

Linear Distance Metric Learning for Large-scale Generic Image Recognition
(線形距離計量学習による大規模一般画像認識)

氏 名 中山 英樹

概要

制約のない実世界の画像を計算機に認識させ、言語により記述させる技術を一般画像認識 (generic image recognition) と呼ぶ。我々人間も、外界の多くの情報を視覚から認識し行動決定を行っているように、実世界で行動する知能システムにとって一般画像認識は必要不可欠な機能の一つであるといえる。しかしながら、一般画像認識はその長い歴史にも関わらず未だ実現されておらず、コンピュータビジョンにおける究極的な目標の一つと捉えられている。

一般画像認識の難しさは、扱う画像の多様性と認識対象となる事物の多さに起因する。このため、汎用性の高い一般画像認識を実現するためには大量の事例データからの学習が鍵となる。しかしながら、従来の手法は学習サンプル数に対するスケーラビリティを欠いていたため、大規模な画像データセットを用いて学習・認識を行うことは著しく困難であった。そこで本研究では、学習サンプル数に対しスケーラブルかつ高精度な一般画像認識 (画像アノテーション) アルゴリズムの開発に取り組みこれを実現した。これを可能としたのは、本研究で提案したサンプル間距離計量の統計的学習手法と、画像特徴量の抽出手法である。これらは互いに密接に影響するプロセスであり、両者の相性を考慮しそれぞれを設計することが極めて重要である。

最終的に、開発した画像アノテーション手法を1,200万枚の画像データセットへ適用し、その有効性を示した。以下、各章の要旨をまとめる。

第1章 : Introduction

本章では、一般画像認識の紹介を行い、その本質的な難しさについて述べると同時に、大規模な学習データを用いるアプローチの有効性について説明した。この際、既存手法のボトルネックが学習サンプル数に対するスケーラビリティの低さであることを示し、この問題について本論文で取り組むことを宣言した。

第2章 : Outline of Image Recognition Method

本章では、最初に一般画像認識の歴史について詳しく述べ、現在の潮流について分析した。この際、学習手法と学習用データセットの両側面から考察を進めた。これを踏まえ、本論文で開発する画像認識手法の設計を行った。まず、一般画像認識の構造を詳しく考察し、いくつかの問題へ分類を行った。このうち、本論文では最も基本的かつ重要な画像アノテーションと呼ばれる課題へ取り組むことを述べた。ここで、low-levelな画像特徴と認識対象となる意味との間には大きな隔たりがあるという、いわゆる semantic gapが本質的な問題となる。本研究ではこの問題に対し、(1)多様な画像特徴の抽出、(2)判別的統計学習、の二つを重要な機能要素と考え、開発に取り組んだ。スケーラブルなシステムを実現するためには、これらの相性を考えてそれぞれを設計することが必要となる。

第3章 : Related Work of Image Annotation

本章では、まず画像アノテーションの先行研究について精査した。本分野ではさまざまなアプローチからなる手法が提案されているため、これらを分類し、各アプローチの代表的な手法について解説を行い、認識精度を比較した。この中で、ノンパラメトリック (事例ベース) なアプローチに基づく手法が近年成功を収めていることに触れ、その理由について分析した。これを受け、本研究においてもノンパラメトリックな方法を念頭に画像アノテーション手法を開発することを述べた。

ノンパラメトリックな画像アノテーションにおいてsemantic gapを緩和するためには、教師ラベルを用いサンプル間の判別的な距離計量を学習する必要がある。これはdistance metric learningと呼ばれる機械学習のテーマである。本章後半では、この話題に関する先行研究をまとめた。まず、いくつかの

代表的な手法について紹介を行った。また、特にスケーラビリティに優れるバイモーダルな線形次元圧縮手法について、実装を含め詳しく説明した。

第4章 : Development of Scalable Image Annotation Method

本章では、第3章での考察を元に、バイモーダルな線形次元圧縮手法である正準相関分析(CCA)を応用した画像アノテーション手法の開発を行った。CCAの確率構造を利用することで、最適なサンプル間の判別的距離計量の導出を行い、これを**Canonical Contextual Distance (CCD)**と名付けた。CCAは、バイモーダルの次元圧縮手法として最も一般的な確率構造を備えており、ガウス分布により定義されるトピックモデルと解釈できる。CCDはトピックの間のKLダイバージェンスをサンプル間の距離計量として用いるものであり、理論的に最も妥当性の高い定式化になっている。CCDをK最近傍識別やカーネル密度推定に利用することで、スケーラブルかつ高精度な画像アノテーションが実行できる。

第5章 : Evaluation of Image Annotation Method

本章では、第4章で開発した画像アノテーション手法について、標準的なベンチマークデータセットを用い先行研究と比較実験を行った。実験により、CCDに基づく画像アノテーション手法は相対的に少ない計算コストで学習・認識が可能であり、かつ先行研究と遜色ない認識精度を達成できることが示された。また、CCDは同種の次元圧縮に基づく距離計量学習手法と比べ、最もよい認識精度となることも確認された。しかしながら、CCDを有効に利用するためには入力となる画像特徴量に関して注意が必要であることも判明した。

CCDは入力特徴空間がユークリッド空間であることを仮定している。すなわち、特徴空間の内積が特徴の生成モデルの類似度を適切に定義していることを前提としている。しかしながら、現存する多くの画像特徴量にはこの前提が成り立たず、線形手法であるCCDを適用すると著しい性能低下につながる場合がある。このような場合、一般的にはカーネル法の利用により陰に問題解決が図られるが、十分な認識精度を得るためには多くの基底サンプルをカーネル化に用いる必要がある。結果として、学習アルゴリズムのスケーラビリティは著しく損なわれ、大規模学習データへの適用は不可能となる。これは、既存の一般画像認識手法がスケーラビリティを欠いていた理由に他ならない。この問題を解決するためには、元の生成モデル間の非線形距離ができるだけ近似されるユークリッド空間上に特徴量を設計することが必要である。

第6章 : Development of Image Feature Extraction Scheme

本章では、第5章での考察を受け、CCDのような線形手法に直接適用可能な画像特徴量を記述する枠組みの開発を行った。本研究では、局所特徴分布を単一のガウシアンによってモデル化する**global Gaussian approach**を提案した。また、ガウシアンを情報幾何の手法により近似的にコーディングした大域的特徴ベクトルである**Generalized Local Correlation (GLC)**の開発を行った。**Global Gaussian approach**は、従来注目されなかった局所特徴分布の低次元統計的情報を活用することを目的としたものであり、3つのシーン認識のベンチマークにおいて最高の認識精度を達成した。その特性は以下のとおりである。

- ・ 画像ごとに固有の表現である。
- ・ 線形近似を行った場合でも、一般的なbag-of-visual-wordsと同程度の高認識精度を得る。
- ・ さらに、bag-of-visual-wordsとは相補的な関係にあるため、両者を併用することでより認識精度が向上する。

最終的な特徴ベクトルであるGLCは、ガウシアンが為す多様体の座標系をとるものであり、以下の特性を持つ。

- ・ CCDなど、特徴空間のアフィン変換に対する不変性を有する線形手法に直接適用可能である。
- ・ 一般的なbag-of-visual-wordsよりも高速に抽出可能である。

特に一つ目の特性により、CCDに基づく画像アノテーション手法にとって理想的な特徴表現になっているといえる。

第7章 : Evaluation of Large-scale Image Annotation

CCDとGLCを利用した画像アノテーション手法を1,200万枚の画像データセットへ適用し、以下の重要な知見を得た。

- ・ 学習サンプル数を増やすほど、入力画像に対し意味的に近い事例が近傍に出現する確率が向上する。この結果、認識可能な語彙数が増え、個々の画像に対するラベルづけの精度も向上する。
- ・ CCDは他の次元圧縮手法と比較して常に高い認識精度を得る。特に、学習サンプル数が増えるほど学習の安定性が向上するため、精度面での優位性が顕著となる。
- ・ 異なる多くの画像特徴量を用いるほど認識精度が向上する。特に、GLCに基づく画像特徴量が有効である。

これらはいずれも、大規模学習データを用いた一般画像認識および本研究で開発した手法の有効性を裏付ける事実である。

第8章 : Conclusion and Future Work

本論文の学術的貢献は、以下のようにまとめられる。

- ・ 従来実現困難であった、大規模学習データを用いた画像アノテーションを行うための方法論として、線形学習手法と線形評価可能な高次元画像特徴量を用いるアプローチの重要性を示した。
- ・ 上記を具体的に実現する新手法の開発を行い、実際に大規模学習データを用いた検証実験によりその有効性を示した。

本論文で開発した手法により、画像認識の精度をデータドリブンに向上させることが可能となった。今後、一連のシステムをより実用化へ近づけるためには、学習データの教示方法が特に重要になると予想される。例えば、質のよい大規模画像データセットの構築や、インタラクティブな逐次学習などが考えられる。