

コンピュータビジョン = コンピュータが『視覚』を持つこと

入力



実世界の様子
(カメラ映像・画像)

コンピュータ
が理解



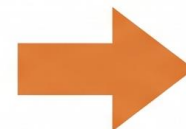
理解

何があるか
(認識)

何が起きているか
(状況把握)

何が起きそうか
(予測)

理解を
活用



活用

現実世界での
判断・行動に
役立てる

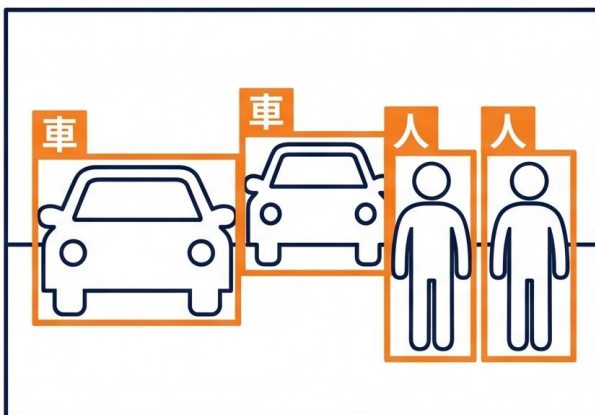
例：自動運転、
検査の自動化

コンピュータビジョンの主要タスク



種類、画素、個体の特定の違い

物体検出



物体の位置と大きさを
四角形で特定する

セマンティック・ セグメンテーション



画素ごとに種類 (クラ
ス) を割り当てる。

インスタンス・ セグメンテーション



種類に加えて、同じ種
類でも一つ一つ
(個体) を区別する

画像分類



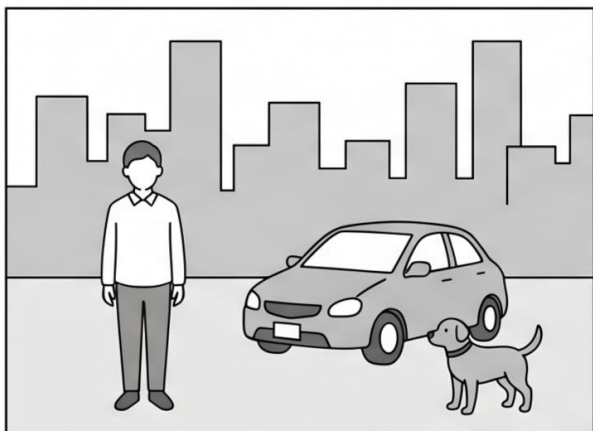
画像分類：画像全体を見て『何が写っているか』を1つのクラスラベルで答える



ポイント：出力は『どこに写っているか』ではなく、『何が写っているか』というラベル
画像全体に対して1つの答えを返す（物体検出のように位置は示さない）

物体検出：画像の中の「何が」「どこに」「どの大きさと」あるかを特定する

入力

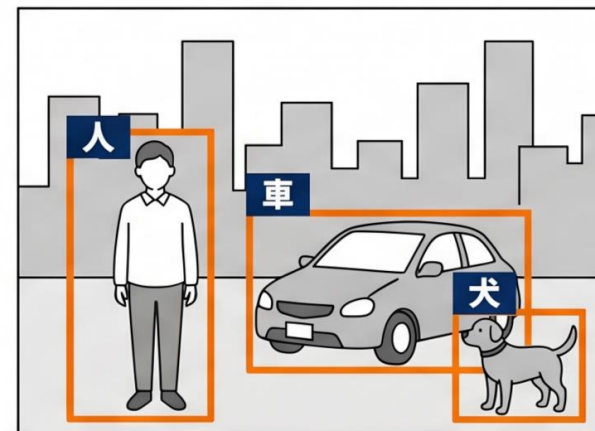


物体検出

位置・大きさ・種類を
同時に特定



出力

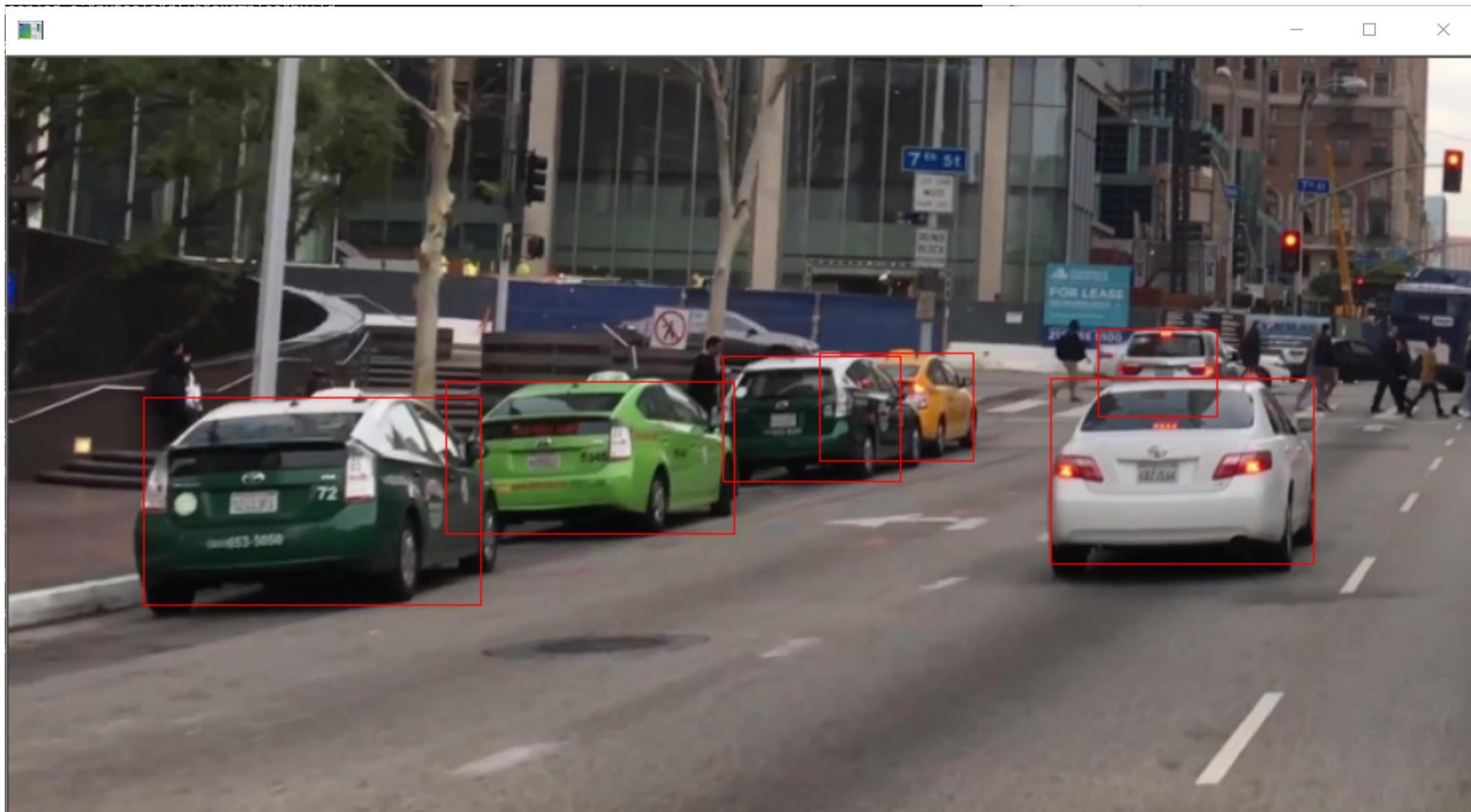


バウンディングボックス＝物体を囲む最小の四角形



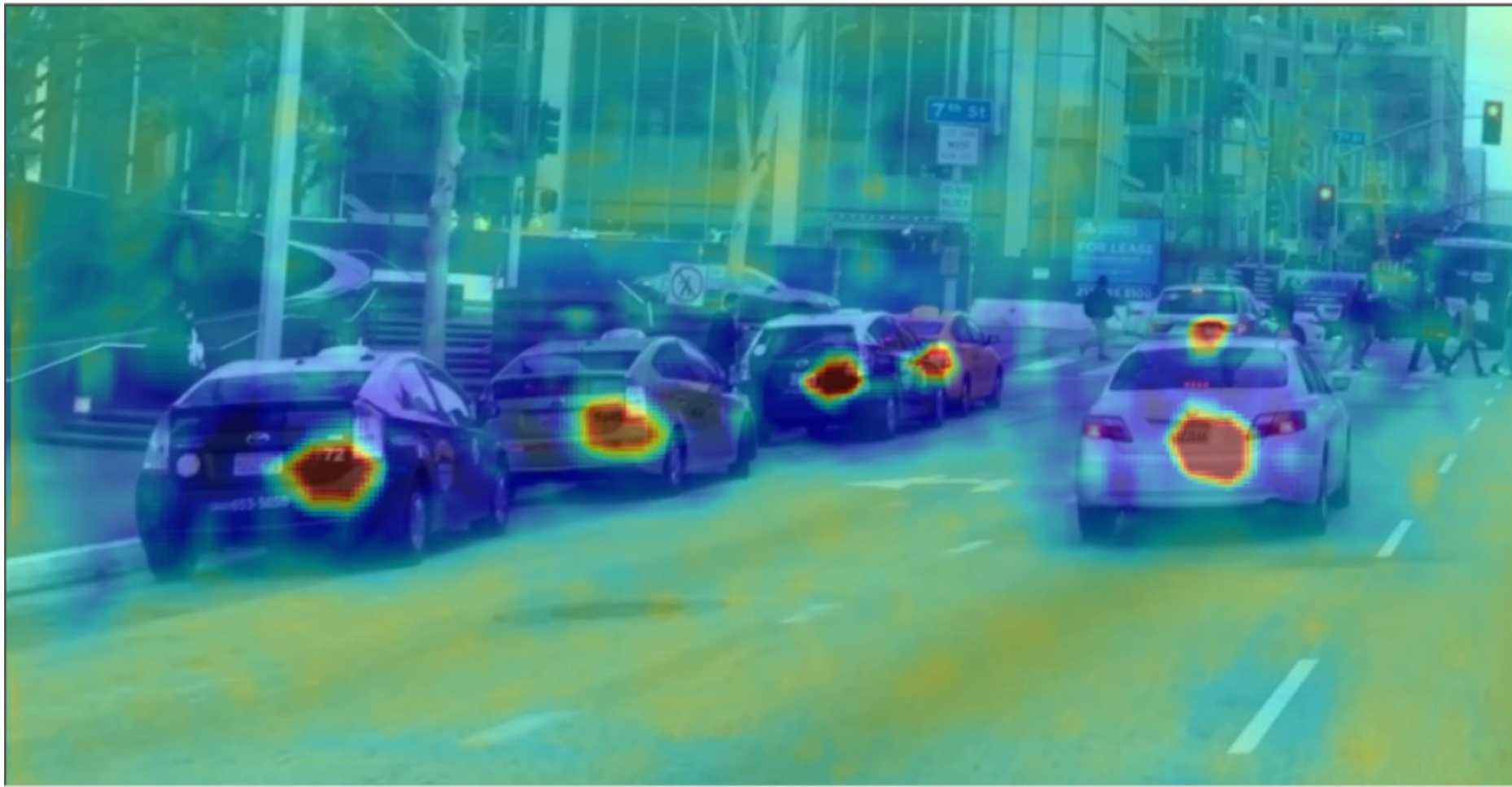
出力されるのは「物体を囲む四角形 (位置と大きさ)」と「ラベル (種類)」
四角形は物体を囲む最小の長方形で、位置と大きさを表す

物体検出



物体検出における特徴マップ

Collapsed detection scores on raw image

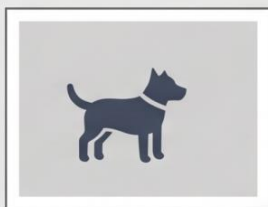


物体検出モデル YOLO のしくみ



"物体の位置" と "種類" を一括で予測する高速な物体検出モデル

画像分類 (既知)



犬

画像全体に1つの種類を予測

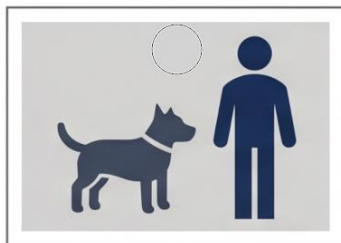
物体検出 (YOLO)

+位置



位置 (枠) + 種類 を同時に予測

入力画像



単一のネットワーク (CNN基盤)

A

バックボーン
特徴抽出



B

ネック
特徴マップの
統合・改善



出力



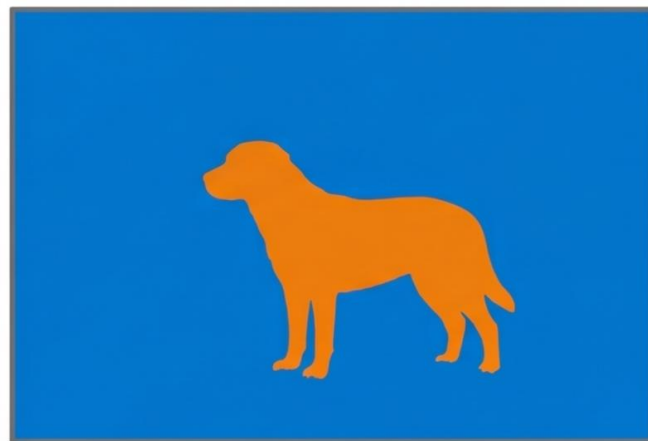
1段階方式：領域の候補を別に求めず，1つのネットワークで一括推論 → 高速

**セグメンテーション：画像を意味のある領域に分割し、
物体の形を詳細に解析する**



入力画像

領域に分割



物体の形を領域として抽出

位置が分かる

物体が画像のどこにあるかを正確に把握できる

形・大きさを数値化

物体の形や大きさを数値として測る基礎になる

セグメンテーションの例



元画像



セグメンテーションの結果

セグメンテーションの種類

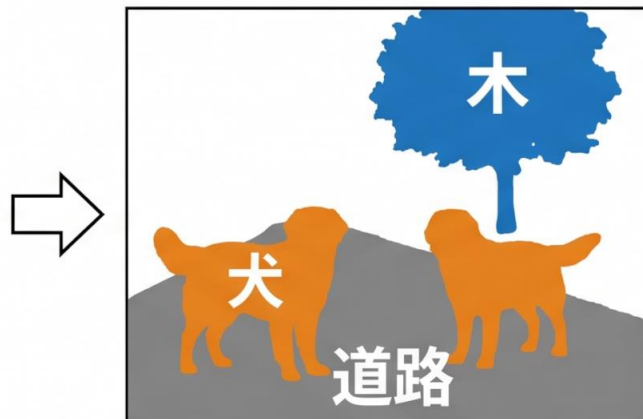


画素ごとにラベルを付け、物体の正確な形を抽出する技術

入力画像



セマンティック・セグメンテーション



同じクラスの物体を区別しない

インスタンス・セグメンテーション



同じクラスでも個々の物体を区別する

どちらも画素ごとにクラスを割り当て、物体の形状を画素単位で正確に抽出する

セマンティックセグメンテーションの例



