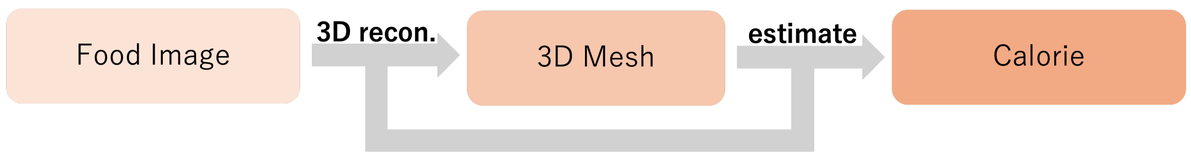
Diffusion-Guided 3D-Aware Calorie Estimation from a Single Food Image

Mayu Ogishi, Hikaru Tanabe, Keiji Yanai
The University of Electro-Communications, Japan

ACM multimedia mm Dublin, Ireland 27-31.10.2025 MMFood'25

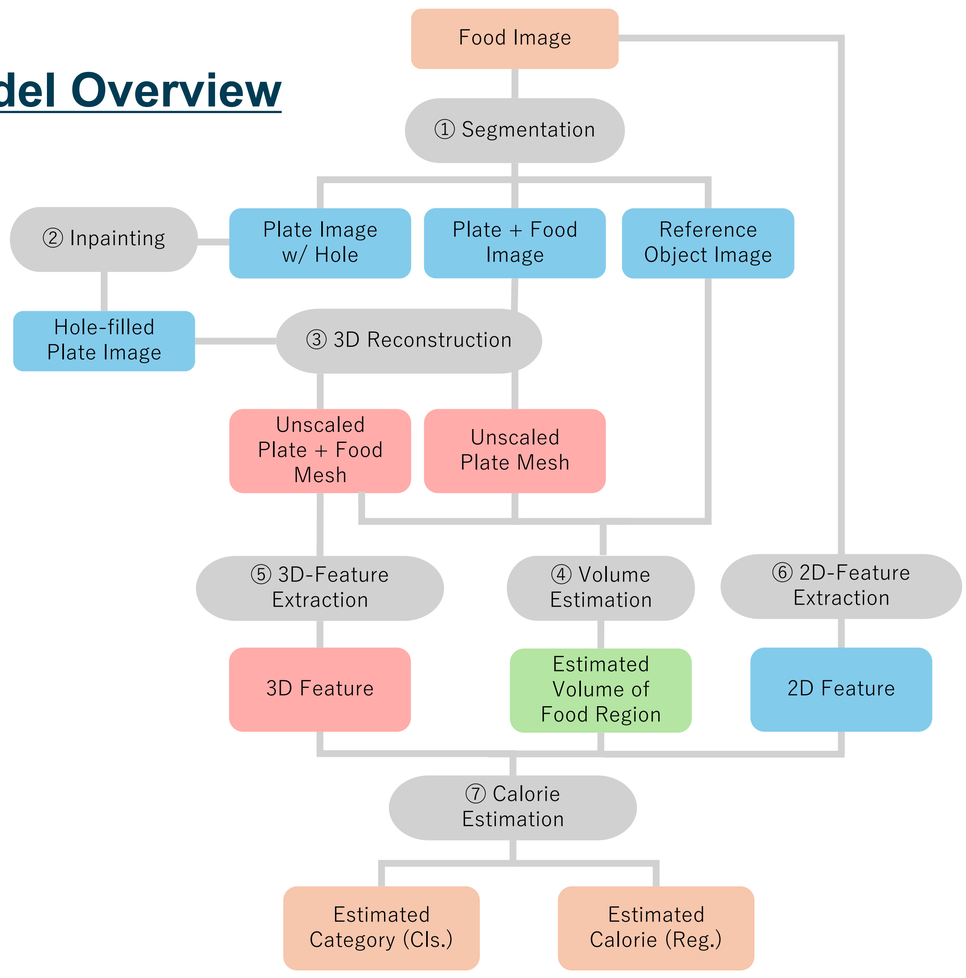
Background

- **Accurately estimating calorie from a single food image** is essential for improving the reliability of dietary records in health management applications.
- Conventional methods estimate calorie by interpreting meals on a 2D plane, **overlooking the 3D structure and volume of the food**.
- We introduce a framework for accurately estimating calorie content from a single food image based on **diffusion-guided 3D reconstruction techniques**.



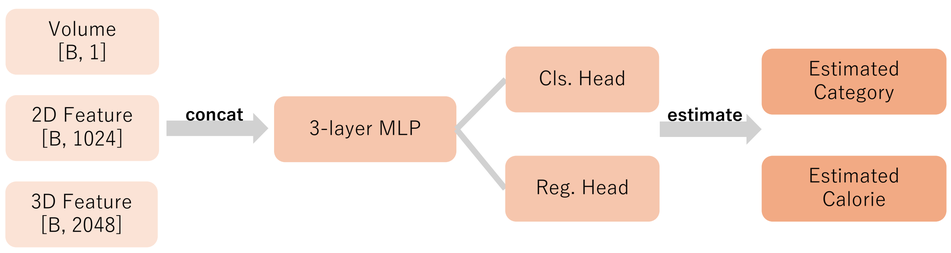
Method

Model Overview



① **Segmentation:** Extract region masks for the food, plate and reference object with Grounded SAM [1] . Based on the masks, obtain three images:
(1) plate (w/o food) (2) plate w/ food (3) reference object
② **Inpainting:** Fill the hole in (1) plate image with fine-tuned SDv2.
③ **3D Reconstruction:** Generate 3D meshes using Wonder3D [2] :
(1) the plate (2) the plate w/ the food
④ **Volume Estimation:** Estimate the actual food volume by taking the volume difference between the two meshes and applying a scale factor derived from the reference object.
⑤ **3D Feature Extraction:** Apply PointGPT [3] to the point cloud obtained from (2) the plate w/ the food mesh to extract 3D features.
⑥ **2D Feature Extraction:** Apply OpenAI ViT-14/L to the original image to extract 2D features.
⑦ **Calorie Estimation:** Volume, 2D, 3D features are concatenated and fed into 2-head MLP to jointly estimate calorie and food category via multi-task learning (figure below)

Calorie Estimation Network



Experiments

Implementation and Dataset

- **MetaFood3D [4] :** A multimodal food dataset containing images, 3D food meshes, masks, volume, and nutrition.
- It consists of 637 food items across 108 categories, split 80:20 for training and evaluation.
- Inpainting with SDv2 occasionally generates incorrect food artifacts when filling plate holes.
  - ➢ We **fine-tuned SDv2 with LoRA on a dataset of 3,000 plate images**, including 1,000 real images collected from public sources and 2,000 synthetic images.

Comparison with Related Approaches

Method	Volume MAE (ml)	Volume MAPE (%)	Calorie MAE (kcal)	Calorie MAPE (%)	Category Accuracy
MFP3D [5]	62.60	**41.43**	**77.98**	**68.05**	-
Ours	**57.53**	269.31	92.74	212.78	0.96

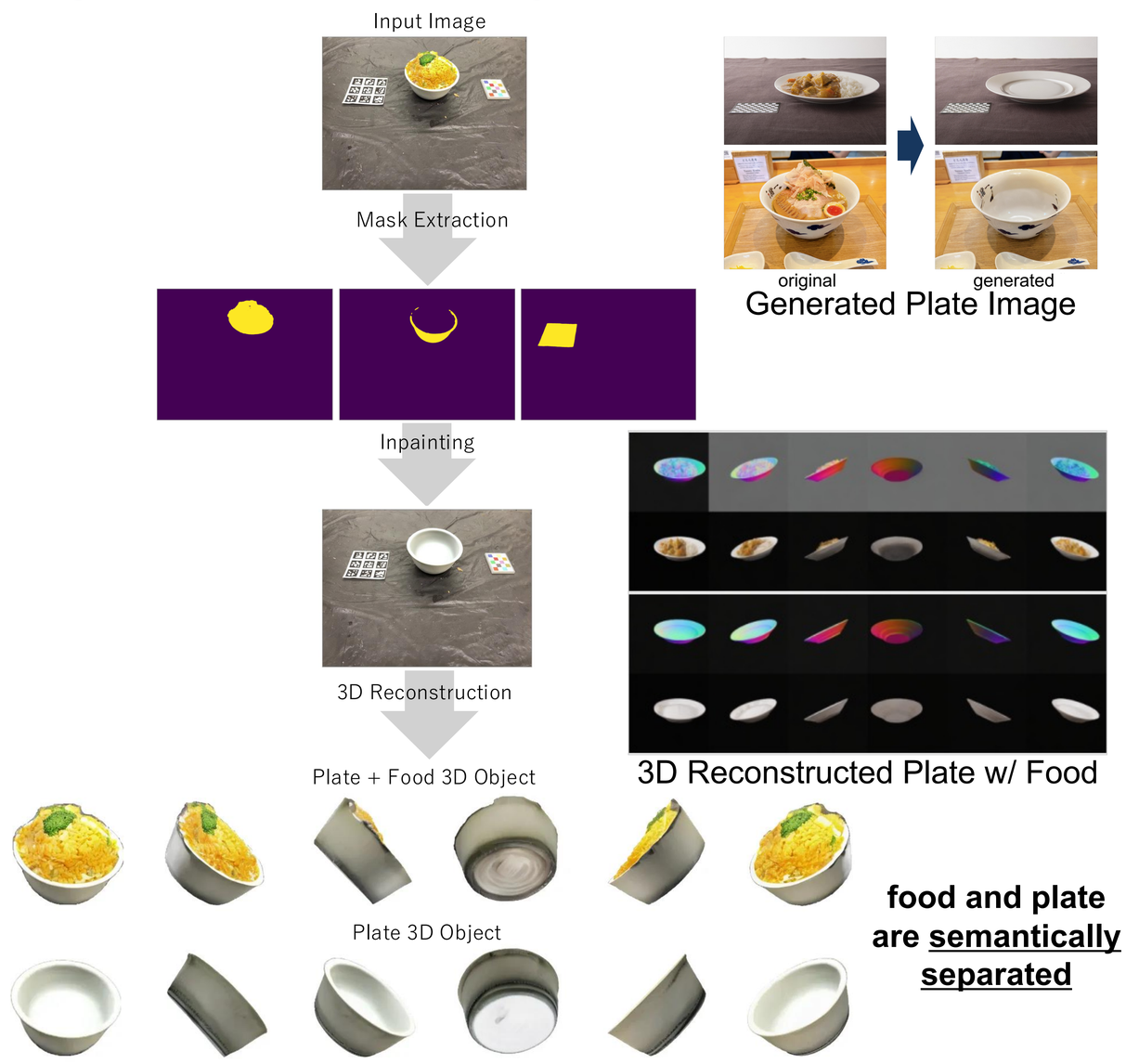
- MAPE was high due to segmentation failures especially on small food items (e.g., nuts, blueberries), leading to large relative errors.

Contribution of the Features

2D	Volume	3D	Calorie MAE (kcal)	Calorie MAPE (%)
✓			133.37	373.23
✓	✓		101.12	254.09
✓	✓	✓	**92.74**	**212.78**

- Incorporating volumetric and 3D features led to more accurate calorie estimation than using only 2D features.

Segmentation / Inpainting / 3D Reconstruction



Conclusion

We introduced a calorie estimation framework accurately capturing 3D structure and volume with 3D reconstruction, achieving more accurate calorie estimation than with 2D features alone.

Limitation and Future Work

- Mitigate segmentation errors and their downstream impact.
- Construct a multimodal food dataset of plated meals for more robust training and evaluation.

[1] Tianhe Ren et al. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks. arXiv preprint arXiv:2401.14159, 2024.
[2] Xiaoxiao Long et al. 2024. Wonder3D: Single Image to 3D using Cross-Domain Diffusion. In Proc. of IEEE Computer Vision and Pattern Recognition (CVPR), 2024.
[3] Guangyan Chen et al. PointGPT: Auto-regressively generative pre-training from point clouds. Advances in Neural Information Processing Systems (NeurIPS), 2023.
[4] Yuhao Chen et al. MetaFood3D: Large 3D Food Object Dataset with Nutrition Values. In Proc. of International Conference on Learning Representations (ICLR), 2024.
[5] Jinge Ma et al. MFP3D: Monocular Food Portion Estimation Leveraging 3D Point Clouds. In Proc. of 27th International Conference on Pattern Recognition (ICPR), 9th International Workshop on Multimedia Assisted Dietary Management, 2024.