4L1-4

Convolutional Neural Network による写真と手描きスケッチの認識

Classification of Photo and Sketch Images Using Convolutional Neural Networks

山川 まどか[*1], 関口 香菜[*1], 佐々木一磨[*1], 尾形哲也[*1]
Madoka Yamakawa Kana Sekiguchi Kazuma Sasaki Tetsuya Ogata

[*1] 早稲田大学基幹理工学部表現工学科
Intermedia Art and Science, School of Fundamental Science and Engineering, Waseda University, Toyko, Japan.

In this study we propose a Convolutional Neural Network(CNN) which can classify hand drawn sketch images. Though CNN is known to be very effective on classification of realistic images, there are few studies on CNN dealing with non-photorealistic images and even more images those types are mixing. Classifying non-photorealistic images is difficult mainly because there are no large datasets. In this paper, to classify sketch images using CNN, we propose the simple method to make training datasets from photo and illustration images. We also made several training datasets and compared the accuracy of the trained CNN models. Our proposed method is shown to be effective on classification task of not only sketch images but also the mixed dataset of photo and sketch images.

1. はじめに

近年,コンテンツベースの検索システムサービスが提供されており, 画像を用いた検索が可能になっている. これにはユーザーが対象の画像を持っていないと検索できないという欠点があった. Mind Finder[1][2][3]というシステムはユーザーのスケッチから写真を検索するシステムである. ユーザーは名前を知らないものでも簡単なスケッチを書くだけで, 画像を用いることなく検索を行うことができる.

また, 画像認識の分野において Convolutional Neural Network(CNN)が注目されている. たとえば, 1000 クラスの画像を判別する実験[4]において, top-5 で 83%の正答率を記録し, 近年は人間の識別性能を上回る性能を発揮している. CNN は多層の畳み込み層とプーリング層からなる大規模な神経回路モデルであり, 大量の画像を学習するにより認識に必要な特徴量を自己組織化し獲得することができる.

検索システムではユーザーからの入力は写真やスケッチなど多種多様な画像が入力されるが, CNN に関する既存研究の多くは実写画像のみを扱っており, 手描きのスケッチ画像含むような多様な入力に対して対応するさせることはできていない. 非実写画像については同じ対象物を描いていたとしても作者によって表現が多様であり, クラス固有の視覚特徴量を獲得することが難しい.

スケッチ画像の認識を実現させるためには多くの画像を学習させる必要があるが, 実写画像に比べてスケッチ画像のデータベースの規模は非常に限られている. たとえば, ImageNet[5]のデータベースには1500 万枚以上もの写真が含まれているが, 手描きスケッチのデータベース[6]は2万枚程度である.

そこで本研究では CNN による写真・スケッチの同時認識を可能とするため, 実写画像, またデータを集めやすいカラーイラスト画像を利用した CNN の学習データセットの構築方法を提案する.

連絡先:山川まどか, 早稲田大学, 〒169-8555 東京都新宿区大久保 3-4-1, TEL: 03-5286-3000、FAX: 03-5286-3500, yamakawa@idr.ias.sci.waseda.ac.jp

2. 実験設定

提案手法の評価を行うため, CNN を用いて20種類の動物の手描きスケッチ画像をクラス判別する実験を行った.

2.1 学習データセットの用意

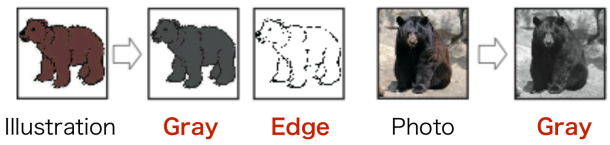

図 1 学習データの構築方法

学習データセットとして, インターネット検索により写真 28,468 枚とイラスト 12,734 枚の画像を用意した. これらの画像をスケッチ画像に視覚的に近づけるためにグレースケール化とエッジ強調を行い, 学習データセットを作成した(図1). これに加えて比較対象として写真, スケッチ画像とそれぞれのグレースケール化・エッジ強調の組み合わせを表1のように用意した.

表 1 学習データセットの内訳

Dataset / Image Type	Photo	Illustration	Photo & Illustration	Proposed Method
Photo	O	-	O	O
Illustration	-	O	O	O
Photo(Gray)	-	-	-	O
Illustration(Gray)	-	-	-	O
Illustration(Edge)	-	-	-	O
Number of Images	12,734	38,468	51,202	115,138

学習には各画像を水平方向に反転した画像も含まれており，画像サイズはそれぞれ 256x256pixel で，CNN に入力される際にランダムで 227x227pixel 分を切り取って入力した．

評価用のデータとして 100 枚の手書きスケッチ画像を用意し，表1のそれぞれのデータセットで学習した CNN を使って認識率を評価した．

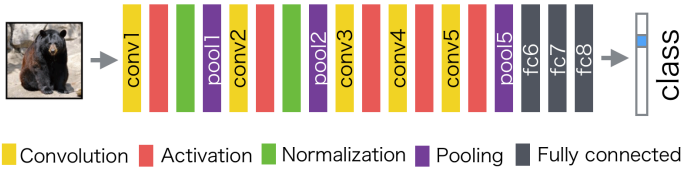

図２　Alex-net の層構造

CNN の構造について本実験では Alex-net[1]というネットワーク構造を用いた．Alex-net は図2のように 5 つの Convolution 層と 3 つの Pooling 層，3 つの Full-connection 層から成る．学習には深層学習フレームワークである Caffe[7]を利用した．学習最適化方法は Stochastic Gradient Descent[8]を用いた．

3.　実験結果

3.1 クラス判別正答率

表2に学習データの内容と，学習結果を示す．写真の正答率ではイラスト以外の学習データセットで 99%であるが，スケッチの正答率は提案手法がもっとも高く，76%だった．また，写真とスケッチ混合のテストデータについてもこのモデルの正答率は 85%である．提案手法によって，写真とスケッチのどちらもクラス判別を行うことができることが確認された．

表２　学習データ別クラス判別正答率

Test data / Train dataset	Photo	Sketch	Mixed
Illustration	26%	41%	33%
Photo	99%	11%	55%
Illustration & Photo	99%	42%	71%
Proposed Method	99%	**76%**	**85%**

3.2 画像特徴量の解析

提案手法で作成したデータセットで学習した CNN モデルについて，主成分分析(PCA)を行い解析を行った．

図 3 に Full-connection 層の2番目の層の値について PCA 解析を行った結果を示す．クラスごとに色分けし，左右で違う角度から表示している．PC1，PC2，PC3 はそれぞれ第 1～第 3 主成分軸であり寄与率はそれぞれ 1.84%，1.06%，0.97%である．データには写真とイラストが含まれているが，クラスごとにまとまりができていることがわかる．

図 4 は「クマ」の画像群についてタイプごとに色分けした画像特徴量の PCA である．pool5 層では写真、イラストごとに大まかなクラスタができている．イラストでは，オリジナルのイラスト，白黒化したイラスト，線画化したイラストの順にクラスタが並んでいる．写真とイラストの大きなまとまりは，白黒化した写真とオリジナルのイラストのクラスタが接している．また，このような pool5 層クラスタ構造は，最終層の full-connection へと層を進めることで徐々にまとまっていくことが分かった．

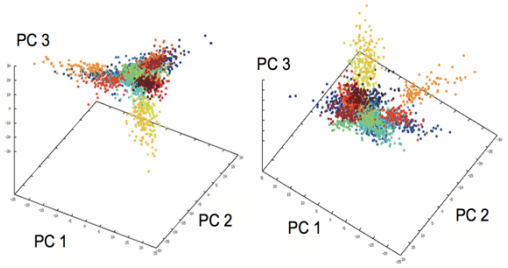

図３　学習データの主成分分析(クラスごと)

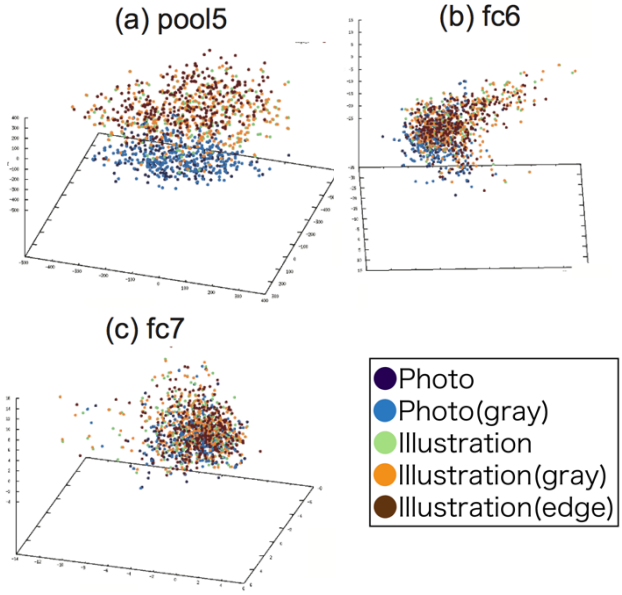

図４　学習データの主成分分析(画像タイプごと)

4.　まとめ

本研究では，ユーザーの簡単なスケッチから検索する，新しいコンテンツベースの検索システムを背景として CNN を用いて写真に加えて非実写的画像であるスケッチ画像を認識する手法を提案した．スケッチは画像の数が少なく，学習が難しいため，写真とイラストの画像を集め，それぞれ色抜きや線画化の加工を行って学習データを用意した．学習した結果，提案手法がスケッチの認識に効果的であることがわかった．また，写真とスケッチ混合のデータセットにも対応できることがわかった．

謝辞

文科省科研費基盤研究(A)(No. 15H01710)の助成を受けた．

参考文献

[1]　[Wang 2010] Changhu Wang, Zhiwei Li, and Lei Zhang :"MindFinder: Image Search by Interactive Sketching and Tagging", WWW.

[2] [Cao 2011] Yang Cao, Changhu Wang, Liqing Zhang, and Lei Zhang. Edgel :"Index for Large-Scale Sketch-based Image Search", CVPR.

[3] [Cao 2010] Yang Cao, Hai Wang, Changhu Wang, Zhiwei Li, Liqing Zhang, and Lei Zhang :"MindFinder: Interactive Sketch-based Image Search on Millions of Images", ACM Multi- media.

[4] [Krizzhevsky 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton: "ImageNet Classification with Deep Convolutional Neural Networks", NIPS.

[5] [Deng 2009] J. Deng, K. Li, M. Do, H. Su, L. Fei-Fei: "Construction and Analysis of a Large Scale Image Ontology". In Vision Sciences Society (VSS)

[6] [Eitz 2012] Eitz., M., Hays., J.,Alexa., M.:"How Do Humans Sketch Objects?": ACMtransactions of Graphics. 31, 1–10.

[7] [Yangqing 2014] Jia Yangqing, Shelhamer Evan, Donahue Jeff, Karayev Sergey, Long Jonathan, Girshick Ross, Guadarrama Sergio, Darrell Trevor: "Caffe: Convolutional Architecture for Fast Feature Embedding", arXiv preprint arXiv:1408.5093.

[8] [Bottou 2012] Bottou, L. :"Stochastic Gradient Descent Tricks. Neural Networks: Tricks of the Trade", Springer, 7700, 421–436.