VolE-Complete: Enhancing Food 3D Reconstruction and Volume Estimation with Symmetry-Guided Point Cloud Completion

Umair Haroon [1,*], Ahmad AlMughrabi [1], Ricardo Marques [2], and Petia Radeva [1,3]

[1] Matemáticas e Informática, Universitat de Barcelona, Barcelona, Spain
[2] Universitat Pompeu Fabra, Grup de Tecnologies Interactives (GTI), Barcelona, Spain
[3] Institut de Neurociències, Universitat de Barcelona, Barcelona, Spain
Email: umairharoon@ub.edu (U.H.); ahmad.almughrabi@ub.edu (A.A.); ricardo.marques@upf.edu (R.M.); petia.ivanova@ub.edu (P.R.)
*Corresponding author

***Abstract*—Accurate estimation of food volume is essential for personalised nutrition and health initiatives. However, the challenge of obtaining high-quality 3D reconstructions of food items, especially in the presence of occlusions or limited observations, remains considerable. Our prior framework, VolE, provided a robust foundation for mobile-driven 3D food reconstruction but encountered difficulties with entirely unseen or incomplete object parts, thereby limiting model accuracy and volume estimates. We present VolE-Complete, an advanced framework that incorporates a cutting-edge point cloud completion technique for 3D reconstruction and volume estimation of food. By effectively inferring and reconstructing missing geometric details, VolE-Complete yields more comprehensive and precise 3D representations of food. valuations carried out on the challenging FoodKit and MetaFood3D datasets demonstrate a Mean Absolute Percentage Error (MAPE) of 0.2% and substantially improved reconstruction quality. This development facilitates dependable, mobile-based, and depth-free food volume estimation, thereby enhancing dietary assessments and enabling broader applications. The source code is available at: https://github.com/GCVCG/VolE-Complete.**

***Keywords*—volume estimation, 3D reconstruction, point cloud completion, food volume, SymmCompletion**

I. Introduction

Accurate and efficient food volume estimation is essential for various critical applications, including personalized dietary monitoring, clinical nutrition management, and large-scale public health initiatives [1, 2]. Traditionally, manual methods and specialised hardware have been employed, but these approaches often lead to inaccuracies, high costs, and scalability challenges. This highlights the urgent need for automated, objective, and accessible solutions [3]. Recent advancements in computer vision and mobile computing have enabled image-based intelligent dietary assessment, enabling the computational analysis of images to automate tasks such as food segmentation, recognition, and volume estimation [3]. However, achieving precise volumetric measurements from visual data remains complex because it relies on robust 3D information [4].

The field of 3D reconstruction from multi-view images has made significant strides with techniques such as Structure from Motion (SfM) [5], photogrammetry (e.g., Multi-view Stereo), Neural Radiance Fields (NeRF) [6], and 3D Gaussian Splatting (3DGS) [7]. Although these methods excel at generating detailed volumetric representations and 3D shapes, they face limitations like scale ambiguity, which requires external calibration or depth information for accurate real-world measurements [8, 9]. Additionally, many approaches are restricted to fixed environments or specific reference objects, which limits their adaptability to dynamic real-world scenarios [10–13].

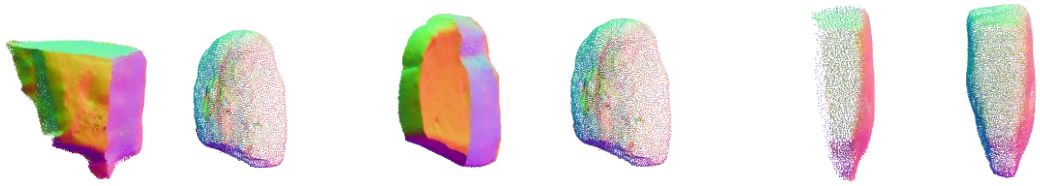
VolE [14] advanced the challenges of 3D reconstruction by utilising AR-capable mobile devices to capture images and camera locations in free motion, enabling reference- and depth-free 3D reconstruction and volume estimation of food items. By integrating an auto-mask generation module (FoodMem [15]) and refining camera locations using COLMAP [5], VolE achieved notable accuracy across multiple datasets. However, a key limitation of mobile-driven reconstruction remains the inability to capture the entirety of an object. Occlusions, limited viewpoints, and sparse feature matching in low-textured or complex food items often result in incomplete point clouds. This compromises the fidelity of the 3D models and can lead to inaccuracies in volume estimates, such as underestimating the volume of a partially reconstructed spherical fruit (Fig. 1). Addressing this challenge is vital for precise food volume measurements across various environments.

To tackle this critical issue, we present VolE-Complete, an advanced pipeline that substantially enhances the capabilities of our original VolE framework [14] by adding a robust point-cloud completion module [16]. VolE-Complete specifically addresses the challenge of

Manuscript received September 1, 2025; revised September 11, 2025; accepted November 17, 2025; published June 12, 2026.

reconstructing unseen or incomplete parts of food items by leveraging a symmetry-guided point cloud completion technique [16]. Our framework intelligently infers and "fills in" missing geometric information, transforming sparse and partial 3D reconstructions into dense, complete, and geometrically consistent models. This enhancement is essential to ensure that subsequent mesh reconstruction and volume estimation steps are based on the most accurate and comprehensive 3D data available. Our contributions are summarized as follows:

- We introduce VolE-Complete, the first-ever food volume estimation framework with an advanced point cloud completion method for reconstructing unseen or incomplete 3D food object parts.
- We conducted extensive experiments on real-world datasets, including FoodKit [14] and MetaFood3D [8], showing that VolE-Complete outperforms existing methods, achieving an impressive improvement in food volume estimation and superior performance in 3D reconstruction.



(a) (b) (c)

Fig. 1. Comparison between incomplete point clouds generated by the initial VolE method (left) and complete point clouds produced by VolE-Complete (right) for three food items from the MTF dataset, highlighting accurate reconstruction of incomplete and unseen 3D structures. (a) Cinnamon bun; (b) French toast; (c) Salmon.

The remainder of this paper is structured as follows: Section II reviews related work, including advancements in 3D reconstruction and food volume estimation. Section III details the proposed VolE-Complete methodology. Section IV presents the experimental setup and results. Finally, Section V concludes the paper and outlines future work.

II. Related Work

Accurate 3D reconstruction and volumetric estimation from visual data are longstanding challenges in computer vision, particularly for dietary assessment and health monitoring. We review 3D reconstruction methods for food items and discuss point cloud completion, which addresses limitations of traditional methods. We start with an overview of 3D Food Computing and its relation to Food Volume Estimation, followed by a focus on point cloud completion techniques and their significance.

A. 3D Food Computing and Volume Estimation

Early methods for estimating food volume often utilised specialised hardware like 3D scanners and depth sensors [2]. While precise in controlled settings, these methods are costly, not portable, and impractical for broad dietary monitoring. Approaches that use fiducial markers to infer scale from 2D images can introduce errors due to incorrect placements [4]. With advancements in computer vision, image-based 3D reconstruction has gained traction. SfM pipelines, such as COLMAP [5], enable the recovery of 3D scene geometry from uncalibrated image sequences. Recent methods, such as NeRF [6] and 3DGS [7], have improved view synthesis and 3D shape estimation, but still struggle with scale ambiguity in the absence of explicit cues [12]. Additionally, they can produce incomplete models due to occlusions and limited viewpoints. VolE [14] addresses these limitations by offering a mobile-driven, reference-free, and depth-free framework. Using AR-capable devices for pose acquisition and an auto-mask module. However, similar to other methods, their point clouds can be incomplete, affecting model fidelity and volume estimation accuracy.

B. Point Cloud Completion

Point cloud completion addresses the widespread issue of incomplete 3D data by inferring and reconstructing missing geometric information from partial scans, resulting in complete and dense point clouds [17]. This task is crucial in domains such as robotics, autonomous driving, and augmented reality, where understanding 3D structures is essential [17–20]. Existing methods generally follow two strategies. The first approach reconstructs the entire point cloud, as in Point Completion Network (PCN) [17], which uses a coarse-to-fine approach followed by refinement. Transformer-based models such as SnowflakeNet [18] and SeedFormer [19] enhance feature perception but may miss fine details when presented with partial inputs. The second strategy focuses on generating only the missing regions, with techniques such as PoinTr [21] and PF-Net [20] ensuring geometric consistency. However, merging new and existing geometries can introduce artefacts. Recent models like GTNet [22] attempt global optimisation but still struggle with capturing fine geometry. A promising direction is to leverage symmetry priors, particularly for objects such as food that exhibit inherent symmetries. Methods such as USSPA [23] and GTNet [22] explore these transformations but often rely on simplistic assumptions, risking loss of local geometric detail. SymmCompletion [16] advances this by combining local and global symmetry guidance for high-fidelity results. It comprises two components: the Local Symmetry Transformation Network (LSTNet), which estimates local symmetry transformations to map geometries into missing areas, and the Symmetry-Guidance Transformer (SGFormer), which refines the point cloud using a dual-path mechanism that integrates symmetry information to preserve structures while filling gaps. This

approach achieves state-of-the-art performance, effectively handling complex geometries, especially in food items with strong local symmetry.

C. VolE-Complete: Bridging Reconstruction and Completion

VolE-Complete innovatively integrates the strengths of the VolE framework [14] with advanced point-cloud completion. While VolE generates scaled 3D point clouds from mobile captures, it faces limitations due to occlusions and narrow viewing angles, resulting in partial reconstructions. To address this, VolE-Complete incorporates a SymmCompletion [16] model specifically trained after the point cloud masking stage. This novel leveraging enables the inference and reconstruction of unseen parts from the accurately scaled, albeit incomplete, point cloud. This synergy ensures that subsequent mesh reconstruction and volume estimation are performed on a complete, high-fidelity 3D model, significantly improving volume measurement accuracy. By leveraging SymmCompletion's ability to generate geometry-consistent missing regions based on learned symmetry, VolE-Complete effectively infers full object geometry, marking a significant advance in robust mobile-driven food volume estimation in unconstrained environments. Our integration presents significant challenges because SymmCompletion must accurately infer missing geometry without compromising the precise scale and alignment established by VolE. This necessitates meticulous coordination to maintain geometric consistency while ensuring metric accuracy. Furthermore, the completion model must function effectively on noisy, partial point clouds captured by mobile devices, which often lack reliable contextual information.

III. Our Proposal: VolE-Complete

Our framework is designed to overcome the inherent limitations of reconstructing entirely unseen or incomplete parts of 3D food objects, a challenge that significantly impacts the fidelity of 3D models and the precision of subsequent volume estimations. VolE-Complete introduces a critical enhancement: a sophisticated point cloud completion module. This section outlines the overall architecture of VolE-Complete, elaborating on each stage with a particular emphasis on the novel integration of symmetry-guided point cloud completion.

A. Overall Architecture

VolE-Complete enhances the original VolE pipeline by adding a point cloud completion stage. As illustrated in Fig. 2, the framework consists of four stages: (a) Data Acquisition, (b) Parameter Extraction, (c) 3D Mesh Reconstruction, and (d) Volume Estimation. The key innovation lies in the 3D Mesh Reconstruction stage, where partial point clouds undergo a novel completion process to enhance 3D representation and improve volume estimation.

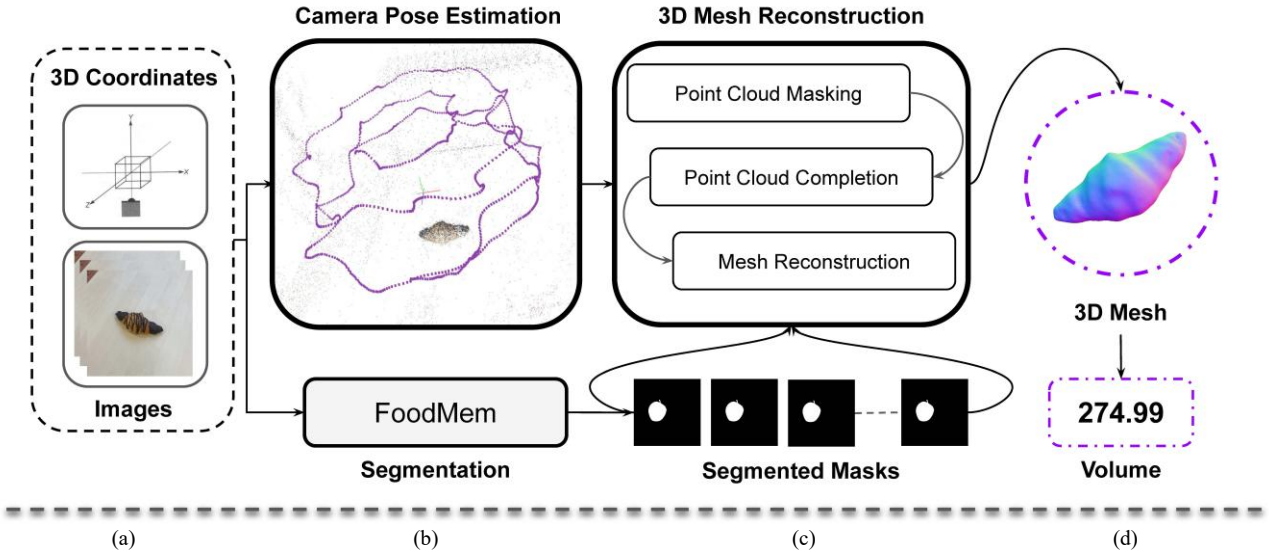


Fig. 2. The VolE-Complete framework enables accurate food 3D reconstruction and volume estimation through a systematic pipeline: (a) Data Acquisition for images and camera coordinates, (b) Parameter Extraction using Structure from Motion and FoodMem for pose estimation and segmentation, (c) 3D Mesh Reconstruction that combines point cloud masking with SymmCompletion and 3D mesh Reconstruction, and (d) Volume Estimation for calculating the volume of the 3D mesh.

B. Data Acquisition

In this stage, we gather input image sequences, denoted as $(\mathcal{I} = \{I_i\}_{i=1}^{N_I})$, along with their corresponding real-time camera locations, represented as $(\mathcal{C} = \{C_i\}_{i=1}^{N_I})$. This free-motion acquisition method is essential to the reference- and depth-free behaviour, ensuring that the captured data is inherently scaled for real-world measurements.

Parameter Extraction: Alongside data acquisition, essential parameters are extracted to facilitate 3D reconstruction: (a) Camera Pose Estimation: The initial camera location $\mathcal{C}$ obtained from the input is refined using robust SfM techniques, specifically COLMAP [5]. This process involves several steps, including feature extraction

($\mathcal{F}_i$), feature matching ($\mathcal{M}_{a\&}$), and geometric verification ($\mathcal{G}_{a\&}$). The result is a set of highly accurate and globally consistent camera poses, which are essential for precise metric 3D reconstruction from multiple views. (b) Food Video Segmentation: A specialized segmentation model, such as FoodMem [14], processes the input images to create precise binary masks $\mathcal{S} = \{S_t\}_{t=1}^{T}$. These masks are crucial for isolating the food item of interest from the background, ensuring that only relevant points are considered for 3D reconstruction.

C. 3D Mesh Reconstruction

This critical stage involves synthesising processed data into a complete and accurate 3D mesh using VolE-Complete. It incorporates point-cloud masking, a novel point-cloud completion module, and final mesh reconstruction, addressing the challenges posed by partial data acquired with mobile devices.

1) Point cloud masking

After refining the camera poses and segmenting the masks, we generate an initial 3D point cloud of the scene using Multi-View Stereo (MVS) techniques. Point cloud masking [14] is then applied to isolate the target food object. The complete point cloud of the scene $P = \{(x_i, y_i, z_i)\}_{i=1}^{N_P}$ is projected onto the segmented masks $\mathcal{S}$ using the camera intrinsic matrix K and the estimated poses C. We retain only the points that consistently fall within the segmented mask across multiple views, resulting in a segmented object point cloud $\mathcal{P} = \bigcap_{j=1}^{N_I} M_j$. Although this method effectively isolates the food item and helps eliminate noisy regions and artefacts/outliers. However, it preserves the object's real-world scale. As a result, the point cloud $\mathcal{P}$ is often incomplete and may exhibit unobserved or sparse areas due to self-occlusions, limited viewpoints, or textureless surfaces. This inherent incompleteness necessitates a subsequent completion step.

2) Point cloud completion with SymmCompletion

The core innovation of our framework lies in addressing the incompleteness of $\mathcal{P}$ by incorporating a high-fidelity point cloud completion module. We utilise Symm-Completion, a cutting-edge method chosen for its ability to generate complete point clouds with high fidelity and consistency, leveraging symmetry guidance, which is particularly effective for food items that often exhibit inherent or approximate symmetries. While most effective for geometrically regular foods, the Local Symmetry Transformation Network (LSTNet) can identify locally symmetric regions even within irregular or asymmetric foods to support completion. The SymmCompletion process consists of two main components: LSTNet: This network takes an incomplete point cloud $\mathcal{P}$ as input, extracting key geometries and features to estimate local symmetry transformations (affine matrices A and translation matrices T). These transformations are applied to generate missing parts P_m, leading to an initial complete point cloud P_{init}. SGFormer: This component refines P_{init} into a final high-fidelity point cloud P_{fine}. It utilises symmetric guidance from partial-missing pairs via cross-attention and self-attention layers, preserving structures while filling gaps.

To further enhance the quality of completions, our framework includes a realistic partial-point-cloud preprocessing strategy. Key preprocessing steps involve centring, scaling, PCA-based orientation alignment, and uniform sampling, which normalise incomplete point clouds for effective learning. To preserve scale consistency, transformation parameters are tracked during preprocessing and applied in post-processing to restore the completed point cloud to its original scale and orientation. This ensures volumetric accuracy and prevents bias introduced by hallucinated geometry. Additionally, implementing Statistical Outlier Removal after point cloud completion eliminates isolated and outlier points, further refining the output and enhancing data quality. This food-specific, scale-consistent, symmetry-guided completion pipeline distinctly improves upon generic point cloud completion methods and is critical for precise food volume estimation, culminating in the generation of a dense and geometrically refined point cloud P_{fine}.

3) 3D mesh reconstruction

The completed, high-fidelity point cloud P_{fine} serves as input for mesh reconstruction. This stage transforms P_{fine} into a surface mesh $\mathcal{M}$ through algorithms such as Delaunay triangulation [5], followed by graph-cut optimization to determine internal and external regions. Finally, we use the marching cubes algorithm to extract the mesh surface [14]. The completeness and accuracy of P_{fine} are crucial, as they directly influence the accuracy and seamless representation of the object's full geometry, significantly improving upon meshes derived from incomplete point clouds.

D. Volume Estimation

The final stage of our framework focuses on deriving precise volumetric measurements of the food item from its generated 3D mesh. This process greatly benefits from the comprehensive, high-fidelity mesh produced by the preceding completion and reconstruction stages. The exact volume of the food object is computed directly from the refined, complete 3D mesh $\widehat{\mathcal{M}}$.

IV. Experimental Results

We present a thorough evaluation of our framework, highlighting its effectiveness in producing high-fidelity 3D reconstructions and accurately estimating the volume of food items. We describe our implementation settings, evaluation protocols, and datasets. Additionally, we provide both quantitative and qualitative comparisons with state-of-the-art methods. Finally, we conclude with a discussion of our findings and the current limitations.

A. Implementation Settings

Our experiments are conducted on a system with an NVIDIA GeForce RTX 3090 GPU (24 GB). Our 3D reconstruction pipeline configuration sets the point cloud masking "max resolution" to 512. Mesh reconstruction was set to "close holes" at 50 and a "smooth" factor of 5

for effective surface regularisation. The point cloud completion module utilised the SymmCompletion network, trained with the Adam optimiser (learning rate of 1×10^{-4}, batch size of 16) using a preprocessing strategy outlined in Section III. The training involved 26 partial views from 30 MTF dataset objects and converged after approximately 500 epochs. Benchmarking shows that the completion step adds 0 m 42.307 s to the baseline VolE (0 m 84.614 s with preprocessing and postprocessing). For 1005 images at 720×960 resolution, VolE requires 9 min; VolE-Complete requires 10 mins, with memory usage under 10% higher. The inclusion of the SymmCompletion module results in an approximate 11% increase in runtime and less than 10% rise in GPU memory consumption compared to the baseline pipeline. Scalability tests demonstrate efficient handling of up to 1000 images while maintaining manageable computational resource demands, making it suitable for offline batch processing. To enhance the effectiveness of our framework for various food items, our point cloud completion model is trained on a subset of the MetaFood3D dataset [8], which includes 30 diverse food objects and 26 distinct synthetic partial views. This training exposed the model to diverse incompleteness patterns, equipping it with robust completion capabilities.

B. Evaluation Protocol

To evaluate our framework, we use two key metrics: MAPE for volume accuracy and Chamfer Distance (CD) for 3D reconstruction fidelity. MAPE measures relative volume error as $\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{V_{\text{est},i}-V_{\text{gt},i}}{V_{\text{gt},i}}\right| \times 100\%$ where $V_{\text{est},i}$ and $V_{\text{gt},i}$ are the estimated and ground truth volumes. Lower MAPE indicates higher accuracy. CD quantifies the geometric difference between reconstructed and ground truth point clouds: $D_{\text{CD}}(S_1, S_2) = \frac{1}{|S_1|}\sum_{x\in S_1} min_{y\in S_2}\left|x-y\right|^2 + \frac{1}{|S_2|}\sum_{y\in S_2} min_{x\in S_1}\left|x-y\right|^2$, where S_1 and S_2 are the reconstructed and ground truth point clouds, respectively. Lower CD implies better geometric fidelity.

C. Datasets

Our evaluation includes two distinct datasets to thoroughly assess the performance of VolE-Complete, focusing on its generalizability and accuracy across a variety of food items.

The Foodkit dataset [14] is a benchmark for food volume estimation in mobile, free-motion settings. It includes 21 diverse food items, with ground-truth volumes (via water displacement) and mass measurements. Data was collected using ARKit-based mobile apps, capturing 360° videos with dense images and accurate 3D camera poses. Notably, FoodKit was not used to train the Symm-Completion module, making it suitable for evaluating our framework's generalizability to unseen foods.

The MTF dataset [8] contains 637 3D food models across 108 categories. We used 30 items to train the SymmCompletion module to learn shape priors. For evaluation, we used the MTF 2024 CVPR challenge subset, comprising 20 scenes: easy (8 scenes, 200 images), medium (7 scenes, 30 images), and hard (single image). Each image includes food masks and depth. We focused on easy and medium scenes that support multi-view 3D reconstruction and used MTF's scaling board for accurate scene scaling.

D. Comparative Analysis

We perform a detailed comparative analysis of our framework against our VolE [14] and other state-of-the-art 3D reconstruction and volume estimation baselines, including methods based on COLMAP [5] and various approaches from the MetaFood3D challenge [8].

1) Qualitative results

Qualitative results clearly demonstrate the superior performance of VolE-Complete. It consistently outperforms other methods in detail and geometric accuracy in 3D reconstructions. Compared to VolE [14] on the Food-Kit dataset, it shows comparable performance with some improvements, effectively handling variations in shape, size, and surface texture, as shown in Fig. 3. A comparison with VolETA [12] and VolE [14] on the MTF dataset further highlights VolE-Complete's superior ability to capture intricate geometries, leading to more precise 3D representations, illustrated in Fig. 4.

2) Quantitative results

Our analysis evaluates our framework against VolE [14] and other state-of-the-art methods using the FoodKit and MTF datasets, as shown in Tables I and II. The Food-Kit dataset results demonstrate that our framework outperforms VolE in volume estimation, achieving an MAPE of 0.72 compared to 0.88 for VolE, resulting in mean accuracies of 99.28% and 99.12%, respectively, which showcases our framework's strong generalizability and precision. Additionally, our framework excels on the MTF dataset, with the lowest MAPE of 2.69%, outperforming VolE (3.08%) and other benchmarks such as VolETA (7.84%) and ININ (14.47%). Moreover, our framework has a mean CD of 0.0042, striking a good balance between geometric completeness and volume estimation, although FoodR (0.0028) and ININ (0.0032) have slightly lower CD values.

To validate the improvements, Table III presents the paired t-test analysis comparing estimation errors and Chamfer Distances between VolE and our proposed framework on the FoodKit and MTF datasets. Our framework demonstrated lower mean estimation errors and Chamfer Distances, with MAPE reduced from 0.88 to 0.72 on FoodKit and from 3.08 to 2.69 on MTF. The mean CD decreased from 0.0044 to 0.0035 on MTF. While these reductions did not reach statistical significance (p = 0.370 for FoodKit MAPE and p = 0.300 for MTF MAPE), the reduction for Chamfer Distance on MTF approached significance (p = 0.057), suggesting improved 3D geometric fidelity. Overall, these results demonstrate that our VolE-Complete framework offers significant improvements in volume estimation accuracy and geometric reconstruction, reinforcing its superiority over existing methods.

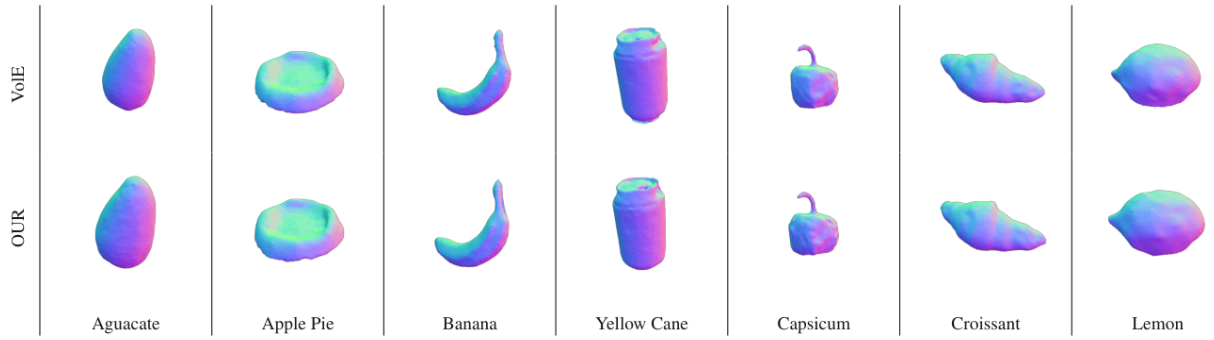


Fig. 3. Qualitative comparison of 3D reconstructions on the FoodKit dataset with VolE [14].

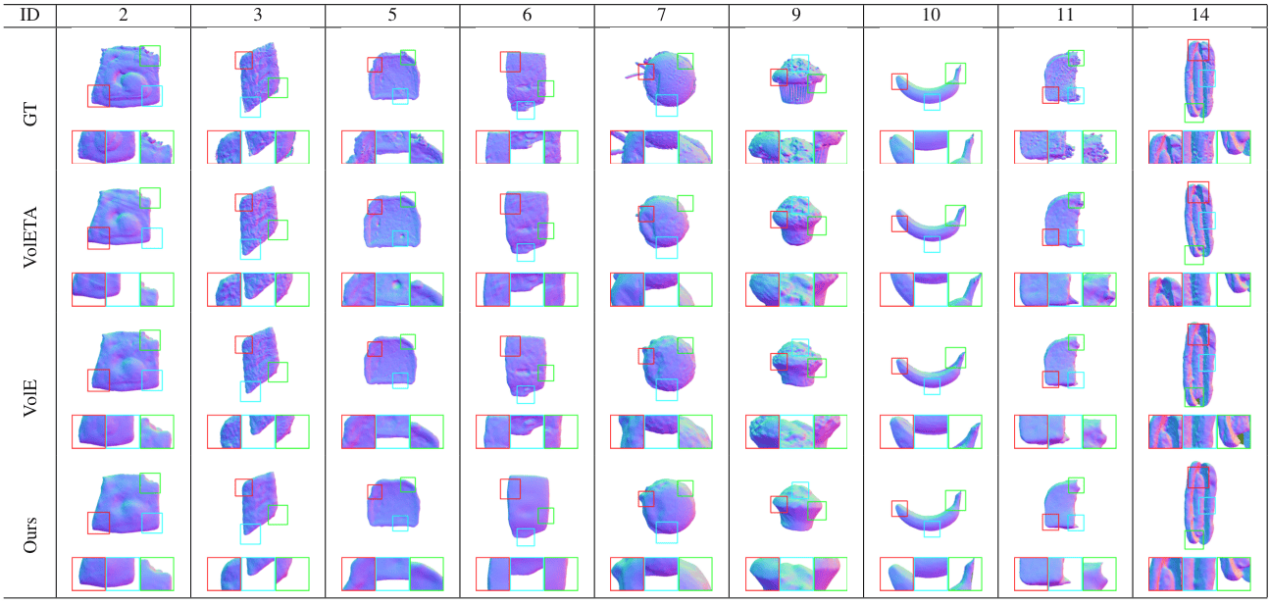


Fig. 4. Our framework 3D reconstruction Visual Results on the MTF dataset in comparison with the Ground Truth (GT), VolETA [12], and VolE [14] 3D reconstruction.

TABLE I. A COMPARISON OF VOLUME ESTIMATION PERFORMANCE BETWEEN VOLE [14] AND OURS USING THE FOODKIT DATASET, INCLUDING ESTIMATED VOLUMES, ABSOLUTE ERRORS, AND ACCURACIES

Items	Images	GT (±5)	Est. Volume		Est. Error ↓		Accuracy ↑	
			VolE	Ours	VolE	Ours	VolE	Ours
Apple	1005	175	176.68	**175.41**	0.96	**0.23**	99.04	**99.77**
Orange	1001	200	201.06	**200.75**	0.53	**0.37**	**99.47**	**99.63**
Aguacate	1078	85	83.11	**84.61**	2.23	**0.46**	97.77	**99.54**
Lemon	887	140	**134.74**	133.34	**3.76**	4.76	**96.24**	95.24
Donut	780	245	242.24	**243.54**	1.12	**0.59**	98.88	**99.41**
Durum	1006	200	200.79	**199.69**	0.40	**0.15**	99.60	**99.85**
Pear	849	170	168.13	**169.93**	1.10	**0.04**	98.90	**99.96**
Choc. Cake	781	195	195.38	**194.68**	0.19	**0.17**	99.81	**99.83**
Choc. Croissant	1122	275	**274.99**	275.09	**0.00**	0.03	**100.00**	99.97
Samosa	848	145	**144.10**	143.60	**0.62**	0.96	**99.38**	99.04
Apple Pie	1201	135	135.52	**134.62**	0.39	**0.28**	99.61	**99.72**
Choc. Bomb	1111	200	197.65	**198.95**	1.17	**0.52**	98.83	**99.48**
Empanadilla	926	95	94.86	**94.96**	0.15	**0.04**	99.85	**99.96**
Falafel	929	48	**47.58**	46.88	**0.87**	2.33	**99.13**	97.67
French Bread	1139	163	**162.49**	160.99	**0.32**	1.24	**99.68**	98.76
Paxoco Mini	911	150	148.08	**149.48**	1.28	**0.35**	98.72	**99.65**
Napolitanas	1071	233	**232.79**	231.49	**0.09**	0.65	**99.91**	99.35
Capsicum	881	320	**318.64**	318.44	**0.42**	0.49	**99.58**	99.51
Choc. Panettone	1209	293	290.79	**291.29**	0.75	**0.58**	99.25	**99.42**
Banana	1156	150	153.03	**151.03**	2.02	**0.69**	97.98	**99.31**
Yellow Cane	715	350	**350.07**	349.37	**0.02**	0.18	**99.98**	99.82
Mean	-	-	-	-	0.88	**0.72**	99.12	**99.28**

TABLE II. COMPARISON OF VOLUME ESTIMATION AND 3D RECONSTRUCTION METHODS ON THE MTF DATASET

ID	Predicted Volume					GT	Error Percentage ↓					Chamfer Distance ↓				
	VolETA	ININ	FoodR.	VolE	Our		VolETA	ININ	FoodR.	VolE	Our	VolETA	ININ	FoodR.	VolE	Our
1	40.06	37.65	44.51	37.47	37.86	38.53	3.97	2.28	15.52	2.74	1.74	0.0016	0.0020	0.0011	0.0028	0.0022
2	216.90	325.44	321.26	275.38	278.05	280.36	22.64	16.08	14.59	1.78	0.82	0.0071	0.0036	0.0031	0.0022	0.0021
3	278.86	473.40	336.11	268.93	262.04	249.65	11.70	89.63	34.63	7.72	4.96	0.0137	0.0049	0.0053	0.0068	0.0060
4	279.02	294.32	347.54	277.56	281.76	295.13	5.46	0.27	17.76	5.95	4.53	0.0020	0.0038	0.0015	0.0046	0.0033
5	395.76	353.66	389.28	394.04	390.92	392.58	0.81	9.91	0.84	0.37	0.42	0.0137	0.0020	0.0040	0.0021	0.0020
6	205.17	237.88	197.82	215.21	214.23	218.31	6.02	8.96	9.39	1.42	1.87	0.0067	0.0038	0.0025	0.0039	0.0031
7	372.93	361.49	412.52	370.69	366.00	368.77	1.13	1.97	11.86	0.52	0.75	0.0047	0.0048	0.0025	0.0036	0.0034
8	186.62	172.32	181.21	176.43	172.56	173.13	7.79	0.47	4.67	1.91	0.33	0.0030	0.0019	0.0010	0.0012	0.0015
9	224.08	253.01	233.79	233.95	230.74	232.74	3.72	8.71	0.45	0.52	0.86	0.0039	0.0029	0.0033	0.0029	0.0026
10	153.76	157.58	160.06	159.20	154.80	163.23	5.80	3.46	1.94	2.47	5.17	0.0027	0.0034	0.0019	0.0118	0.0061
11	80.40	76.46	86.00	82.75	83.00	85.18	5.61	10.24	0.96	2.85	2.56	0.0034	0.0015	0.0015	0.0021	0.0020
12	363.99	246.60	334.70	297.09	298.58	308.28	18.07	20.01	8.57	3.63	3.15	0.0052	0.0026	0.0041	0.0055	0.0051
13	535.44	495.10	517.75	541.58	543.58	589.82	9.22	16.06	12.22	8.18	7.84	0.0043	0.0044	0.0046	0.0082	0.0065
-	7.84	14.47	10.26	3.08	2.69	S.D. ↓	6.36	23.47	9.48	2.63	2.40	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	0.0720	0.0416	0.0364	0.0576	0.0459
-	-	-	-	-	-	-	-	-	-	-	-	0.0055	-	0.0028	0.0044	0.0035

TABLE III. PAIRED T-TEST RESULTS COMPARING ESTIMATION ERROR (MAPE) AND CHAMFER DISTANCE (CD) BETWEEN BASELINE VOLE AND OUR FRAMEWORK ON FOODKIT AND MTF DATASETS

Dataset	Metric	VolE (Mean ± SD)	Ours (Mean ± SD)	*t*-value	df	*p*-value
FoodKit	MAPE	0.88 ± 0.90	0.72 ± 1.06	0.896	20	0.370
MTF	MAPE	3.08 ± 2.63	2.69 ± 2.40	1.041	12	0.300
	CD	0.0044 ± 0.0030	0.0035 ± 0.0018	2.020	12	0.057

E. Discussions

The experimental results highlight the significant advantages of our framework over previous methods, particularly in addressing incomplete 3D reconstructions. By integrating the SymmCompletion module, it effectively resolves challenges faced in 3D scanning, such as occlusions and limited viewpoints. Our approach intelligently infers and fills in missing geometries, leading to more complete and geometrically accurate 3D models, as reflected in improvements in CD. This enhanced accuracy results in 0.2% MAPE in volume estimation, which is vital for nutrition applications where small inaccuracies can have major implications. The strong generalizability demonstrated across both the MTF (i.e., unseen objects) and FoodKit datasets underscores the robustness of our integration of the SymmCompletion model. The effective preprocessing strategy, exposing the model to diverse partial views, facilitated this generalisation, allowing our framework to operate reliably across various food items without needing object-specific retraining. Our findings indicate that utilizing learned shape priors for point cloud completion is an effective strategy for improving 3D reconstruction in real-world scenarios. Note that our benchmarks do not include amorphous, transparent, or mixed-dish foods; evaluating these categories is reserved for future work to further delineate framework boundaries and enhance adaptability.

F. Limitations

Despite its strengths, our framework still has some limitations. While the LSTNet can exploit local symmetry in moderately irregular foods, performance naturally declines for highly amorphous or globally asymmetric shapes outside our learned priors. Both FoodKit and MetaFood3D contain only solid, opaque items; amorphous foods, transparent foods, and mixed dishes are not represented. Highly symmetric foods achieve under 1% MAPE, while moderately asymmetric items show 2–5% MAPE. The pipeline also assumes stable capture conditions; issues such as motion blur or reflective surfaces can degrade SfM performance. The SymmCompletion module increases approximately 1 min of processing time per item, which includes preprocessing and postprocessing, resulting in about an 11% increase in batch runtime. While this may limit real-time mobile applications, it remains practical for offline and batch processing. Additionally, the current system is not designed to handle complex real-world scenarios such as mixed food portions, utensil occlusions, or partially eaten items, which can introduce noise and potentially misleading symmetry cues to the completion model. The SymmCompletion framework incorporates local symmetry estimation that enables partial adaptation when global symmetry cues are weak or misleading by emphasising local geometric features; however, completion quality degrades in the absence of consistent symmetry, representing an area for future enhancement through semantic or shape-based priors integration.

V. CONCLUSIONS

In this paper, we introduced VolE-Complete, an advanced framework for improving food volume estimation from mobile captured data. Utilising a specialised SymmCompletion model, we tackled the issue of incomplete 3D reconstructions, converting partial point clouds into detailed and geometrically consistent models. Evaluations of various food objects demonstrate the versatility and reliability of our proposal, making it a promising tool for dietary assessment and public health. Future work will focus on optimising the computational pipeline for real-time use and expanding the framework by integrating semantic priors and shape models. This

enhancement will combine semantic segmentation with symmetry guidance to better handle irregular and asymmetric food shapes, while also enabling lightweight and scalable deployment for mobile-based dietary assessments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

U.H. developed the code, built the framework, conducted experiments across multiple pipelines, and wrote the manuscript. A.A. contributed innovative ideas, provided critical feedback, and assisted with writing. R.M. and P.R. supervised the overall research process, offered guidance throughout the study, and performed the final review to enhance the paper's quality. All authors reviewed and approved the final version.

FUNDING

This work was partially funded by the EU project MUSAE (No. 01070421), 2021-SGR-01094 (AGAUR), Icrea Academia'2022 (Generalitat de Catalunya), Robo STEAM (2022-1-BG01-KA220-VET000089434, Erasmus+ EU), DeepSense (ACE053/22/000029, ACCIO), CERCA Programme/Generalitat de Catalunya, and Grants PID2022141566NB-I00 (IDEATE), PDC2022-133642-I00 (DeepFoodVol), and CNS2022-135480 (A-BMC) funded by MICIU/AEI/10.115039/501100 011033, by FEDER (UE), and by European Union NextGenerationEU/PRTR. A. AlMughrabi acknowledges the support of FPI Becas, MICINN, Spain. U. Haroon acknowledges the support of FI-SDUR Becas, MICINN, Spain.

REFERENCES

[1] A. Rouhafzay, G. Rouhafzay, and J. Jbilou, "Image-based food monitoring and dietary management for patients living with diabetes: A scoping review of calorie counting applications," *Frontiers in Nutrition*, vol. 12, 1501946, 2025.

[2] W. Jia, B. Li, Q. Xu *et al.*, "Image-based volume estimation for food in a bowl," *Journal of Food Engineering*, vol. 372, 111943, 2024.

[3] A. Phalle and D. Gokhale, "Navigating next-gen nutrition care using artificial intelligence-assisted dietary assessment tools—A scoping review of potential applications," *Frontiers in Nutrition*, vol. 12, 1518466, 2025.

[4] J. Dehais, M. Anthimopoulos, S. Shevchik, and S. Mougiakakou, "Two-view 3D reconstruction for food volume estimation," *IEEE Trans. Multimedia*, vol. 19, no. 5, 2016.

[5] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. CVPR*, 2016, pp. 4104–4113.

[6] B. Mildenhall, P. P. Srinivasan *et al.*, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[7] B. Kerbl, G. Kopanas, T. Leimkuhler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, 139, 2023.

[8] J. He, Y. Chen, G. Vinod *et al.*, "Metafood CVPR 2024 challenge on physically informed 3D food reconstruction: Methods and results," arXiv preprint, arXiv:2407.09285, 2024.

[9] U. Haroon, A. AlMughrabi, R. Marques, and P. Radeva, "MVSBoost: An efficient point cloud-based 3D reconstruction," arXiv preprint, arXiv:2406.13515, 2024.

[10] A. AlMughrabi, U. Haroon, R. Marques, and P. Radeva, "VolTex: Food volume estimation using text-guided segmentation and neural surface reconstruction," in *Proc. the Computer Vision and Pattern Recognition Conference*, 2025, pp. 450–457.

[11] U. Haroon, A. AlMughrabi, R. Marques, and P. Radeva, "Vole++: A text-guided point-cloud framework for food 3D reconstruction and volume estimation," in *Proc. International Conference on Computer Analysis of Images and Patterns*, 2025, pp. 386–397.

[12] A. AlMughrabi, U. Haroon, R. Marques, and P. Radeva, "Voleta: One-and few-shot food volume estimation," arXiv preprint, arXiv:2407.01717, 2024.

[13] A. AlMughrabi, U. Haroon, R. Marques, and P. Radeva, "Pre-NeRF 360: Enriching unbounded appearances for neural radiance fields," arXiv preprint, arXiv:2303.12234, 2023.

[14] U. Haroon, A. AlMughrabi, T. Zoumpekas, R. Marques, and P. Radeva, "VolE: A point-cloud framework for food 3D reconstruction and volume estimation," arXiv preprint, arXiv:2505.10205, 2025.

[15] A. AlMughrabi, A. Gal´an, R. Marques, and P. Radeva, "FoodMem: Near real-time and precise food video segmentation," arXiv preprint, arXiv:2407.12121, 2024.

[16] H. Yan, Z. Li, K. Luo, L. Lu, and P. Tan, "SymmCompletion: High-fidelity and high-consistency point cloud completion with symmetry guidance," in *Proc. the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 9094–9102.

[17] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "PCN: Point completion network," in *Proc. 2018 International Conference on 3D Vision (3DV)*, 2018, pp. 728–737.

[18] P. Xiang, X. Wen, Y.-S. Liu, Y.-P. Cao, P. Wan, W. Zheng, and Z. Han, "SnowflakeNet: Point cloud completion by snowflake point deconvolution with skip-transformer," in *Proc. the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5499–5509.

[19] H. Zhou, Y. Cao, W. Chu, J. Zhu, T. Lu, Y. Tai, and C. Wang, "SeedFormer: Patch seeds based point cloud completion with upsample transformer," in *Proc. European Conference on Computer Vision*, 2022, pp. 416–432.

[20] Z. Huang, Y. Yu, J. Xu, F. Ni, and X. Le, "PF-Net: Point fractal network for 3D point cloud completion," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7662–7670.

[21] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou, "PointR: Diverse point cloud completion with geometry-aware transformers," in *Proc. the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12498–12507.

[22] S. Zhang, X. Liu, H. Xie, L. Nie, H. Zhou, D. Tao, and X. Li, "Learning geometric transformation for point cloud completion," *International Journal of Computer Vision*, vol. 131, no. 9, pp. 2425–2445, 2023.

[23] C. Ma, Y. Chen, P. Guo, J. Guo, C. Wang, and Y. Guo, "Symmetric shape-preserving autoencoder for unsupervised real scene point cloud completion," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13560–13569, 2023.