

Web画像の分布に基づく 単語概念の視覚的な多様性の推定

カスタナー マークアウレル, 井手 一郎
川西 康友, 平山 高嗣, 出口 大輔, 村瀬 洋

名古屋大学大学院情報学研究科

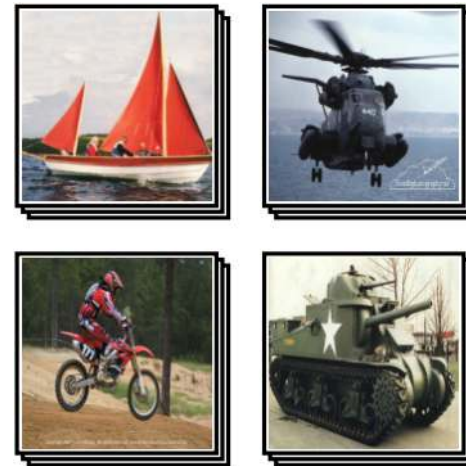
2018年3月1日CVIM

視覚的多様性

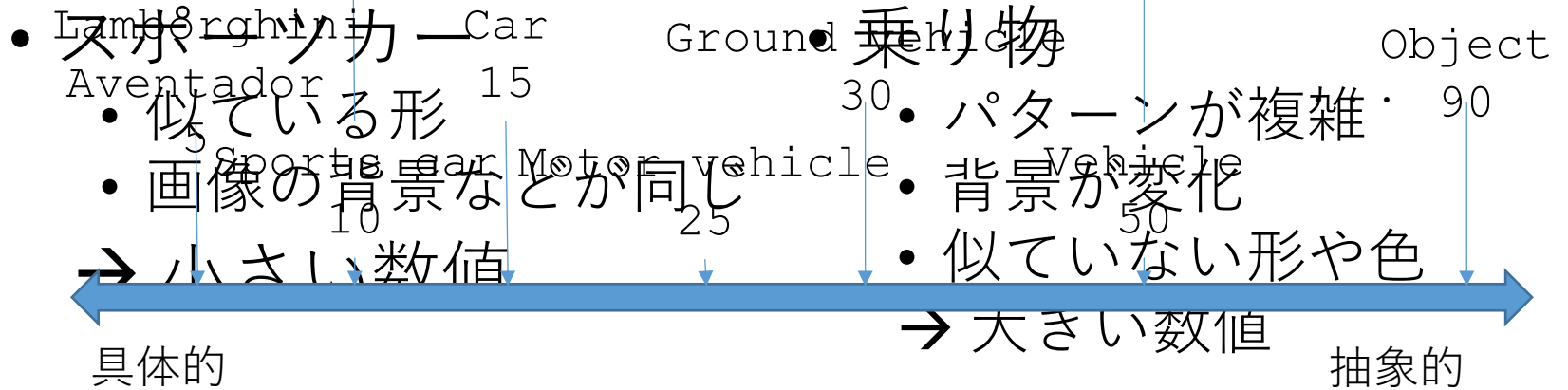
単語概念の広さは？



$vv = 10$

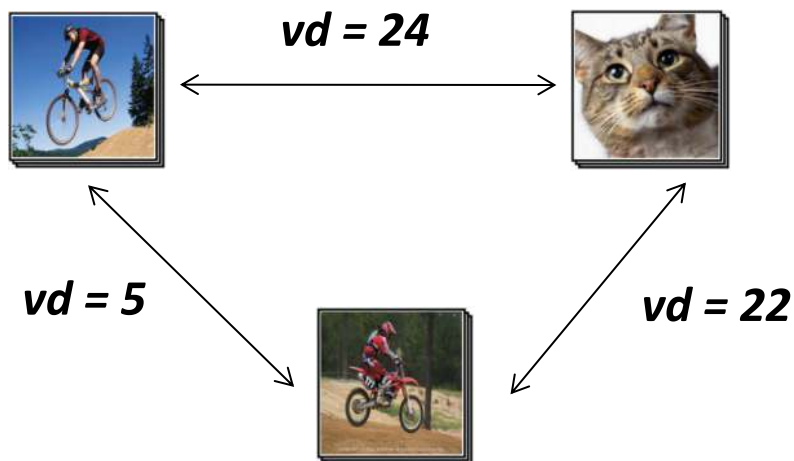


$vv = 50$



関連研究 (1/2)

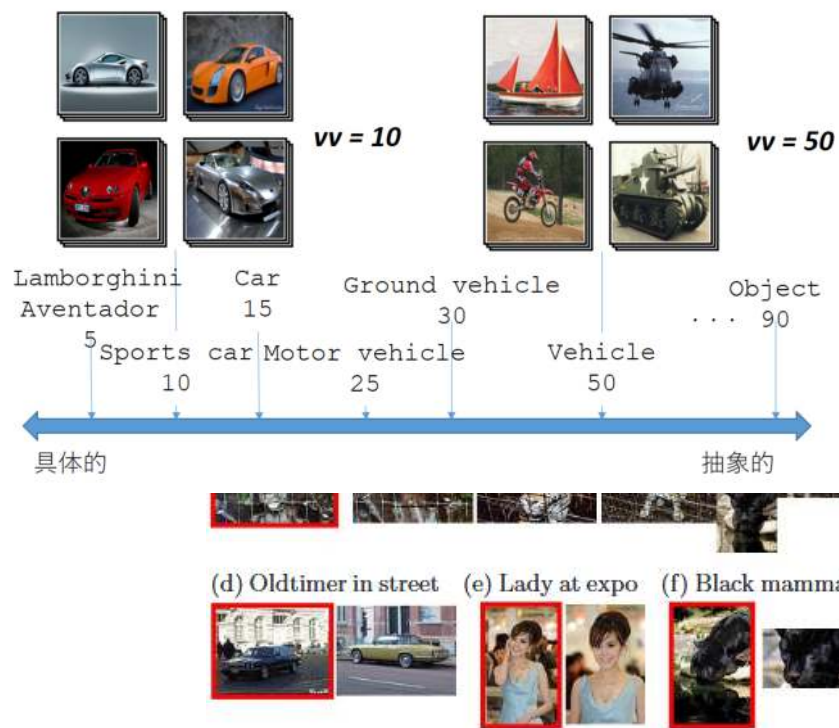
- 概念間の視覚的な距離の推定[1]
 - 画像特徴に対する概念間距離を測定
 - いくつかの研究が存在
 - ノイズの画像の影響なども分析



1: K. Nakamura and N. Babaguchi. Inter-concept distance measurement with adaptively weighted multiple visual features. ACCV 2015

関連研究 (2/2)

- 視覚的多様性の推定[?]
 - ない!
- 最も近い研究: 視覚的多様性
 - 画像検索結果を改良するため
概念の画像を縮退させる研
 - 視覚的多様性は推定しない
 - 評価実験
 - 人間が作成した多様化と比較
 - 多様化の定量的な評価はない



本研究の目的

- 画像と自然言語の間に存在するセマンティックギャップを克服するために
 - まずはセマンティックギャップを定量化したい

⇒単語概念の視覚的多様性の度合を推定
- 応用
 - セマンティックギャップが小さい語彙の選択
 - 機械翻訳
 - 説明文の自動作成
 - 視覚的多様化[2]の定量的評価

視覚的多様性の推定

1. 画像の特徴量を抽出
2. 特徴空間のクラスタリング
3. 多様性の度合 = クラスタ数

$vv = 10$



視覚的に多様な概念ではクラスタ数が多くなる

視覚的多様性の推定



- 画像特徴空間のクラスタリング
 - 異なる概念でも似た状況の画像→同じクラスタ
 - 同じ概念でも異なる状況の画像→別のクラスタ
- データセットに含まれる画像分布が推定に強く影響
 - 人間の感覚に近い画像分布で構成したデータセットが必要

理想的な分布

- 主観的なもの
- 各単語概念に含まれる下位概念の画像数の現実的な分布
 - 珍しい概念は少なめ
 - 日常的な概念は多め
- したがって,
 - 仮定 1 : クラウドソーシングにより決定した分布が理想
 - 仮定 2 : Web上の画像分布は理想的な分布に近い

画像コーパスの構築

- 各単語概念について...

1. 下位概念を収集

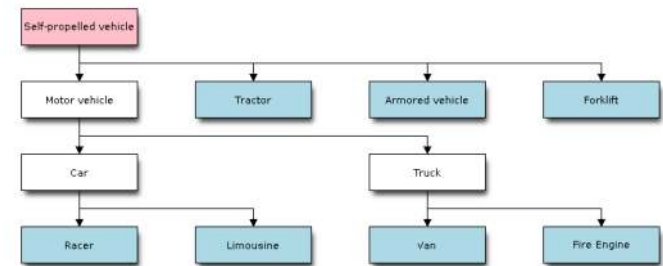
- WordNetの構造を利用

2. 下位概念の画像を収集

- ImageNet中の画像 + 不足分はWebクロール

3. Web上の画像分布に基づいて下位概念の重みを決定

4. 重みに従った枚数分、下位概念の画像を選択し、単語概念の画像セットを構成



Web上の画像分布

アイデア



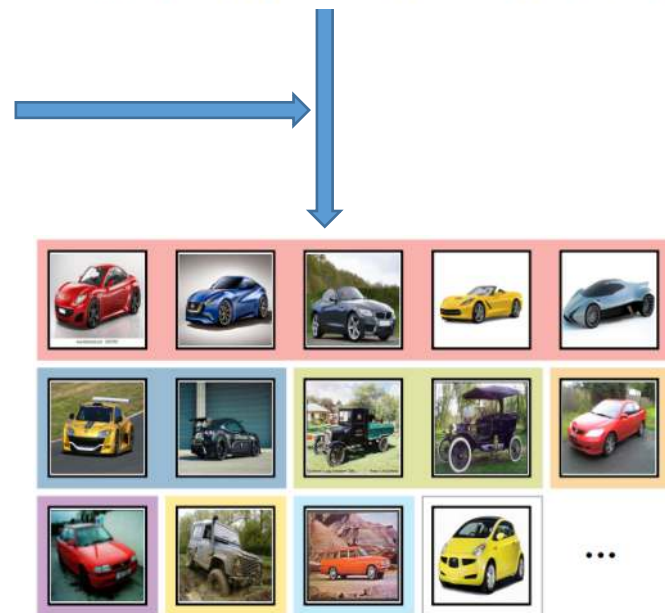
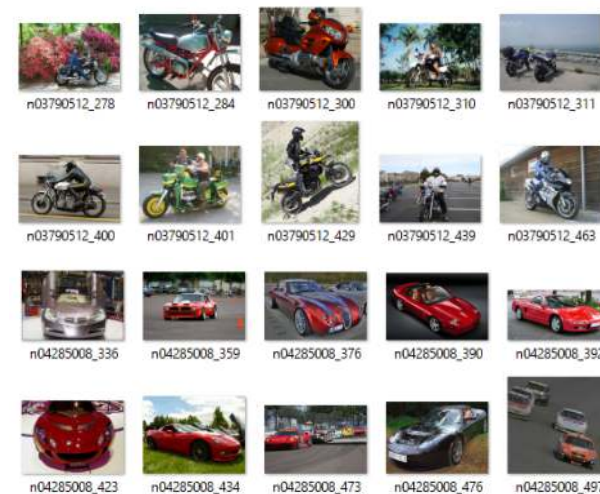
- Google APIで収集可能
 - Google画像検索のヒット数
- ヒット数に応じて、各下位概念の重みを決定

画像コーパスの再構成

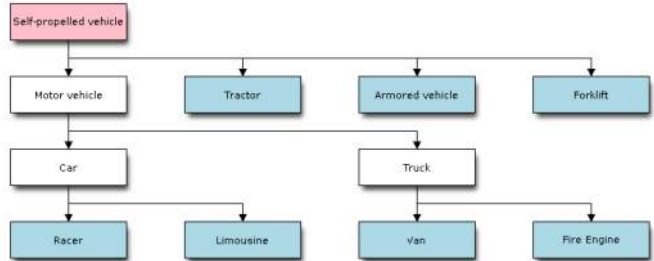
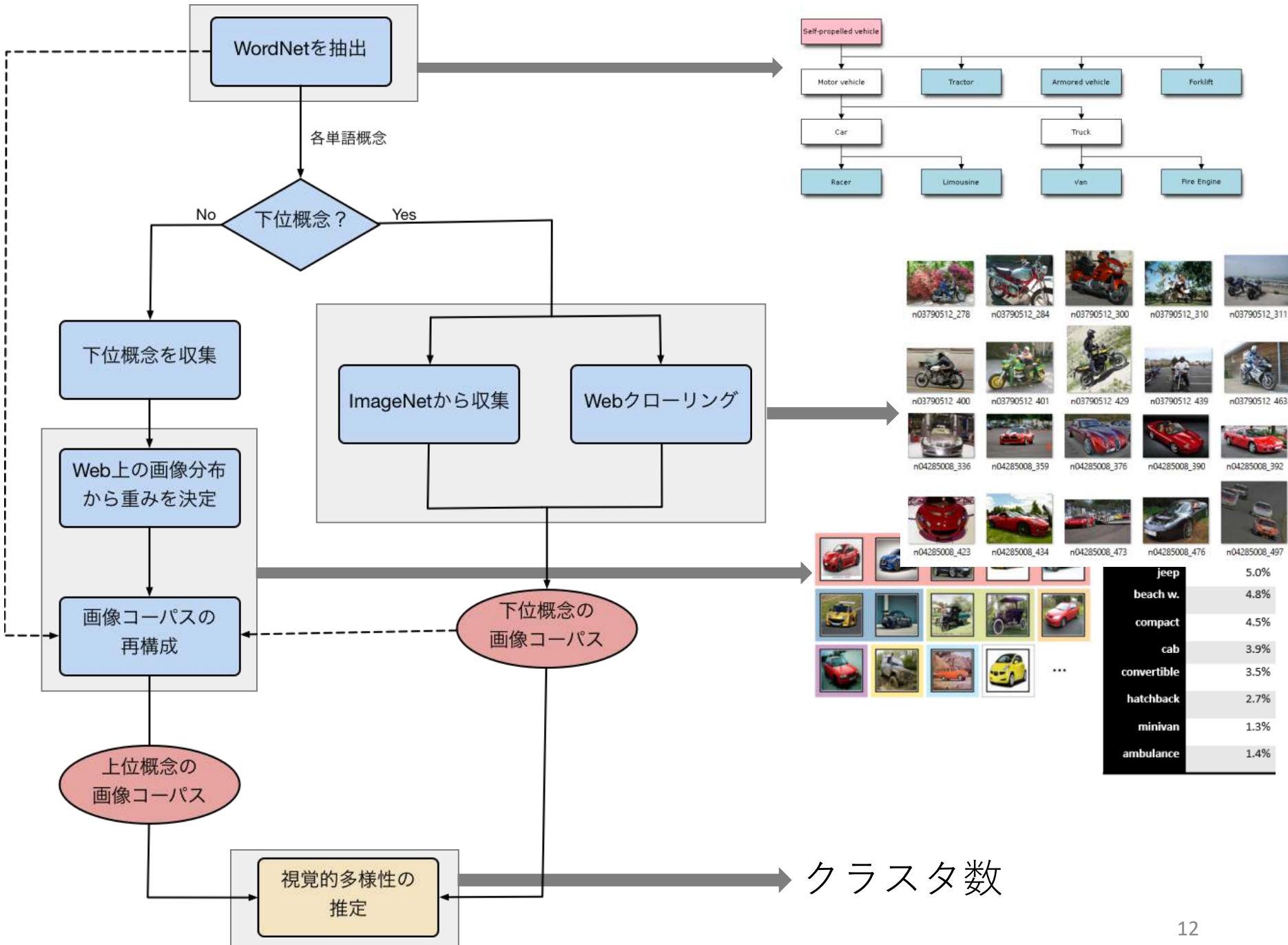
- 各単語概念の画像セットを再構成

→理想的な画像分布に近いデータセット

	Google Image
sports car	27.4%
racer	9.2%
Model T	8.8%
coupe	6.9%
used-car	6.7%
jeep	5.0%
beach w.	4.8%
compact	4.5%
cab	3.9%
convertible	3.5%
hatchback	2.7%
minivan	1.3%
ambulance	1.4%



Re-composited dataset for "car"



jeep	5.0%
beach w.	4.8%
compact	4.5%
cab	3.9%
convertible	3.5%
hatchback	2.7%
minivan	1.3%
ambulance	1.4%

真値の決定

- 評価実験を行うために真値を決定
 - Thurstoneの1対比較法[3]
 - 乗り物に関する15個の単語概念
「Sports car」, 「Airplane」, 「Boat」 など
- クラウドソーシングにより
 - Facebook, Twitter, Redditなど (SNS)
 - 世界中の75人の男女から2,139対の比較結果を収集

被験者実験

How variant are these words?

Let's create a mental image for them, and think about it for a second...

Survey

Which concept **related to vehicles** has more visual variety?

jeep

A small, sturdy motor vehicle with four-wheel drive, especially one used by the military

jeep has more variety

sailing vessel

a vessel that is powered by the wind; often having several masts

sailing vessel has more variety

I don't know

About equal

Submit



A lot of different looking animals exist.



Similar animal in different situations...

評価実験

- 3つのデータセット

- ImageNetの画像コーパスそのまま（ベースライン手法）
- 均一な重みで再構成した画像コーパス（比較手法）
- Web上の画像分布に基づく重みで再構成した画像コーパス（提案手法）



評価実験

- アプローチ

- 画像特徴：SURF特徴量のBag-of-Words表現
- クラスタリング：Mean-Shift法

→ クラスタ数を単語概念の視覚的多様性とする

- 真値

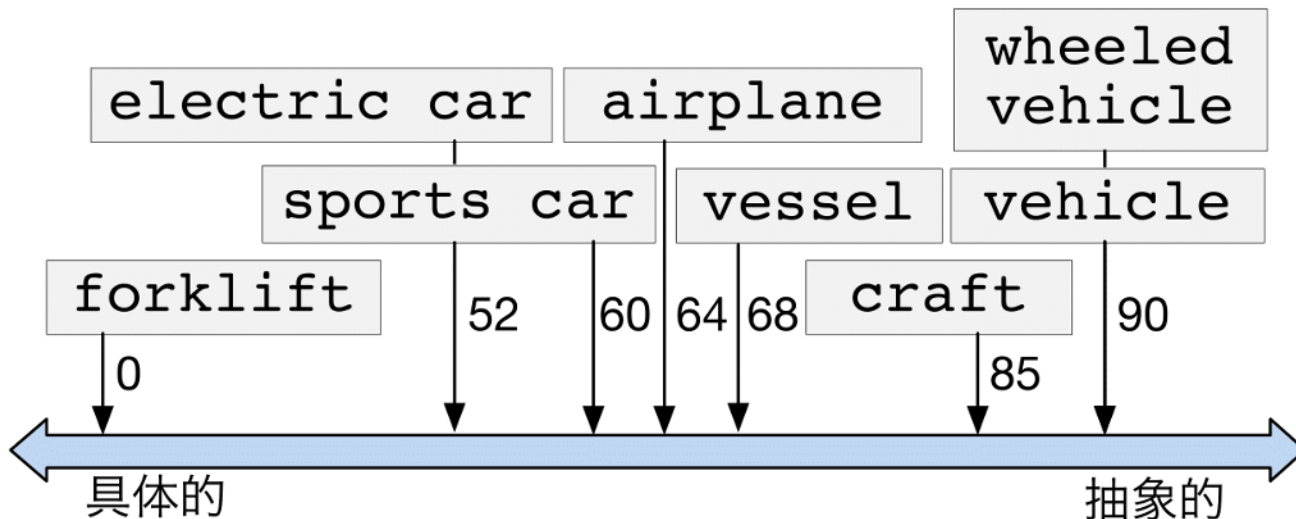
- 被験者実験の対比較結果を最尤推定した間隔尺度値

- 評価指標

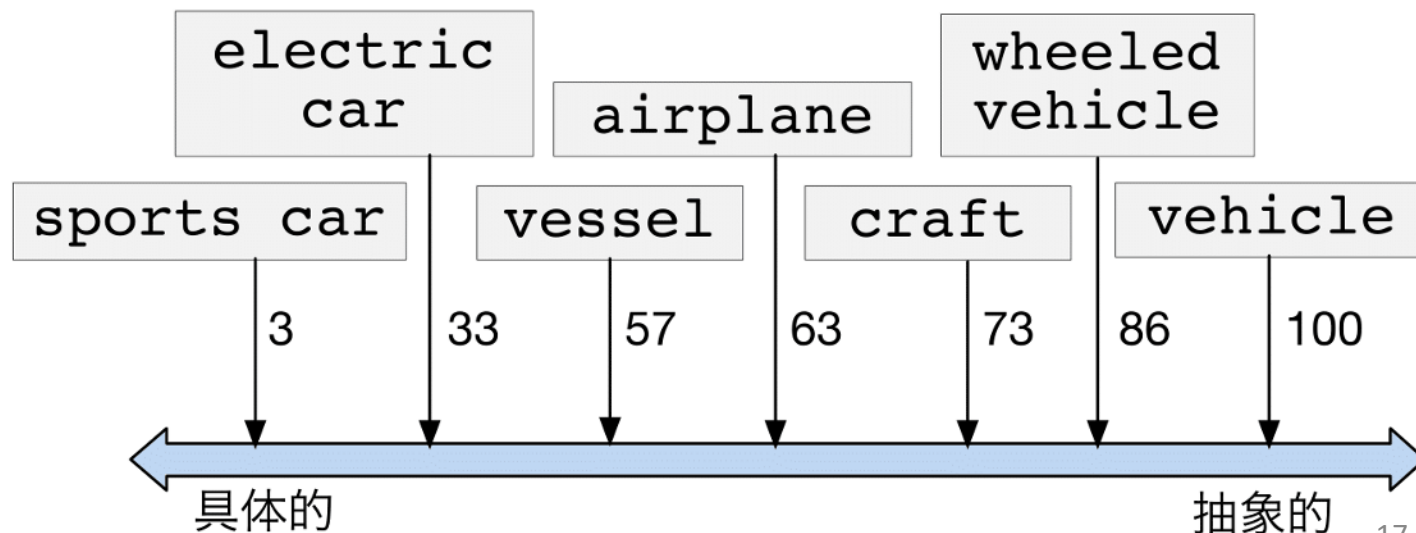
- Spearmanの順位相関係数
- 平均2乗誤差

評価実験

真値



提案
手法



評価実験

データセット	順位相関係数	平均2乗誤差
ベースライン手法 (ImageNetそのまま)	0.45	9.08
比較手法 (均一な重み)	0.73	6.11
提案手法 (Web上の画像分布に基づく重み)	0.80	4.56

- ベースライン手法は真値とほとんど一致しない
- 順位相関は比較手法で約60%，
提案手法でさらに約10%も向上！

むすび

- 単語概念の視覚的多様性を推定する手法を提案
 - 理想的な画像分布に近づくように既存なデータセットを再構成
 - クラウドソーシングにより被験者実験で真値を決定
 - ベースライン手法より約**80%**も向上し，視覚的多様性をより正確に推定できることを確認
- 今後の課題
 - WordNetにない単語概念に対応
 - 画像のメタデータなどの利用
 - 語彙選択のアプリケーションに適用して効果を確認