第2章 特定人物を追従しながら学習させる re-identification

学習済みモデルを使った半自動アノテーション

横山 裕樹

第3次AIブームが起こってから久しいですが，近年は個人でもさまざまな計算モデルを利用したり，自分で学習したりできるようになってきました．特に，YOLO（You Only Look Once）[7]のような物体検出モデルが公開されてから，実問題に適用できるケースが増えてきているように感じます．物体検出モデルを学習させてみた方も，かなりいらっしゃるのではないでしょうか．

トライすること

● トライ1…物体検出のための学習済みモデルを作る

深層学習は，問題を解くためのアルゴリズムを設計する必要がない反面，大量の訓練データを必要とします．典型的な物体検出モデルは教師あり学習であるため，画像に加えて，物体が写っている位置や物体の種類などのアノテーションが必要となります．このアノテーションが特に手間のかかる作業です．

ウェブ上で簡単に入手できる学習済みモデルを使って，この手間を少しでも軽減できないだろうか，というのが本稿のテーマです．もちろん，検出したい物体を検出できるモデルが手に入るのであれば，それを使えば良いです．ここでは，他の用途のモデル，または，他の物体の検出モデルを転用する形で，アノテーションの補助に使おうというアイデアです．

● トライ2…検出した物体をトラッキングするついでにアノテーションしてしまう

▶課題…追加学習するとて学習データは要る，アノテーションが面倒

物体検出モデルを学習するためには，対象となる物体をさまざまな背景で，さまざまな角度から撮影する必要があります．これは物体やカメラを動かしながら動画を撮って，フレームを切り出すことである程度簡単に実現できます．最も手間がかかるのは，こうして得られた画像に対象物体のバウンディング・ボックスを与える作業ではないでしょうか．物体をターンテーブルに乗せ，カメラを固定して撮影したり，物体にはっきりとした色がついていたりする場合は，簡単な画像処理で解決するかもしれません．そうではない場合，動画の各フレームから目視で対象物体を見つけて，それを囲むように矩形を描かなければなりません．

▶解決法…隣接するフレーム間で似ている領域を探す

この手間を少しでも軽減できないでしょうか．動画の隣接するフレームは，ほとんど似たような画像です．そのため，最初のフレームに手動でアノテーションをすれば，その情報をそれ以降のフレームのアノテーションに役立てることができそうです．つまり，あるフレームで指定した領域と似ている領域を他のフレームから探し出し，そのフレームのアノテーションとすることが考えられます．

ここで「似ている」というのをどのように定義するのかが問題となります．ピクセル値をそのまま比較するのは得策ではありません．物体が少しカメラから離れれば，ピクセルの配置は大きく変わってしまいます．また，物体の位置や向きによって光の当たり具合が違うかもしれません．このような変化にロバスト（堅牢）な特徴量を定義する必要があります．

▶特徴量の抽出には既存の学習済みモデルを利用する

そこで，この特徴量として，画像処理のタスクで学習済みのディープ・ニューラル・ネットワークを使うことを考えます．ここで画像処理のタスクとは，クラス分類など，物体検出以外を含む何らかのタスクですが，写真などの自然画像を使ったタスクでは，上述したような問題が常に付随してきます．これを乗り越えて一定の精度を達成したモデルの能力を少し使わせてもらおう，ということです．

モデルの出力形式は，タスクに依存します．例えば，10クラスの画像の分類問題を解くモデルの場合，10個の小数値を出力することが多いです．そしてこれらの値は当然，学習済みのタスクの解であるため，汎用性はありません．しかし，その中間層の値は，外乱にロバストな出力をするために集められた，画像のさまざまな特徴を表しているはずです．