

拡散モデルを用いた画像データ拡張による メロン等級判定モデルの提案

海老沢源^{†1} 島田拓人^{†2} 平原健太郎^{†2} 小池誠^{†3}
小川晋^{†4} 野村祐一郎^{†5} 峰野博史^{†5†6}

静岡大学情報学部^{†1} 静岡大学大学院総合科学技術研究科情報学専攻^{†2}
静岡大学大学院自然科学系教育部^{†3} 株式会社大和コンピューター i 農業開発部^{†4}
静岡大学大学院情報学領域^{†5} グリーン科学技術研究所^{†6}

1. はじめに

農業における品質管理は熟練農家の判断に依存している。特に高級メロンの選別作業では、果実の形状、色調、表皮の模様、傷や汚れの有無などの外観を目視で評価する。この方法は新規就農者には時間と経験が要求され、一貫した基準の確立が難しい課題が存在する。これに対し、機械学習を用いて自動選別を行うメロン等級判定モデル[1]は、メロンの全周映像から特徴を抽出し、各等級92玉のメロン全周映像の学習によって4段階の等級判定に対し、82.1%の精度を実現している。表 1にメロン等級判定システムで判断している4等級の条件を示す。しかし、網目の均一性といった詳細な網目特徴を捉えることが重要なA等級は63.7%、B等級は76.7%と、判定性能が低く、改善の余地がある。

一方、潜在拡散モデル (LDM) を用いた画像生成モデルの発展が著しく、画像から画像を生成する画像拡張技術の研究が注目を集めている。本研究では、LDMベースのStable Diffusionを拡張したControlNet[2]を用いて詳細な網目特徴を保持したメロン表面画像を生成し、データ拡張を行い、メロン等級判定モデルの精度向上を目指す。入力画像の構図を維持したまま画像を生成できるControlNetを用いて、小さい傷や汚れ、網目の集合などの詳細な網目特徴を含んだ画像を適切に生成し、効果的な学習が可能か分析を行う。

2. 関連研究

2.1 工業製品の表面欠陥画像の分類

工業製品の表面欠陥画像の分類問題では、表面欠陥の画像が少なく、画像拡張手法[3]が検証されている。画像拡張には古典的なStyleGANと最新のStable Diffusion, LoRA, ControlNetを組み合わせた手法が比較された。分類精度は元データのみで71.3%、StyleGANで81.4%、最新手法で

表 1 メロン等級判断概要

等級	網目の特徴	輪郭の特徴
A	A_m (密度が均一)	A_c (歪みが少ない)
B	B_m (網目の乱れ)	A_c
C	A_m, B_m	B_c (歪みがある)
D	C_m (傷や汚れ)	A_c, B_c

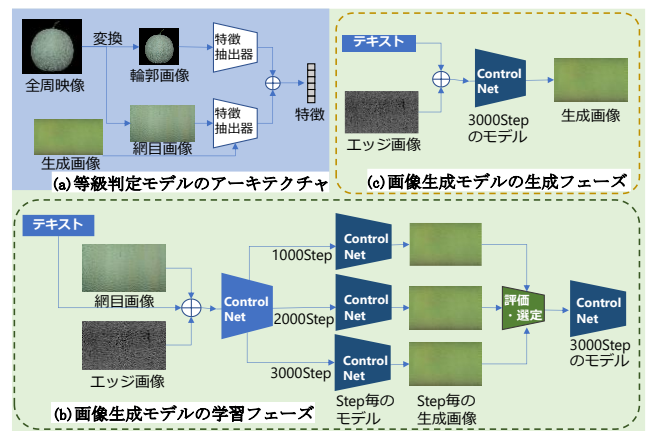


図 1 提案手法の全体像

85.1%と報告され、ControlNetの有効性が検証された。

2.2 大腸がんの病理組織領域画像の分類

大腸がんの病理組織領域画像の分類問題では画像拡張手法[4]が検証された。画像拡張には古典的なCycleGAN2と、LDMをベースとして元の大腸がんの組織画像に対して転移学習を行ったHistoDiffusionが比較された。分類精度は元データのみで85.5%、StyleGAN2で89.5%、提案手法で91.9%と報告されLDMベースの画像生成を行う手法の有効性が検証された。

3. 提案手法

3.1 メロン等級判定モデルのアーキテクチャ

本研究では、LDMベースのStable Diffusionを拡張したControlNet[2]を用いて、詳細な網目特徴を保持した多量のメロン表面画像を生成し、ファインチューニングを行うことで、メロン等級判定モデルの精度向上を目指す。図 1(a)にメロン等級判定モデルのアーキテクチャを示す。既存のメロン等級判定モデル[1]では、メロンの全周映像を輪郭画像と網目画像に変換し、それぞれの特徴抽出を行う。抽出された特徴から等級別の特徴を算出し、その結果を基に未知の全周映像の等級判定を行う。提案手法では網目画像を

Proposal for the Melon Grading Model Using Image Data Augmentation with Diffusion Model

^{†1} HAJIME EBISAWA, Faculty of Informatics, Shizuoka University
^{†2} TAKUTO SHIMADA, KENTARO HIRAHARA Graduate School of Science and Technology, Shizuoka University
^{†3} MAKOTO KOIKE, Graduate School of Science and Technology, Shizuoka University
^{†4} SUSUMU OGAWA, Daiwa Computer Co., Ltd
^{†5} YUHICHIRO NOMURA, HIROSHI MINENO, College of Informatics, Academic Institute, Shizuoka University
^{†6} HIROSHI MINENO, Research Institute of Green Science and Technology

基に作成した生成画像を網目画像の特徴として抽出し、新たなメロン等級判定モデルを作成する。

3.2 画像生成モデルの学習

画像生成モデルの学習では、画像にテキストを付与したデータセットを用いて ControlNet の追加学習を行う。図 1(b)に画像生成モデルの学習フェーズを示す。

ControlNet の学習に使用する画像には、生成したい画像と画像の構図を指示するための画像が必要である。生成したい画像である網目画像に対して、網目特徴などの構図を指示するエッジ検出を施した画像を作成する。それぞれの画像に対し、テキスト情報を付与して生成時に傷や汚れなどの網目特徴の誘導を行う。等級判断の条件となる網目特徴のラベルに加えて網目画像の詳細な説明を追加する。

画像にテキストを付与したデータセットを用いて ControlNet の学習を行う。学習済みのモデルに基づいて、100Step 毎の学習でモデルを作成し、各モデルから各等級の網目画像を生成する。生成した画像の元画像との類似性、指定された等級に適合しているかを可視化し、画像生成モデルを選定する。

3.3 等級判定モデル用画像の作成

画像生成モデルを使用して、メロン等級判定モデル用画像を生成する。図 1(c)に画像生成モデルの生成フェーズを示す。テキスト情報として生成したい等級の網目画像の詳細な情報を使用し、画像としてその等級の網目画像に基づいて作成したエッジ画像を使用する。生成した画像は、既存の等級判定システムによる評価を経て選別し、メロン等級判定モデルに使用する生成画像として決定される。

4. 実験

4.1 生成画像の評価

画像生成モデルの選定時とメロン等級判定モデルに使用する生成画像の決定時に、生成画像に対して評価を行った。目的とする生成画像は、元の網目画像と類似した特徴を含むこと、付与される等級ラベルと一致する必要がある。熟練農家の等級判定を自動化するシステムに適用するため、元画像との類似度は人間の知覚的類似度と、傷や汚れなどの画像情報からの評価が求められる。したがって、学習済み画像分類ネットワークを用い、人間の知覚的類似度を評価する LPIPS[5]と、CLIP, OpenCLIP, DINO の各モデルが生成する特徴表現を組み合わせ、画像の意味内容や文脈、テキストとの関連性を評価する DreamSim[6]を使用して、元画像と生成画像を比較した。指定された等級の画像が適切に生成されているかは、既存の等級判定システムを使用して評価した。図 2 に以上の評価に基づいて生成した網目画像を示す。

4.2 メロン等級判定モデルの評価

簡易評価では、メロンの全周映像を用いて各等級に学習データ $N=92$ 玉、テストデータ 20 玉を使用する。学習デー

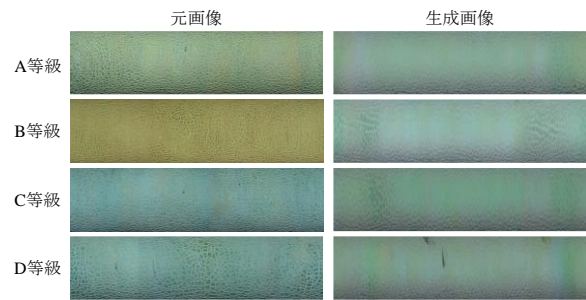


図 2 生成した網目画像

表 2 各等級 20 玉の等級判定結果

	Accuracy (%)	F-Score
生成画像なし ($N=92$)	81.3	79.7
生成画像あり ($N=92, N_g=92$)	82.5	81.4

タ 92 玉と画像生成モデルで生成した同数の網目画像 $N_g=92$ を用いて等級判定モデルを構築し、テストデータに対して等級判定を行った。表 2 に各等級 20 玉の等級判定結果を示す。提案手法で Accuracy が 1.2 ポイント、F-Score が 1.7 ポイント上昇し、詳細な網目特徴を保持した生成画像の追加による判定精度の向上を確認した。さらに k 分割交差検証を行い、モデルを選出し、提案手法の性能評価を行う予定である。

5. おわりに

ControlNet[2]を用いて詳細な網目特徴を保持したメロン表面画像を生成し、生成画像の特徴を組み込んだメロン等級判定モデルを構築した。第 4 章では提案手法による判定精度向上の可能性が示された。

今後の予定としては、元画像と生成画像の差異を分析し、類似度を近づけ、より高精度な網目画像の生成を目指す。また、提案手法の評価に基づいて分析を行い、等級判断における重要な特徴を生成し、分類に特化した画像生成を検討している。

謝辞

本研究の一部は、JST 創発的研究支援事業ならびに静岡大学グリーン科学技術研究所プロジェクト研究支援を受けたものである。

参考文献

- [1] 小池誠, 平原健太郎, 渡辺隆一ほか: 画像局所特徴の類似度を用いたメロン等級判定システムの開発, 情報処理学会研究報告コンシューマ・デバイス&システム, Vol.13, No.1, pp12-25 (2023).
- [2] Zhang, L., Rao, A. and Agrawala, M.: Adding conditional control to text-to-image diffusion models, Proc. *IEEE/CVF ICCV*, pp. 3836-3847 (2023).
- [3] Zhong, X., Zhu, J., Liu, W., et al.: An Overview of Image Generation of Industrial Surface Defects, Proc. *Sensors*, Vol.23, No.8160 (2023).
- [4] Ye, J., Ni, H., Jin, P., et al.: Synthetic Augmentation with Large-scale Unconditional Pre-training, Proc. *MICCAI*, pp.754-764 (2023).
- [5] Zhang, R., Isola, P., Efros, A. A., et al.: The unreasonable effectiveness of deep features as a perceptual metric, Proc. *IEEE CVPR*, (2018).
- [6] Fu, S.: Learning New Dimensions of Human Visual Similarity using Synthetic Data, Diss. MIT (2023).