

# 1. 画像認識の概要

---

## 1.1 画像認識のさまざまなタスク

画像認識とは、画像の内容を理解することです。コンピュータに画像認識をさせる技術をコンピュータビジョンといいます。

画像認識と一口に言っても、さまざまなタスクがあり

- 目的変数がインスタンスの固有名詞かクラス名か
- 目的変数の予測のほかに位置を特定するタスクなどがあるか？

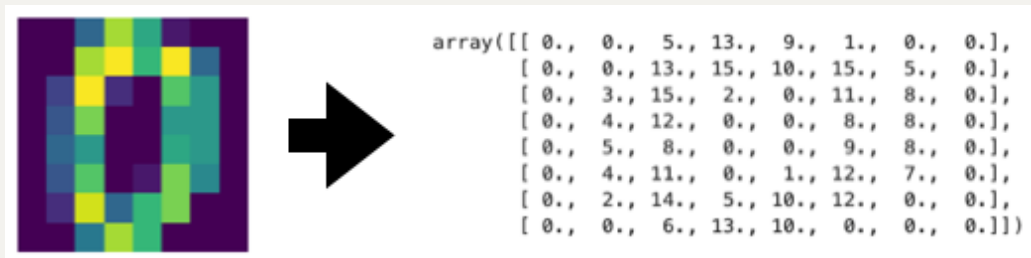
という2つの視点で分類されています。例えば、インスタンスのクラス名を予測するタスクを**クラス分類**、インスタンス名を予測するタスクを**インスタンス分類**といい、これらをまとめて**画像分類**といいます。また、画像分類に加えて対象物体が写っている部分を枠で囲むようなタスクを**物体検出**、物体と背景の境界を切り分けるようなタスクを**セマンティックセグメンテーション**といいます。



他にも、AさんとBさんが走っている姿が写っている画像から「AさんがBさんを追いかけている」という文章を作る画像キャプション生成や、実際には存在しない人の顔画像を作るような画像生成があります。

## 1.2 画像データの形式

画像データは小さな長方形に細かく分けられています。各長方形を **ピクセル** といい、ピクセルにはその画像の部分の情報を表す数値 (**標本値**) が記録されています。



色の表現方法には、**グレースケール** と言ってピクセルの標本値として明るさのみを準備する形式と、**カラー** と言って赤・青・緑という3つの原色をどれくらい混ぜ合わせるかをピクセルの標本値として準備する形式があります。また、標本値のとりうる値の範囲を **色深度** といい、例えば  $0$  以上  $2^8 - 1 = 255$  以下の整数で標本値をとる場合、**8bit** といいます。