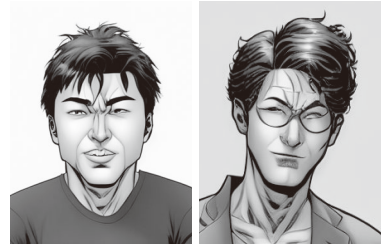


環境構築から始める! 最新物体検出アルゴ DETR



河崎 太郎, 服部 篤樹

● トライすること

自動運転や無人店舗, 医療画像診断や顔認証システムなどの画像認識技術は, 従来, 人が目で判断していたタスクの自動化を行い, われわれの生活を大きく変革しつつあります. 本章では, 画像認識の中でも, 物体検出を実践的に活用できるように,

- 物体検出モデル選定
- 学習環境構築
- 学習データ作成
- 学習方法

について解説を行い, 明日から実際に皆さんが物体検出の学習と推論を行えるようになることを目標としています.

物体検出モデル選定

● YOLOがこれまでは注目を集めていた

物体検出は昔から研究されてきた有名なタスクの1つです. 特にYOLO (You Only Look Once) と呼ばれるCNNベースのネットワークは, 従来のアルゴリズムと比較して大幅な高速化を行うことに成功し, 世間でも大きな注目を集めました. このYOLOの登場によって物体検出タスクは, より研究が盛んになり, 今に至るまでさまざまなアルゴリズムが提案されてきました.

● CNNやRNNに代わるTransformerがベースのDETR

その中でも, 今回はDETR (DEtection TRansformer) というアルゴリズムを試してみます. DETRは, 2020年5月にFacebook AI (現Meta AI) が発表したアルゴリズムで, その名の通りTransformerを物体検出に応用したアルゴリズムです.

Transformerは, CNNやRNNに代わる新しいニューラル・ネットワークであり, 2017年に自然言語処理分野で生まれました. そしてこのTransformerを画像処理タスクにも応用することが, 最新の画像処理研究ではトレンドとなっています.

● DETRの派生形が複数上位にランクイン

最新論文の比較を行うサイトのPapers With Code上で, COCOデータセットの物体検出タスクに対するSOTA (State Of The Art ≡ 最高精度) を見てみると, DETRの派生形が複数, 上位にランクインしています. このようにDETRは現在でもホットな研究対象の1つと言えるでしょう.

DETRの特徴

● CNNベースの物体検出で必要だったもの

DETRの特徴について説明します. DETR登場以前のCNNベースの物体検出アルゴリズム (YOLOやFaster R-CNN) では, 以下の処理が必要でした.

▶ Anchorボックスの指定

Anchorボックスとはバウンディング・ボックスの初期値に相当する矩形です. 例えば, Faster R-CNNでは, 9つのサイズが用意され, 各Anchorボックスからの位置とサイズの修正量を推定します.

▶ NMS (Non-maximal Suppression) 処理

同じ物体に対して重複したバウンディング・ボックスを検出してしまおうので, それらを除去する後処理です.

これらの処理にはAnchorボックスのサイズなどのハイパ・パラメータを手動で設定する必要があり, 精度に大きな影響を及ぼします. そのため, 可能な限りハイパ・パラメータは減らし, 手動で設定する項目が少ないほうが望ましいです.

● Anchorボックスを不要としたDETRが登場した

上記処理を不要とした下記特徴を持つDETRが提案されました.

▶ End to Endに物体検出を行うシンプルさ

NMSなどの前処理, 後処理が不要です.

▶ Transformerの利用

Transformerによって, CNNよりも大局的な特徴を捉えることが可能です. DETRはEnd to Endに物体検出を行うシンプルさを持ちながら, 従来の