

知っておきたいキーワード

Attribute

～画像特徴表現における新たな潮流～

梅田崇之[†], (正会員) 入江 豪[†]

[†] NTTメディアインテリジェンス研究所

"Attribute" by Takayuki Umeda and Go Irie (NTT Media Intelligence Laboratories, Yokosuka)

キーワード: Attribute, 画像認識, 画像検索

Attributeとは

Attribute (視覚的屬性)とは、画像認識・画像検索のための比較的新しい特徴表現の一つです。明確な定義はありませんが、多くの論文で共通して言及されるのは、「人間が理解可能」、「機械が識別可能」、「複数のカテゴリ間で共有される」という三つの特性です。わかりやすい例を挙げれば、

車やバイクにおける「タイヤ」や「金属」、犬や馬における「尻尾」や「鼻」などが挙げられます(図1)。色や形状、テクスチャといった、従来のLow-levelな特徴量による画像表現とは異なり、Attributeは、それ単体として意味を持つ「部品」や「素材」などを要素とした、意味を持つ特徴表現であることが最大の特徴です。

2009年にFarhadi¹⁾やLampert²⁾に

よって提案されて以来、コンピュータビジョン、マルチメディアを中心とした分野で一大ブームを巻き起こしています。図2はコンピュータビジョン系の主要国際会議 (IEEE Conference on Computer Vision and Pattern Recognition, International Conference on Computer Vision, European Conference on Computer Vision) にて発表された、タイトルに

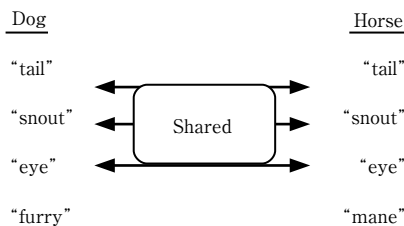
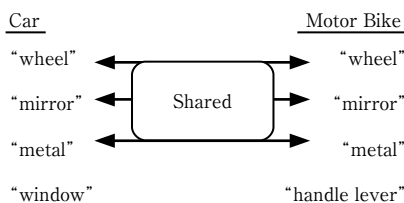
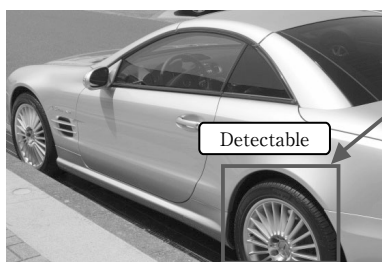


図1 Attributeの例

🔍 “Attribute” を含む論文数の推移です。その数は年々増加の一途を辿り、2013年には30本にも昇っていることから、その注目度の高さがうかがえるでしょう。

本稿では、これまでのAttributeに関する研究動向を、具体的な事例やその利点とともに概観していきたいと思えます。

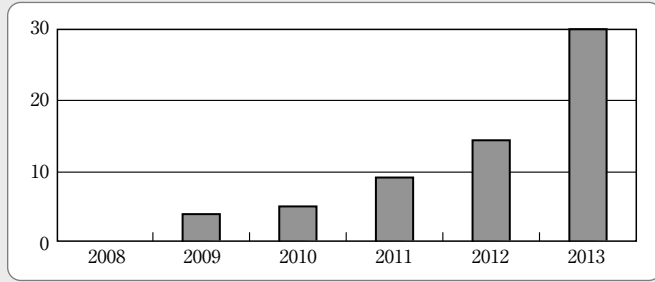


図2 タイトルに“Attribute”を含む論文数の経年変化

Attributeによる画像認識・画像検索

Attributeは、主として画像認識や画像検索のための特徴表現として利用されています。ここでは、具体的な事例とともに、そのメリットを紹介していきたいと思えます。

(1) 一般画像認識

最も基本的な利用の仕方は、画像認識のための特徴表現としてAttributeを用いるものです。従来のLow-level特徴量に基づく画像認識(図3上)とは異なり、物体の“部品”や“素材”を表すAttributeを用いることで、認識結果の根拠や物体間の関係が見えやす

くなるという大きなメリットがあります。Attributeを用いた画像認識では、まず事前に各Attributeの識別器を学習しておきます。次に、カテゴリを学習する際には、学習画像に含まれるAttributeを認識した後、Attributeとカテゴリの関係性を学習します(図3下)。文献1)では、Low-level特徴量を用いる場合に比少数の学習データで高いカテゴリ認識精度が得られることを確認しています。

(2) 詳細画像識別

Fine-Grained Visual Categorization (FGVC) と呼ばれるタスクへの適用もなされています。FGVCとは、近年、特に注目されている新しい画像認識の

タスクで、従来の一般画像認識が対象としていた「鳥」や「車」など、いわば大分類の識別を対象とするのではなく、さらに詳細な中/小分類(鳥の品種、車種による違い)の識別を目的としています。このような詳細な分類をしようとする、物体全体を見てしまってもはや区別できないことも多く、従来一般画像認識で成功を取ってきたような特徴表現(例えば、局所特徴量によるBag-of-Wordsなど)といった、画像全体を記述する特徴量ではうまくいかないことが知られていました。そこで、くちばしの色やバンパーの形といった、認識対象の微小な違いを記述できるAttributeを🔍

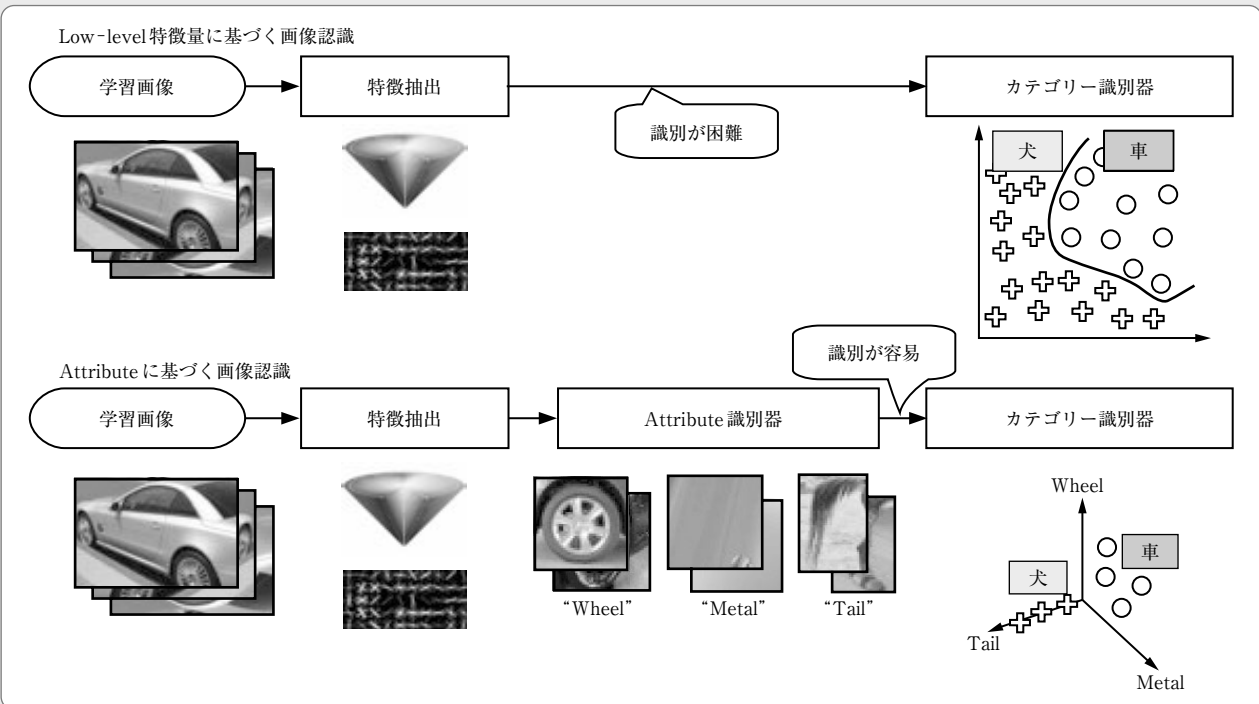


図3 画像認識のパイプライン

特徴表現として使用する試みがなされています。文献4)では、Low-levelな特徴量を用いる場合に識別が難しいカテゴリーに対して、これを効果的に識別できることを示しています。

(3) Zero-Shot Learning

「複数のカテゴリー間で共有される」というAttributeの特性を活かした、新たな問題の解決も試みられています。Zero-Shot Learningは、その名の通り、学習画像ゼロ、すなわち、認識したいカテゴリーの学習画像が存在しない場合であっても、適当な事前知識を与えることによって、これを学習・認識可能にすることを目的としています。例えば、「キリン」そのもの

の学習画像が一切なかったとしても、「キリン」は、「長い首」、「茶色」、「斑点」といった特性を持つという事前知識さえ得られていれば、このようなAttributeを備える物体を「キリン」と認識することができるようになります。文献2)では、学習したいカテゴリーとAttributeの関係性のみから、学習画像のないカテゴリーを、一定の精度で認識可能であることを示しています。

(4) 画像検索

画像認識だけでなく、画像検索に用いる試みもなされています。Attributeは、それ自体人間に理解可能な意味を持つ特徴表現ですので、画像に対して意味を持つキーワードを割当てること

ができます。これによって、これまでのキーワード検索同様、Attributeを直接クエリとして指定した画像検索が実行できる点が魅力です。さらに、文献3)では、Attribute間の関連性(相関など)を事前に学習しておくことで、クエリとして入力されなかったAttributeの情報も利用した画像検索を試みています。例えば「アジア人」の「女性」というAttributeをキーワードクエリとして与えた時、「アジア人」と関連性の低い「ブロンドの髪」というAttributeを持つ画像は除外することが可能になり、より精度の高い画像検索ができることを示しています。

Attributeの獲得

Attributeがさまざまなアプリケーションに対して有用な特徴表現であることは、ここまでご紹介してきた通りです。一方で、タスクやカテゴリーによって、有効なAttributeが異なるであろうことは容易に想像できるかと思えます。例えば、車種を識別したいのに、「鼻」や「足」など、動物に関連するAttributeばかりを揃えては意味がありません。そこで、有効なAttributeをいかに効率的・効果的に獲得するかが重要な課題となっています。

従来の取組みでは、人間が理解可能で、カテゴリーをうまく識別できるAttributeの獲得を目指し、さまざまな試みがなされてきました。大枠として、前者を達成するためには、人手をいかにうまく活用するか、後者を達成するためには、いかにカテゴリー間の差異となるような特徴量を発見するかがカギとなります。

最もオーソドックスな手法は、人手でAttributeの種類を決定してしまう方法です。文献1)~3)などで用いられています。しかし、このやり方では、Attributeの決定とAttribute識別器を学習するための学習データの準備の双

方に多大な人手を要するという問題点があります。また人手で決めたAttributeは必ずしもカテゴリー識別に有効であるとは限らない点も大きな問題です。

前者の問題に対しては、Web画像を利用することでこれを回避する試みがなされています。文献5)では、商品分類を目的としたAttributeの獲得にショッピングサイトを活用しています。商品画像にある豊富な商品説明を利用し、この説明文からAttributeとして備えるべき項目を自動発見しま

す。さらに、このように発見されたAttributeと商品画像とのペアを、そのまま学習データとして利用することで、人手を介さずにAttribute識別器を学習しています。

このように獲得されたAttributeは、「人間が理解可能である」という特性を満たすものの、必ずしも良好な識別性能を与えるとは限りません。そこで、識別性能に特化したデータドリブなAttributeを学習する手法が文献6)などで提案されています。認識したいカテゴリーの学習画像から得られた

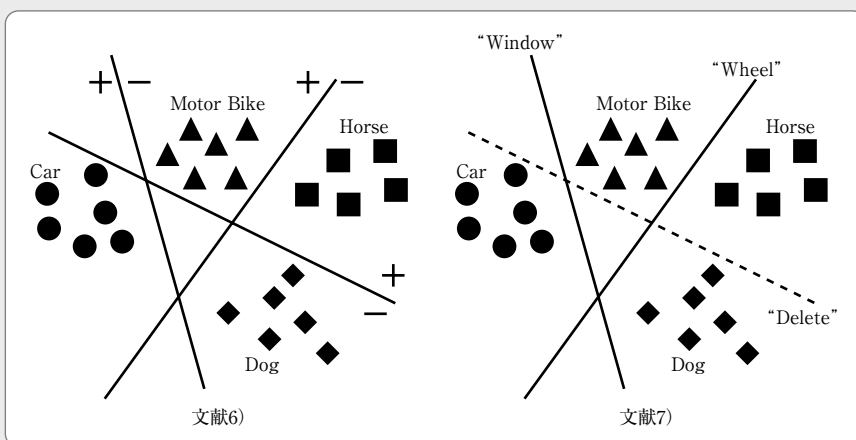


図4 文献6)7)で獲得するAttributeのイメージ

☞ Low-level特徴空間を、各カテゴリーが区別できるように分割していくことでAttributeを自動で発見します(図4左)。

さらに、人間が理解可能で、識別性能も高いAttributeの獲得を目指した

方法も模索されています。文献7)に提案されている手法では、まず、識別性能に特化したAttributeを獲得した後、これらを候補として、ヒューマンインタラクションを通じて人間に理解可能なAttributeへ更新します(図4

右)。車とバイクにあって、馬と犬にはないAttributeは例えば「タイヤ」と名付けることができます。一方で図中の破線のような、バイクと馬にあって、車と犬にはないというAttributeは考えにくいので候補から削除します。

むすび

画像特徴表現の一つの潮流となっているAttributeについて、その基本的な特徴とこれまでの取組みについて紹介してきました。Attributeは、概念的にはシンプルでありながら、さまざま

なメリットを持ち、同時に新しい問題を切り拓いてもいる興味深い話題です。画像認識・画像検索における一般的な特徴表現の一つとして根付いていくであろうことは想像に難くありません。今後、Attributeの持つ人間に理解可能であるという特性を活かした新

しい画像認識アプリケーションの創出、および、カテゴリー間で共有されるという特性を活かした超多クラスのカテゴリへの展開など、ますますの発展が期待されます。(2014年5月30日受付)

参 考 文 献

- 1) A. Farhadi, I. Endres, D. Hoiem and D. Forsyth: "Describing objects by their attributes", In CVPR (2009)
- 2) C.H. Lampert, H. Nickisch and S. Harmeling: "Learning to detect unseen object classes by between-class attribute transfer", In CVPR (2009)
- 3) B. Siddiquie, R. Feris and L. Davis: "Image ranking and retrieval based on multi-attribute queries", In CVPR (2011)
- 4) D. Kun, D. Parikh, D. Crandall and K. Grauman: "Discovering localized attributes for fine-grained recognition", In CVPR (2012)
- 5) T.L. Berg, A.C. Berg and J. Shih: "Automatic attribute discovery and characterization from noisy web data", In ECCV (2010)
- 6) M. Rastegari, A. Farhadi and D. Forsyth: "Attribute discovery via predictable discriminative binary codes", In ECCV (2012)
- 7) D. Parikh and K. Grauman: "Interactively building a discriminative vocabulary of nameable attributes", In CVPR (2011)



梅田 崇之 (うめだ たかひろき) 2010年、名古屋大学工学部卒業。2012年、同大学大学院工学研究科修士課程修了。同年、NTTサイバースペース研究所入社。現在、NTTメディアインテリジェンス研究所研究員。主に画像認識に関する研究開発に従事。



入江 豪 (いりえ ごう) 2004年、慶應大学理工学部卒業。2006年、同大学大学院修士課程修了。同年、NTT(株)入社。以来、画像検索、マルチモーダル解析と応用の研究に従事。2012年～2013年、米コロロンビア大学客員研究員。現在、NTTメディアインテリジェンス研究所研究員。博士(情報理工学)。正会員。

キーワード募集中

この企画で解説して欲しいキーワードを会員の皆様から募集します。ホームページ (<http://www.ite.or.jp>) の会員の声より入力可能です。また電子メール (ite@ite.or.jp)、FAX (03-3432-4675) 等でも受け付けますので、是非、編集部までお寄せください。(編集委員会)