

IS1-180 拡散モデルベースの3次元復元手法を用いた 単一食事画像からのカロリー量推定

大岸 真由, 田邊 光, 柳井 啓司 (電気通信大学)



概要

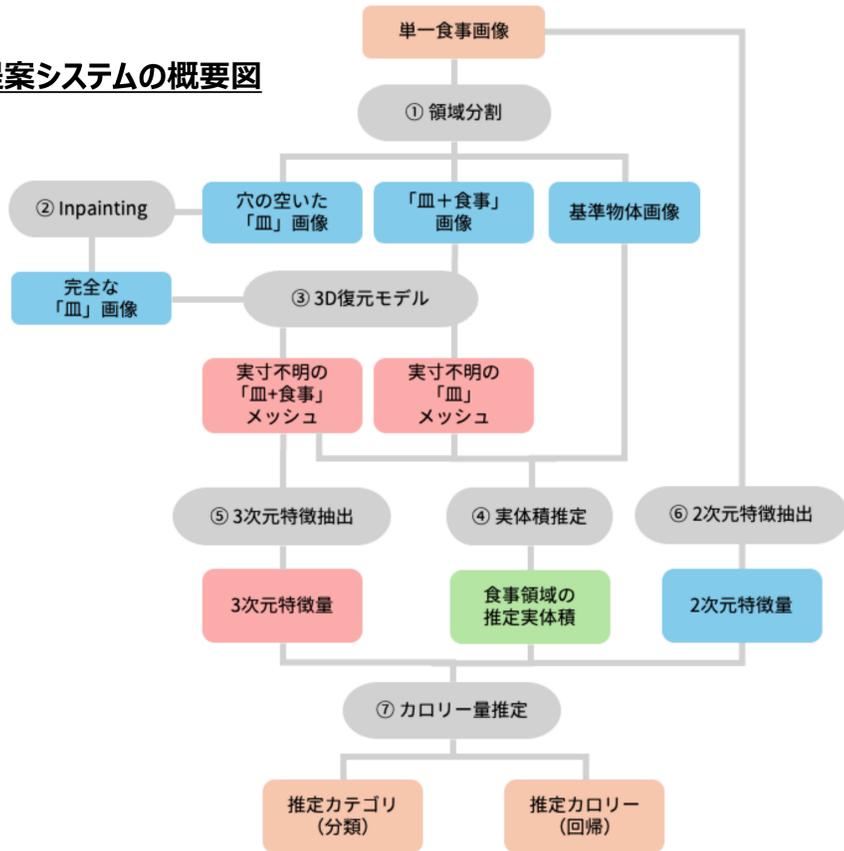
単一食事画像からカロリー量を正確に推定することは、健康管理アプリにおける食事記録の信頼性向上に重要
従来手法は食事を2次元平面で捉えてカロリー量を推定しているが、**食事の立体構造や体積を軽視している**

目的 3次元再構成技術を用いて
単一食事画像からカロリー量を正確に推定する



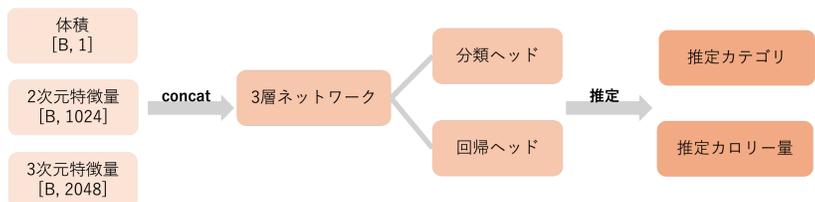
手法

提案システムの概要図



- ① 領域分割: Grounded SAM [1] で食事・皿・参照物体の領域マスクを抽出しそれらを用いて 1.食事が無い皿画像 2.皿+食事画像 3.基準物体画像 を得る
- ② 皿のみ画像生成: LoRA 訓練済 SDv2 で食事が無い皿画像を復元
- ③ 3次元再構成: Wonder3D [2] で 1.皿 2.皿+食事 のメッシュを生成
- ④ 実体積推定: 両メッシュの体積差をとり、参照物体から推定したスケール因子を適用して食事の実体積を推定
- ⑤ 3次元点群特徴抽出: 皿+食事のメッシュから得た点群に PointGPT [3] を適用し点群特徴抽出
- ⑥ 2次元特徴抽出: 元画像から OpenAI ViT-14/L で画像特徴抽出
- ⑦ カロリー量推定: 体積・2次元特徴・3次元特徴を結合して2ヘッド MLP に入力し、カロリー量と食品カテゴリを推定しマルチタスク学習 (下図)

カロリー量推定ネットワーク



実験

データセット・実装詳細

- MetaFood3D [4]: 食事画像・食事3次元メッシュ・領域マスク・体積・栄養素情報などの食事マルチモーダルデータセット
- 108種類 637件の食品からなり 80:20で訓練と評価に使用
- SDv2による inpainting は皿の穴修復時に食品を誤生成
 - 公開ソースから収集した実画像 1k枚 + 合成画像 2k枚からなる計 3k枚の皿画像データセットで SDv2 を訓練 (LoRA)

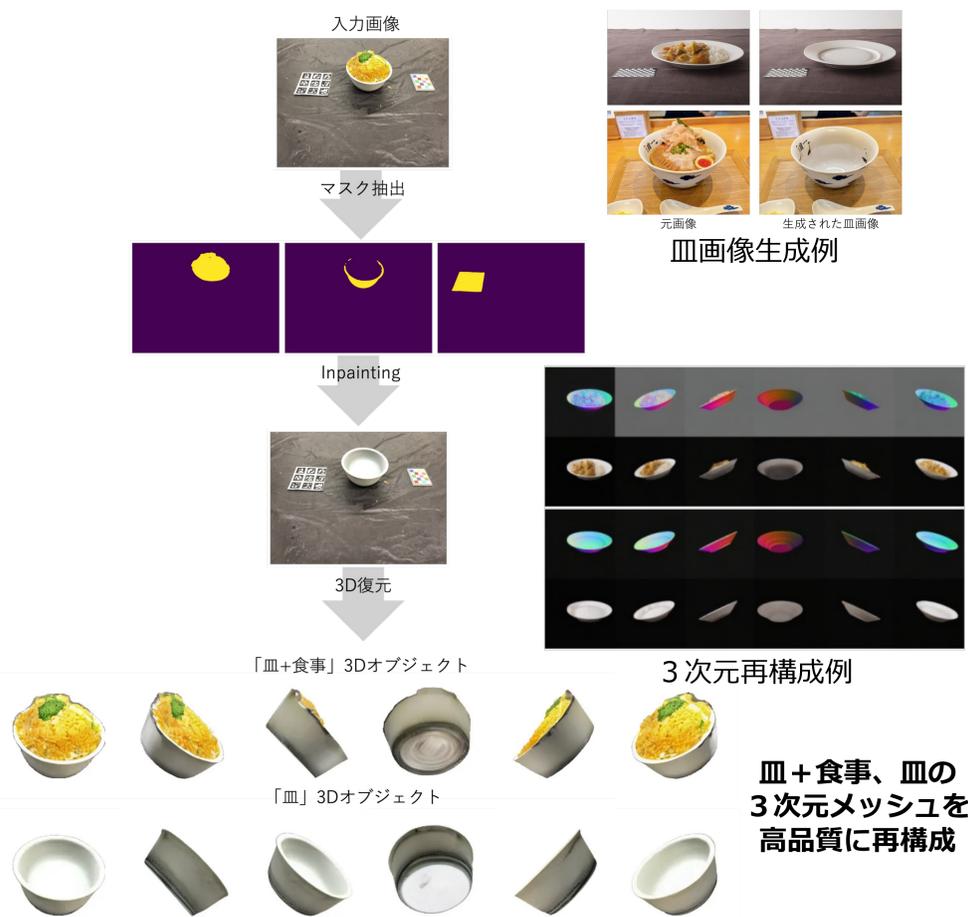
関連手法との比較

手法	体積 MAE (ml)	体積 MAPE (%)	カロリー量 MAE (kcal)	カロリー量 MAPE (%)	カテゴリ正解率
MFP3D [5]	62.60	41.43	77.98	68.05	-
Ours	57.53	269.31	92.74	212.78	0.96

特徴量の追加による貢献

2D	体積	3D	カロリー量 MAE (kcal)	カロリー量 MAPE (%)
✓			133.37	373.23
✓	✓		101.12	254.09
✓	✓	✓	92.74	212.78

マスク抽出・Inpainting・3次元再構成の結果



結論・今後の方針

単一画像からの3次元再構成で食事の形状や体積を正確に把握
➢ 2次元特徴のみによる推定よりも正確なカロリー量推定を実現

課題

- 領域分割モデルの改善・分割ミス時における後段への影響緩和
- 皿に盛られている食事のマルチモーダルデータセット構築

[1] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, and Lei Zhang. 2024. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks. arXiv preprint arXiv:2401.14159.
[2] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, and Wenping Wang. 2024. Wonder3D: Single Image to 3D using Cross-Domain Diffusion. In Proc. of IEEE Computer Vision and Pattern Recognition. 9970-9980.
[3] Guangyan Chen, Meiling Wang, Yi Yang, Kai Yu, Li Yuan, and Yufeng Yue. 2024. PointGPT: Auto-regressively generative pre-training from point clouds. Advances in Neural Information Processing Systems.
[4] Yuhao Chen, Jiangpeng He, Chris Czarnecski, Gautham Vinod, Talha Ibn Mahmud, Siddeshwar Raghavan, Jing Ma, Dayou Mao, Saejith Nair, Pengcheng Xi, Alexander Wong, Edward Delp, and Fengqing Zhu. 2024. MetaFood3D: Large 3D Food Object Dataset with Nutrition Values. In Proc. of International Conference on Learning Representations.
[5] Jing Ma, Xiaoyan Zhang, Gautham Vinod, Siddeshwar Raghavan, Jiangpeng He, and Fengqing Zhu. 2024. MFP3D: Monocular Food Portion Estimation Leveraging 3D Point Clouds. In 2024 27th International Conference on Pattern Recognition (ICPR), 9th International Workshop on Multimedia Assisted Dietary Management.