

少データ状況下における敵対的生成ネットワークを用いた 果実識別 AI の性能向上法

1255120 三本 拓未 【 知能情報学研究室 】

Improvement of Fruit Image Recognition AI Models Under Small Data Situations Using Generative Adversarial Networks

1255120 MIMOTO, Takumi 【 Intelligent Informatics Lab. 】

1 はじめに

農業分野における出荷予測・収量予測などの農作物に対する予測を行うことが重要になっている。経験に基づく予測は新規参入者には障壁が高く、AIによる予測が期待されている。予測の方法として、気温や湿度といった環境情報のみでの予測に対して、画像による生育状況の情報があればさらに良い予測ができると考える。生育状況把握の一つとして着果数把握があるが、画像内の果実を自動でカウンティングを行うことで、より詳しく予測することができる。深層学習を用いたカウンティングが有効と考えられるが、そのためには画像内にある対象物体の位置情報を付与した教師ありデータを作成する必要がある。農業用データセットを作成するには大量の作物に付与するため、コストが高いという問題がある。

そこで本研究では、敵対的生成ネットワーク (GAN) の一つである pix2pix[1] を用いて、人工的に学習データを生成することを提案する。

2 関連技術

2.1 pix2pix

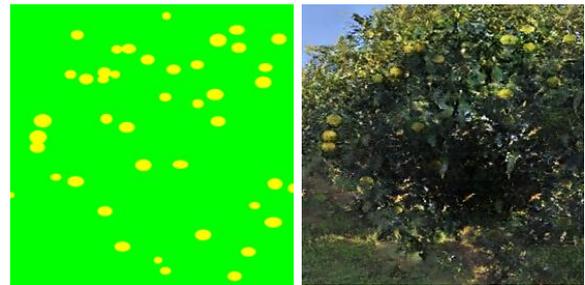
画像から画像への変換を行うことが可能である。変換を行うために、色のみが異なる画像や、ラベル画像等のペア画像を用いて学習を行う必要がある。

2.2 Adaptive Discriminator Augmentation

識別器に入力するデータをランダムにデータ拡張を行うことで、少量のデータセットでも過学習を起さずに学習することが可能となる [2]。

3 提案手法

本研究では、pix2pix を用いた物体検出用データ拡張を提案する。pix2pix の学習データとしてセグメンテーション画像を作成する必要があるため、物体検出用のラベルデータを用いて簡易的なセグメンテーション画像を作成し学習を行う。また、生成器に Attention 機構を追加した学習モデルと、識別器の入力にデータ拡張を行う Adaptive Discriminator Augmentation (ADA) を行ったものでそれぞれ学習を行う。物体検出用に生成するデータとして、pix2pix の学習で使用した学習用デー



(a) 条件画像 (果実の位置) (b) 生成画像

図 1: pix2pix による果実識別用疑似訓練画像の生成

タから、ランダムに位置情報を取り除いたものを用いる。図 1(a) の果実の位置を示す条件画像を入力し、図 1(b) の画像を生成し既存のデータに加えて学習を行う。

4 実験

4.1 実験設定

本研究では、高知県農業技術センター果樹試験場で撮影された画像 101 枚と高知県の農家で撮影された画像 13 枚を用いる。果樹試験場の画像を学習用データ 81 枚、検証用データ 21 枚、農家で撮影された画像 13 枚をテスト用データに分けて使用する。

pix2pix の構造として生成器に U-Net、識別器に Spectral Normalization を加えた 3 層からなる CNN を用いる。予測モデルとして、Yolov5 の COCO2017 データセットを用いた学習済みモデルのサイズ S、学習時に、拡大縮小、左右フリップ、画像移動、画像の組み合わせ、の 4 種類のデータ拡張を用いる。評価指標として予測領域と正解領域の重なりが 50%以上のものを正解とした、Average Precision (AP@.5) を用いる。

4.2 事前実験

ADA で用いるデータ拡張の種類と適用率を決定するために事前に実験を行った結果、色を変更するデータ拡張を用いた際に生成される画像が全体的に白くなった。そのため、拡大縮小、左右フリップ、回転、画像移動の形を変えるもののみを用いる。また、適応率を上げると元

表 1: 各条件での結果と生成データを増量した結果

	既存データ	pix2pix	Attention	ADA	Attention+ADA	P_1	P_{10}	P_{20}
AP@.5	0.702	0.709	0.698	0.708	0.711	0.711	0.72	0.721
Precision	0.872	0.883	0.869	0.882	0.891	0.891	0.838	0.9
Recall	0.506	0.51	0.498	0.508	0.503	0.503	0.52	0.516

表 2: Attention 機構 + ADA での学習データを減らした結果

	50%				25%				12.5%			
	BASE	P_1	P_{10}	P_{20}	BASE	P_1	P_{10}	P_{20}	BASE	P_1	P_{10}	P_{20}
AP@.5	0.577	0.67	0.678	0.694	0.494	0.606	0.647	0.65	0.153	0.28	0.589	0.585
Precision	0.736	0.851	0.882	0.887	0.662	0.837	0.86	0.854	0.261	0.431	0.827	0.786
Recall	0.494	0.458	0.457	0.478	0.429	0.366	0.417	0.421	0.215	0.183	0.331	0.369

データと大きく異なる画像が生成されるため、適応率は 0.1 を用いる。

4.3 実験手順

pix2pix の学習を行った後に、Attention 機構、ADA、両方を加えたもので再度学習を行い、学習データで果実数が平均以上のデータから 10%、20%、30% の果実領域を除いた 3 パターンで学習データの生成を行う。作成した生成画像セットを P_1 とし、さらに条件を変えて生成し 10 倍したものを P_{10} 、20 倍したものを P_{20} とする。生成した画像を既存の学習データに加えて予測モデルの学習を行い評価の比較を行う。最も評価が良かった条件で、 P_{10} 、 P_{20} を用いて学習を行う。また、評価が良かった条件で学習データの数を 50%、25%、12.5% に減らした状態で上記の手順で実験を行う。

5 結果・考察

10% の果実領域を取り除いたもので、pix2pix での画像生成の際に異なる条件での実験結果を表 1 に示す。Attention 機構と ADA を行った条件が他の条件よりも良い結果となった。 P_{10} 、 P_{20} と生成画像を増やして学習した結果、既存データのみでの学習結果から 1.9 ポイント向上した。また、ADA を加えた条件のみでも精度の向上が見られた。学習時に ADA を行うことで学習データが少ない場合でも一定の学習ができていたため、ADA を加えた条件の結果が良いと考えられる。また、Attention 機構と ADA の条件で用いる学習データを減らした結果を表 2 に示す。提案手法を使用しない通常の学習結果から、学習データを 50% に減らした場合で 0.8 ポイントの低下、学習データを 25% に減らした場合で 5.2 ポイントの低下となった。この結果から、精度を維持しつつ学習データを 25% にまで減らすことが可能となった。

全体の結果から、提案手法を用いて予測モデルの学習

を行った場合、Precision が向上しており、Recall はほとんど変わらない結果となった。提案手法では、果実部分を再現した画像を生成しているため、果実以外の検出を抑えることができていると考えられる。しかし、重なっているものや葉に隠れたものものを再現した生成データが含まれていないため、Recall が向上しなかったと考えられる。また、今回使用した予測モデルでは、学習データを 50% に減らした場合と減らさなかった場合で、Recall の値が殆ど変化がなかった。そのため、人手での学習データを作成し増加させてもこれ以上は向上しないと考えられる。そのため、Recall を上げるためには、学習データを増加させる以外の方法でアプローチをする必要がある。

6 おわりに

本研究では、pix2pix を用いて学習用データの作成を提案した。提案手法を用いてデータ拡張を適用した場合と、適用しなかった場合で比較した際に 1.9 ポイント向上することができた。また、既存データのみでの学習結果から 5.2 ポイントの低下したが、学習データを 25% 削減することができた。

参考文献

- [1] Isola, *et al*, "Image-to-Image Translation with Conditional Adversarial Networks," Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, 2017.
- [2] T.Karras, *et al*, "Training Generative Adversarial Networks with Limited Data," NeurIPS, 2020.