

果実位置検出における半教師あり深層学習の適用

1210378 三本 拓未

【 知能情報学研究室 】

1 はじめに

Convolutional Neural Network (CNN) を用いた位置検出とクラス分類を目的とした物体検出が提案されている。検出モデルを構築する際に、画像に位置情報とクラスラベルを付与した学習データが必要である。人手での位置情報の入力には時間がかかるため、学習データを増やすことは困難である。

本研究の目的は、位置情報を付与していない画像を含めて学習を行うことである。位置情報を付与していない画像を利用する方法として、CNN を用いた半教師あり学習の一手法である Noisy Student[1] を基に改善手法を提案する。

2 Noisy Student(NS)

画像 $\mathcal{I}$ と対応する位置ラベル $\mathcal{L}$ で構成されるラベルありデータを $\mathcal{X} = \{\mathcal{I}, \mathcal{L}\}$ とし、学習した検出モデルを教師とする。ラベルなしデータ $\tilde{\mathcal{I}}$ に対する教師モデルの予測結果を疑似ラベル $\tilde{\mathcal{L}}$ とし、構成される $\tilde{\mathcal{X}} = \{\tilde{\mathcal{I}}, \tilde{\mathcal{L}}\}$ と $\mathcal{X}$ にデータ拡張を用いて生徒モデルを学習する。上記の学習手順を図 1 に示す。

3 改善手法

3.1 データのスケール拡張

検出対象のサイズ変化が大きい場合、小規模データによる学習ではサイズ変化への対応が困難である。画像の一部を切り取ることによって、データ数を増やし、スケール変化の改善を目的とする。画像一辺の分割数を s とすると、切り取り画像の一辺は $1/s$ となる。画像の左上を基準点として、垂直水平方向に $1/2s$ ずつ移動し、切り取る。各切り取り画像中に検出対象が含まれているものを学習データとする。 $s = 2$ の場合、図 2 に示すように最大 $3 \times 3 = 9$ 枚となる。

3.2 推論時データ拡張を用いた疑似ラベルの生成

教師モデルが与える疑似ラベルは生徒モデルの精度に関係しているため、疑似ラベルの質が向上すれば生徒モデルの精度が向上すると考える。疑似ラベルの質を向上させるために、推論時データ拡張 (Test Time Augmentation: TTA) を用いる。

3.3 教師モデルの重みを用いた生徒モデルの転移学習

転移学習 (Transfer Learning: TL) は、事前学習モデルの重みを初期値とし、新たなデータで学習することで学習時間の短縮と精度向上が可能である。

NS における生徒モデルの初期重みはランダムに生成されるため、教師モデルの学習重みから開始することで改善を試みる。

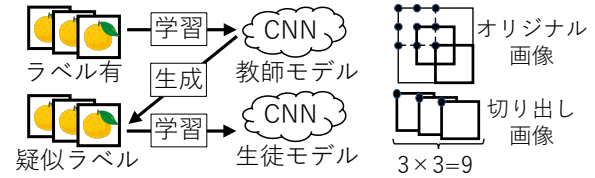


図 1 NS の学習手順

図 2 $s = 2$ の例

表 1 mAP による NS の改善手法の結果

実験条件	モデル S	モデル M	S to M
Yolov5	38.1	57.7	—
+Scale	66.1	67.4	—
+NS	70.6	69.0	70.1
+NS, Scale	68.0	68.6	67.8
+NS, TTA	68.8	68.9	69.2
+NS, TL	70.0	70.7	—

4 実験

4.1 実験設定

学習用 8 枚、検証用 5 枚のラベルありデータ合わせて 13 枚とラベルなしデータ 168 枚を用いる。検出モデルとして Yolo v4[2] の PyTorch 実装 (Yolo v5) のモデル S と M、評価指標として mAP を用いる。また、事前実験として教師モデルにデータ拡張 ($s = \{2, 3\}$) を用いた場合、mean Average Precision (mAP) が 3.2% から 3.8% に改善し、事前学習モデルを用いることで、3.8% から 38.1% に改善したため、データ拡張と事前学習を導入した教師モデルをベースラインとする。

4.2 結果・考察

教師モデルに対してスケール拡張を用いると、モデル S では 28.0%、モデル M では 9.7% 改善した。生徒モデルでは、疑似ラベルに用いると 70.6% から 68.0% となった。これは、疑似ラベル生成後にスケール拡張を用いると、検出ミスを含む画像が増え、学習時のノイズとなった。TTA, TL では過学習により精度が低下した。

5 おわりに

Yolo v5 を用いて NS を基に果実の位置検出を行った。教師モデルにおいてスケール拡張は効果があり、モデル S では 30%、モデル M では 10% 程度改善が見られた。生徒モデルでは、過学習しやすい傾向にあるため、今後の課題として正則化を用いて対応する。

参考文献

- [1] Q. Xie, *et al*, "Self-training with Noisy Student improves ImageNet classification," CVPR, 2020.
- [2] A. Bochkovskiy, *et al*, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv, 2020.