

## 少データ・少ラベル学習のためのデータ拡張法と農業画像認識への応用

1245134 野田 峻平 【 知能情報学研究室 】

## A Data Augmentation Method for Few-Shot and Few-Annotation Learning and Its Application to Agricultural Object Detection

1245134 NODA, Shumpei 【 Intelligent Informatics Laboratory 】

## 1 はじめに

農業において、AI・IoTの活用による省力化、生産性向上が重要である。AI・ディープラーニングを活用した画像認識において、近年、物体検出は畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) モデルをベースとした手法の精度が向上している。CNNでは、タスク (処理する問題) に対するモデルへの入力値と期待する出力値のペアを作成し、教師データのデータセットとしてモデルが学習することで、画像認識モデルを構築する。しかし、ディープラーニングモデルの学習には多くの質の良い教師データを必要とするため、データセットの作成に多くの労働力を要する。さらに、物体検出タスクにおいて出力値は物体の位置情報とクラスであるため、他の画像認識タスクと比較してアノテーション (教師情報付与) を行うコストが高い。本研究では、物体検出用の学習データセットの画像数を減らすことを少データ (Few-shot) 設定、アノテーション数を減らすことを少ラベル (Few-annotation) 設定として、両方の設定を導入した農業用の物体検出データセットによる学習から畑の農作物を検出するモデルの提案をする。

## 2 関連研究

## 2.1 Few-shot 物体検出

Few-shot 設定を導入したデータセットから物体検出モデルを構築する手法である。普通の物体検出手法と異なり、追加データを画像 [1] や自然言語で与えることで、少ないデータからの学習を可能にしている。

## 2.2 データ拡張

少ないデータでモデルを構築するために、既存のデータから人工的に拡張データを作成し、拡張データと既存のデータの両方でモデルを学習させる手法である。物体検出では、物体領域画像をランダムに画像に貼り付け、アノテーションが増加した様々なコンテキストのデータを人工的に作成する手法 [2] がある。

## 3 提案手法

## 3.1 パッチワーク拡張

Few-annotation 設定を導入したデータセットでモデルを学習させると、アノテーションのない物体領域が

ノイズとなり、過学習またはアンダーフィットする可能性が高いため、ノイズを取り除いた新しいデータセットを作成する方法を提案する。アノテーションのある物体矩形領域のパッチ画像を Positive パッチとし、画像内からランダムに抽出した矩形領域のパッチ画像を Negative パッチとする。Negative パッチにはアノテーションのない物体領域が含まれる可能性があるため、k-means によって Positive パッチと同じクラスに分類される Negative パッチを削除する。Positive パッチと残りの Negative パッチをランダムにグリッド状に並べることでパッチワーク画像を作成する。

## 3.2 Few-shot モデル

農作物は種類ごとに分けて育てるため、検出対象の農作物を1種類に限定できることから、1つのクラスのみを検出する Few-shot 設定用モデル [1] を参考にモデルを構築する。

物体を検出する場面を表す  $W_T \times H_T$  の画像をターゲット画像  $T \in \mathbb{R}^{3 \times W_T \times H_T}$  とし、検出対象物体を表す  $W_Q \times H_Q$  の例画像をクエリ画像  $Q \in \mathbb{R}^{3 \times W_Q \times H_Q}$  とする。まず、正方形かつ同じサイズ  $S_b$  にリサイズした  $T$  と  $Q$  から1つの特徴抽出器でチャンネル数  $C_b$  の特徴量  $\phi(T) \in \mathbb{R}^{C_b \times S_b \times S_b}$  と  $\phi(Q) \in \mathbb{R}^{C_b \times S_b \times S_b}$  を抽出する。そして、 $\phi(Q)$  から  $\phi(T)$  に、 $\phi(T)$  から  $\phi(Q)$  に注意 (図1 白矢印) をかけた特徴量  $F(T) \in \mathbb{R}^{C_b \times S_b \times S_b}$  と  $F(Q) \in \mathbb{R}^{C_b \times S_b \times S_b}$  を生成する。さらに、 $F(Q)$  から  $F(T)$  への注意と  $F(Q)$  に自己注意 (図1 黒矢印) をかけた特徴量  $\tilde{F}(T) \in \mathbb{R}^{C_b \times S_b \times S_b}$  と  $\tilde{F}(Q) \in \mathbb{R}^{C_b \times S_b \times S_b}$  を生成する。次に、 $F(T)$  から Region Proposal Network (RPN) で物体らしい  $N_r$  個の候補領域  $R \in \mathbb{R}^{N_r \times 4}$  を推定し、 $R_r$  に対応する  $\tilde{F}(T)$  の特徴量  $\tilde{R}_r(T) \in \mathbb{R}^{N_r \times C_1 \times S_r \times S_r}$  を固定サイズ  $S_r$  で抽出する。また、 $\tilde{F}(Q)$  を  $r \in \mathbb{R}^{C_1 \times S_r \times S_r}$ ,  $r \in \tilde{R}_r(T)$  と同様のサイズの特徴量  $\tilde{R}_r(Q) \in \mathbb{R}^{1 \times C_1 \times S_r \times S_r}$  に変換する。最後に最終予測器で  $\tilde{R}_r(T)$  と  $\tilde{R}_r(Q)$  から  $N_f$  個の物体の位置  $R_f \in \mathbb{R}^{N_f \times 4}$  とクラス  $L_f \in \mathbb{R}^{N_f \times 1}$  を予測する。

RPN と最終予測器は位置推定を回帰タスク、クラス推定を分類タスクとして学習する。また、 $\tilde{R}_r(T)$  と  $\tilde{R}_r(Q)$  の間の距離を近くするタスクを学習する。

表1 実験条件と結果一覧. ①は Few-shot 設定. ②は Few-annotation 設定. ③は SSD のデータ拡張. ④は バッチワーク拡張. 結果は 500 エポックでの指標 AP50.

モデル	ケース名	Easy(%)	Normal(%)	Hard(%)	①	②	③	④
Faster R-CNN	Case1	24.0	14.7	10.7	✓			
	Case2	45.8	22.1	11.1	✓		✓	
	Case3	3.1	1.5	1.0	✓	✓	✓	
Few-shot モデル	Case4	52.6	40.9	26.0	✓			
	Case5	3.1	1.5	1.0	✓	✓		
	提案手法	51.6	40.9	26.0	✓	✓		✓

表2 データセットの情報. 対象ラベルはデータセットが含む最小のラベルサイズを示す.

データセット	アノテーション数	対象ラベル
Easy	16	192 × 192 以上
Normal	41	96 × 96 以上
Hard	84	1 × 1 以上 (全て)

## 4 実験

### 4.1 農業用データセット

検出対象をナスとして, ナス畑の画像を 98 枚用意し, アノテーションを手作業によって作成する. Few-shot 設定を導入した場合, 学習用画像を 5 枚, テスト用画像を 93 枚にする. Few-annotation 設定を導入した場合, 学習用データに含まれるアノテーションの数を 5 つにする. また, 物体の大きさを基準に異なる難易度のデータセット Easy, Normal, Hard を作成する. 各難易度のデータセットの情報を表 2 に示す.

### 4.2 モデル学習

表 1 に示す条件で実験し, 比較する. Faster R-CNN と Few-shot モデルは事前にデータセット MS-COCO の train2017 によって学習を行う. それぞれの手法で学習回数 (エポック) 500 回で学習し, その予測結果による AP50 を評価指標とする. AP50 は間違った予測をせずに漏れなく予測を行うことができたかどうかを示す指標であり, 予測領域と正解領域の重なりが 50% 以上の時を正解と判定する. また, パッチワーク画像の生成やターゲット画像とクエリ画像のペアの生成で発生するランダム性を考慮して 10 試行の平均 AP50 を算出する.

## 5 結果と考察

それぞれのモデルの AP50 を表 1 に示す. また, Hard で 500 エポック学習した提案手法の予測結果を図 1 に示す. Easy の 500 エポックでは, 提案手法は Case5 と比較して AP50 は 0.4 ポイント低く, アノテーションの数は 69% 減. Hard の 500 エポックでは, 提案手法は

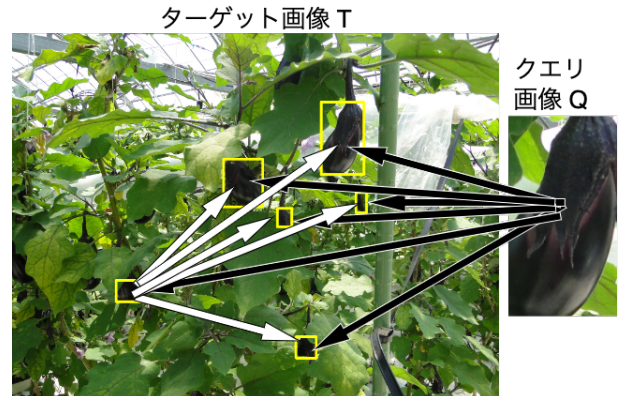


図 1 自己注意 (白矢印) と注意 (黒矢印) と Case5 をデータセット Hard で 500 エポック学習したときの予測結果 (矩形).

Case5 と比較して AP50 は 4.8 ポイント低く, アノテーションの数は 94% 減. Hard の 500 エポックにおいて, 1 データあたりの AP50 は提案手法では 5.2 ポイントであり, Case5 では 0.3 ポイントであるため, 提案手法の方がアノテーションの数に対する AP50 が高効率である.

## 6 まとめ

農作物の検出にはニューラルネットワークを用いた手法の使用が考えられるが, データセットを作成するためのコストが大きい. そこで, 学習に必要な画像とアノテーションを減らして学習させる方法を提案した. 提案手法を用いてデータセットを減らした場合と減らさない場合で比較したとき, AP50 は 4.8 ポイント下がったが, アノテーションの数は 95% 削減することができた.

## 参考文献

- [1] Hsieh Ting-I, et al., "One-Shot Object Detection with Co-Attention and Co-Excitation," NeurIPS, 2019.
- [2] Hsieh Ting-I, et al., "Patch-Level Augmentation for Object Detection in Aerial Images," ICCVW, 2019.