IS3-134

Real-CalAgent: Agentic Calorie Estimation from Single Real-World Food Images

電気通信大学　LONGHAO SUN　YITIAN MA　KEIJI YANAI

Introduction

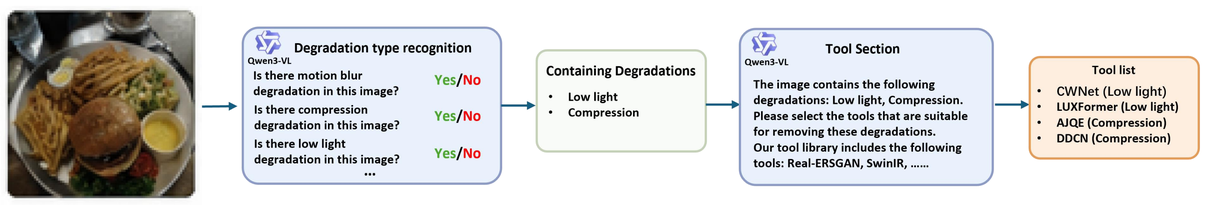
Problem: Existing single-image calorie estimation methods usually assume clean inputs and are vulnerable to blur, low light, and compression artifacts.

Objective: To robustly estimate calories from a single real-world food image with unknown and mixed degradations.

Method

1. Greedy iterative CoT image restoration

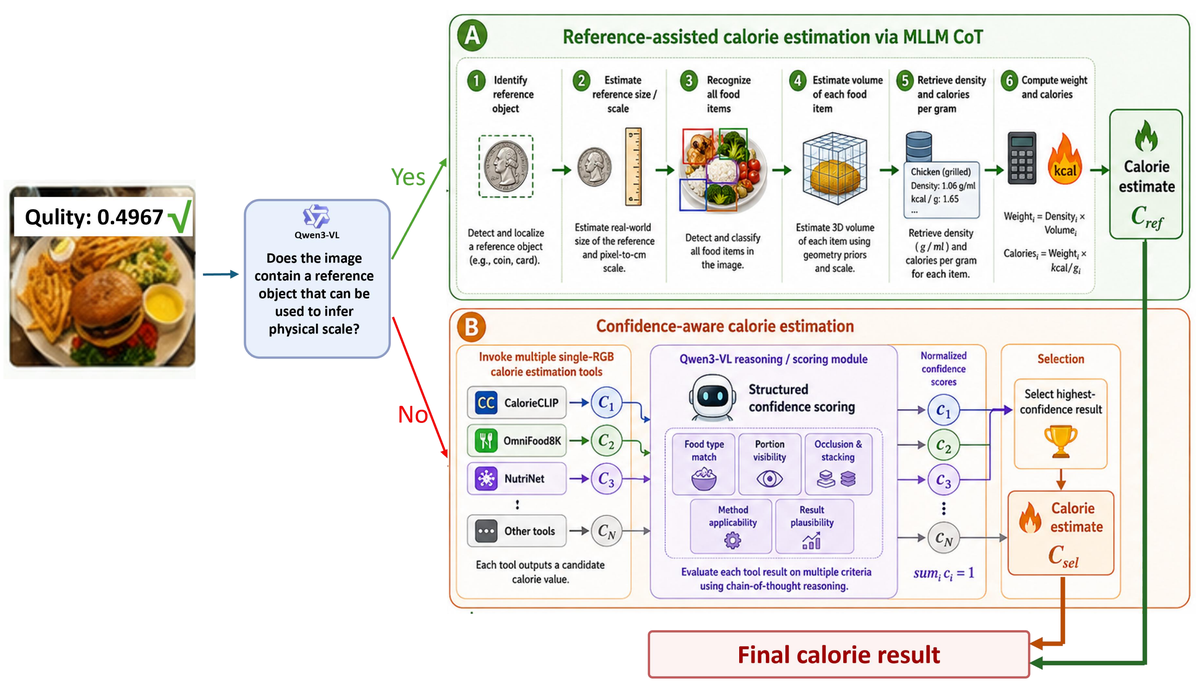
We use CoT to guide MLLM in identifying degradation types.



After the degradation type recognition is completed, call tool-centric greedy recursive image restoration method to restore the image



2. CoT-based Calorie Estimation via MLLM



Determine whether there is a reference object in the image through MLLM, and then call the corresponding method to calculate calories.

Experiments

1. Quantitative results

Compared with other calorie estimation methods, using PMAE matric. Our method achieve the best performance.

Method	OmniFood8K [1]	Deg-OmniFood8K	Nutrition5k	Deg-Nutrition5k
CalorieCLIP	10.9	18.7	15.6	21.6
Swin-Nutrition	11.7	22.1	16.2	25.9
DPF-Nutrition	12.5	23.6	14.7	23.1
FoodLMM	24.2	25.8	26.6	28.5
OmniFood	8.8	21.9	14.1	24.2
Real-CalAgent (Ours)	**8.7**	**10.6**	**13.2**	**14.9**

Comparison with other image restoration methods, our method also achieve the best performance.

Method	PSNR↑	SSIM↑	LPIPS↓	CLIPIQA↑
Q-Agent	26.91	0.824	**0.181**	0.657
AgenticIR	26.43	0.812	0.196	0.632
SwinIR	24.86	0.765	0.245	0.557
Real-CalAgent (Ours)	**27.48**	**0.839**	0.187	**0.661**

2. Visualization results



GT: 293.8 kcal

Now, as a large model designed to estimate food calories based on reference objects, This is the first step: (1) Please identify what the reference object is.

There are two reference objects in the image that can assist with scale estimation:
The black and white checkerboard reference card in the lower left corner;
The black dot matrix calibration board beneath the food.
The more suitable for scale reference is the black-and-white checkerboard card in the lower left corner, as it has clear boundaries and regular grid patterns.

This is the second step: (2) Please identify what the reference object is.

Based on the form of the checkerboard card in the image, it is approximately 5 × 4 small squares. Referring to the checkerboard reference card in common food images, I assume: Reference card size≈7.5 cm×6.0 cm. Each small grid is approximately: 1.5 cm×1.5 cm. The reference card itself is a flat paper card, and its thickness has virtually no impact on estimating food volume.

This is the thirdth step: (3) Please carefully and thoroughly identify all food types in the image.

Chinese food appears to be a dessert pastry with thick white frosting/cream, which could be: iced pastry / frosted cake square / cinnamon-roll-like pastry.
The overall estimation should be based on high sugar and high-fat desserts and pastries, rather than regular bread.

This is the fourth step: (4) Based on the estimated volume and size of the reference object and the relative scale in the image, estimate the volume of each food item.

From the scale of the image, the top size of this dessert is approximately the width of the reference card, but due to the food being positioned further back and having a perspective effect, the actual size is slightly larger. Estimated as: length≈7.0cm, width≈6.0cm, height≈3.0cm. The food as a whole is approximately a rectangular prism: V≈7.0×6.0×3.0≈126. Considering that the top frosting is relatively thick, it can be disassembled into: Pastry body: 100 cm ³; Top frosting/cream cream: 21 cm ³; A small amount of interlayer/debris/filling: 5 cm ³.

This is the fiveth step: (5) For each food item, look up the calories per gram and its density.

Common estimated values are as follows:
Food Components：
Sweet bread/cake sponge, 0.50–0.60 g/cm³, 3.2–3.8 kcal/g;
Frosting/buttercream,0.90–1.10 g/cm³, 3.8–4.5 kcal/g;
Small amount of filling/crumbs, 0.70–1.00 g/cm³, 3.0–4.0 kcal/g;
Copy Table Median values are adopted here: ρcake=0.55 g/cm3, ρfrosting=1.0 g/cm3

This is the sixth step: (5) Calculate the weight of each food item based on its volume and density.

Calculation of Weight and Calories
Cake Base
Wcake=100×0.55=55 g
Calorie density adopted: 3.5 kcal/g
Ccake=55×3.5≈193 kcal
Wfrosting=21×1.0=22 g
Calorie density adopted: 4.1 kcal/g
Cfrosting=21×4.1≈86 kcal
Small Amount of Filling/Crumbs:
Wfilling≈5 g. Cfilling≈5×3.5=18 kcal
Total Calorie Estimation: Ctotal=193+86+18
Ctotal≈297 kcal.
Therefore, the estimated calorie content of this dessert is:
297 kcal
Given uncertainties in frosting thickness, cake density and actual size, a reasonable calorie range is approximately: 280-330 kcal
My final estimation will be stated as:
Approximately 297 kcal.

Conclusion

- Real-CalAgent combines image restoration and calorie estimation in one MLLM agent.
- It achieves strong robustness on degraded food images.
- Future work will improve inference speed and tool-selection reliability.

[1] Yu, Dongjian, et al. "OmniFood8K: Single-Image Nutrition Estimation via Hierarchical Frequency-Aligned Fusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2026.
[2] Thames, Quin, et al. "Nutrition5k: Towards automatic nutritional understanding of generic food." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.