

データサイエンス教育支援のための リアルタイムインスタンスセグメンテーションモデルの設計ツール

永田研究室 F122010 大津 海渡

1. 目的

近年、ディープラーニングを基本とした AI の研究が世界的に行われており、ChatGPT 等の所謂、生成 AI が話題となっている。政府の掲げた「AI 戦略 2019」においては数理、データサイエンス、AI についての素養を身に付け、社会で活躍できる人材を育成する必要があるとして「数理・データサイエンス・AI 教育プログラム認定制度」が創設された。本学の数理・データサイエンス・AI 教育プログラムは「薬工連携による教育、地方自治体との連携による課題解決型文理融合教育」といった特色のある取組みを実施しているとして「応用基礎レベル」に認定されている。本研究では、演習を含む実験系講義の時間の中でインスタンスセグメンテーション (IS) モデル SOLO の設計体験を可能とするために必要となる、汎化性が発揮される「最小限のラベリング画像から構成された G-Truth データ」について考察する。

2. 研究内容

当初、代表的な IS モデルである Mask R-CNN と SOLOv2 を併用しながら実験を進めていた。Mask R-CNN が物体検出→セグメンテーションという 2 ステージの過程を要していたが、SOLOv2 ではこれを 1 ステージで行っていることもあり、訓練時間だけでなく訓練後の汎化性能の点でも SOLOv2 の方が優れていることが確認できたため、その後の実験は SOLOv2 を基本モデルとしている。SOLOv2 の訓練時の損失関数はマスク損失の Dice Loss と分類損失の Focal Loss の和である。実験では、person, desk, chair, door, alcohol, wheel_chair, wheel_person, vending_machine の計 8 クラスを認識するための SOLOv2 モデルをそれぞれ構築することとした。ラベリング画像の枚数を 8 クラスそれぞれ、10, 15, 20, 25 枚と増やしながら訓練を行い、それぞれの SOLOv2 モデルの汎化性能の向上を観察し、評価した。評価にはラベリングに使用していないテスト画像をクラスごとに 10 枚ずつ用意し、SOLOv2 の出力である確信度スコアをもとに平均値を比較した。なお、訓練画像とテスト画像の解像度は 608×800 に調整している。これは SOLOv2 の訓練に使用できる画像解像度は縦横ともに 32 の倍数であることが指定されていること、画像はスマートフォンのカメラで撮影するため縦横比を変更すると画像が歪んでしまうこと、さらに将来本学のデータサイエンス関連の演習講義に取り入れる場合 GPU 搭載の高速 PC でも高解像度の画像では収束までに多くの訓練時間を要することなどを考慮した結果である。なお、今回の SOLOv2 は大規模な画像認識データセット COCO (Common Objects in Context) で事前学習された resnet50_coco をベースにしている。

3. 結果

図 1 には今回実施した多くの実験結果から得られた「訓練に用いたラベリング画像数と訓練後の SOLOv2 がオブジェクト検出したときのスコア(自信度, 確信度)」の関係を示している。この図から読み取れるように、訓練画像枚数が増えるにつれて、SOLOv2 のセグメント領域に対するカテゴリ予測の確信度が徐々に向上できていることがわかる。今回取り扱った全てのカテゴリにおいてこの傾向が観察されたことで、この程度のラベリング画像数でも汎化性能を発揮できることが確認されたため、将来的には実験的講義への導入が可能であるものと思われる。

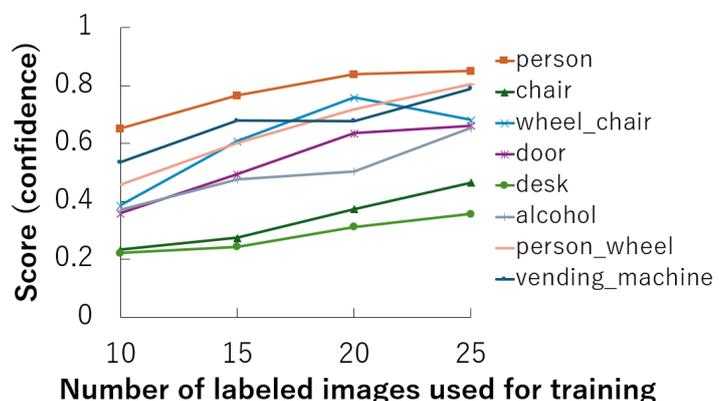


Fig. 1 訓練に用いたラベリング画像数と訓練後の SOLOv2 がオブジェクト検出したときのスコアの関係