画像理解システムの演習

https://www.kkaneko.jp/ai/ae/index.html

金子邦彦

BY NC SA

アウトライン

番号	項目
	復習
1	画像理解の応用例
2	演習

各自、資料を読み返したり、課題に取り組んだりも行う

コンピュータによる画像理解

- コンピュータが画像を理解する

画像が何であるか，物体がどこに，どういう大きさ，形で，いくつあるか，をコンピュータが説明できる能力を持つ

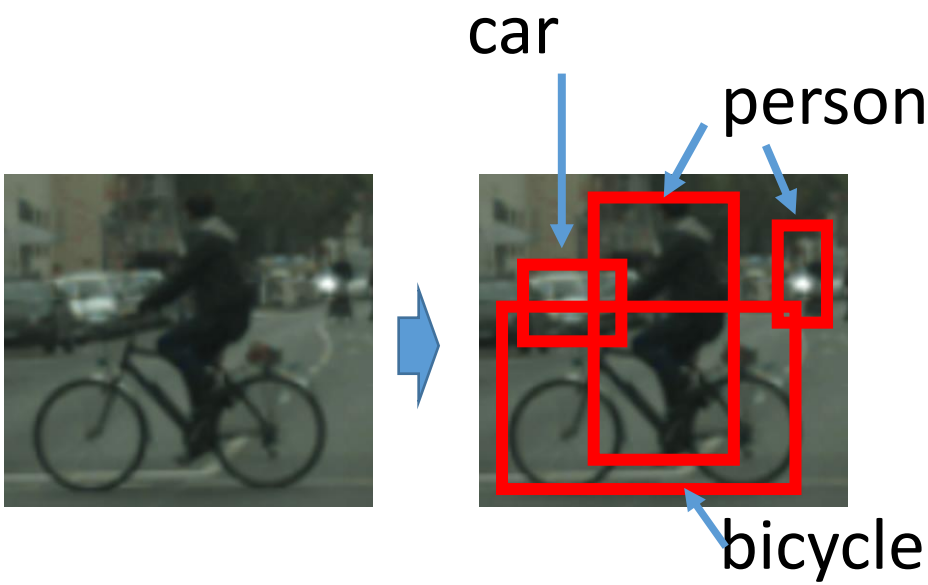
① 画像分類

```
Score 0.9827020168304443, Label lab_coat
Score 0.0030872616916894913, Label syringe
Score 0.0024311079178005457, Label beaker
Score 0.0016609227750450373, Label stethoscope
Score 0.00037950885598547757, Label plate
```

画像分類の結果は，ラベルと**確率**

※ 5 つの候補 (top 5) が表示されている

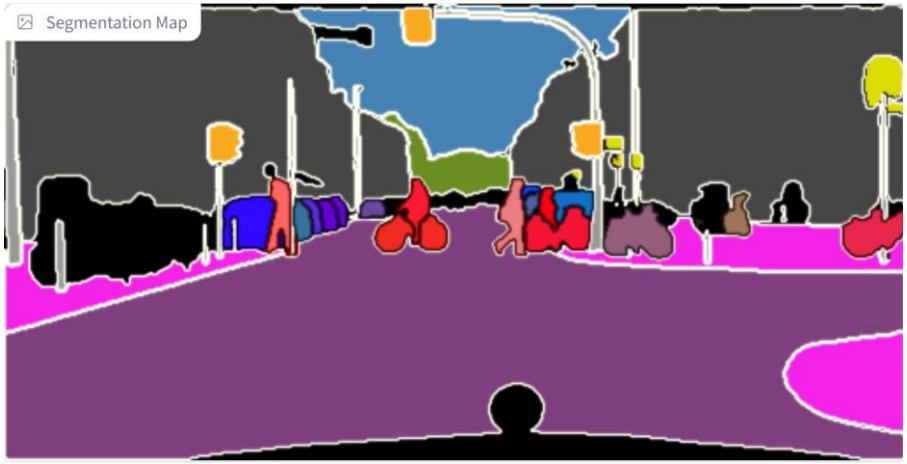
② 物体検出



バウンディングボックス，
ラベルを得る

バウンディングボックスは，
物体を囲む最小のボックス（四角形）

③ セグメンテーション



物体の形を画素単位で抜き出し

ラベルを得ることもできる

画像理解の主な種類

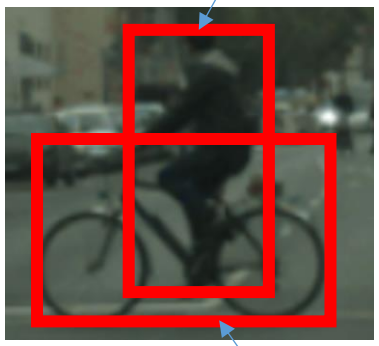
① 画像分類

「何があるか」を理解

person
bicycle

② 物体検出

場所と大きさも理解

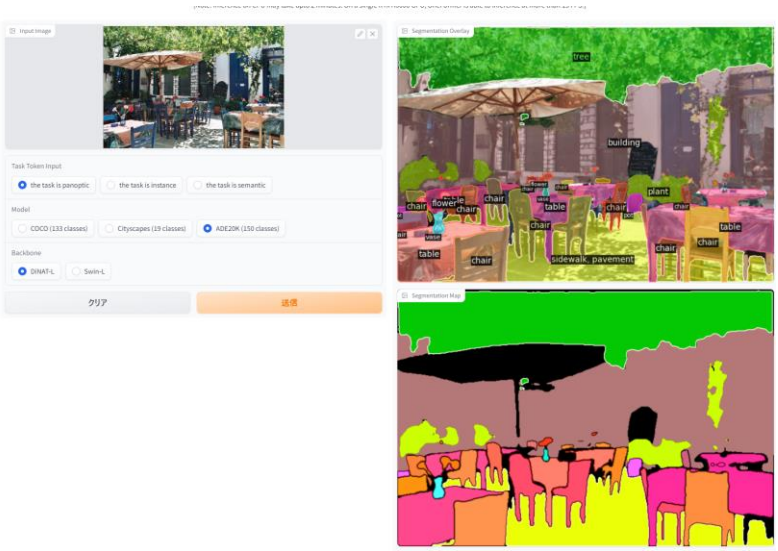


③ セグメンテーション

画素単位で領域を理解

セグメンテーションを試すことができるオンラインのサイト

- OneFormer のデモサイト
- URL: **https://huggingface.co/spaces/shi-labs/OneFormer**
- セグメンテーションの種類：**パノプティック，インスタンス，セマンティック**
- データセット：**COCO**（133 クラス），**Cityscapes**（19 クラス），**ADE20K** （150クラス）



文献: Jitesh Jain, Jiachen Li, MangTik Chiu, Ali Hassani, Nikita Orlov, Humphrey Shi, OneFormer: One Transformer to Rule Universal Image Segmentation, arXiv:2211.06220, 2022.

訓練データにより結果が変わってくる

元画像

訓練データは Cityscapes

訓練データは COCO

訓練データは ADE20K

- OneFormer のデモサイトを使用
- URL: **https://huggingface.co/spaces/shi-labs/OneFormer**
- パノプティック・セグメンテーションを実行
- バックボーンは DiNAT-L を使用

ADE20K

- アノテーション済みの画像データ
- オブジェクト(car や person など) も，背景領域も (grass, sky など) ，画素単位でアノテーションされている
- 画像数： 30,574
- クラス数: 3,688

ADE20K データセットの URL:
http://groups.csail.mit.edu/vision/datasets/ADE20K/

文献: Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, Antonio Torralba,
Scene Parsing Through ADE20K Dataset,
CVPR 2017, also CoRR, abs/1608.05442, 2017.

CityScapes

- アノテーション済みの画像データ
- 50都市の数ヶ月間（春，夏，秋）の日中，良好な/中程度の天候のもとで撮影，計測
- 画像数： 24,998
- クラス数: 30

road, sidewalk, parking, rail track, person, rider, car, truck, bus, on rails, motorcycle, bicycle, caravan, trailer, building, wall, fence, guard rail, bridge, tunnel, pole, pole group, traffic sign, traffic light, vegetation, terrain, sky, ground, dynamic, static

CityScapes データセットの URL:
https://www.cityscapes-dataset.com/

文献: Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Frank¥
e, Stefan Roth, Bernt Schiele,
The Cityscapes Dataset for Semantic Urban Scene Understanding,
CVPR 2016, also CoRR, abs/1604.01685, 2016.

COCO

- 画像データ，人体のランドマーク，人体姿勢のデータ
- ラベル付け済みの画像数： 200,000以上
- オブジェクトのクラス数: 80
- ランドマーク：左目、鼻、右腰、右足首などの 17 のキーポイント

COCO データセットのURL: https://cocodataset.org/

文献: Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollr,
Microsoft COCO: Common Objects in Context,
CoRR, abs/1405.0312, 2014.

1. 画像理解の応用例

計測（寸法，面積，角度など）

数の数え上げ

異物の発見

動画での利用（動きの把握，監視，変化の発見）

2. 画像理解システムの構築

① 前準備

- **既存の技術の利用・評価**
- **学習済みモデルの利用・評価**

② 応用・展開

- ①で「どこまで役に立つか（性能，機能）」を考察，十分に検証
- 不足があれば改良
- 既存の記述の適切な利用（利用条件，著作権等）の再確認

演習

① 目的

画像理解システムの「前準備」の部分.

- **既存の技術の利用**
- **学習済みモデルの利用**

について，実際に各自で体験すること.

画像理解システムの構築について理解を深め，スキルを高めること

② 注意点

1. どのような用途で役立てるかは，各自で，自由に想像すること
2. **画像は各自で準備すること**
3. 必ずしも完璧な精度で結果が得られるわけではない．**誤りや誤差を含む**ものである．

4. **得られた結果で何に役に立つか**の考察も大切である．

③実験の基礎

セグメンテーションの種類：

パノプティック (panoptic)

セマンティック (semantic)

インスタンス (instance)

学習済みモデルのバリエーション：

COCO, Cityscapes, ADK20K など

④ 手順

（１）次のページで公開されているページを利用

- OneFormer のデモサイト
- URL: **https://huggingface.co/spaces/shi-labs/OneFormer**

（２）画像は各自で準備

（３）このサイトで,

panoptic, instance, semantic（3通り）

COCO, Cityscapes, ADK20K（3通り）の組み合わせ

計**9通り**を試す

（４）Backbone のところは「DiNAT-L」で実施

⑤各自で考察して欲しいこと

- panoptic, instance, semantic（3通り）での違い
- COCO, Cityscapes, ADK20K（3通り）での違い
- 実行は簡単だったか，難しかったか
- 「どのような用途で役立てるか」を想像してください．役に立ちそうですか？　AIを全く使わずに，人間による作業だけで行った場合と比べて，良いところはありそうですか？