

Wikipedia からの画像アトリビュート抽出に関する検討

福井 翔貴[†]桂井 麻里衣[†][†] 同志社大学理工学部情報システムデザイン学科

1. はじめに

物体のアトリビュート(例:「丸い」、「タイヤ」、「羽」など)の認識結果を中間表現とすることで、物体カテゴリの識別性能が向上することが示されてきた[1]. しかしながら、各カテゴリに固有なアトリビュートを手動で定義するには多大な労力が必要となる. そこで本稿では、Wikipedia の記事を用いたアトリビュート候補の自動抽出手法を提案する. 提案手法では、各単語に対する画像特徴の分散に着目することで、視覚性の高い単語のみを抽出する.

2. 提案手法

提案手法の概要を図1に示す. まず、与えられた物体カテゴリに該当する Wikipedia の記事より、名詞および形容詞を全て選択する. このとき、全てのカテゴリで頻出する単語を予め除去する. 次に、抽出した単語をクエリキーワードとして用いてウェブ画像を収集する. このとき、単語の視覚性を定量化するために、画像特徴空間における分散を算出する. 本稿では、各画像を 8 層 Convolutional Neural Network (CNN) [2] に入力し、7 層目の出力を画像特徴として用いる. これらの特徴ベクトルの分散を視覚的抽象スコアとして定義する. スコアが大きいほど単語の視覚性が低く、スコアが大きいほど単語の視覚性が高いことを意味する. 最終的に、最もスコアの小さい単語をカテゴリのアトリビュート候補として選択する.

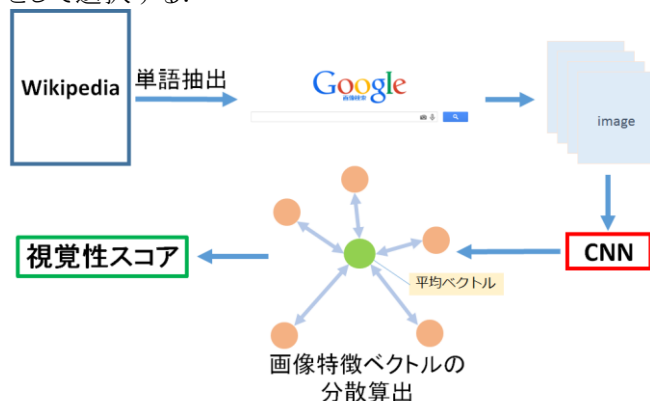


図1. 提案手法の概要.

3. 実験

本実験では、fish, car, bird, bag の4つのカテゴリに対し、アトリビュート候補となる単語を抽出する. データセット中の各単語に対し、Google 画像検索を用いて、80 枚の画像を収集した. 画像特徴空間における分散

算出にはユークリッド距離を用いた.

各カテゴリに対応する Wikipedia の記事から抽出した単語のうち、ランダムに選出した 10 個の単語に対するスコア算出結果を表1に示す. 「head」や「seat」などはそのカテゴリのアトリビュートに相応しいと思われる単語でありスコアも高くなっているが、本来、視覚性が高いと思われる「tail」のスコアが低くなった. 今後は利用する画像枚数や画像特徴量について検討する.

表1. 各カテゴリに対応する Wikipedia 記事中の単語とその視覚的抽象スコア.

fish	スコア	car	スコア
head	26.3951	seat	21.2426
skeletal	29.1348	passenger	29.0786
water	29.3836	engine	29.1746
permeable	29.4825	wheel	30.5971
circulatory	30.0119	battery	30.8263
visible	31.8622	safety	31.6219
cartilage	32.2046	automation	31.7878
spherical	33.3525	motor	32.3255
tail	34.3903	mechanical	32.3400
structure	36.0808	large	36.0236

bird	スコア	bag	スコア
hind	30.2603	leather	21.5066
narrow	30.3258	zipper	29.5357
smooth	30.6485	container	30.2440
limb	30.7085	simplicity	30.8930
adaptation	32.5619	light	31.1253
wing	32.6989	material	31.4770
feather	34.3700	cotton	31.6465
tail	34.3903	tool	32.5060
unique	34.8907	small	32.8569
beak	41.6253	basket	34.2929

4. まとめ

本稿では、物体カテゴリに対応する Wikipedia 記事から、アトリビュート候補となりうる視覚性の高い単語を自動抽出する手法を提案した. 今後は、様々な画像特徴量を検証するほか、カテゴリ間識別・カテゴリ内識別のためのアトリビュート候補の分類に取り組む.

参考文献

- [1] A. Farhadi et al., “Describing objects by their attributes,” in Proc. CVPR, 2009, pp. 1778–1785.
- [2] Y. Jia, et al., “Caffe: Convolutional architecture for fast feature embedding,” in Proc. MM, 2014, pp. 675–678.