNN と SVM

特徴抽出部は物体画像で学習済みの CNN(Alex-Net) を用いる [2][6]。入力画像は 227x227 画素に正規化される。図3に 23 層をもつ Alex-Net の第 2 層の重み係数マップを示す。第 19 層の 4096 次元の特徴ベクトルを SVM に入力して、学習及び識別を行う [6]。

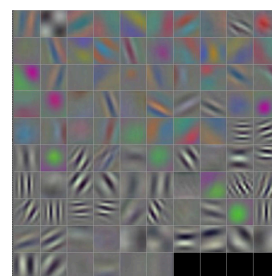


図3 CNN 第2層の重み係数(96セット)

Common Category Classification for Object Images and Character Patterns using Deep CNN and SVM

<sup>†</sup>Yoshihiro Shima · Meisei University, School of Science and Engineering

<sup>‡</sup>Yumi Nakashima, Michio Yasuda · Meisei University, School of Information Science

(2) 個別の分類結果

表 1 に物体画像の分類成功率を示す。物体画像 10 カテゴリに対して各 100 枚を学習に用いた。学習には文字パターンを含めていない。評価は物体画像のカテゴリ毎に 25 枚を用いた。物体画像単独での分類成功率は 93.6%である。

表 2 に文字パターンの分類成功率を示す。文字パターンとして手書き数字 10 字種に対して各 100 枚を学習に用いた。学習には物体画像を含めていない。評価は手書き数字の字種当たり 25 枚である。手書き数字パターン単独での分類成功率は 95.6%である。

表 1 物体画像の分類成功率  
(学習：物体画像 10 カテゴリ×各 100 枚)

評価サンプル数	成功数	成功率	誤分類数
250 枚	234 枚	93.6 %	16 枚

表 2 文字パターンの分類成功率  
(学習：文字パターン 10 字種×各 100 枚)

評価サンプル数	成功数	成功率	誤分類数
250 枚	239 枚	95.6 %	11 枚

(3) 物体画像と文字パターン混在の分類結果

表 3 に物体画像と文字パターン混在の場合のカテゴリ別の分類成功率を示す。学習には物体画像 10 カテゴリと文字パターン 10 カテゴリを用いた。各カテゴリの学習サンプル数は 100 枚であり、ランダムに選択する。カテゴリ別の分類成功率の最小値は手書き数字'3'の 76%、最大値は bathtub 等 8 種のカテゴリで 100%であった。図 4 に誤分類サンプルの例を示す。

物体画像と文字パターン混在の分類成功率は、表 4 に示すように、94.0%であった。物体画像単独の分類精度や文字パターン単独の分類精度と比べると物体画像と文字パターン混在の場合の分類成功率は同程度であった。物体画像と文字パターン間で交差するような誤分類は無かった。

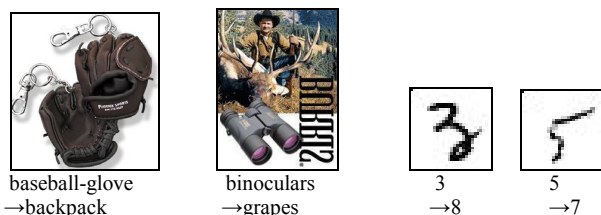


図 4 誤分類サンプルの例(truth→error)

表 3 物体画像と文字パターンのカテゴリ別分類成功率(評価:20 カテゴリ×各 25 枚)

	Category	Training #	Test #	Success Rate(%)
1	backpack	100/126	25	92
2	baseball-bat	100	25	96
3	baseball-glove	100/123	25	88
4	bathtub	100/207	25	100
5	billiards	100/253	25	80
6	binoculars	100/191	25	84
7	breadmaker	100/117	25	96
8	computer-monitor	100/108	25	96
9	elephant-101	100/106	25	96
10	grapes	100/176	25	92
11	digit:'0'	100	25	100
12	digit:'1'	100	25	100
13	digit:'2'	100	25	100
14	digit:'3'	100	25	76
15	digit:'4'	100	25	100
16	digit:'5'	100	25	92
17	digit:'6'	100	25	100
18	digit:'7'	100	25	100
19	digit:'8'	100	25	92
20	digit:'9'	100	25	100

表 4 物体画像と文字パターン混在の分類成功率  
(学習：物体画像と文字パターン計 2000 枚)

評価サンプル数	成功数	成功率	誤分類数
500 枚	470 枚	94.0 %	30 枚

4. おわりに

物体画像と文字パターンが混在する対象に対する共通のカテゴリ分類器を提案した。20 カテゴリ、学習用計 2000 枚、評価用計 500 枚に対して分類成功率 94.0%であった。今後の課題は、カテゴリ数および学習数を増やし信頼性の評価を行うことである。

参考文献

- [1] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE (Volume:86 ,Issue:11), pp. 2278-2324, Nov 1998.
- [2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Neural Information Processing Systems, 2012.
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, June 2009.
- [4] Yann LeCun, Corinna Cortes, Christopher J.C. Burges, THE MNIST DATABASE of handwritten digits, <http://yann.lecun.com/exdb/mnist/>.
- [5] Gregory Griffin, Alex Holub, and Pietro Perona, Caltech-256 Object Category Dataset (2007), <http://authors.library.caltech.edu/7694/>
- [6] MatConvNet, <http://www.vlfeat.org/matconvnet/pretrained/>.
- [7] 北村浩治,二階堂真理恵,中島由美,安田道夫,振動相関法と svm 法,明星大学研究紀要[情報学部]第 20 号,pp.45-56,2012 年