異なる撮影環境に対応したブドウの房計数手法の検証

和坂 優佑[1,a)]　内海 ゆづ子[1,b)]　三輪 由佳[2,c)]　岩村 雅一[1,d)]

概要：ブドウ栽培における摘房は，単位面積あたりのブドウの房の数が一定の数になるように房を間引く作業である．現状の摘房作業では，ブドウの房の数は正確には数えておらず，作業者の経験や感覚で摘房している．そこで本研究では正確な房数の把握を目的とし，全方位カメラで撮影した画像を用いた房の計数システムを構築する．既存研究では，撮影環境が変化した場合の計数モデルの精度低下が課題であった．そのため本研究ではドメイン一般化手法を適用し，撮影環境に依存しない計数の実現を目指す．撮影した年が異なる 2 つのデータセットを用いて実験した結果，一部の結果では異なるデータセットに対しても精度を維持できたが，全体的に実用化できるほどの精度に達していないことが分かった．また撮影時の全方位カメラとブドウ棚の距離が計数精度に影響を与えていることが示唆された．

1. はじめに

ブドウ栽培における作業の 1 つに，成長途中でブドウの房を間引く摘房がある．摘房を行うことで，残された房に十分な養分を供給することができ，出荷可能な品質のブドウを収穫することができる．そのため，摘房はブドウ栽培においてブドウの品質を左右する重要な作業である．

ブドウ栽培では単位面積あたりに存在する房の数に基準があり，摘房ではその基準の数になるように房を間引く．そのため，間引く房の数を知るためには，作業者は単位面積あたりの房の数を知る必要がある．しかし，現状作業者はブドウの房の数を正確には数えておらず，作業者の経験や感覚で摘房している．また，一般に摘房には大きな労力が必要であることが知られている．ブドウを栽培している棚の高さが 1.5 m から 2 m ほどであり，作業者は中腰で作業をするため，作業負荷が大きい．さらに，摘房の時期の房は未熟なため，房の色が葉の色と似ている．葉と区別がつきにくい房を数える作業もまた，作業者への負担が大きい．

赤井ら [1] は，摘房における作業者の負担軽減を目的として，作業者の代わりに房を自動的に計数するシステムを構築した．システムの概要を **図 1** に示す．システムの流れとしては，まずブドウ棚を全方位カメラで撮影し，全方位カメラの画像形式の一種であるステレオ投影画像に変換する．その後，変換した画像から計数領域を切り出し，それを計数モデルに入力することで画像中に含まれる房を計数する．このシステムを使用すると，作業者は圃場内を撮影するだけで房の数を調べることができ，手作業で房の数を数える必要がなくなる．また摘房に慣れていない作業者であっても，このシステムで房の数を容易に知ることができ，摘房すべき房の数を正しく把握できると考えられる．

赤井らの手法ではブドウの計数に成功したものの，1 つの課題がある．赤井らが実験に用いたブドウ棚の画像データセットは 2019 年に大阪府立環境農林水産総合研究所のブドウ圃場で撮影された画像のみで構成されている．そのため撮影する年や太陽の位置，天候などの撮影条件が変わった場合に計数精度がどのように変化するか検証されていない．

撮影環境が変化すると生じる問題として，ドメインシフトがある．ドメインシフトとは，学習データと推論データ間のドメインの違いによって，モデルの性能が低下してしまう問題である．農業分野においては，撮影環境の違いをドメインの違いとし，ドメインシフトを解決するための研究が多く行われている [2], [3]．ドメインシフトを回避する方法の一つに，撮影環境が変化する度に計数モデルを再度学習させることが挙げられるが，これは非現実的である．なぜなら計数モデルの学習に用いる学習データの作成には大きなコストがかかるためである．以上のことから，本システムを実用化するには，撮影環境が変化した場合でも計

[1] 大阪公立大学大学院情報学研究科
Graduate School of Informatics, Osaka Metropolitan University
[2] 大阪府立環境農林水産総合研究所
Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture
[a)] sp25281t@st.omu.ac.jp
[b)] yuzuko@omu.ac.jp
[c)] MiwaY@knsk-osaka.jp
[d)] masa.i@omu.ac.jp

数精度が変わらず高精度なモデルの作成が必要不可欠である．

そこで本研究では計数手法にドメイン一般化を適用し，撮影環境に依存しない計数モデルの作成を行う．ドメイン一般化とは，ドメインシフトを解決する手法の一つである．ドメイン一般化では，あるドメインからドメインに依存しない特徴を抽出し，任意のドメインで高い精度が得られるようにモデルを学習する．ドメイン一般化手法には，複数のドメインを用いてドメイン一般化を行うもの [4] や，単一のドメインを用いてドメイン一般化を行うもの [5] がある．本研究ではラベル付けのコストを少なくするため，単一のドメインでドメイン一般化可能な手法である MPCount [5] を用いる．

実験では，2019 年と 2022 年に大阪府立環境農林水産総合研究所のブドウ圃場で撮影された 2 つのデータセットを用いて MPCount の学習および精度評価を行った．学習データとテストデータの組み合わせを変えて複数実験したところ，一部の結果では撮影環境が異なるデータセットに対しても精度を落とすことなく計数できたが，全体として高い精度で計数することはできなかった．また計数モデルが推定した密度マップを可視化したところ，撮影時のカメラとブドウ棚の距離が計数精度に影響を与えていることが示唆された．

2. 関連研究

本章では，まずセンシングや情報技術を用いてブドウ栽培を支援する研究例を紹介する．続いて，本研究で取り組む問題である物体計数およびドメイン一般化について，特に農業分野で使用されている例を紹介する．

2.1 ブドウ栽培支援システム

近年 Internet of Things (IoT) 技術を活用したブドウ栽培支援の研究が盛んである．例えば，ブドウ圃場内に温度や土壌の水分量などの環境データを取得するセンサを設置し，ブドウの病気の発生リスクを分析するシステム [6] や，ドローンで撮影した画像を用いてブドウの品質を予測するシステム [7] がある．また，ワイン用ブドウ栽培における，画像処理技術を用いたブドウの収量予測などの研究 [8], [9] も行われている．

しかし，ブドウの摘房を支援することを目的としたシステムは，我々の知る限り本研究で扱う赤井らの研究 [1] しか存在しない．赤井らのシステムを実用化することは作業の効率化や作業者の負担軽減につながると考える．

2.2 物体計数

物体計数とは，画像中に含まれる対象物体の数を数えるタスクのことである．物体計数手法は検出ベースの手法 [10], [11] と回帰ベースの手法 [12], [13] に分けられる．検出ベースの手法は，物体検出を用いて対象物体を検出し，検出された数を数える手法である．物体検出を用いているため，対象物体の画像上での位置情報を知ることができるという利点があるが，回帰ベースの手法と比べると計数精度が低い．一方で回帰ベースの手法では，入力画像から対象物体の密度マップを推定し，密度マップを積分することで計数する手法である．検出ベースの手法と比べて精度は高いが，正確な位置情報を取得できないという問題がある．本研究では，ブドウの房の位置情報は必要ではなく，高い精度で計数できることが重要であるため，回帰ベースの手法を用いる．

農業分野における物体計数は様々な作物に適用されており，イネの苗 [14] や青リンゴ [15] を計数する例がある．しかし，ブドウ棚のステレオ投影画像を用いた物体計数についての研究は，本研究で扱う赤井らの研究しか存在しない．

2.3 ドメイン一般化

ドメイン一般化とは，ドメインに依存しない特徴を抽出してモデルの学習を行う手法のことである．ドメイン一般化を適用した計数モデルを用いることで，未知のドメインに対しても精度を低下させることなく計数することができる．

ドメイン一般化に似た手法にドメイン適応 [16], [17] がある．ドメイン適応とは，あるドメインで学習したモデルを，別の特定のドメインでも効果的に使用できるように適応させる手法である．ドメイン一般化手法を適用したモデルは，様々なドメインで高い精度が得られるのに対し，ドメイン適応を適用したモデルは，適応させた特定のドメインでしか高い精度を得ることができないという点で異なる．そのため様々な撮影環境が想定される農業分野では，ドメイン一般化を用いてモデルを作成するのが望ましい．

ドメイン一般化が農業分野に適用されている例として，様々な作物に対応可能なセグメンテーション手法 [18] がある．しかし，物体計数におけるドメイン一般化が農業分野に適用されている例は我々が知る限り存在しない．

3. 手法

本研究ではドメイン一般化を適用した計数モデルを用いて，ブドウ棚の撮影環境が変化しても高い精度で計数できるかを検証する．本章では，赤井らのブドウの房計数手法 [1] と，ドメイン一般化モデルである MPCount [5] について説明する．

3.1 赤井らのブドウの房計数手法

赤井らのブドウの房計数手法の流れは 図 1 に示す通りである．この手法では，全方位カメラで圃場を撮影し，ステレオ投影した画像を入力とする．まず，ブドウ棚のステレオ投影画像から計数領域のみを切り出す．そして，切り

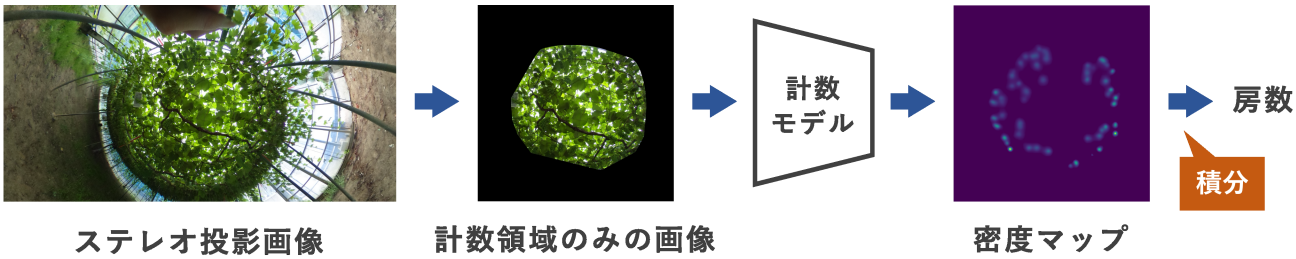


図 1: 計数システムの概要

出した画像を計数モデルに入力することで，計数モデルが入力画像に対するブドウの房の密度マップを推定する．最後に推定した密度マップを積分することで，画像中に含まれるブドウの房の数を求める．

計数モデルの学習では，ブドウ棚のアノテーションデータをもとに作成した密度マップを用いる．ただし，ステレオ投影画像は一般的に用いられる透視投影画像と異なる歪みを持つため，ステレオ投影画像の歪みに対応した密度マップの作成手法を用いる必要がある．またブドウ棚のアノテーション付きデータの数が少ないため，データ拡張手法を適用して学習データの数を増やす必要がある．ここでは赤井らの手法である計数モデル，密度マップ作成手法およびデータ拡張手法について説明する．

3.1.1 計数モデル

赤井らの手法では，密度マップの推定に S-DCNet [19] を用いる．S-DCNet は，入力画像中の密度が高い部分を分割して計数するため，物体数が多くても高い精度で予測することができる．また S-DCNet では推定する物体の数を離散化することで，計数問題を分類問題に帰着させて計数している．回帰問題では，物体数が多い場合に誤差が大きくなり，学習が安定しないという欠点がある．しかし分類問題は比較的安定した学習が可能で，高精度な計数が可能になるという利点がある．S-DCNet は群衆計数手法として提案された [19] が，トウモロコシの穂の計数においても良好な精度が確認されており [19]，植物の計数に有効であると考えられる．

3.1.2 密度マップ作成

本研究で扱う回帰ベースの計数モデルの学習には，計数物体のアノテーションをもとに生成した密度マップを用いる．一般的な密度マップの生成手法としては，画像中の各計数物体の中心をガウスカーネルの平均として 2 次元のガウスカーネルを発生させ，それらを重ね合わせることで各画素の密度を計算する [13]．ここで，学習データセットの i 番目の画像における画素 p の密度関数 $F_j^0(p)$ を考える．I_i は学習データセットの i 番目の画像，$c(i)$ は画像 I_i でアノテーションした物体の数，$\boldsymbol{P}_i = \left\{ P_i^j \middle| j = 1, 2, ...c(i) \right\}$ は画像 I_i 中でアノテーションされた物体の位置とする．また，$\mathbf{1}_{2\times2}$ は 2×2 の単位行列であり，σ^2 はその係数である．σ^2 は密度マップ作成時に定数として与える．$\mathcal{N}(x;\mu,\Sigma)$ を，平均 μ，共分散行列 Σ の 2 次元ガウスカーネルとすると，$F_i^0(p)$ は，以下の式で表現される．

$$F_i^0(p) = \sum_{j=1}^{c(i)} \mathcal{N}\left(p; P_i^j, \sigma^2 \mathbf{1}_{2\times2}\right) \tag{1}$$

しかし，式 (1) で作成する密度マップは，ガウスカーネルの分散 σ^2 を一定とするため，画像中心からの距離に応じて変化するステレオ投影画像の歪みを考慮できない．ステレオ投影画像は，画像中心からの距離が大きくなるほど，物体は画像上で小さく表示されるという特性をもつ．そこで，画像中心から離れるほど，ガウスカーネルの分散を小さくすることで，ステレオ投影画像の歪みを考慮する．赤井らはガウスカーネルの分散 σ^2 を固定値ではなく，画像中心からの距離に反比例するように設定する適応的ガウスカーネルを提案し，歪みを考慮した密度マップを生成した．適応的ガウスカーネルは，式 (1) のガウスカーネルを以下で置き換えることで計算できる．

$$\mathcal{N}\left(p; P_i^j, \left(\sigma_{adap}\left(P_i^j\right)\right)^2 \mathbf{1}_{2\times2}\right) \tag{2}$$

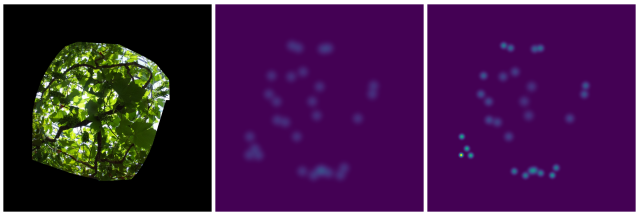
ただし，$\sigma_{adap}\left(P_i^j\right)$ は，画像中心と P_i^j とのユークリッド距離 $D(P_i^j)$ および定数 σ_α を用いて，

$$\sigma_{adap}\left(P_i^j\right) = \frac{1}{D\left(P_i^j\right)}\sigma_\alpha \tag{3}$$

と表される．適応的ガウスカーネルを用いて作成した密度マップの例を **図 2(c)** に示す．

3.1.3 データ拡張

物体計数に用いられる学習データは，正解データを付与するコストが高いことから，規模が小さい．そのため，データ拡張により学習データ数を増やして用いることが一般的である．本研究で用いるブドウ棚のデータセットもデータ数が少ないため，データ拡張手法を用いて学習データを増やす．一般的なデータ拡張手法としては，元の学習データ画像の一部の領域をランダムに切り取るランダムクロップという手法があるが，この手法はステレオ投影画像に適さない．これは，ステレオ投影画像が画像の中心からの距離に応じた歪みを持つため，一部の領域をランダムに切り取



(a) 元画像　(b) σ 固定のガウスカーネル　(c) 適応的ガウスカーネル

図 2: 密度マップの作成例．(a) は密度マップの作成に用いる元画像．(b) は分散 σ が固定されたガウスカーネルを用いて作成した密度マップ．(c) は適応的ガウスカーネルを用いて作成した密度マップ．密度マップのカラーマップには viridis を用い，各画素の値が大きくなるにつれて青，緑，黄となるように色付けしている．

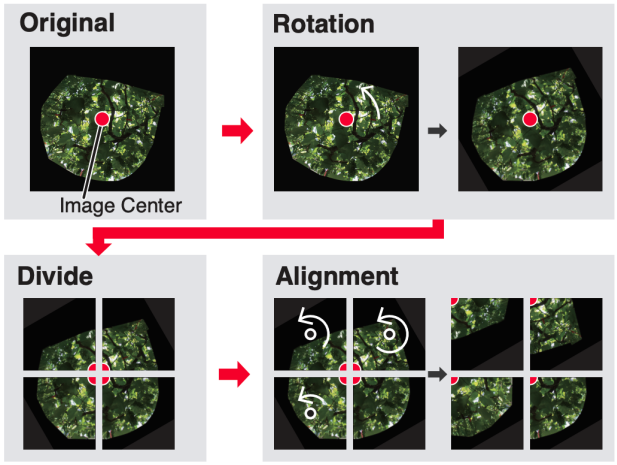


図 3: ステレオ投影画像の歪みを考慮したデータ拡張手法 [1]

ると切り取った画像ごとに歪みの度合いが異なってしまうからである．そのため，ステレオ投影画像の歪みを考慮したデータ拡張手法を用いる必要がある．

本研究では赤井ら [1] が提案するデータ拡張手法を適用する．この手法の概要を **図 3** に示す．この手法では，まず元画像の画像中心を中心にランダムに回転させる．そして回転させた画像を，縦方向と横方向へ半分に分割することで，サイズが等しいの 4 つの画像を得る．最後に，分割前の画像の画像中心が左上になるように分割後の画像をそれぞれ回転させる．この手法を用いることで，拡張後の画像が全て同じような歪みを持つようにデータを拡張することができる．

3.2 MPCount によるドメイン一般化

本研究ではドメイン一般化を適用した計数モデルとして，Peng らが提案する MPCount [5] を用いる．MPCount は，単一のドメインを用いてドメイン一般化が可能な手法であるため，複数のドメインを含むデータセットを構成する必要がない．1 で述べた通り，本研究では学習データの作成に大きなコストがかかることから，複数のドメインを含むデータセットを作成するのは困難である．そのため，学習データ作成時のコストが比較的小さい MPCount を本研究で用いる．

MPCount の構造を **図 4** に示す．MPCount では，学習データ画像に対して輝度変換や色変換を行った画像を仮想的な別のドメインと考え，変換前の元画像と比較することで，ドメインに依存する特徴と依存しない特徴を分離する．そして，ドメインに依存しない特徴のみを用いて密度マップを推定することで，ドメインが変化した場合でも高い精度で対象物体を計数することができる．

MPCount の学習は，図 4 左に示す学習データの画像 I^{ori} とその画像に対して輝度変換や色変換を行った画像 I^{aug} の 2 枚を 1 組として用いる．まず 図 4 の VGG Encoder と Decoder 部分には VGG16-BN [20] を用いて，2 枚の画像の特徴量 F^{ori}，F^{aug} をそれぞれ抽出する．次に， 図 4 中央の CEM で，画像のドメインに依存しない特徴とドメインに依存する特徴を分けるインスタンス正規化を適用する．これにより，特徴量 F^{ori}，F^{aug} からドメインに依存しない特徴量 $\mathrm{IN}(F^{ori})$，$\mathrm{IN}(F^{aug})$ を得る．これを用いて，ドメインに依存しない特徴量のみを取り出すマスクを求める．ここでは，$\mathrm{IN}(F^{ori})$ と $\mathrm{IN}(F^{aug})$ の差が小さい場合，その特徴量はドメインに依存しないとみなす．特徴マップ上の画素 (i,j) の k 次元目のマスク M_{ijk} は，α をどの程度の特徴量の差をドメインの違いとするかを判別する閾値とすると，式 (4) で求められる．

$$M_{ijk} = \begin{cases} 1 & \text{if} |\mathrm{IN}(F^{ori})_{ijk} - \mathrm{IN}(F^{aug})_{ijk}| \leq \alpha \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

そして得られたマスク M と特徴量 F^{ori}，F^{aug} を用いてドメインに依存する特徴を除去した特徴量を得る．この特徴量を 図 4 中央の Attention Memory Bank (AMB) に入力する．AMB は F^{ori}，F^{aug} に対し，AMB に保存されている類似のパターンを出力することで，ドメインに依存しない特徴を再構成する．そして再構成された特徴量 $\tilde{F}^{ori}$，$\tilde{F}^{aug}$ を密度マップの推定器 Density Head に入力することで，物体計数の密度マップ D^{ori}，D^{aug} を出力する．

また計数精度を高めるため，画像を分割し，パッチ単位でそのパッチに対象物体が存在するか存在しないかの分類を 図 4 の PC Head で行う．分類には VGG16-BN による特徴抽出で得られる特徴量 Z^{ori}，Z^{aug} を用いる．得られた分類結果 C^{ori}，C^{aug} は，生成した密度マップの中から対象物体が存在しない領域をフィルタリングするために使用する．対象物体が存在しないと分類されたパッチは，密度マップの値を 0 とすることで，より正確な計数を行う．そして D^{ori}，D^{aug} にフィルタリングを適用した後の密度

マップ D'^{ori}，D'^{aug} を積分することで，計数結果を求める．

また MPCount では 3 種類の損失関数が存在する．AMB でドメインに依存しない特徴を抽出するための $\mathcal{L}_{con}$，PC Head で分類を行うための $\mathcal{L}_{cls}^{ori}$，$\mathcal{L}_{cls}^{aug}$，特徴量から密度マップを生成するための $\mathcal{L}_{den}^{ori}$，$\mathcal{L}_{den}^{aug}$ である．

4. 実験

本研究では，MPCount をブドウの房の計数手法に適用し，撮影環境が異なる場合の計数精度の変化を検証した．本章では実験データと実験条件，実験結果について説明する．

4.1 実験データ

実験には撮影した年が異なる 2 つのデータ A，B を用いた．A は 2019 年，B は 2022 年にそれぞれ大阪府立環境農林水産総合研究所のブドウ圃場で撮影された画像である．撮影に使用したカメラはリコー社製の全方位カメラ RICOH THETA S である．

撮影後，まず全方位カメラで撮影された画像をステレオ投影画像に変換した．その後，計数領域の四隅を手動で指定し，計数領域の境界を描画した．計数領域が描画された画像を用いて，計数領域内に存在するブドウの房を 1 つずつ囲うようにバウンディングボックスを付与することでアノテーションを行った．そして計数領域の境界とバウンディングボックスの凸包を切り出すことで計数領域を切り出した画像を得た．以上の処理により得られた画像例を **図 5**，**図 6** に示す．枚数は A，B ともに 241 枚である．データセット A には，2 つの異なるブドウ棚で撮影した画像が存在したため，一方のブドウ棚で撮影した画像 121 枚を学習データ，もう一方のブドウ棚で撮影した画像 120 枚をテストデータとした．データセット B は 241 枚の画像を学習データ 121 枚，テストデータ 120 枚となるようにランダムに分割した．

またデータセット B のステレオ投影画像には，図 6 に示すように画像上での計数領域の大きさが比較的大きいものと小さいものの 2 種類が存在した．そこで画像上での計数領域の大きさが計数結果に与える影響を調べるため，データセット B を目視により画像上での計数領域の大小で 2 つに分類した．画像上での計数領域が比較的大きいデータセットを B_large，比較的小さいデータセットを B_small とした．B_large と B_small の学習データ枚数はそれぞれ 61 枚，60 枚であり，テストデータ枚数はともに 60 枚である．

全てのデータセットに対して，各画像の密度マップを作成し，学習データのデータ拡張を行った．まず密度マップの作成については，各バウンディングボックスの重心位置を中心とした式 (2) の適応的ガウスカーネルを発生させることで密度マップを作成した．このとき，式 (3) における定数 σ_α の値は 18 とした．次に学習データに対して赤井らのデータ拡張手法を適用した．学習データ画像を $0 < \theta < \frac{\pi}{2}$ の範囲でランダムに回転させ，4 分割する操作を 2 回行い，左右反転させた画像に対しても同様の処理を行うことでデータを拡張した．その結果，元の学習データ 1 枚から 24 枚の学習データを作成した．以上の処理で作成した 4 つのデータセットのデータ枚数を **表 1** に示す．また全てのデータセットにおいて，学習データの解像度は 512×512 pixels，テストデータの解像度は 1024×1024 pixels である．

表 1: データ拡張後の各データセットの枚数

データセット	学習データ（枚）	テストデータ（枚）
A	2904	120
B	2904	120
B_large	1464	60
B_small	1440	60

4.2 実験条件

撮影した画像からステレオ投影画像への変換，計数領域の切り出し，データ拡張，密度マップの作成については，OpenCV を用いて実装した．MPCount の学習は，[5] の著者である Peng らの実装[*1]を用いた．MPCount で行う Photometric augmentation は，Pytorch の Transforms を用いて，明るさ，コントラスト，彩度，色相の変換および先鋭化，平滑化を施した．また，この学習では AdamW [21] による最適化を行った．最大学習率は 0.001，weight decay は 0.0001 であり，学習率スケジューラとして OneCycleLR [22] を用いることで学習率を動的に調節した．エポック数は 180 である．また式 (4) における閾値 α は 0.5 とした．使用した計算機は，MPCount の学習および推論以外は NVIDIA TITAN Xp (3840 コア，クロック周波数 1.58 GHz，メモリ 12 GB) であり，MPCount の学習および推論では NVIDIA TITAN RTX (4608 コア，クロック周波数 1.35 GHz，メモリ 24 GB) である．

モデルの精度評価指標には，ブドウの房の個数に対する平均絶対誤差 (Mean absolute error; MAE)，平均二乗誤差 (Mean square error; MSE) を用いた．テストデータの画像枚数を N，i 番目のテストデータ画像に含まれる房の数を c_i，i 番目のテストデータ画像に含まれる房の数の推定値を $\hat{c}_i$ とすると，MAE，MSE はそれぞれ式 (5)，(6) で計算される．

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|c_i - \hat{c}_i| \tag{5}$$

$$\mathrm{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - \hat{c}_i)^2} \tag{6}$$

MAE，MSE ともに値が小さいほど，モデルの計数精度が

[*1] https://github.com/Shimmer93/MPCount

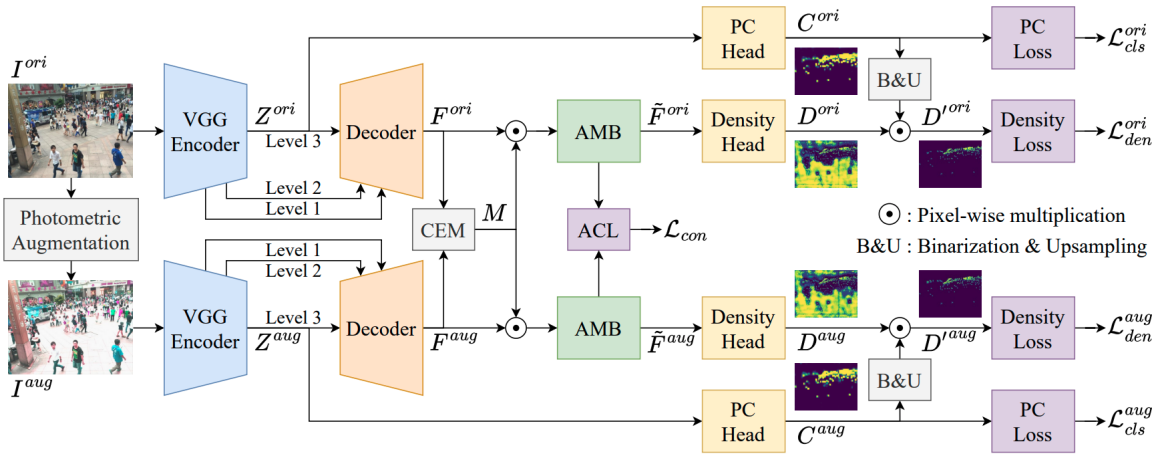
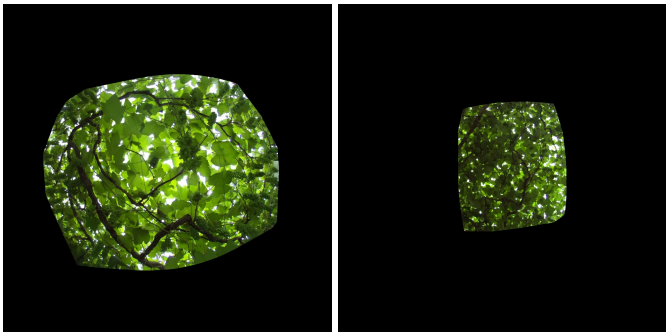


図 4: MPCount の構造 [5]

図 5: データセット A の画像例



(a) 計数領域が大きく表示されているもの　(b) 計数領域が小さく表示されているもの

図 6: データセット B の画像例

良いことを表している．

また Matplotlib を用いて密度マップの正解データおよび推定結果を可視化した．密度マップのカラーマップには viridis を用いた．

4.3 実験結果

まずデータセット A，B を用いた場合の実験結果を **表 2** に示す．データセット A で学習したモデルを用いてデータセット A をテストした場合が MAE，MSE ともに最も小さく，他の場合と比べても著しく小さかった．一方でデータセット A で学習したモデルを用いてデータセット B をテストした場合，他の場合と比べて MAE，MSE が最も大きくなった．本実験ではドメイン一般化手法を用いたため，適切なドメイン一般化が行われていればテストに用いるデータセットを変えても MAE，MSE は大きく変化しないはずである．しかしこの 2 つの結果に大きな差があることから，データセット A を用いたドメイン一般化がうまく行えていない．

表 2: データセット A，B をそれぞれ用いてドメイン一般化した結果

学習データ	テストデータ	MAE	MSE
A	A	3.12	4.15
A	B	21.28	25.24
B	B	9.80	11.91
B	A	8.54	11.71

次に， 表 2 のデータセット B で学習したモデルに対するテスト結果を見ると，テストにデータセット A を用いた場合とデータセット B を用いた場合で MAE，MSE に大きな差は見られないことがわかる．このことから，データセット B を用いたドメイン一般化は上手く行われたと考えられる．

最後に 表 2 全体を見ると，学習データやテストデータにデータセット B を用いた場合，学習データとテストデータともにデータセット A を用いた場合と比べて MAE，MSE がかなり大きくなったことがわかる．また一枚の画像に含まれる房の数の平均はデータセット A が 31.38 個，データセット B が 40.17 個であった．そのため，これほどの計数誤差が生じると適切な数だけ摘房することが困難である．よってまだ実用化できる程の精度には達していない．

データセット B_large，B_small を用いて実験した結果を **表 3** に示す．表を見てわかるように，テストデータに

表 3: データセット B をデータセット B_large，B_small に分けた結果

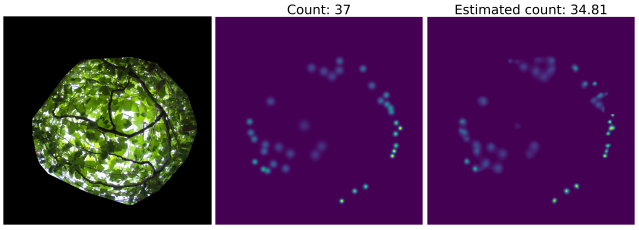
学習データ	テストデータ	MAE	MSE
A	B_large	10.86	12.24
A	B_small	31.70	33.54
B	B_large	8.24	9.56
B	B_small	11.36	13.86
B_large	B_large	6.65	7.97
B_large	B_small	23.83	25.87
B_large	A	16.21	19.07
B_small	B_large	10.28	12.13
B_small	B_small	11.47	13.77
B_small	A	12.07	15.61

B_large を用いた場合の方が，B_small を用いた場合と比べて MAE，MSE が小さくなっている．また，データセット A で学習済みのモデルを用いてデータセット B_large，B_small を推定し，密度マップを可視化した例をそれぞれ **図 7**，**図 8** に示す．この密度マップは各画素の値が大きくなるにつれて青，緑，黄となるように色付けした．表 3 および 図 7，図 8 より，画像上での計数領域の大きさが比較的大きいと精度が良く，比較的小さいと精度が悪い傾向があることがわかった．

以上の結果を踏まえると，表 2 においてデータセット B を学習データに用いた場合，上手くドメイン一般化できたように見られたのは，データセット B には画像上での計数領域の大きさが異なる 2 種類のデータが含まれていたためだと考えられる．そのためデータセット B で学習したモデルは画像上での計数領域の大きさが 2 種類あっても対応できたが，画像上での計数領域の大きさが全て大きいデータセット A で学習したモデルでは，画像上での計数領域の大きさが小さい場合に対応できなかったと考えられる．画像上での計数領域の大きさは，撮影時の全方位カメラとブドウ棚の距離に依存する．そのため画像上での計数領域の大きさの違いによる計数精度への影響を軽減するためには，撮影時に全方位カメラとブドウ棚の距離を一定に保ちながら撮影をする必要がある．

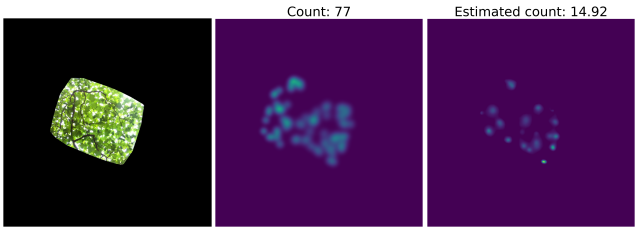
5. 結論

ブドウ栽培において，摘房は手作業でブドウの房を数えて間引く必要があるため，大きな労力のかかる作業である．摘房時の作業者の負担軽減を目的として，ブドウ棚を撮影するだけで自動で房を計数するシステムの構築が行われている．システムの実用化では，ある 1 つの撮影環境で学習済みの計数モデルを様々な圃場で使用できることが重要である．しかし，現状のシステムでは，ある時期に撮影された画像のみを学習データに用いてシステムを構築しているため，別の圃場や環境で撮影された画像への適用が困難である．そこで本研究では，撮影された時期や環境が異なる



(a) 元画像　(b) Ground truth　(c) 推定結果

図 7: データセット A で学習済みのモデルでデータセット B_large を推定した例．(a) はモデルに入力したテストデータの画像．(b) は (a) とアノテーションデータをもとに作成した密度マップの正解データ．Count は (a) の画像中に含まれるブドウの房の数を表す．(c) はモデルが出力した密度マップの推定結果．Estimated count は推定した密度マップを積分して得られたブドウの房の数を表す．



(a) 元画像　(b) Ground truth　(c) 推定結果

図 8: データセット A で学習済みのモデルでデータセット B_small を推定した例．(a)，(b)，(c) はそれぞれ 図 7 の (a)，(b)，(c) に対応．

場合でもシステムの利用が可能になることを目的として，ブドウの房計数システムにドメイン一般化を適用した．実験の結果，ドメイン一般化を適用しても，学習と推定に用いるデータセットが異なると計数精度が低下する現象が一部の結果で見られた．また多くの実験結果では，実用化できる程の精度に達していなかった．さらに密度マップの推定結果の可視化により，画像上での計数領域の大きさが小さいほど，計数精度が悪くなる傾向にあることがわかった．

今後の課題としては，まず MPCount 内の Photometric augmentation で使用する augmentation 手法を変化させることが挙げられる．元画像の明るさやコントラストなどをどの程度変化させるかによって，モデルの精度が変化するため，適切な augmentation 手法を調べる必要がある．次に，他のドメイン一般化手法を使用することが挙げられる．本研究ではドメイン一般化手法として MPCount しか用いていない．そのため，他のドメイン一般化手法も用いて精度評価を行うことで，ブドウの房計数に適した手法を調べる．

謝辞　この研究は JSPS 科研費 JP24K03020 の支援を受けて実施された．

参考文献

[1] Akai, R., Utsumi, Y., Miwa, Y., Iwamura, M. and Kise, K.: Distortion-Adaptive Grape Bunch Counting for Omnidirectional Images, *Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 599–606 (2021).

[2] Gogoll, D., Lottes, P., Weyler, J., Petrinic, N. and Stachniss, C.: Unsupervised Domain Adaptation for Transferring Plant Classification Systems to New Field Environments, Crops, and Robots, *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2636–2642 (2020).

[3] Magistri, F., Weyler, J., Gogoll, D., Lottes, P., Behley, J., Petrinic, N. and Stachniss, C.: From one field to another—Unsupervised domain adaptation for semantic segmentation in agricultural robotics, *Computers and Electronics in Agriculture*, Vol. 212, No. 108114 (2023).

[4] Du, Z., Deng, J. and Shi, M.: Domain-general crowd counting in unseen scenarios, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 1, pp. 561–570 (2023).

[5] Peng, Z. and Chan, S.-H. G.: Single Domain Generalization for Crowd Counting, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28025–28034 (2024).

[6] Pérez-Expósito, J. P., Fernández-Caramés, T. M., Fraga-Lamas, P. and Castedo, L.: VineSens: An Eco-Smart Decision-Support Viticulture System, *Sensors*, Vol. 17, No. 3 (2017).

[7] García-Fernández, M., Sanz-Ablanedo, E. and Rodríguez-Pérez, J. R.: High-Resolution Drone-Acquired RGB Imagery to Estimate Spatial Grape Quality Variability, *Agronomy*, Vol. 11, No. 4 (2021).

[8] Nuske, S., Achar, S., Bates, T., Narasimhan, S. and Singh, S.: Yield estimation in vineyards by visual grape detection, *Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2352–2358 (2011).

[9] Liu, S., Zeng, X. and Whitty, M.: 3DBunch: A Novel iOS-Smartphone Application to Evaluate the Number of Grape Berries per Bunch Using Image Analysis Techniques, *IEEE Access*, Vol. 8, pp. 114663–114674 (2020).

[10] Chen, S., Fern, A. and Todorovic, S.: Person count localization in videos from noisy foreground and detections, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1364–1372 (2015).

[11] Li, M., Zhang, Z., Huang, K. and Tan, T.: Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection, *Proceedings of 2008 19th International Conference on Pattern Recognition*, pp. 1–4 (2008).

[12] Lempitsky, V. and Zisserman, A.: Learning To Count Objects in Images, *Proceedings of Advances in Neural Information Processing Systems*, Vol. 23 (2010).

[13] Cohen, J. P., Boucher, G., Glastonbury, C. A., Lo, H. Z. and Bengio, Y.: Count-ception: Counting by Fully Convolutional Redundant Counting, *Proceedings of IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 18–26 (2017).

[14] Wu, J., Yang, G., Yang, X., Xu, B., Han, L. and Zhu, Y.: Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network, *Remote Sensing*, Vol. 11, No. 6 (2019).

[15] Linker, R., Cohen, O. and Naor, A.: Determination of the number of green apples in RGB images recorded in orchards, *Computers and Electronics in Agriculture*, Vol. 81, pp. 45–57 (2012).

[16] Zhang, C., Li, H., Wang, X. and Yang, X.: Cross-scene crowd counting via deep convolutional neural networks, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 833–841 (2015).

[17] Cai, Y., Chen, L., Guan, H., Lin, S., Lu, C., Wang, C. and He, G.: Explicit invariant feature induced cross-domain crowd counting, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 1, pp. 259–267 (2023).

[18] Angarano, S., Martini, M., Navone, A. and Chiaberge, M.: Domain Generalization for Crop Segmentation with Standardized Ensemble Knowledge Distillation, *Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5450–5459 (2024).

[19] Xiong, H., Lu, H., Liu, C., Liu, L., Cao, Z. and Shen, C.: From Open Set to Closed Set: Counting Objects by Spatial Divide-and-Conquer, *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, pp. 8361–8370 (2019).

[20] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *Proceedings of International Conference on Learning Representations* (2014).

[21] Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, *International Conference on Learning Representations* (2017).

[22] Smith, L. N. and Topin, N.: Super-convergence: very fast training of neural networks using large learning rates, *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, Vol. 11006, No. 1100612 (2019).