

令和 4 年度
修士学位論文

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

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

1255120 三本 拓未

指導教員 吉田 真一

2023 年 2 月 28 日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要 旨

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

三本 拓未

近年では、出荷予測・収量予測などの農作物に対する予測は省略化、高付加価値に重要であり、AI（人工知能）を用いた予測が期待されている。AI を用いた予測への応用の一つとして、深層学習を用いた画像からの着果数把握があるが、果実の位置検出を行うためには位置情報を付与した教師データが必要となる。しかし、教師データを作成するには時間と労力を要するためコストが高い問題がある。そこで本研究では、pix2pix を用いて既存の学習データから人工的に学習データを生成することで、精度向上の手法を提案する。pix2pix に用いる条件画像として物体検出用のデータセットから果実領域を表現した簡易的なセグメンテーション画像を用いる。pix2pix に対して Attention 機構を追加と ADA の追加、両方を追加した条件で学習を行い、学習データから果実領域を 10%, 20%, 30%, 取り除いた条件画像を用いて学習データの生成を行う。生成した画像と既存データを組み合わせて学習を行い提案手法の有効性の確認を行う。実験結果より、Attention 機構と ADA を追加した条件での生成画像を学習データとして用いた場合 AP が 1.9 ポイント向上したため提案手法の有効性が示された。また、意図的に学習データを削減したデータセットを作成することで少データに対しての提案手法の効果を確認する。Attention 機構と ADA を追加した生成モデルの結果、学習データ 25%で提案手法を用いない場合で 20.8 ポイント低下するのに対し、提案手法を用いると 4.5 ポイント低下するだけであった。提案手法を用いることで精度を維持したまま学習データを 25%にまで削減することが可能となった。

キーワード 物体検出, 少データ, pix2pix, データ拡張

Abstract

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

MIMOTO Takumi

In recent years, prediction for agricultural production is important for work-saving and high added value, and predictions using artificial intelligence (AI) are expected to be used. One of the applications of AI for agriculture is to counting the number of fruit from images using deep learning, which requires training data with location information to detect the location of fruits. However, there is a problem of high cost due to the time and effort required to create the training data. In this study, we propose a method to improve the accuracy by artificially generating training data from existing training data using pix2pix under low number of labels condition. As a conditional image for pix2pix, we use a simple segmentation image that represents a fruit region from a dataset for object detection. For pix2pix, training is performed with the addition of the attention mechanism, the addition of ADA, and the addition of both, and training data is generated using conditioned images in which fruit regions are removed 10%, 20%, 30%, and from the training data. Combining the generated images with existing data, the proposed method is validated by training. The experimental results show that the proposed method is effective because the AP improved by 1.9 points when the training data are images generated under the condition that the Attention mechanism and ADA are added. In addition, we will also create a dataset with intentionally reduced training

data to confirm the effectiveness of the proposed method for small data sets. The results of the generative model with the addition of the Attention mechanism and ADA, show that the proposed method reduces the average precision by only 4.7 points, while it reduces the average precision by 20.8 points without the proposed method at 25% of the training data. Using the proposed method, it was possible to reduce the training data to 25% while maintaining accuracy.

key words object detection, small data, pix2pix, data augmentation

目次

第 1 章	序論	1
第 2 章	関連技術	3
2.1	畳み込みニューラルネットワーク (Convolutional Neural Network)	3
2.2	U-Net	3
2.3	敵対的生成ネットワーク (Generative Adversarial Network)	4
2.4	pix2pix	5
2.5	データ拡張 (Data Augmentation)	7
第 3 章	提案手法	8
3.1	pix2pix を用いたデータ拡張	8
3.2	Attention 機構を追加	9
3.3	ADA を用いた学習	9
3.3.1	生成画像の多様性維持のためのデータ拡張用損失関数の提案	10
第 4 章	実験設定	11
4.1	物体検出用のデータセット	11
4.2	pix2pix 用のデータセット	12
4.3	学習データ生成に使用する条件画像	12
4.4	pix2pix モデル構造	13
4.4.1	生成器	13
4.4.2	識別器	14
4.5	予測モデル	14
4.6	評価指標	15
4.7	実験手順	15

目次

第 5 章	事前実験	18
5.1	予測モデルに対するデータ拡張	18
5.2	ADA に用いるデータ拡張と適応率	19
第 6 章	結果と考察	20
6.1	画像生成を行うための学習	20
6.1.1	pix2pix を用いた画像生成	20
6.1.2	Attention 機構を追加したモデルで生成された画像	21
6.1.3	ADA を追加したモデルで生成された画像	21
6.1.4	Attention 機構, ADA を追加したモデルで生成された画像	22
6.2	生成画像を加えた学習結果	22
6.3	Attention 機構と ADA の生成画像を増加	23
6.4	学習データを減らした場合	24
第 7 章	結論	27
	謝辞	28
	参考文献	29

目次

2.1	U-Net の構造	4
2.2	GAN の構造	5
2.3	学習用データ生成までの手順	6
3.1	学習用データ生成までの手順	9
3.2	Attention 機構を追加したモデル	10
4.1	pix2pix に用いるペア画像	13
4.2	生成器のモデル構造	14
4.3	識別器のモデル構造	15
4.4	予測結果の例の PR 曲線	17
5.1	画像変形と色変換	19
5.2	適応率 0.2 の画像変形	19
5.3	適応率 0.1 の画像変形	19
6.1	pix2pix で生成された画像	20
6.2	Attention 機構を追加したもので生成された画像	21
6.3	ADA を追加したもので生成された画像	21
6.4	Attention 機構と ADA を追加したもので生成された画像	22
6.5	Attention 機構と ADA を追加したもので生成された画像	25

表目次

4.1	使用するデータセット	12
4.2	Train 用を減らした場合のデータ数	12
4.3	学習データ生成に使用する条件画像の数	13
4.4	Yolov5 のハイパーパラメータ	16
4.5	予測結果の例	16
4.6	pix2pix に用いるハイパーパラメータ	17
5.1	予測モデルに対するデータ拡張の条件	18
5.2	予測モデルに対するデータ拡張の条件ごとの結果	19
6.1	pix2pix と Attention 機構での生成画像を加えた結果	23
6.2	ADA と Attention 機構 + ADA での生成画像を加えた結果	23
6.3	Attention 機構と ADA の生成画像を増加させた学習結果	24
6.4	提案手法なしでの実験結果	25
6.5	学習データ 50%	25
6.6	学習データ 25%	25
6.7	学習データ 12.5%	26

第 1 章

序論

日本の少子高齢化により農業従事者の数が減少しており農業現場において人手不足が問題であり、平成 27 年に 175.7 万人であった農業従事者の数も令和 4 年には 122.6 万人にまで減少している [1]。一方、農業従事者の数が減少傾向である中で、データを活用した農業を行っている農業経営者の数が増加しており、IT やデータ活用を用いて省力化を進めていくことで、人手不足への問題解決が期待されている。

使用されている IT 技術として、IoT を用いて温度や湿度を計測しハウス内の環境を調整するものや、ドローンによる広範囲に農薬散布を行うなどの農作業を補助するものがある。しかし、作業を補助する技術では既存の農業従事者にとっては有用であるが、新規参入者にはそれだけでは不十分である。また、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 [2]。経験や勘に基づく予測は新規参入者には困難であり、AI を用いた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センサー類から取得される環境情報を用いた予測方法があるが、画像による生育状況の情報があればさらに良い予測が可能となる。画像による生育状況把握の方法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。自動でのカウンティングとして深層学習を用いて対象となる果実の位置検出行う方法がある。しかし、果実の位置検出を行うためには、画像内にある対象物体の位置情報を付与した教師データを作成する必要がある。教師データは人手で位置情報を付与する必要があり、時間と労力を要するためコストが高い問題がある。そのため、少ない教師データを用いて精度を上げる方法が必要となる。

本研究では、ユズを位置検出するための農業用データセットを作成し、このデータセット

用いて画像生成モデル学習させ学習用データを生成する．生成モデルとして一般的に敵対的生成ネットワーク (GAN)[3] があるが，位置情報を付与していない画像を使用するため画像の生成後に位置情報の付与をする必要がある．位置情報を付与したデータから画像生成が可能となる GAN の一つである pix2pix[4] を用いて学習データの生成を行う．新たに生成画像と既存データを組み合わせてデータセットを作成し，精度比較の検証を行った結果 1.9 ポイント増加したため，提案手法によるデータ拡張が有効であることを示す．また，学習データ数を 50%，25%まで減らした場合で提案手法を用いて検証を行った結果，提案手法を用いないとそれぞれ精度が 12.5 ポイント，20.8 ポイント低下するのに対して，提案手法を用いると 0.8 ポイント，4.5 ポイント低下となり，学習データを 1/2, 1/4 程度まで減らしても精度が維持できることを示す．

第 2 章

関連技術

2.1 畳み込みニューラルネットワーク (Convolutional Neural Network)

畳み込みニューラルネットワーク (Convolutisotonal Neural Network, CNN) は, 主に画像認識の分野において優れた効果を発揮することができることで知られている. CNN モデルは畳み込み層とプーリング層を組み合わせて構築され, 畳み込み層では入力された画像から特徴量を計算し特徴マップを作成する. プーリング層では, 得られた特徴マップから最大値や平均値を取ることでより重要な値を抽出する. 本研究では, 画像生成モデルを学習させるためにセグメンテーション画像を用いるため, セマンティックセグメンテーション向けに提案された U-Net を用いて画像の生成を行う.

2.2 U-Net

Olaf Ronneberge らによって医学分野に用いるために 2015 年に提案された, セマンティックセグメンテーション用のモデルの一つでありモデル構造が図 2.1 で示すように, U の字をしているため U-Net と呼ばれている [5]. モデルの構造としてエンコーダとデコーダで構成されており, エンコーダでは通常の CNN と同様に複数回畳み込みを行い特徴を抽出し, デコーダでは逆畳み込みを行い入力画像と同じサイズの出力を行う. 通常の逆畳み込みで特徴マップを拡大させると物体の位置情報を正しく得ることができない. そのため, Skip Conection と呼ばれる, エンコーダの特徴マップを逆畳み込みをする際にを連結させる処理

2.3 敵対的生成ネットワーク (Generative Adversarial Network)

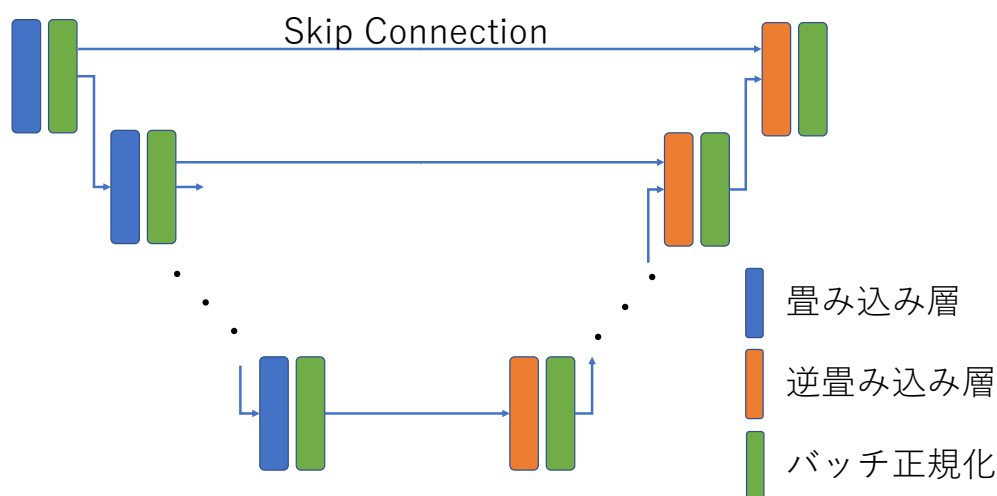


図 2.1 U-Net の構造

を用いることで、位置情報を得ることが可能になっている。また、3次元画像を入力とした 3D U-Net[6] などの様々な種類のモデルが提案されている。

2.3 敵対的生成ネットワーク (Generative Adversarial Network)

Generative Adversarial Network(GAN) とは、データから特徴を学習し、学習した特徴を持った画像を生成するモデルである。通常の学習では1つのモデルを用いて学習を行うが、GAN では生成器 (Generator) と識別器 (Discriminator) の2つで構成され、基本的な GAN の構造を図 2.2 に示す。Generator ではランダムな値が入力されたものから新たなデータを出力し、Discriminator では偽物のデータと本物のデータを判別する。

GAN の学習では、Generator は Discriminator を騙せるように、Discriminator は本物か偽物のデータなのか判断できるように学習を行い、式 (2.1) の目的関数を用いる。

2.4 pix2pix

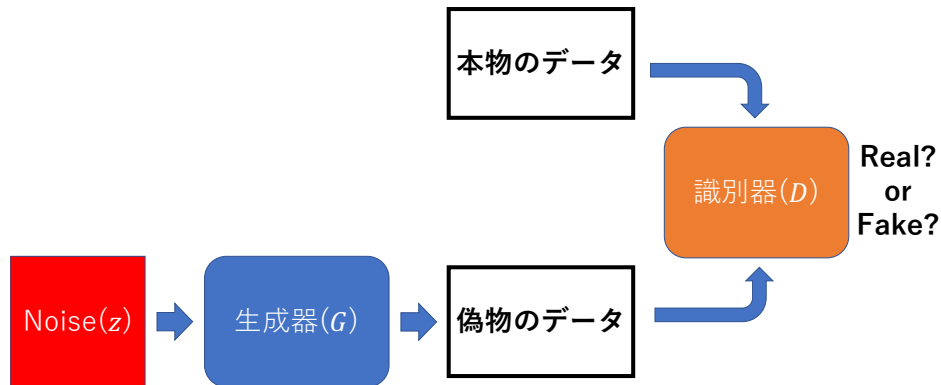


図 2.2 GAN の構造

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (2.1)$$

Generator を学習する際には最小化, Discriminator を学習する際には最大化するように学習を行う。2つのモデルの関係が敵対的であるため、敵対的生成ネットワークと呼ばれる。また、CNN を用いて生成画像の品質を上げる Deep Convolutional GAN[7] や、主に教師なしデータを用いて GAN は学習を行うが、教師ありデータを用いた Conditional GAN(CGAN)[8] がある。

2.4 pix2pix

pix2pix とは、ある画像を別の画像に変換する GAN の一つである。学習に用いるデータとして、変換したい画像と変換後の画像のペアを用いる。画像変換の例として図 2.3 を用いて説明する。図 2.3 の条件画像は黄緑色の背景に、黄色の果実領域を描画したものであり、黄色の部分に果実を自然に描くように生成器は画像を生成する。識別器は生成した画像と本物の画像の真偽を判定し、生成器は識別器を騙せるように画像を生成するように学習していく。pix2pix では、式 (2.2) に示すように、通常の GAN に対して条件画像を加えた目的関数

2.4 pix2pix

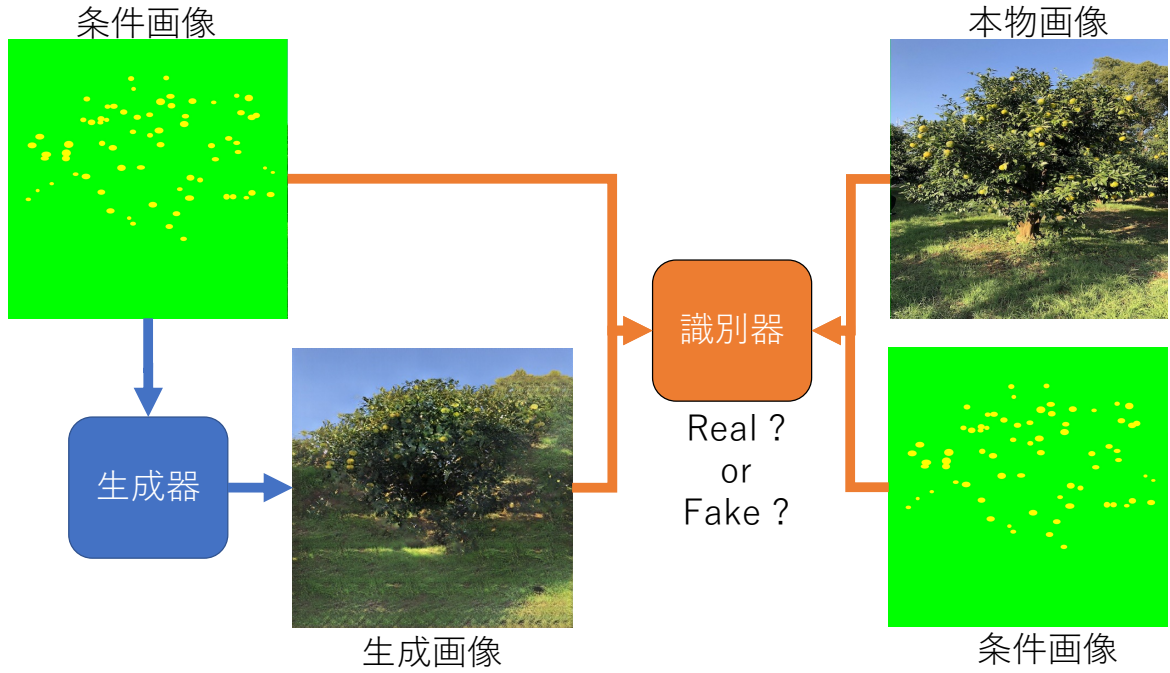


図 2.3 学習用データ生成までの手順

を用いる。しかし、通常の GAN では生成画像がぼやけた画像を生成する傾向があるため、式 (2.3) に示す L1 ノルム損失関数を用いて対処している。2 つの関数を組み合わせて式 (2.4) に示す目的関数を用いて学習を行い、重み変数 λ を用いて L1 ノルム損失関数の値を制御している。

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log (1 - D(x, G(x, z)))] \quad (2.2)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1] \quad (2.3)$$

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (2.4)$$

また、ペア画像を用いない変換方法として、2 つの大量のデータセットを用いてデータセット間の関係を学習させる CycleGAN[9] がある。本研究では、条件画像と本物画像のペア画像を用いて学習を行う pix2pix を用いる。

2.5 データ拡張 (Data Augmentation)

一般的に精度高めるためには、大量のデータセットを用いて学習を行う必要がある。しかし、教師ありデータといった人手で学習データを作成するデータセットが必要となる場合、大量に作成するには非常にコストが高いといった問題がある。データ拡張では、既存の学習データを用いて人工的に学習データを生成する手法である。主に使用されるものとして、Geometric Transformation の回転やフリップ、Photometric Transformation の輝度やコントラストなどの画像全体を変換するものがある。また、画像にマスクキングを行い隠れた学習対象が隠れた状態でも判別できるように CutOut[10] や Random erasing[11] など様々な拡張方法が提案されている。また、データセットには含まれていないものを生成することができるため、ロバスト性を高める効果がある一方、既存の学習データと似た画像や大きく異なる画像を加えて学習することで、過学習や精度の低下を引き起こす可能性がある。

第 3 章

提案手法

少データでは学習が困難であることから, pix2pix を用いた学習データの生成を行うことで精度を向上させるためのデータ拡張を提案する. また, Generator に対して, Self-attention GAN (SAGAN) [12] に用いられる Attention 機構を追加したものと, 少量のデータセットでも学習を可能とさせる StyleGAN2-ADA(Adaptive Discriminator Augmentation)[13] に用いられる ADA を導入したもので学習を行う. 新たな学習データを生成するために用いる入力データとして, 学習用データのラベルの一部を取り除いたものをも用いる. 学習済みの Generator を用いて生成した画像を既存のデータセットと組み合わせて使用することで精度の向上が可能であるかの検討を行う.

3.1 pix2pix を用いたデータ拡張

画像データに位置情報の付与が必要ないデータ画像分類のタスクにおいて, データ拡張の方法として GAN を用いたものがある.[14] しかし, 物体検出において画像データに位置情報の付与が必要であるため, 画像生成を行うだけであれば学習データとして使用が難しい問題がある. そのため, クラスラベルに位置情報を付与して入力し, 画像を生成することが可能である pix2pix を用いて新たに学習データを生成する.

pix2pix に用いる学習データとしてセグメンテーション画像と本物画像のペアを用いて学習を行う. セグメンテーション画像を一から作成するには多くの時間を要するため, 簡易的に作成する. 作成の方法として物体検出用のラベルデータを用いて, 位置情報に対して楕円を描画し果実部分の領域表したものをセグメンテーション画像として用いる. 作成した学習

3.2 Attention 機構を追加

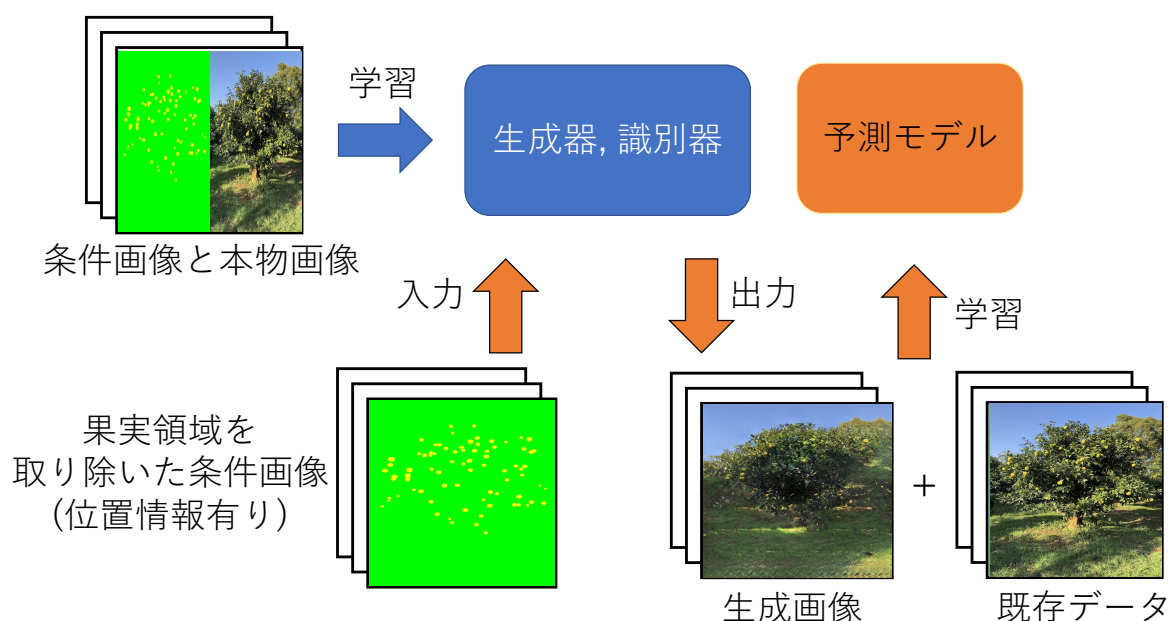


図 3.1 学習用データ生成までの手順

データを既存のデータと組み合わせて学習を行うことで精度の向上を検討する。学習データ作成までの流れを図 3.1 に示す。

3.2 Attention 機構を追加

pix2pix の Generator に使用している U-Net に対して Attention 機構を追加する。Attention 機構の構築として、Self-attention GAN (SAGAN) に用いられている構成を用いる。Attention U-Net[15] を参考に、使用している U-Net に Attention 機構を追加し Generator のモデルを作成する。図 3.2 に構築した Generator のモデルを示す。

3.3 ADA を用いた学習

一般的な GAN の学習では数千枚規模のデータセットを用いて学習を行う必要がある。少量のデータセットでの学習を可能とする手法として、学習時にデータ拡張が提案されている。StyleGAN2-ADA は Discriminator にデータを入力する際に、一定の確率でデータ拡張を行

3.3 ADA を用いた学習

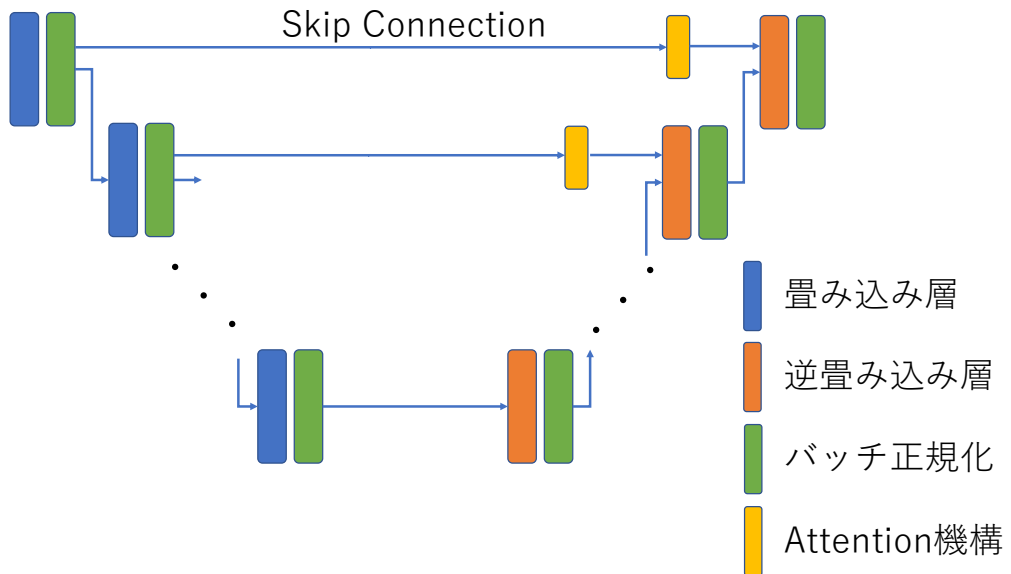


図 3.2 Attention 機構を追加したモデル

うことで過学習を防ぐことが可能となる。StyleGAN2-ADA では、セグメンテーション画像などのクラスラベルを必要としない教師なし学習である。本研究で使用する pix2pix は教師あり学習のため、クラスラベルであるセグメンテーション画像に対してもデータ拡張後に対応するように変更を行うことで、教師ありデータに対して ADA を適用させる。

3.3.1 生成画像の多様性維持のためのデータ拡張用損失関数の提案

学習データを用いた生成画像を物体検出用のデータセットに組み合わせるため、同じ画像が生成されるのを防ぐために Generator の損失関数に変更を加えた式 (3.1) に示す L_1 ノルム損失関数を用いる。

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G'(x, z)\|_1] \quad (3.1)$$

$G'(x, z)$ は生成画像に対してデータ拡張を適用した画像であり、 y は本物画像を表す。

第 4 章

実験設定

本章では研究で行った実験について示す.

4.1 物体検出用のデータセット

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. 画像内には必ず対象物体が含まれており, 撮影時間や日光、天候などは考慮しないとする. 画像サイズは果樹試験場が縦 4032 ピクセル, 横 3024 ピクセルで, 高知県の農家では画像サイズが一定ではないが学習時にリサイズを行うため問題はないとする. 撮影された画像に対して, 人手で対象物体に位置情報を扶養するアノテーション作業を行う. アノテーション作業は複数人で行うと個人差が生じる可能性があるため一人で行う. 位置情報を付与する対象物体の基準としては, 葉に隠れたものや重なったものなどを含めて付与を行う.

データセットの分け方として表 4.1 に示すように, 実環境に近い高知県の農家で撮影された画像を Test 用に使用し, 果樹試験場で撮影された画像を Train 用と Valid 用とする. また, 学習データを減らした条件での実験を行うために, Train 用データを 50%, 25%, 12.5% に減らしたもので新たにデータセットの作成を行う. 減らした後のデータ数を表 4.1 に示す.

4.2 pix2pix 用のデータセット

表 4.1 使用するデータセット

高知県農業技術センター果樹試験場		高知県の農家
Train(枚)	Valid(枚)	Test(枚)
81	20	13

表 4.2 Train 用を減らした場合のデータ数

Train(枚)			
100%	50%	25%	12.5%
81	40	20	10

4.2 pix2pix 用のデータセット

物体検出用のデータセットを用いてセグメンテーション画像の作成を行う。物体検出用のラベルデータを用いて果実のある場所に楕円を描画することで、果実領域を表したセグメンテーション画像とする。セグメンテーション画像と本物画像のペアを pix2pix 用の学習用データとして用いる。データセットの分け方としては物体検出用のデータセットと同じ条件とする。

4.3 学習データ生成に使用する条件画像

学習データ生成に使用する条件画像として、pix2pix の学習用データを用いて作成を行う。学習用データの変更を加えずに生成を行った場合、ペアとなる本物画像が生成される。そのため、学習データ内で果実領域が平均値以上のデータを用いて、果実領域を 10%, 20%, 30%, 取り除いた条件画像 $R1$, $R2$, $R3$ を作成し、取り除いた果実領域の座標を物体検出用のラベルデータからも取り除く。生成画像に使用する条件画像の枚数を表 4.3 に示す。

4.4 pix2pix モデル構造



図 4.1 pix2pix に用いるペア画像

表 4.3 学習データ生成に使用する条件画像の数

学習データ	100%	50%	25%	12.5%
果実数が平均値以上	28	14	7	4

4.4 pix2pix モデル構造

本研究で使用した pix2pix の生成器と識別器のモデル構造について記述する。

4.4.1 生成器

pix2pix に用いる生成器として, Olaf Ronneberge らが提案した U-Net を用いる. 図 4.2 は生成器のモデル構造を表したものであり, エンコーダ部分に 7 層, デコーダ部分に 7 層からなる U-Net を構成する. Convolution 層を表す「Conv, i, o, k, s」は, 「畳み込み, 入力チャンネル, 出力チャンネル, カーネルサイズ, ストライド」を表しており, ConvTranspose 層も同様に, 「逆畳み込み, 入力チャンネル, 出力チャンネル, カーネルサイズ, ストライド」を表している. 活性化関数としてエンコーダ部分で LeakyReLU, デコーダ部分では ReLU, 出力層に Tanh を用いる. 特徴マップをダウンサンプリングする方法として, プーリング層を使用せず, Stride= 2 用いて行う.

4.5 予測モデル

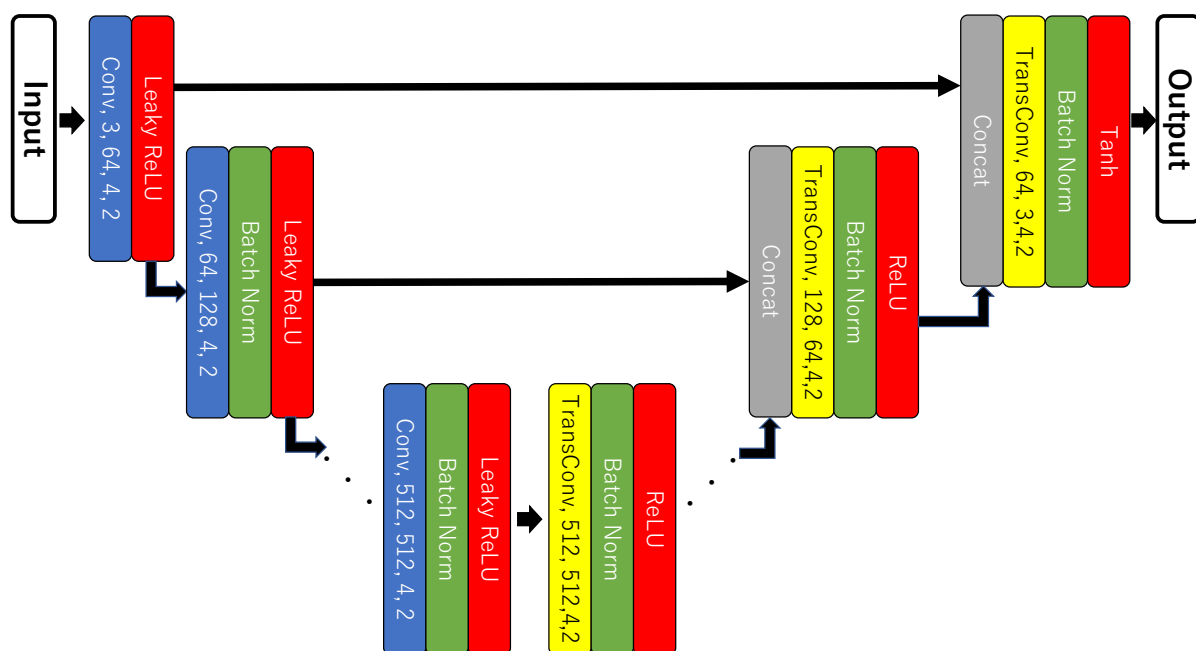


図 4.2 生成器のモデル構造

4.4.2 識別器

pix2pix に用いる識別器として、単純なモデル構造の CNN を用いる。図 4.3 は識別器のモデル構造を表したものであり、活性化関数として Leaky ReLU を用いる。また、GAN における学習を安定化させるために、Batch Normalization の代わりに Spectral Normalization を用いる。

4.5 予測モデル

果実の位置検出に用いる予測モデルとして、Yolov5[16] の COCO2017 データセットでの学習済みモデル S を用いる。モデルサイズは Efficient Net[17] の考え方を基に、CNN の幅と深さのスケーリングを行いサイズを決定している。モデル S では、Depth=0.33, Width=0.50 でスケーリングを行なっている。学習に使用するパラメータを表 4.5 に示す。

4.6 評価指標

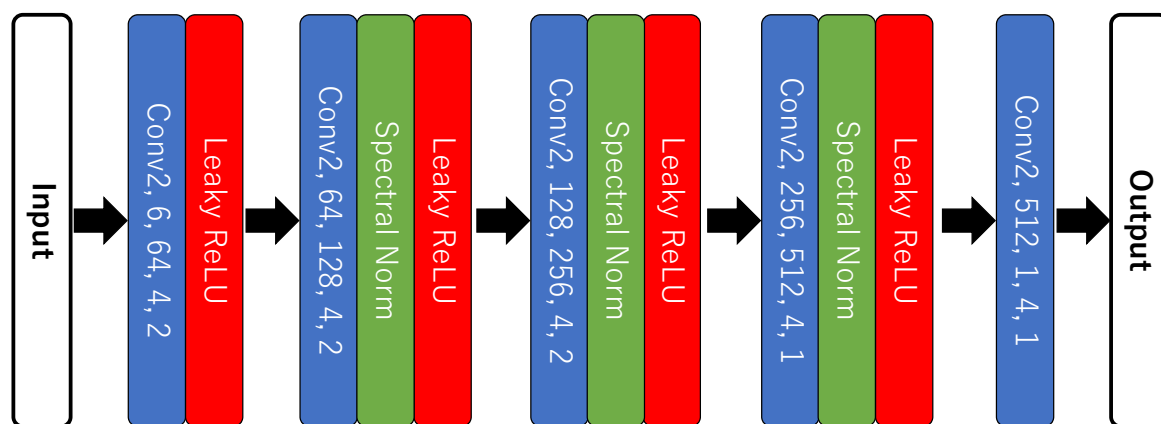


図 4.3 識別器のモデル構造

4.6 評価指標

評価指標として、正解領域と予測領域の重なりが 50%を正解とする、Average Precision(AP@.5) を用いる。AP 計算方法を例を用いて説明する。表 4.5 は検出対象が 3 つの場合の Precision と Recall を示す。AP の計算では予測スコア (Confidence) の降順に並べ、Precision と Recall を計算する。予測スコアが 0.87 の時、正しく予測た数が 2 個のため、Precision は 0.50 となり、Recall は 0.66 となる。この計算を Recall が 1 なるまでか予測結果全てに行う。Precision と Recall から図 4.4 に示した青色の曲線を作成する。作成した青色の曲線を補完した図 4.4 に示したオレンジ色の曲線を作成する。作成したオレンジ色の曲線の面積が AP の値である。

4.7 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

4.7 実験手順

表 4.4 Yolov5 のハイパーパラメータ

Name	Value	Name	Value
lr0	0.01	cls_pw	1.0
lrf	0.01	obj	1.0
momentum	0.937	obj_pw	1.0
weight_decay	0.0005	iou_t	0.2
warmup_epochs	3.0	anchor_t	4.0
warmup_momentum	0.8	fl_gamma	0.0
warmup_bias_lr	0.1	epoch	300
box	0.05	batch-size	16
cls	0.5	img-size	512

表 4.5 予測結果の例

Confidence	Result	Precision	Recall
0.98	True	$1/1 = 1.00$	$1/3 = 0.33$
0.90	False	$1/2 = 0.50$	$1/3 = 0.33$
0.89	True	$2/3 = 0.66$	$2/3 = 0.66$
0.87	False	$2/4 = 0.50$	$2/3 = 0.66$
0.82	True	$3/5 = 0.60$	$3/3 = 1.00$

($R1, R2, R3$) 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。

また, pix2pix を学習を行うにあたり, 表 4.6 に示すハイパーパラメータを用いる。

4.7 実験手順

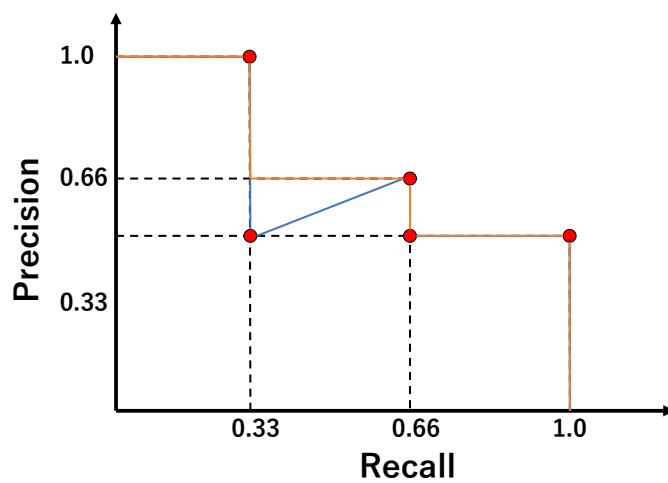


図 4.4 予測結果の例の PR 曲線

表 4.6 pix2pix に用いるハイパーパラメータ

data size	epoch	batch size	optimizer	learning rate
100%	500	1	Adam	2×10^{-4}
50%	1000	1	Adam	2×10^{-4}
25%	1500	1	Adam	2×10^{-4}
12.5%	2000	1	Adam	2×10^{-4}

第 5 章

事前実験

本章では, 本実験を行う前に予測モデルと ADA に用いるデータ拡張の適用結果を示す.

5.1 予測モデルに対するデータ拡張

予測モデルに対して, 画像の変形を行うデータ拡張を用いるが, 明るさや輝度などの色を変更するデータ拡張を用いるかどうかの実験を行う. 実験条件として表 5.1 に示す 4 つの条件で学習を行う. 条件ごとの結果を表 5.1 に示す. 結果から, 明るさなどの色を変える変換を適用すると少しではあるが, 精度の低下が確認された. そのため予測モデルに対するデータ拡張として条件 1 の画像の変形を適用する.

表 5.1 予測モデルに対するデータ拡張の条件

	拡大縮小	回転	左右フリップ	画像移動	結合	明るさ	輝度	HSV
条件 1	✓	✓	✓	✓	✓			
条件 2	✓	✓	✓	✓	✓	✓		
条件 3	✓	✓	✓	✓	✓		✓	
条件 4	✓	✓	✓	✓	✓			✓

5.2 ADA に用いるデータ拡張と適応率

表 5.2 予測モデルに対するデータ拡張の条件ごとの結果

	条件 1	条件 2	条件 3	条件 4
AP@.5	0.73	0.726	0.723	0.72
Precision	0.854	0.804	0.839	0.822
Recall	0.612	0.627	0.615	0.621

5.2 ADA に用いるデータ拡張と適応率

ADA に用いるデータ拡張として、拡大縮小、左右フリップ、回転、画像移動の画像変形の拡張と明るさ、輝度値、色相の色変換の拡張の 2 種類を用いる。判別器に入力するデータにランダムにデータ拡張を行うための異なる適応率で学習を行なった後に画像生成を行う。図 5.1 は 2 種類の拡張を適用した結果、図 5.2 は画像変形の拡張のみを適用した結果を示す。図 5.1, 図 5.2 を比較した際に色変換の拡張を適用している画像は全体的に白い画像が生成された。予測モデルで色の変換を用いた場合精度の低下したため、元の画像から色合いが変わるような生成画像になっているため色の変換のデータ拡張は行わないこととする。図 5.3, 図 5.2 を比較した際に、適応率を増加することで元画像と大きく異なる画像が生成されるため実験で用いる適応率は 0.1 とする。

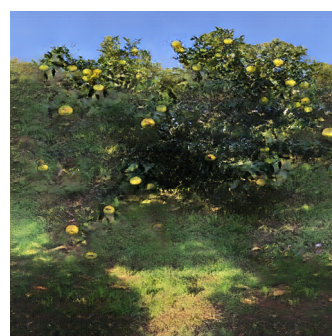


図 5.1 画像変形と色変換 図 5.2 適応率 0.2 の画像変形 図 5.3 適応率 0.1 の画像変形

第 6 章

結果と考察

本章では, 本実験における結果を示す.

6.1 画像生成を行うための学習

6.1.1 pix2pix を用いた画像生成

pix2pix を用いた結果として図 6.1 に示すような画像が生成された. 学習データを変換し条件画像としているため, R1, R2, R3 共に元の学習データに酷似しているが, 条件画像に含まれていない果実領域は再現されておらず, 本物画像から果実を消したような画像が生成された. 条件画像から取り除く果実領域が増えるほど元画像とは異なるような画像が生成され, 地面等に空のような領域が増えている.



(a) R3

(b) R2

(c) R1

図 6.1 pix2pix で生成された画像

6.1 画像生成を行うための学習



(a) R3

(b) R2

(c) R1

図 6.2 Attention 機構を追加したもので生成された画像



(a) R3

(b) R2

(c) R1

図 6.3 ADA を追加したもので生成された画像

6.1.2 Attention 機構を追加したモデルで生成された画像

Attention 機構を追加したモデルで図 6.2 に示すような画像が生成された。pix2pix での結果と同様な画像だが、取り除いた果実領域の場所に対して空の色を生成している画像になった。

6.1.3 ADA を追加したモデルで生成された画像

ADA を追加したモデルで図 6.3 に示すような画像が生成された。学習データを変換し条件画像を使用しているが、R1, R2, R3 共に元の学習データと異なる画像を生成している。

6.2 生成画像を加えた学習結果



(a) R3

(b) R2

(c) R1

図 6.4 Attention 機構と ADA を追加したもので生成された画像

6.1.4 Attention 機構, ADA を追加したモデルで生成された画像

Attention 機構と ADA を追加したモデルで図 6.4 に示すような画像が生成された. ADA を追加したモデルでは, 生成した果実部分が強調されているが, 図 6.4 では, 果実領域に緑が混じった色で生成されており, 本物画像に近い色合いをしている.

6.2 生成画像を加えた学習結果

pix2pix と Attention 機構での生成画像を既存データに加えた結果を表 6.1, ADA, Attention + ADA の結果を表 6.2 に示す. pix2pix, ADA, ADA + Attention 機構での結果では, 果実領域を取り除く割合が小さくなるほど, AP の値が増加しており精度が向上している. 一方, Attention 機構での結果では果実領域を取り除く割合が小さくなるほど AP の値が減少しており, 精度の低下が見られる. Attention 機構を加えることで, 他の結果と比べて元の学習データに近いデータを生成している可能性が考えられる. そのため, 取り除く果実領域が少ないほど学習データに酷似したデータを入力しているため精度の低下したと考えられる. ADA を用いた場合の結果では, 本物画像と異なる画像を用いて学習を行っているが精度の向上が見られた. 本物画像と見た目は異なるが, 学習に必要な特徴が存在していると考えられる. 精度が良かった Attention + ADA の結果では, 元画像に近い特徴を持った画像を生成していると考えられるため, 本物画像と異なる画像でも精度が向上したと考えられる.

6.3 Attention 機構と ADA の生成画像を増加

表 6.1 pix2pix と Attention 機構での生成画像を加えた結果

		pix2pix			+Attention 機構		
	提案手法なし	R3	R2	R1	R3	R2	R1
AP@.5	0.702	0.686	0.708	0.709	0.706	0.7	0.698
Precision	0.872	0.895	0.882	0.883	0.898	0.895	0.869
Recall	0.506	0.458	0.506	0.51	0.491	0.481	0.498

表 6.2 ADA と Attention 機構 + ADA での生成画像を加えた結果

		+ ADA			+ Attention 機構, ADA		
	提案手法なし	R3	R2	R1	R3	R2	R1
AP@.5	0.702	0.689	0.708	0.708	0.703	0.708	0.711
Precision	0.872	0.873	0.873	0.882	0.877	0.894	0.891
Recall	0.506	0.48	0.515	0.508	0.497	0.498	0.503

また, Recall の値は変わらず Precision の値が増加傾向にあることが確認できる. 生成モデルの学習データを用いて画像の生成を行っているため, 類似した画像をデータセットに加えているため誤検出が減少し Precision が増加したと考えられる. そのため, Recall の値を改善するには葉や果実に隠れた状態を再現した画像を生成することが必要となる.

6.3 Attention 機構と ADA の生成画像を増加

AP の精度が良かった条件で生成画像を P_{10} , P_{20} と増加させ, 学習を行った結果を表 6.3 に示す. 学習結果から, 提案手法なしと比較すると Precision の値が増加しており, AP が 1.9 ポイント向上した. この結果より, 提案手法を用いることで精度の向上した.

6.4 学習データを減らした場合

表 6.3 Attention 機構と ADA の生成画像を増加させた学習結果

		Attention 機構 + ADA		
		P_1	P_{10}	P_{20}
mAP@.5	0.702	0.711	0.72	0.721
precision	0.872	0.891	0.838	0.9
recall	0.506	0.503	0.52	0.516

6.4 学習データを減らした場合

学習データの数 を 50%, 25%, 12.5% の条件で, Attention 機構と ADA を用いた生成モデルで画像生成を行い, 生成画像は図 6.5 に示す. また, 提案手法なしの実験結果を表 6.4 に示し, 提案手法を用いた各条件の実験結果を表 6.5, 表 6.6, 表 6.7 に示す. 学習データを 50% に減らした条件では, 提案手法を用いることで P_{20} で 11.7 ポイント向上し, 学習データを減らしていない結果と比較すると 0.8 ポイント低下した. 学習データを 25% に減らした条件では, 提案手法を用いることで P_{15} で 16.3 ポイント向上し, 学習データを減らしていない結果と比較すると 4.5 ポイント低下した. 学習データを 12.5% に減らした条件では, 提案手法を用いることで P_{10} で 43.6 ポイント向上し, 学習データを減らしていない結果と比較すると 11.3 ポイント低下した. 結果から, 提案手法を用いることで 4.7 ポイント低下するが, 精度を維持したままでデータ数を 25% 削減することが可能である. また, 表 6.4 の結果より, 学習データを減らさない場合と 50% に減らした場合で比較した際に, Recall の値が 1.2 ポイントしか差がない結果となった. このことより, 同じような画像に対してアノテーションを行った場合 Recall の値が改善されないと考えられる. そのため, 学習データを増加させる以外の方法でアプローチする必要があると考える.

6.4 学習データを減らした場合

表 6.4 提案手法なしでの実験結果

	100%	50%	25%	12.5%
AP@.5	0.702	0.577	0.494	0.153
Precision	0.872	0.736	0.662	0.261
Recall	0.506	0.494	0.429	0.215



(a) 50%

(b) 25%

(c) 12.5%

図 6.5 Attention 機構と ADA を追加したもので生成された画像

表 6.5 学習データ 50%

	提案手法なし	P_1	P_5	P_{10}	P_{15}	P_{20}
AP@.5	0.577	0.67	0.669	0.678	0.688	0.694
Precision	0.736	0.851	0.865	0.882	0.879	0.887
Recall	0.494	0.458	0.452	0.457	0.474	0.478

表 6.6 学習データ 25%

	提案手法なし	P_1	P_5	P_{10}	P_{15}	P_{20}
AP@.5	0.494	0.606	0.638	0.647	0.657	0.65
Precision	0.662	0.837	0.84	0.86	0.854	0.854
Recall	0.429	0.366	0.412	0.417	0.44	0.421

6.4 学習データを減らした場合

表 6.7 学習データ 12.5%

	提案手法なし	P_1	P_5	P_{10}	P_{15}	P_{20}
AP@.5	0.153	0.28	0.539	0.589	0.558	0.585
Precision	0.261	0.431	0.789	0.827	0.789	0.786
Recall	0.215	0.183	0.286	0.331	0.302	0.369

第 7 章

結論

本研究では, GAN の一つである pix2pix を用いて物体検出用の学習データの生成を行い, 既存の学習データと組み合わせることで精度の向上と学習データを減らすための手法を提案し, ユズを検出するためのデータセットを用いて実験を行った. 実験では pix2pix に Attention 機構, ADA を追加した条件で学習し画像生成を行った. 画像生成に用いる条件画像として学習用データの果実領域を一部取り除いた画像を用いた. Attention 機構を追加すると本物画像に近い画像が生成され, ADA を追加すると本物画像と異なる画像が生成された. 生成画像と既存画像を用いて予測モデルを学習した結果, Attention 機構 + ADA で 1.9 ポイント高いことを示した. ADA での生成画像を用いた学習結果から, 精度の向上が示されたため, 本物画像と大きく異なるが検出対象と同じような特徴を有していると考えた.

また, Attention 機構 + ADA の条件で意図的に学習データを 25%に減らした実験では, 提案手法を用いない場合で 20.8 ポイント低下するのに対し, 提案手法を用いると 4.7 ポイント低下するだけであった. この結果より, 25%まで学習データを削減した状態でも精度を維持することが可能であり, 提案手法を用いることで学習データが少ない場合でも効果があることを示した. しかし, 提案手法を用いた学習において Recall の値が変わらず, Precision の値が増加する傾向が示された. そのため, データ拡張で精度向上を目指すだけでなく, 別の方法でアプローチする必要があると考えた.

謝辞

本研究を行うにあたって、高知工科大学情報学群 吉田真一教授にはご指導いただき心から感謝いたします。また、本研究の副査をお受け頂きました高知工科大学情報学群 福本 昌弘教授並びに、栗原 徹教授に深く感謝いたします。

参考文献

- [1] 農林水産省, “農業労働力に関する統計,”
<https://www.maff.go.jp/j/tokei/sihyo/data/08.html>, 2023 年 2 月 1 日.
- [2] 高知県農業技術センター, “ユズの収量予測への取り組み,” センターニュース, 102 号,
p. 3, 2021.
- [3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative adversarial networks,” In Proc. NIPS, 2014.
- [4] Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A, “Image-to-Image Translation with Conditional Adversarial Networks,” CVPR, 2017.
- [5] Olaf Ronneberger, Philipp Fischer, Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” MICCAI, pp 234–241, 2015.
- [6] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” MICCAI, pp 424–432, 2016.
- [7] Alec Radford, Luke Metz, Soumith Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” ICLR, 2016.
- [8] Mehdi Mirza, Simon Osindero, “Conditional Generative Adversarial Nets,” arXiv:1411.1784, 2014.
- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” ICCV, 2017.
- [10] Terrance DeVries, Graham W. Taylor, “Improved Regularization of Convolutional Neural Networks with Cutout ,” 2017.
- [11] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang, “Random Erasing

- Data Augmentation,” AAAI, 2020.
- [12] Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena, “Self-Attention Generative Adversarial Networks,” PMLR, 2019.
- [13] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, Timo Aila, “Training Generative Adversarial Networks with Limited Data,” NIPS, 2020.
- [14] 福岡 知隆, 南 貴大, 浦田 渡, 藤生 慎, 高山 純一, “深層学習による橋梁点検のための Pix2Pix による疑似訓練データ生成,” 土木学会論文集 F4, 75 巻 2 号 p. I.27-I.35, 2019.
- [15] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert, “Attention U-Net: Learning Where to Look for the Pancreas,” MIDL, 2018.
- [16] Glenn Jocher, YOLO v5, <https://github.com/ultralytics/yolov5>, 2023 年 2 月 1 日.
- [17] Mingxing Tan, Quoc V Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” In Proceedings of International Conference on Machine Learning, 2019.