



奥村 義和

物体検出モデルの大進化

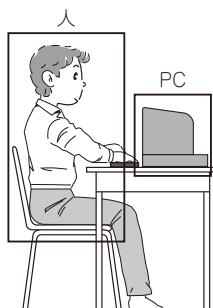


図1 物体検出の例

物体検出は深層学習登場以前からあるタスクです。深層学習の登場以来、非常にホットな分野です。2019年ころからは、物体検出のうちでも1ステージ系と呼ばれる手法が大きく進化しました。本記事では深層学習による物体検出について概要を説明した後、この大進化について説明します。

物体検出の基礎知識

● どんなタスクか…囲み枠に対してクラスとスコアを出力する

物体検出は画像の「どこに何がある」を求めるタスクです。図1のように矩形で物体を囲み、各矩形に対してクラスとスコアを出力するのが一般的です。

● 精度を競うコンペティションがある

物体検出の精度を競うコンペティションが開催されており、PASCAL VOC^{注1}やCOCO^{注2}などがあります。データが公開されているため、論文でもよく利用されています。

● 性能指標mAPとは

物体検出器の性能指標としてmAP (Mean Average Precision) があります。これは次の指標です。

注1: <http://host.robots.ox.ac.uk/pascal/voc/>

注2: <https://cocodataset.org/#home>

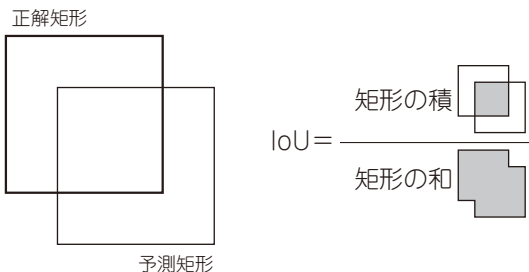


図2 IoUの定義

- 正解矩形と予測矩形の重なり具合をIoUと呼びます。定義より、0.0～1.0の値を取り、値が高いほどよく合致しています(図2)
- IoUのしきい値を決めると、予測矩形の適合率(precision, 正解率)や、正確矩形の再現率(recall, 検出率)を求めることができます
- IoUのしきい値を0.5～0.95まで0.5刻みに変えて適合率の平均を取ります。これがmAPです

物体検出と深層学習

● 先用語…1ステージ系と2ステージ系

深層学習による物体検出モデルは大きく1ステージ系と2ステージ系に分類できます。1ステージ系は物体の場所と種類を一度に計算するのにに対し2ステージ系は、

- 物体がありそうな場所の抽出
- その場所について物体認識

という2段階の処理をします。

一般に1ステージ系の方が速度は速いが精度は低い、2ステージ系は速度は遅いが精度は高いというトレードオフがあります(図3)。

● 1ステージ系は2015年から2019年までさまざまなモデルが登場した

物体検出への深層学習の適用は2013年11月にarXiv^{注3}に投稿されたR-CNN (Regions with CNN features)⁽¹⁾が最初とされています。物体検出のコンテストである