

at-6.物体検出

(ディープラーニングのシステムとプログラミング)
(全12回)

<https://www.kkaneko.jp/ai/at/index.html>

金子邦彦



1. **物体検出は、画像内の物体の種類、位置、サイズを特定するタスクである。**
2. **物体検出では、スケールや特徴の多様性に対応する必要がある。バックボーンとネックからなる多層ニューラルネットワークを用いられている。**
3. **YOLOは高速かつ高精度な物体検出モデルで、畳み込みニューラルネットワークをベースとする。**
4. **COCOなどの大規模データセットを用いた学習によって、物体検出の精度が向上。**

6-1. イントロダクション

画像理解の主な種類

① 画像分類

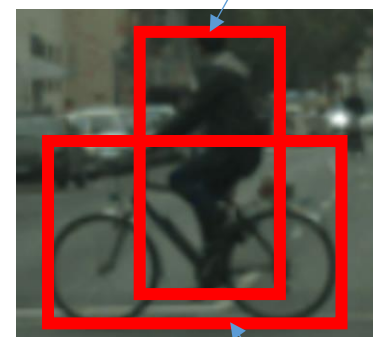
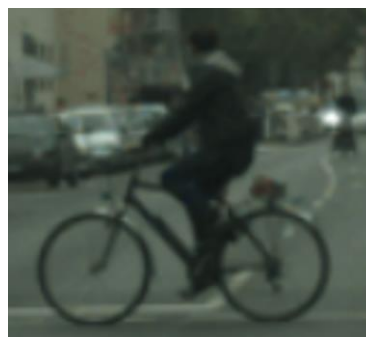
「何があるか」を理解



person
bicycle

② 物体検出

場所と大きさも理解

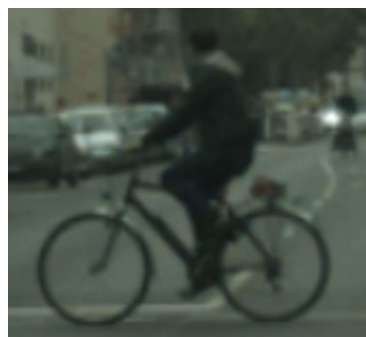


person

bicycle

③ セグメンテーション

画素単位で理解



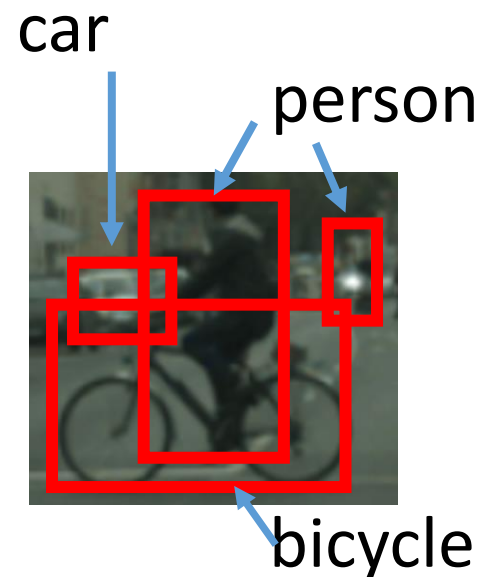
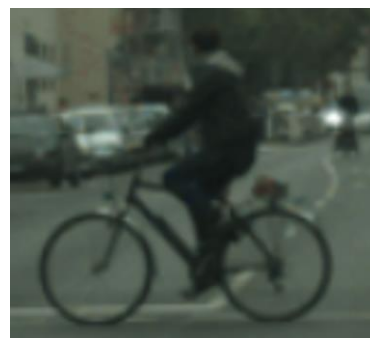
① 画像分類



```
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

ここまでのまとめ

画像理解の主な種類

• 画像分類

「何があるか」を理解。結果は「ラベル」として識別
各ラベルに対する「確率」も提供

• 物体検出

物体の種類、場所、大きさを理解。場所と大きさについての結果は、物体を囲むバウンディングボックス。

セグメンテーション

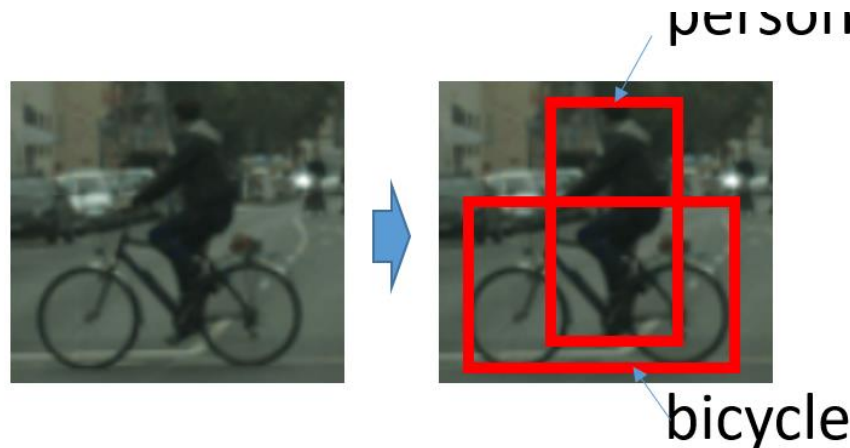
画素単位で理解。

物体の「形」を詳細に抽出。

6-2. 物体検出と YOLO

物体検出

場所と大きさも理解

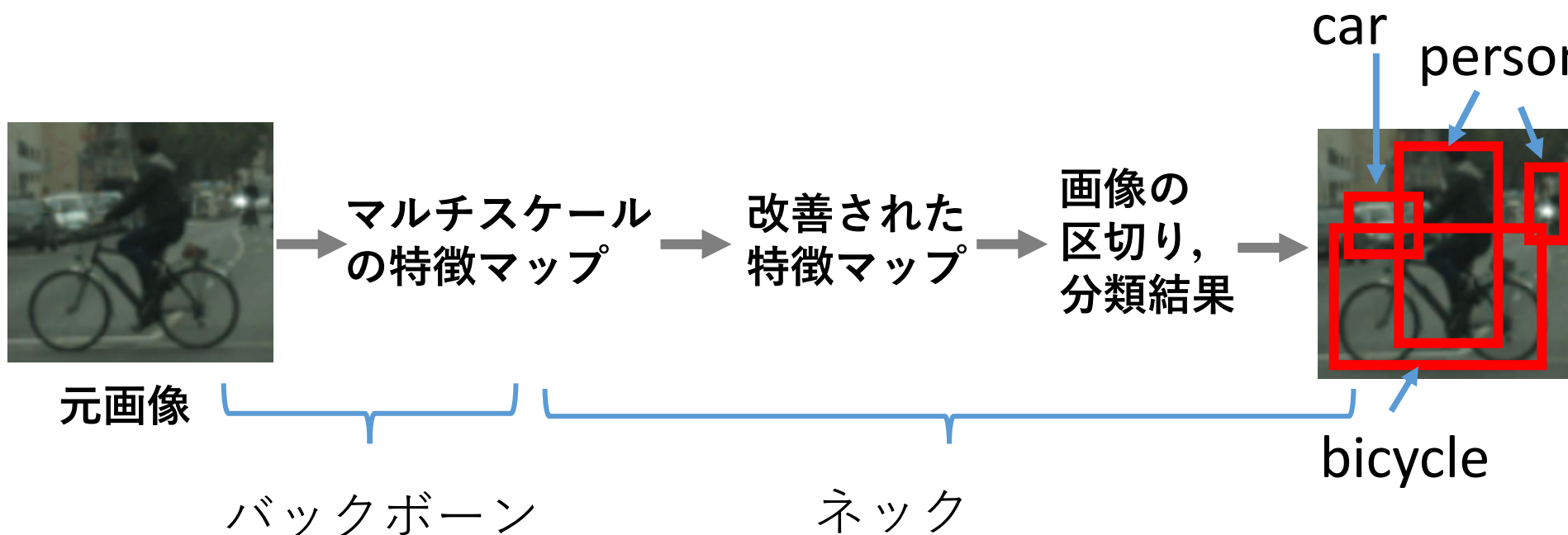


- **スケールの多様性**：画像内には大きささまざまな物体が存在
- **不要な特徴やパターンの排除**：検出に不要な特徴やパターンを取り除く必要がある。
- **特徴の多様性**：検出対象の物体は、多岐にわたる特徴を持つ。

→ バックボーン、ネック

バックボーンとネック

- **バックボーン**： 画像全体から**多様なサイズと形状の特徴**（マルチスケールの**特徴マップ**）を抽出
- **ネック**： バックボーンで得られた**特徴マップ**を**改善、統合し、さまざまなサイズの物体を検出**する



バックボーンとネック

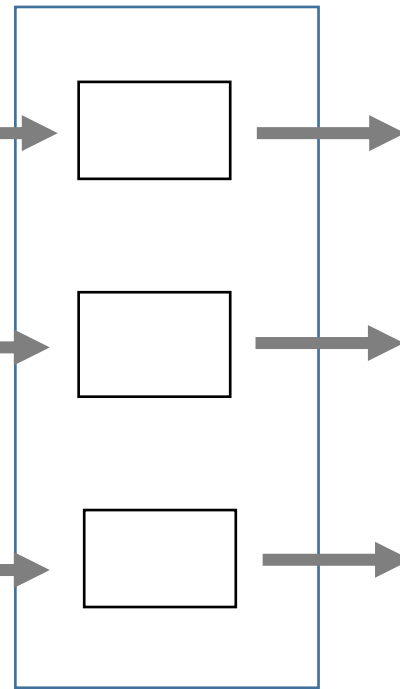
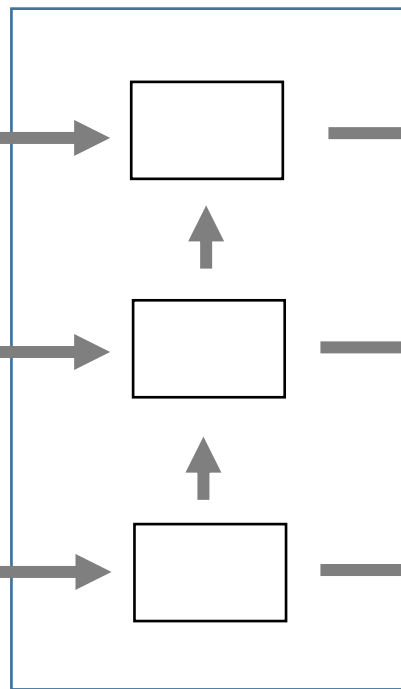
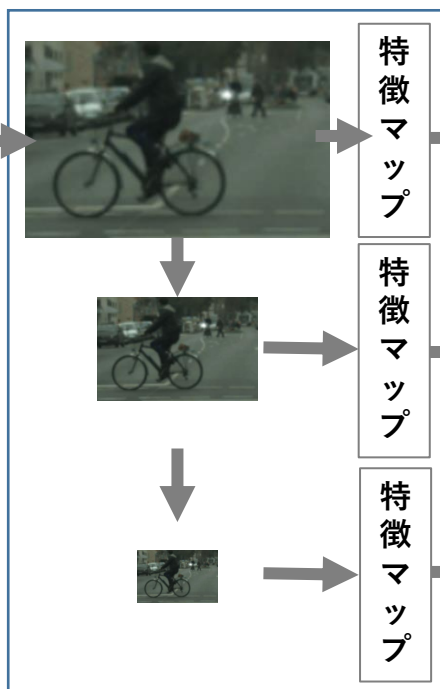
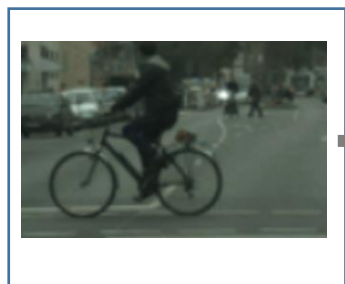
特徴マップ：画像の中の特定のパターンがどこにあるかなど示したマップ

マルチスケール化,
特徴マップ
(backbone)

各スケールの特徴
マップを上位に
反映させるなど
(neck)

画像を区切る処理,
画像分類など

物体検出の結果



元画像

マルチスケールの
特徴マップ

改善された
特徴マップ

画像の区切り,
分類結果

全体で多層のニューラルネットワーク

YOLO

- YOLO (You Only Look Once) は物体検出のモデル
- 畳み込みニューラルネットワークを基盤とする
- 高速に動作
- 高い検出精度
- バックボーンとネックの構造

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 $\text{\textcircled{e}}$, 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.

まとめ



物体検出の特質

- スケールの多様性
- 不要な特徴やパターンの排除
- 特徴の多様性

バックボーンとネック

- バックボーン: 画像からマルチスケールの特徴マップを抽出
- ネック: 特徴マップを改善、統合し、さまざまなサイズの物体を検出

YOLO (You Only Look Once)

- 物体検出のモデル
- 畳み込みニューラルネットワークを基盤とする
- 高速に動作
- 高い検出精度
- バックボーンとネックの構造

COCO データセット

- 画像理解で広く使用される
- 80カテゴリ、数十万枚の画像、その他
- 現実シーンを模倣

授業の学ぶ意義と満足感

- 最先端の物体検出技術を知る。
- 情報工学への理解が深まり、満足感が得られる。
- 物体検出に関する実践的知識が身につき、スキルが向上する
- 物体検出技術の進展と課題解決の工夫を理解。知的好奇心が満たされる。
- 追加の自習にチャレンジすることにより、学習済みモデルの活用法や追加学習を知り、将来の応用への道を拓くことができる。