

可視・熱画像を用いた緑色果実検出

大藪 亜美 【画像情報工学研究室】

Green Pepper Detection Using Visible and Thermal Images

OOZONO, Ami 【Image Processing and Informatics Lab.】

1 はじめに

日本では近年、農業従事者が減少しており、食料の安定的な確保に向けた農業の省力化が期待されている。果実検出は果実の収穫や収量予測に必要な要素であり、農業の省力化において重要な技術である。果実検出は通常、カラー（可視）画像のみを用いて行われることが多い。しかし可視画像では、検出対象と背景の色が類似している場合、対象が背景に溶け込んでしまうことがある。緑色果実は、果実が茎や葉と同じく緑色であるため、色に特徴のある果実と比べ可視画像のみでの検出難易度が高いと考えられる。そのため、緑色果実の検出に焦点を当てた研究も行われている。H.Gan らの研究[1]では、緑色果実である未熟な柑橘果実検出に可視画像のみでなく熱画像も用いることでより優れた結果が得られることを示したが、果実の形状が円形であることを利用しているため円形以外の果実での応用が難しい。そこで本研究では、果実の形状特徴に依存しない、熱画像および可視画像を利用した緑色果実検出手法を提案する。

2 提案手法

2.1 可視・熱融合および特徴抽出ネットワーク

図1は本研究で使用した可視画像および熱画像であるが、熱画像は可視画像と比べてテクスチャ情報などの詳細情報が少ないことがわかる。また、それぞれの画像はセンサの位置が同じでないため、情報の本質の違いから厳密なピクセル単位の位置合わせは困難である。このような熱画像の特性を考え、熱画像と可視画像の融合は、ピクセル単位やそれに近い高解像度で行わず、中から低解像度で行う。さらに、融合には、単純な演算やCNNよりも大域を紐づけられる Attention 機構を用い、画像上で位置がずれていても現実世界で同じ位置にあるもの同士を探索できるようにする。

2.2 可視事前学習重みからの知識転移

本研究では、限定的な可視・熱画像リソース下で高い精度を確保するため、可視画像の大規模データセットで学習させた事前学習モデルの重みを可視画像のみでなく熱画像ストリームの初期重みとして用いる。熱画像ストリームで可視画像事前学習重みを用いるため、1ch



(a) 可視画像 (b) 熱画像
図1: 本研究で撮影した実際の画像

の熱画像をモデル内で単純な連結によって拡張させる。

3 実験方法

3.1 データの取得方法

本研究では、緑色果実としてピーマンを検出対象とし、データは2023年と2024年の10月から4月の正午に高知工科大学の施設園芸ハウスで撮影された。カメラはFLIR E8-XTを用いた。可視・熱モデルに入力する画像形式はRGBAとする。RGBA画像のRGB部分に可視画像のデータ(RGB)を与え、透明度(A)に熱画像を与えた。熱画像は、カメラから得られた温度情報を撮影時のハウス内の気温 $\pm 15^{\circ}\text{C}$ で正規化し、正規化後のデータを8bitにすることで作成した。また、可視画像と熱画像をできるだけ対応させるために、手動で画像の位置合わせをし、ピーマンのずれが8pix程度までのデータを用いた。データは全部で268枚得られた。

3.2 アーキテクチャ

モデルのHeadにはMaskDINO[2]を用い、backboneにはSwin-V1-Tiny[3]を用いた。SwinTransformer[3]は解像度の異なる4つの特徴マップを作成するが、本実験では3つ目の特徴マップを作成する段階から可視・熱融合を行った。融合にはMulti-Head Attentionを用いた。最初の融合ではQに熱特徴量、K、Vに可視特徴量を与えて可視・熱特徴量を作り、その後の融合ではK、Vに可視・熱特徴量を与えて融合を行った。backboneから出力される特徴マップは、解像度が高い順に、上から2つは可視画像のみから得られた特徴マップで、下の2つでは可視・熱融合された特徴マップとなる。具体的には図2のような操作で特徴マップを計算した。

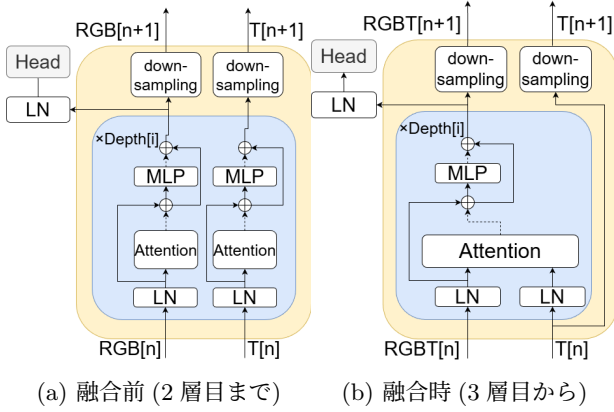


図 2: 融合前と融合時の特徴計算方法

3.3 モデル・比較手法

比較するモデルは3種類で、可視画像のみで検出を行うモデル1種と可視・熱画像を用いて検出を行うモデル2種である。可視・熱画像を用いて検出を行うモデルでは、モデル内で拡張された熱画像の正規化に用いる画素値の平均と標準偏差の値に、ImageNet[4]の(可視)画像の統計量を使用したとき、気温の値の127.5(熱画像は気温±15℃で正規化しているため)と学習データセットの熱画像の標準偏差を使用したときの2種類の設定で実験を行う。各モデルの作成は3-foldクロスバリデーションで行い、201枚画像を学習セット134枚と検証セット67枚に分けてクロスバリデーションする。また、テストセットはクロスバリデーションに用いていない67枚の画像で構成される。すべてのモデルで可視事前学習モデルを用いたファインチューニングを行い、学習は100epoch行う。結果比較には、学習時epoch毎に保存したモデルの中で、バリデーションデータでのmaskAPが最も高かったモデルを用いる。すべてのモデルはテストデータによるmaskAP(IoU=.50:.05:.95)で評価される。

4 結果・考察

表1に各モデルの検出精度を示す。可視・熱画像を入力とする提案手法では、モデル内で拡張した熱画像の正規化に用いる値によって、ベースラインである可視画像のみの入力と比べ、精度が上下する結果となった。

表 1: 各モデルの結果

入力画像	拡張熱画像正規化値	平均 AP(%)	標準偏差
可視のみ	-	55.09	0.61
可視・熱	熱画像	54.83	0.44
可視・熱	可視画像	55.72	1.76

モデル内で拡張した熱画像の正規化を可視画像の統計量に基づいて行った場合には平均APがベースラインよりも0.63%高くなった一方で、拡張熱画像の正規化を学習データの熱画像の統計量に基づいて行った場合には平均APがベースラインよりも0.26%低くなった。モデル内で拡張した熱画像の正規化に用いた実際の値を表2

に示す。標準偏差に注目すると、可視画像の統計量に基づくときは値が57から59であるのに対し、熱画像の統計量に基づくときは値が26から28と、約1/2倍の値となっていることがわかる。つまり、モデル内の最初の学習パラメータが受け取るデータのピクセル値の範囲は、どちらの統計量を用いて正規化するかで約2倍異なる。そのため、2.2節にあるように、熱画像ストリームの重みの初期値に可視画像事前学習重みを用いたとき、熱画像ストリームに当初想定しているよりも大きな値を持つデータが渡されることで、モデルがうまく学習できなくなった可能性があり、それが熱画像統計量を用いて拡張後の熱画像正規化を行った場合の精度低下につながったと考えられる。

表 2: 拡張熱画像正規化値

熱画像によるもの			可視画像によるもの		
学習データ	平均	標準偏差	カラー	平均	標準偏差
set1	127.500	27.951	Red	123.675	58.395
set2	127.500	27.343	Green	116.280	57.120
set3	127.500	26.876	Bule	103.530	57.375

5 おわりに

本研究では、可視および熱画像を用いる緑色果実検出において、中から低解像度でのAttention機構を用いた特徴融合と可視画像による大規模データセットでの事前学習重みを熱画像ストリームに用いる知識転移により、可視画像のみでの検出と比較して最大0.63%高い平均APを達成することができた。しかし、パラメータによっては可視画像のみでの検出結果を下回ることがあったため、精度が上下する要因について、さらなる分析が必要である。また、今回の実験では熱画像を単純な連結で拡張したため、他の拡張方法でも本手法が有効であるか検証したい。

参考文献

- [1] H.Gan et al.: "Immature green citrus fruit detection using color and thermal images", Computers and Electronics in Agriculture, Vol.152, pp.117-125, (2018)
- [2] Feng Li et al.: "Mask DINO: Towards A Unified Transformer-based Framework for Object Detection and Segmentation", arXiv:2206.02777 (2022)
- [3] Liu et al.: "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
- [4] J. Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database", IEEE Computer Vision and Pattern Recognition (CVPR) (2009)