

研究報告書

「大規模 web 情報とライフログによる実世界認識知能の構築」

研究タイプ: 通常型

研究期間: 平成21年10月～平成25年3月

研究者: 原田 達也

1. 研究のねらい

画像から一般的な物体やシーンを認識することはコンピュータビジョンの究極の目標の一つである。一般的な物体・シーンを認識する分野は「一般物体認識」と呼ばれコンピュータビジョンの中でも競争が熾烈であり、その進化のスピードは目を見張るものがある。しかしながら、実世界に存在する物体は無数であり、これらの多様性を適切に記述できるほどの性能に至っていないのが現状である。

そこで、本研究の目的は、真に実世界で利用可能な視覚を中心としたマルチモーダルデータを認識する知能を構築することにある。実世界を認識する知能の構築には知識の獲得が必要不可欠であるが、この知識構築のために大規模 web データとライフログの双方を活用する。また、大規模データから概念構造やコンテキストを抽出し、これらを用いることでセマンティックギャップを乗り越えた実世界認識アルゴリズムを開発する。

さらに本研究課題では、実世界データをより深く認識するために、データを単純な単語として理解するだけでなく、実世界データを人が理解可能な文章として記述することで、より深い認識アルゴリズム構築の試みを行う。

本研究により、実世界とサイバー世界をシームレスに接続する実世界認識知能のブレークスルーを狙う。この実世界認識知能の構築により、情報学のみならず人の認知プロセスの解明や、超高齢化社会で活躍するロボットの知能としても将来的に活用可能である。

2. 研究成果

(1) 概要

本研究で得られた研究成果は以下の3点にまとめられる。

- 1) 画像認識に適切な、効率的かつ豊かな表現が可能な画像特徴量の設計
- 2) 画像を認識し、適切な文章を付与する文章自動生成システムの開発
- 3) 世界最高レベルの画像認識システムの構築

(2) 詳細

1) 画像認識に適切な、効率的かつ豊かな表現が可能な画像特徴量の設計

局所特徴群から画像特徴に変換する表現手法は、画像の識別性能を左右する重要なモジュールである。デファクトスタンダードの画像表現手法は Bag of Words であり、多くの手法はコードブックを基盤とした手法となっているが、コードブックを基盤とした手法はコードワードへの割り当ての計算コストが高い。そこで本研究では、コードブックや局所特徴のマッチングを利用しない効率的な画像表現手法を提案する。提案手法は局所特徴群をガウシアンマルコフ確率場でモデル化することで局所特徴間の空間情報を効率的に表現する。さらにこのモデルに対し情報幾何の知見に基づいてモデルに適切な特徴ベクトルを定義する。さらに特徴ベクトル間の適切な距離計量の特徴量に埋め込む。提案する特徴は線形識別機の利用にもかかわらず、最新の手法と同等の識別性能が出ることが確認できた。

2) 画像を認識し、適切な文章を付与する文章自動生成システムの開発

画像を説明する自然文の自動生成は、画像認識における最終目標の 1 つと言える。既存の説明文生成手法では、画像それぞれに対して「主体、動作、光景」などの属性を表すラベルなど、セマンティックな知識を手作業で与える必要があり、多様で大量な画像データへの拡張が困難であった。そこで本研究では「画像の内容はいくつかのキーフレーズで表現でき、これらを単純な文法モデルで正しく繋げば画像説明文を生成できる」という仮説をたて、画像と説明文のみからなるデータセットを用いて入力画像のキーフレーズを推定するマルチキーフレーズ問題を新たに提唱し、これを効率よく解くアルゴリズムを開発した。このアルゴリズムを用い、実際に画像と文章のペアのみから入力画像に対して適切な文章を生成することが可能であることが分かった(図 1 参照)。



図 1: 画像から自動的に生成された文章の例

さらに、文章生成のための画像認識システムの中核となる複数のラベルと画像との関係性

を効率よく学習するオンライン学習機の開発を行った。画像認識において、多種多様な対象を認識するには大量の学習データが必要となる。大量の学習データを一気に学習するには大量の主メモリを必要とするため実際に解くことは困難である。そこでデータを1つずつ読み込みそれまでの学習結果を更新するオンライン分類学習手法が提案されている。通常、オンライン学習手法は1つのサンプルに複数のクラスラベルが付与されていても問題ないが、1つの訓練サンプルごとに正しいラベルと誤ったラベル1つずつのみを扱うため、学習がなかなか進まないラベルや、ほとんど学習されないラベルが生じる可能性がある。そこで本研究では、画像に付随する複数のラベルから得られる pairwise loss 全ての平均を用いる Passive-Aggressive with Averaged Pairwise Loss (PAAPL) を提案し、複数ラベルと画像を学習する画像アノテーション問題において最新手法を越える性能が獲得できることを確認した。

3) 世界最高レベルの画像認識システムの構築

上記1)、2)から得られた知見を活用することで大規模データを用いた画像認識システムを構築し(図2参照)、国際的なコンペティションにおいて、画像識別部門で2位、動物の種類を細かく分類する詳細な画像識別部門で1位の成績を上げた。

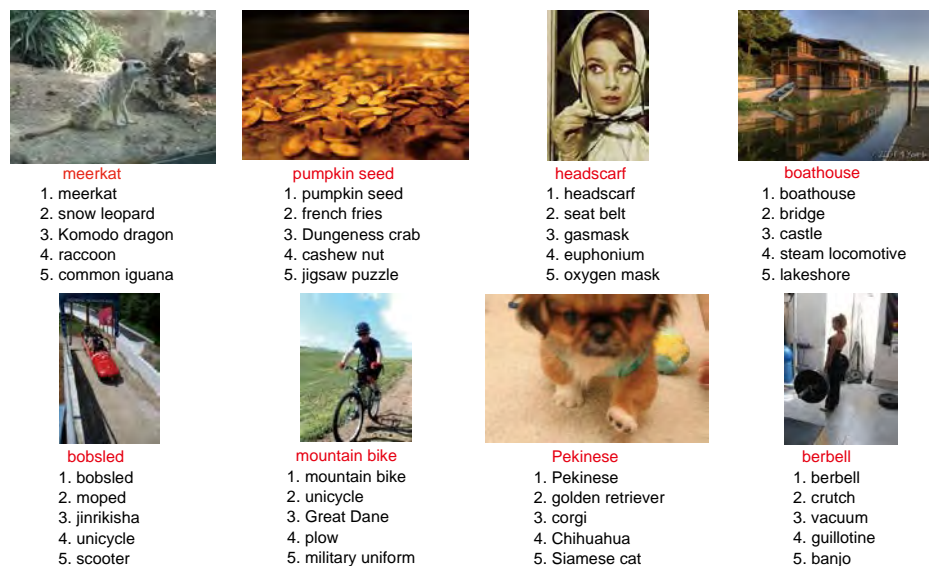


図2: 実現した画像認識システムによる画像の認識結果例

3. 今後の展開

高性能な実世界認識システムを構築するためには高品質なデータセットが必要となる。現状では、全てのデータが人手によってラベル付与されている状況を仮定しているが、多種多様な物体を認識するためには、この仮定は人的コストが高くつくため現実的でない。ラベルが明示的に与えられていないデータであっても、人的コストを最小限にしながら自動的に学習するシステムの構築が必須である。

また、Web に存在する画像は人間が意識的にアップロードしたものであり、人間という高度なフィルタを通して。つまり Web 上の画像認識と真の実世界画像認識とは依然として大きな隔りがある。この人間が持つようなフィルタ(特徴抽出)構築は根本的問題であり、ブレイクスルーが必要となる。例えば、眼鏡に搭載されたカメラと人の視線情報を組み合わせることで人の意図に沿った注視対象を抽出し、現在までに構築した高性能な画像認識システムと組み合わせることで、本当の意味での実世界認識に適応する試みを進める必要がある。

4. 自己評価

画像を中心とした実世界認識知能を目指して、大規模なデータから効率的に高性能な認識システムを獲得するアルゴリズムを構築した。最終的に構築された画像認識システムは世界最高レベルのものであり、この点に関して高く評価できる。しかしながら、真の実世界認識知能には道半ばであり、この目的を実現するためには曖昧模糊とした実世界の中から重要な情報を発見する高度な知能の構築に注力する必要がある。

5. 研究総括の見解

本研究は、情報世界と物理世界をシームレスに繋ぐシステムと、それを支える高速アルゴリズムに関する研究である。本研究の優れた点は、大量のアノテーション情報などを利用する、圧倒的な性能を誇る画像認識アルゴリズムを開発したことである。まず、コードブックや局所特徴のマッチングを利用しない効率的な画像表現手法を提案し、線形識別機の利用にもかかわらず、最新の手法と同等の識別性能が出ることを確認している。次に、画像に付随する複数のラベルを用いる学習アルゴリズムを提案し、ラベルと画像の対応を学習する画像アノテーション問題で最新手法を越える性能を確認している。さらに、こうしたアルゴリズムを用いて、大規模データに基づく画像認識システムを構築し、国際的な画像識別コンペティションにおいて、複数部門で1位、2位の成績を上げている。ここまででも十分な研究成果と言えるが、加えて、画像認識における最終目標である、画像を説明する自然文の自動生成に取り組んでいる。具体的には入力画像のキーフレーズを推定するマルチキーフレーズ問題を提唱し、これを効率よく解くアルゴリズムを開発している。本研究が目指す画像情報を認識し言葉で表現する技術は、ライフログなどに応用すると、膨大に記録されるデータに意味づけを行い検索することができる。近未来の情報環境を見据えた、世界にアピールできる切れ味のよい研究成果である。

6. 主な研究成果リスト

(1) 論文(原著論文)発表

1) 画像認識に適切な、効率的かつ豊かな表現が可能な画像特徴量の設計

1. Tatsuya Harada, and Yasuo Kuniyoshi. Graphical Gaussian Vector for Image Categorization. In The Twenty-Sixth Annual Conference on Neural Information Processing Systems (NIPS



2012), 2012.
2. Tatsuya Harada, Yoshitaka Ushiku, Yuya Yamashita, and Yasuo Kuniyoshi. Discriminative Spatial Pyramid. In The Twenty-Forth IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2011), pp.1617-1624, 2011.
3. Tatsuya Harada, Hideki Nakayama, and Yasuo Kuniyoshi. Improving Local Descriptors by Embedding Global and Local Spatial Information. In The 11th European Conference on Computer Vision (ECCV 2010), 2010.
4. Hideki Nakayama, Tatsuya Harada, and Yasuo Kuniyoshi. Global Gaussian Approach for Scene Categorization Using Information Geometry. In The Twenty-Third IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010), 2010.

2) 画像を認識し、適切な文章を付与する文章自動生成システムの開発

1. Yoshitaka Ushiku, Tatsuya Harada, and Yasuo Kuniyoshi. Efficient Image Annotation for Automatic Sentence Generation. In The 20th Annual ACM International Conference on Multimedia (ACMMM 2012), pp.549-558, 2012.
2. . Yoshitaka Ushiku, Tatsuya Harada, and Yasuo Kuniyoshi. Understanding Images with Natural Sentences. In The 19th Annual ACM International Conference on Multimedia (ACMMM 2011), pp.679-682, 2011.
3. Yoshitaka Ushiku, Tatsuya Harada, and Yasuo Kuniyoshi. Automatic Sentence Generation from Images. In The 19th Annual ACM International Conference on Multimedia (ACMMM 2011), pp.1533-1536, 2011.

3) 世界最高レベルの画像認識システムの構築

1. 郡司直之, 樋口貴之, 安本晃基, 村岡宏是, 牛久祥孝, 原田達也, 國吉康夫. 複数の画像特徴と Passive-Aggressive を用いた fine-grained classification. PRMU, to appear, 2013.
2. Yoshitaka Ushiku, Hiroshi Muraoka, Sho Inaba, Teppei Fujisawa, Koki Yasumoto, Naoyuki Gunji, Takayuki Higuchi, Yuko Hara, Tatsuya Harada, and Yasuo Kuniyoshi. ISI at ImageCLEF 2012: Scalable System for Image Annotation. CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, 17-20 Rome, Italy.

(2) 特許出願

なし。

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. First place in the fine-grained classification task, second place in the classification task, Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)
Tatsuya Harada, Naoyuki Gunji, Takayuki Higuchi, Koki Yasumoto, Hiroshi Muraoka, Yoshitaka Ushiku, Yasuo Kuniyoshi. The PASCAL Visual Object Classes Challenge 2012 (VOC2012), 12th October 2012, ECCV 2012, Florence, Italy



2. ACM Multimedia 2011 Grand Challenge Best Application of a Theory Framework Special Prize
Yoshitaka Ushiku, Tatsuya Harada, and Yasuo Kuniyoshi. A Understanding Images with Natural Sentences. the 19th Annual ACM International Conference on Multimedia (ACMMM 2011), pp.679-682, 2011.
3. Third place in the classification task, second place in the detection task, Large Scale Visual Recognition Challenge 2011 (ILSVRC2011)
Tatsuya Harada, Asako Kanezaki, Yoshitaka Ushiku, Yuya Yamashita, Sho Inaba, Hiroshi Muraoka, Yasuo Kuniyoshi. The PASCAL Visual Object Classes Challenge Workshop 2011, 7th November 2011, ICCV 2011, Barcelona, Spain.
4. IEEE ICME2011 Student Travel Grant
Yuya Yamashita, Tatsuya Harada, and Yasuo Kuniyoshi. Causal Flow. 2011 IEEE International Conference on Multimedia and Expo (ICME 2011), paper_356, 2011.
5. 第14回画像の認識・理解シンポジウム(MIRU2011)インタラクティブセッション賞
牛久祥孝, 原田達也, 國吉康夫. 画像・文章間の類似度学習による画像説明文の自動生成. 画像の認識・理解シンポジウム (MIRU2011), pp.365-372, 2011.
6. 2011PRMU 研究奨励賞.
牛久祥孝, 山下裕也, 井村純, 中山英樹, 原田達也, 國吉康夫. 複数画像特徴とクラスラベルの相関に着目した距離計量による大規模画像分類. パターン認識・メディア理解研究会 (PRMU), pp.1-8, 2011.
7. Third place, Large Scale Visual Recognition Challenge 2010 (ILSVRC2010)
Tatsuya Harada, Hideki Nakayama, Yoshitaka Ushiku, Yuya Yamashita, Jun Imura, and Yasuo Kuniyoshi. The PASCAL Visual Object Classes Challenge Workshop 2010, 11th September 2010, ECCV 2010, Crete, Greece.