Deep Learning Enhanced Non-destructive Examination of Dry Pet Food Kibble Microstructure Using 3D X-ray Tomography

Haozhe Liu[1,2†], Hongshan Liu[1,3†], Xueshen Li[1†], Kyle Lickteig[2], Long Pan[2], Melissa Vanchina[2], Nick Rozzi[2], Lingyan Kong[3*], Nan Yao[4*], Yu Gan[1*], Shiyou Xu[2*]

[1]Department of Biomedical Engineering, Stevens Institute of Technology, Hoboken, 07030, NJ, USA.
[2]Colgate Technology Center, 909 River Road, Piscataway,08854, NJ, USA.
[3]Department of Human Nutrition & Hospitality Management, The University of Alabama, Tuscaloosa, 35487, AL, USA.
[4]Princeton Materials Institute, Princeton University, Princeton, 08544, NJ, USA.

*Corresponding author(s). E-mail(s): lkong@ches.ua.edu; nyao@princeton.edu; ygan5@stevens.edu; shiyou_xu@colpal.edu;
†These authors contributed equally to this work.

Abstract

The microstructural characterization of dry pet food kibbles remains a pivotal yet underexplored domain in food science, hindered by the limitations of conventional imaging techniques that are often destructive and limited to small sample sizes. To overcome these constraints, we present an integrated framework that synergistically combines X-ray tomography with deep learning-based super-resolution and segmentation. Our approach enables non-destructive, micro-scale visualization and quantification of pet food kibbles to an unprecedented level. Central to our methodology are two neural network models: a super-resolution model that reduces acquisition time while maintaining image quality, and an automated segmentation model that facilitates rapid and high-accuracy spatial analysis. Together, these models reduce manual analysis time from $\sim$ 1.5 days to under 1 minute per volume and reduce total imaging and processing duration from $\sim$ 12 hours to under 1 hour per volume, with minimal loss in accuracy (Dice

coefficient reduction $< 1.5\%$). This advancement provides insights into porosity and fat distribution, offering a transformative tool for high-speed, high-quality structural analysis in pet food and nutrition research, and an innovative approach to communicate with professionals and consumers.

Keywords: Kibble Component Analysis, X-ray Tomography, Segmentation, Deep Learning, Super-resolution

1 Introduction

Dry pet food, commonly referred to as kibble, represents the predominant dietary format for pets. Kibble is small, bite-sized pieces of pet food, containing meat, grains, vitamins, and minerals, specifically formulated and manufactured. Spatial analysis of the microstructure in kibble has emerged as a significant area of interest within pet food science [1, 2]. A comprehensive understanding of kibble's internal microstructure is crucial, as it directly influences nutrient bioavailability, provides valuable insights into its ingredients, production, and potential for innovation. For instance, structural features within kibbles govern digestion kinetics and nutrient absorption [3]. In addition, the kibble microstructure determines the texture and palatability of the food, which affects how well pets absorb nutrition, and how pets feel the chewability [4, 5]. Furthermore, the production of kibbles involves a multistage process including ingredient selection, preparation, extrusion, enrobing, and packaging. Each process contributes to the final microstructural profile. With a deeper understanding of the internal structures of kibbles, manufacturers can optimize these processes, leading to more efficient manufacturing, improved quality control, and a longer shelf life [6].

Despite its importance, the complex internal structure of kibble, which is characterized by heterogeneous pore networks and embedded microparticles such as fat, remains difficult to quantify. Conventional porosity measurement techniques, such as the Brunauer-Emmett-Teller (BET) surface area analysis [7], though widely recognized in food science, are unsuitable for kibble due to their destructive nature and reliance on vacuuming materials and filling gas, which compromises structural integrity and yields porosity values without spatial resolution. Similarly, imaging method like scanning electron microscopy (SEM) [8–10] requires invasive sample preparation and offer a limited field of view, restricting their utility for spatial analysis. There is an unmet need for non-destructive, high-resolution techniques capable of capturing porosity and structure in intact kibble matrices.

Recently, advances in X-ray tomography have enabled non-destructive imaging of material microstructures, offering clarity and volumetric insights [11, 12]. Widely adopted in fields such as polymer science [13, 14], biomedical imaging [15, 16], and other materials engineering [17, 18], X-ray remains underutilized in pet food and nutrition science. One of the primary challenges in applying X-ray tomography to dry pet food lies in the analysis of large-scale 3D volumetric data. Manual analysis, while accurate, is prohibitively time-consuming and labor-intensive. For instance, processing of a single volume of $1024 \times 1024 \times 1024$ voxels typically requires over 12 hours for data

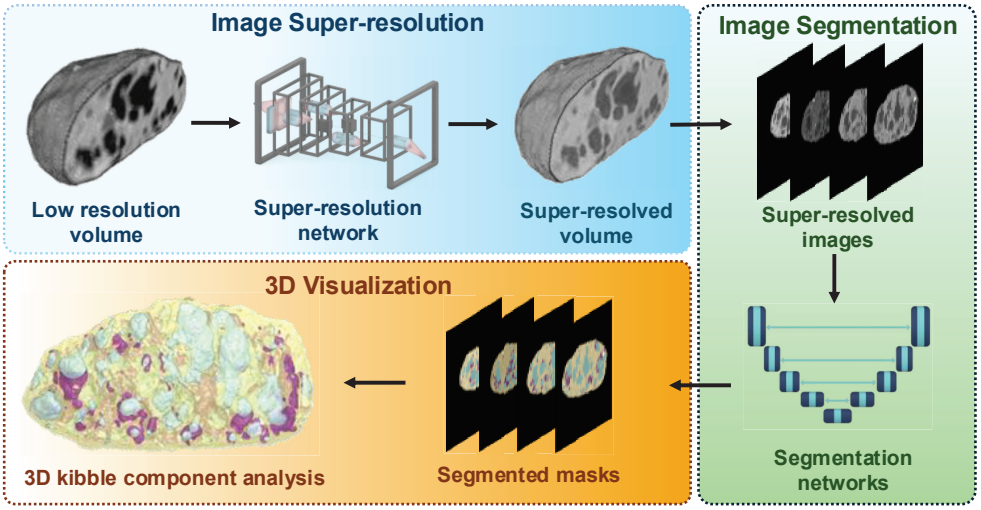

Fig. 1 Generated kibble analysis in this study using low-resolution X-ray tomography with fast data acquisition (∼1 hour). The volumes are processed through two AI modules, a super-resolution network and a segmentation network, to produce a 3D kibble component analysis within about 10 minutes. comparable to a 3D kibble component analysis. This automated pipeline achieves results comparable to conventional analysis, which typically requires approximately 12 hours for data acquisition and an additional day for manual annotation.

acquisition and more than an additional full day for manual analysis. There is thus a need for automated approaches that can efficiently extract meaningful structural and compositional information from high-resolution X-ray datasets.

Deep learning has transformed image analysis by enabling efficient and highly accurate interpretation of complex visual data. Convolutional neural networks (CNNs), in particular, have demonstrated exceptional performance in tasks such as classification, object detection, and segmentation [19], owing to their ability to learn from both local and global feature representations directly from raw pixels. In food science, these techniques have facilitated advancements in quality control, formulation optimization [20–22], and compositional analysis [23, 24]. Despite these successes, the application of deep learning to pet food research, especially for analyzing the dry kibble microstructure, remains limited. Commercial platforms such as Amira-Avizo [25] and Dragonfly [26] offer basic segmentation capabilities such as thresholding and standard learning models. However, these tools lack the flexibility and precision required to accurately identify and quantify critical components within kibble. This gap highlights the need for highly accurate approaches tailored to the unique structural and compositional challenges of pet food analysis.

To address the analytical limitations in the underexplored domain of pet food microstructure, we present a novel deep learning framework tailored for the compositional analysis of kibble. As illustrated in Fig. 1, our approach employs a two-stage neural network design for rapid and accurate volumetric analysis. The first stage integrates an adaptive image super-resolution algorithm that reconstructs images from

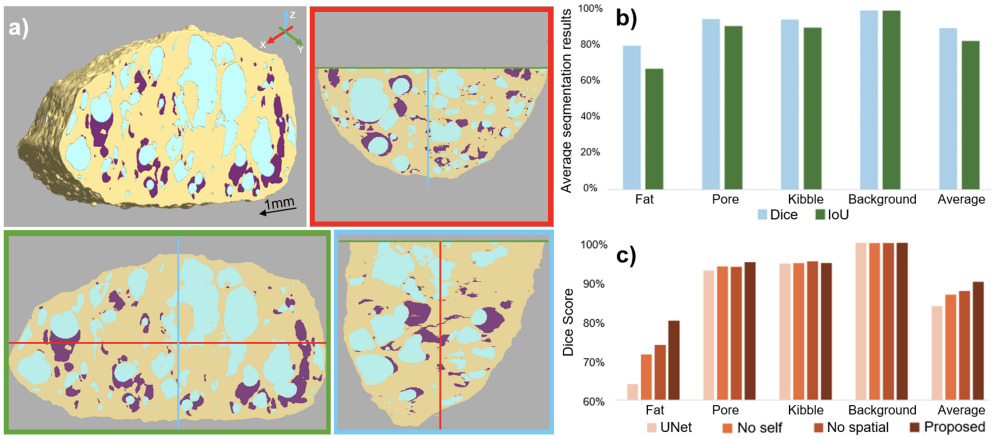

Fig. 2 3D visualization on a Kibble sample. a) Cross-sectional segmentation of kibble, with yellow indicating kibble, blue indicating pores, and purple indicating fat; b) Quantitative analysis on the segmentation performance over manual annotation; c) Ablation study which demonstrates the importance of the proposed segmentation approach.

sparsely acquired tomographic data, significantly accelerating acquisition without compromising image quality. In the second stage, we introduce a Multilevel Attention Fusion (MultiAF) segmentation model that accurately delineates key structural component, including fat, kibble matrix, and porosity. Our framework addresses a critical gap in the spatial analysis of microstructural features, enabling automated, high-speed processing of large 3D kibble datasets. Compared to conventional manual workflows, which require approximately 1.5 days per volume, our methods reduce processing time to under one minute per volume. When combined with super-resolution, our framework significantly reduces total data acquisition and processing time from ∼12 hours per volume to under 1 hour per volume, with minimal loss in accuracy (i.e., only 1.5% drop in Dice coefficient). This advancement facilitates systematic, efficient, and reproducible analysis of kibble microstructure, offering new opportunities for product optimization and innovation in pet food science.

2 Results

2.1 Segmentation and 3D Visualization of Kibble Structure

We conducted a segmentation using the workflow and methodology described in the **Method** Section. To evaluate the performance of the proposed MultiAF model, we conducted experiments on a dataset comprising 2D images extracted from four X-ray tomography scans of kibble. Our dataset includes two kibble types: high vacuum enrobed kibbles, containing three compositional classes (kibble matrix, pore, and fat), and non-enrobed kibbles, which have fat and were annotated with two composition classes (kibble matrix and pores). The segmentation outputs were visualized in 3D

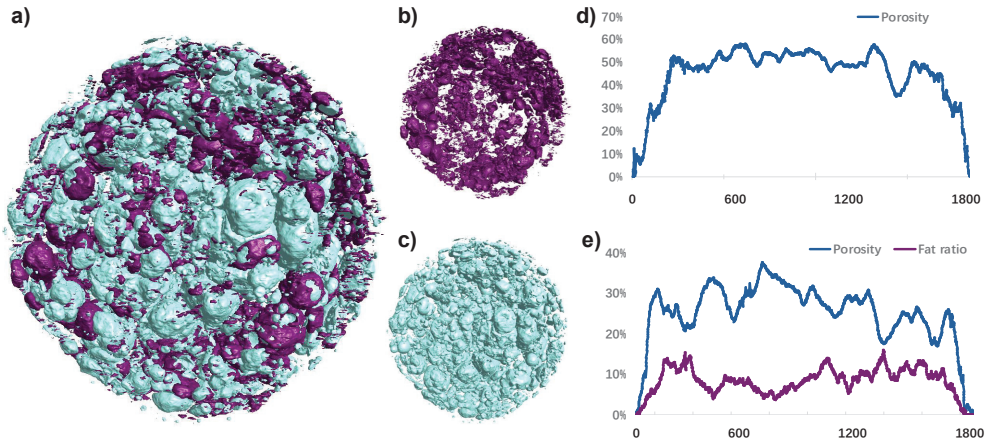

Fig. 3 3D Visualization and Distribution Analysis. a) Visualization of fat (purple) and pore (blue) in 3D space; b) Visualization of fat composition; c) Visualization of pore region; d) The distribution of porosity among scan index; e) The distribution of porosity and fat among scans.

and presented in Fig. 2(a). Our model supports cross-sectional views, with the XZ-plane shown in the green box, the XY-plane in the red box, and the YZ-plane in the blue box. Our segmentation demonstrated clear delineation of structural boundaries, enabling special analysis of key nutritional components.

Quantitative performance was assessed using the Dice coefficient and Intersection over Union (IoU), averaged across test samples. As depicted in Fig. 2 (b)-(c), MultiAF outperformed baseline models, including U-Net model [27], and variants incorporating only self-attention or global-attention mechanisms. Comparative analysis with five well-established deep learning models [28–32] further demonstrated the superiority of our approach, as shown in **Supplementary Fig. 2& Fig. 3**.

2.2 Analysis on 3D Distribution of Kibble Composition

Our MultiAF model significantly advances 3D visualization in X-ray tomography by enabling automated, high-resolution analysis of compositional distributions within kibble volumes. As depicted in Fig. 3 (a), the MultiAF accurately segments and classifies two critical regions, fat and pores, assigning distinct color labels to each class. This facilitates intuitive visualization and enables targeted extraction of individual for further analysis as shown in Fig. 3 (b) and (c). The model's ability to resolve fine-scale features in 3D provides a powerful tool for investigating the internal structure of pet food, offering new insights into its structural complexity.

Analysis of porosity revealed consistent spatial trends across both enrobed and non-enrobed kibble samples. As illustrated in Fig. 3 (d) and (e), porosity remained relatively uniform across the depth of the kibble, with comparable overall porosity between the two groups. Notably, fat content was concentrated in the central regions of the enrobed kibble, suggesting the high-vacuum enrobing process facilitates deep

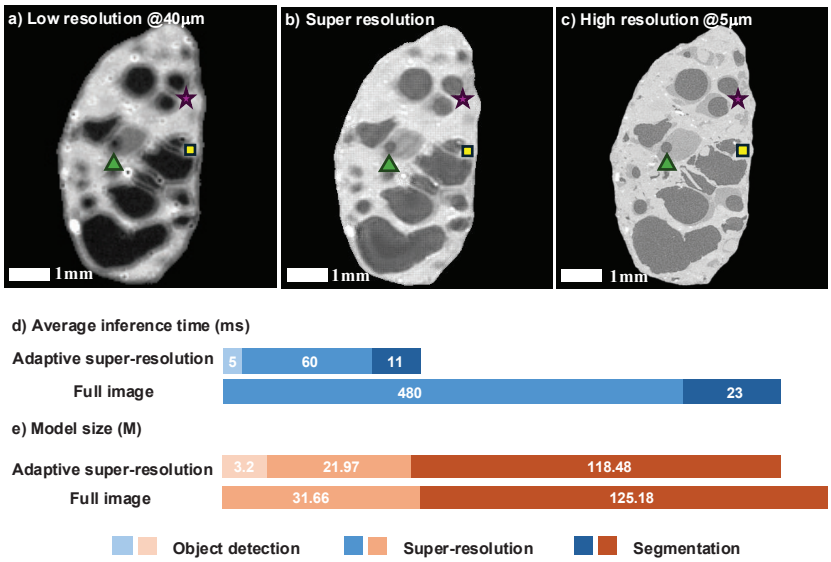

Fig. 4 Super-resolution results and comparison of runtime. a) Low resolution image; b) Super-resolved image processed by neural network. ; c) High resolution images; d) A comparison of runtime and model size between adaptive super-resolution approach and conventional approach to super-resolve the entire image for segmentation.

infusion of fat into porous regions rather than surface-level coating. This observation highlights the critical role of porosity in enabling ingredient penetration during processing, without compromising the structural integrity of the kibble.

Besides visualization, our framework enables quantitative assessment of how formulation and processing influence internal structure. Notably, those capabilities not achievable with conventional destructive methods such as BET surface area analysis. By leveraging learned features from X-ray data, MultiAF performs fully automated segmentation in under one minute per volume, reducing the need for manual annotation, which typically requires over 1.5 days per sample. This capability supports reproducible analysis of pet food microstructure, with direct implications for product design and optimization.

2.3 Super-resolution and Time Efficiency

To enhance the efficiency of X-ray tomography analysis, we implemented a super-resolution algorithm trained on paired low-resolution (LR) and high-resolution (HR) image datasets, as described in the Method Section. Notably, acquiring LR data requires (∼1 hour) of the acquisition time needed for HR data (∼12 hours). We employ a unified, multi-scale super-resolution neural network [33] to enhance the quality of low-resolution images, substantially improving the efficiency without compromising image quality. As demonstrated in Fig. 4, the super-resolved (SR) images exhibit marked improvements in structural details compared to the original LR inputs, with enhanced clarity in key regions (highlighted by markers). Segmentation performance

on SR images closely approximates that of HR images, with only a minor reduction in overall Dice coefficients from 98.6% to 97.1%. Importantly, SR reconstruction significantly improves segmentation accuracy over LR inputs, yielding Dice coefficient gains of 7.59% for kibble, 5.7 % for pores, and 17.07% for fat. These results underscore the necessity of super-resolution for reliable compositional analysis. Detailed comparisons are provided in **Supplementary Fig. 4**. Furthermore, our adaptive super-resolution framework selectively enhances regions of interest rather than the entire field of view, optimizing computational efficiency. As presented in Fig. 4(d), this approach reduces image processing time by 85% for a single 2D scan, enabling rapid and scalable analysis of large volumetric datasets.

3 Discussion

The integration of X-ray tomography with deep learning presents a transformative approach for non-destructive, high-resolution analysis of dry pet food microstructure, an area that remains largely underexplored despite the nutritional and commercial significance of kibble. While previous efforts to characterize kibble structure have been limited by the resolution and sample size of traditional imaging and analysis methods, the advent of advanced computational tools now enables micro-scale visualization and quantification at unprecedented speed and accuracy.

In this study, we introduce a novel deep learning framework that combines a MultiAF segmentation model with an adaptive super resolution model to address key limitations in volumetric data analysis. This deep learning-based analysis demonstrates expert-level accuracy in identifying and segmenting components (i.e., fat, pores, kibble matrix) across diverse sample types. Once trained, the model performs rapid, automated segmentation of large 3D datasets in under one minute per volume, a substantial improvement over manual annotation workflows that typically require over a day per sample.

To further reduce the time and resource burden associated with high-resolution data acquisition, we designed and implemented a super-resolution algorithm capable of reconstructing high-quality images from low-resolution scans. This approach reduces acquisition time by over 90% – from $\sim$12 hours to $\sim$1 hour while maintaining segmentation accuracy within 1.5% of high-resolution benchmarks.

Notably, the super-resolution module selectively enhances regions of interest, optimizing both computational efficiency and image quality in visualizing the kibble structure. Relying upon a deep learning framework, our system visualizes and analyzes high-resolution images from a small amount of kibble data, thus saving unnecessary workload in data acquisition. The framework successfully generalizes to previously unseen kibble types, including both enrobed and non-enrobed samples, and its performance is expected to improve further as additional annotation data are incorporated. This adaptability is critical for supporting diverse kibble formulations and processing conditions in the pet food industry.

Overall, the combination of super-resolution and MultiAF-based segmentation constitutes a robust, high-throughput pipeline conception for structural analysis of dry pet food. This integrated approach not only accelerates research and development but

also enables systematic evaluations of how formulation and processing influence kibble internal structure. We expect these insights will provide valuation guidance for optimizing pet food delivery. By combining advanced X-ray imaging and deep learning, this work lays the groundwork for a potential new approach to data-driven innovation in pet food science.

4 Method

4.1 Data acquisition using X-ray Tomography

Material Production. The vacuum enrobing procedure involves using cooled kibble (at approximately 85°F), applying a vacuum of 200 mbar, enrobing with grease, releasing the vacuum, and then applying a dry palatant. In contrast, the control enrobing process, used for comparison, involves using hot kibble straight from the dryer, blending it with grease under atmospheric conditions, and then enrobing with dry palatant.

Image Acquisition. We conducted evaluations on the kibbles. The images were acquired by the Zeiss Xradia 520 Versa High-resolution 3D X-ray tomography Microscope. A source voltage of 80 kV and source power of 7 W were used for imaging. Both low-resolution image and high-resolution image measure the same field of view of $10,415\ \mu m \times 10,415\ \mu m$. The low-resolution image has a pixel size of 40 μm with an exposure time of 1 second. The high-resolution image has a pixel size of 5 μm with an exposure time of 15 seconds.

Data preprocessing. Images are preprocessed using histogram equalization. Then, we utilized the YOLOv12 [34] object detection model to accurately detect the region of interest (ROI) and remove redundant background. The ROI for the super-resolution process is reduced, improving computational efficiency. This allows the downstream tasks to focus solely on the kibble ROI, enhancing both efficiency and accuracy. The augmented images exhibit enhanced brightness and contrast, revealing finer details more clearly. These enhancements result in a significant improvement in segmentation quality, with the processed images clearly demonstrating successful border removal and enhanced kibble representation.

4.2 Adaptive, Multi-scale, Super-resolution

We developed an adaptive super-resolution to improve the image quality of the area of interest. For this purpose, the ROI detected by YOLOv12 is used to compute a scale factor, determined by the ratio of the ROI size to the full image size, and serves as input for scale factor determination. The scale factor and cropped LR image are sent to a multi-scale super-resolution network [33].

As shown in Fig. 5, we design a generative training scheme to produce high-quality images by learning the inherent relationship between LR images and HR images. Let $I_{\text{LR}} \in \mathbb{R}^{H \times W \times 3}$ be the input low-resolution image, and $I_{\text{SR}} \in \mathbb{R}^{\lfloor mH \rfloor \times \lfloor mW \rfloor \times 3}$ be the output super-resolved image with an arbitrary scale factor $m \in \mathbb{R}$. The feature extraction network Residual Channel Attention Network [35] is applied and is denoted as

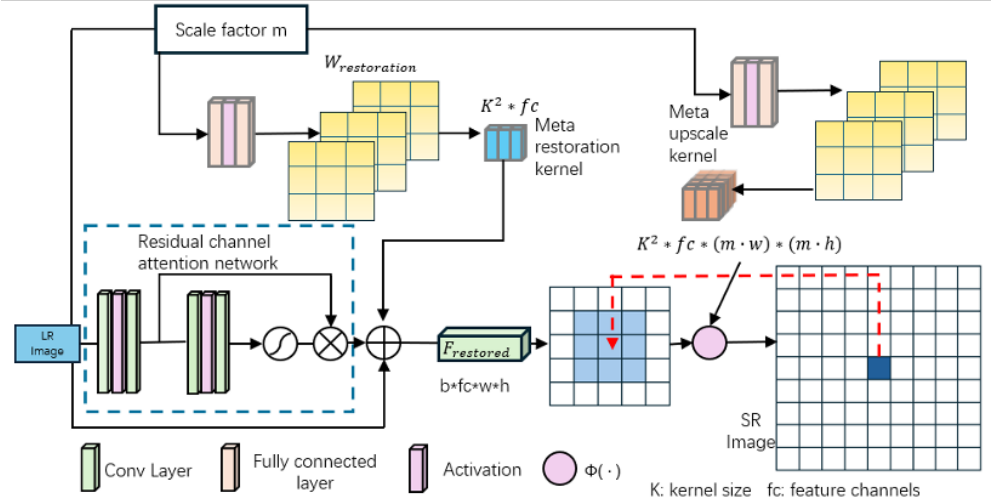

Fig. 5 Adaptive super-resolution framework that selectively enhances regions of interest using a scale-aware meta-upscale network. The network dynamically generates position-specific convolution kernels based on feature offsets to produce high-quality SR outputs at arbitrary scales.

$\phi(\cdot)$, producing feature map $\mathcal{F} = \phi(I_{\text{LR}}) \in \mathbb{R}^{H \times W \times f_c}$, where f_c is the number of feature channels. The meta-upscale network is represented as $W_{\text{restoration}} = \phi_{\text{restore}}(m) \in \mathbb{R}^{K^2 \times fc}$, which generates pixel-wise dynamic weight kernels with size K^2. To generate SR output of arbitrary size, the meta-upscale network uses position-specific filters by dynamically predicting a unique convolution kernel for each output pixel. For a given pixel (i, j) in the SR image, its corresponding location in the LR feature map is obtained by projecting back using the transformation $(i', j') = \left(\lfloor \frac{i}{m} \rfloor, \lfloor \frac{j}{m} \rfloor\right)$, where m is the desired scale factor. The fractional offsets are computed as $\Delta x = \frac{i}{m} - \lfloor \frac{i}{m} \rfloor$ and $\Delta y = \frac{j}{m} - \lfloor \frac{j}{m} \rfloor$. These are used to form the input vector to the Meta-Upscale vector: $v_{ij} = (\Delta x, \Delta y, \frac{1}{m})$. This vector is passed into the Meta-Upscale network $\phi_{\text{upscale}}(\cdot)$ to generate a position-specific filter $W_{ij} = \phi_{\text{upscale}}(v_{ij}) \in \mathbb{R}^{K^2 \times fc}$, where K is the kernel size and fc is the number of feature channels. This dynamically generated kernel is then applied to the feature at location (i', j') to compute the output pixel value in the SR image as $I_{SR}(i, j) = F_{\text{refined}}(i', j') \cdot W_{ij}$. This operation is performed for each pixel in the SR image $I_{SR} \in \mathbb{R}^{(mw) \times (mh)}$, where the dot product corresponds to the computation visualized as $\phi(\cdot)$ in the diagram. The detailed pseudo code is provided in **Supplementary 4, Algorithm 1**.

4.3 Segmentation

We developed the Multilevel Attention Fusion (MultiAF) model, an advanced deep learning framework integrating self-attention and global attention mechanisms across multiple depths of feature representations. This approach enables precise segmentation of kibble compositions, such as pores and fats in X-ray tomography scans, regardless of their size or shape. This model streamlines the analytical process by

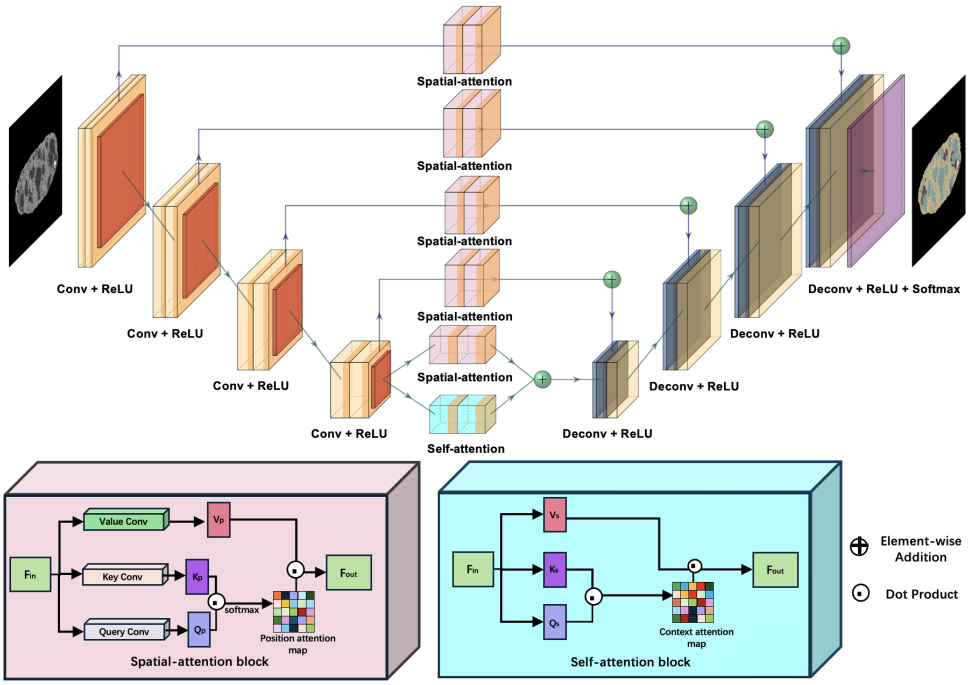

Fig. 6 Multi-attention fusion-based segmentation network. Spatial-attention modules are embedded at the third, fourth, and fifth encoder levels to capture broader contextual dependencies. Both spatial-attention and self-attention modules are jointly applied at the bottleneck to enhance global feature interpretation. This dual attention integration enables refined feature representation and significantly improves segmentation accuracy, particularly for analyzing complex kibble compositions.

combining advanced attention mechanisms, hierarchical feature extraction, and 3D reconstruction, allowing for an efficient and scalable solution in the segmentation of X-ray tomography data.

Our approach consists of a training stage and an inference stage, with details provided in the Methods section and illustrated in Fig. 6. During training, the model follows a supervised learning paradigm, requiring expert-annotated labels. It converts expert-identified 2D coordinates of kibble compositions into segmentation masks, using them alongside X-ray tomography scans as paired inputs for learning. This enables the model to accurately differentiate and classify various kibble components.

During inference, the trained deep neural segmentation model first classifies pixels into distinct categories representing different kibble compositions. A 3D reconstruction software (i.e. Dragonfly) then processes these segmentation maps to create a three-dimensional visualization of the entire kibble volume. This automated workflow enables detailed structural analysis and measurement, significantly reducing the need for manual input while ensuring robust and high-accuracy segmentation.

Attention modules integrated throughout the expansive path enhance the model's focus and accuracy. The self-attention module introduces a novel circular positional

encoding, applied element-wise to feature maps. This encoding incorporates positional information to improve feature representation, particularly in X-ray microscopy where samples often lie near the image's center. The circular positional encoding is determined based on the distance from the center of the feature to other feature positions, using the formula:

$$I = \sqrt{(i - x_{center})^2 + (j - y_{center})^2}, \tag{1}$$

where i and j are feature coordinates, and x_{center} and y_{center} are the center coordinates of the feature. This method enriches the model with both absolute and relative positional awareness.

For the self-attention mechanism, the encoder features are projected into three matrices: queries (Q), keys (K), and values (V). These undergo operations to produce a contextual attention map (A) derived from the softmax-normalized scaled dot-product of Q and the transpose of K, multiplied by V. This process aggregates values weighted by attention, emphasizing relevant features within the global context:

$$SA(Q, K, V) = softmax(QK^T)V \tag{2}$$

.

The global attention module enhances the model's ability to capture broader contextual information. Encoder features F undergo two convolutional operations, producing feature maps F_g and F'_g. F_g is reshaped and transposed into matrices M and N, while F'_g transforms into W. Matrix multiplication between M and N produces positional attention maps (PA), calculated as:

$$PA_{i,j} = \frac{exp(M_i, N_j)}{\sum_{i=1}^{n} exp(M_i, N_j)} \tag{3}$$

where $PA_{i,j}$ measures the influence of the i-th position on the j-th position. PA is then multiplied by W to yield the refined feature representation:

$$GA(F_p) = \sum_{q=1}^{h \times w} W_q PA_{p,q} \tag{4}$$

.

This formulation enables the global attention module to refine feature representations by integrating positional influences across the entire feature map, leading to a more coherent intra-class representation.

During downsampling and feature extraction in the contraction path, the global attention module is applied at the third, fourth, and fifth levels to incorporate broader contextual insights. At the fifth level, the self-attention module is deployed in conjunction with the global attention module. This dual application at the highest feature abstraction level enriches the model with and worldwide awareness, refining attention and optimizing feature interpretation, thereby improving overall performance and accuracy. This sophisticated attention integration significantly contributes to the model's effectiveness in precise kibble composition analysis.

In this study, we adopt a composite loss function that combines multi-class cross-entropy loss and Dice loss to ensure both pixel-wise classification accuracy and overall segmentation overlap. To further improve boundary segmentation, we incorporate an edge-enhanced loss component M_{edge}. This map is generated by computing the Euclidean distance from each pixel to the nearest pixel of a different class. It applies increased penalties for misclassifications near object boundaries to reduce the influence of boundary ambiguity caused by the Partial Volume Effect. By integrating these losses, our framework aims to achieve more robust and anatomically consistent segmentation results. The loss functions can be expressed as:

$$L_{total} = L_{CE}(1 + M_{edge}) + L_{Dice} + ||W||_2^2, \tag{5}$$

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) \tag{6}$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{i=1}^{N} p_{i,c}\cdot y_{i,c} + \epsilon}{\sum_{i=1}^{N} p_{i,c} + \sum_{i=1}^{N} y_{i,c} + \epsilon} \tag{7}$$

where N denote the number of pixels or samples, and C the total number of classes. For each pixel i, $y_{i,c}$ represents the binary ground truth indicator, which equals 1 if the true class of pixel i is class c, and 0 otherwise. The predicted probability for class c at pixel i is denoted as $p_{i,c}$, typically obtained via a softmax activation. In the case of Dice loss, both $y_{i,c}$ and $p_{i,c}$ can be interpreted as one-hot and softmax probabilities, respectively. A small constant ϵ is added to the denominator to avoid division by zero. The detailed pseudo code is provided in **Supplementary 3, Algorithm 2**.

4.4 Deep Learning Network Training

In implementing our deep learning model, we trained a segmentation network using 2D X-ray tomographic images of individual kibbles. The source dataset consisted of 2037 cross-sectional scans, with dimensions of 2008 × 2047 pixels. To construct an exhaustive training dataset, we meticulously selected scans at regular intervals, resulting in a total of 220 labeled images. These were then strategically split into training and testing datasets, maintaining a 4:1 ratio. The dataset covered three primary classes: kibble, fat, and pore, in addition to background elements, enriching the training data's diversity. To prepare the 2D slices for processing, we scaled them down by a factor of 4 of their original dimensions. This step was followed by our specified image preprocessing techniques. Utilizing the Adam optimizer [36] with a set learning rate of 0.001, our network underwent 300 training epochs. This extensive training, executed on an RTX A6000 GPU, ensured the model's effective convergence. Our MultiAF model demonstrated exceptional capability in learning from the X-ray tomographic slices, adeptly using pixel-level annotations to categorize each pixel into one of the predefined classes (i.e., kibble, pore, or fat) with the rest marked as background. To enhance the validation of our model's versatility and broad applicability, we enriched our dataset with an extra 40 sets of labels for randomly chosen kibbles, designated for external testing.

Comparable studies were conducted across various state-of-the-art deep learning models on fat, kibble, hole, and background classes, demonstrating the superior segmentation accuracy of MultiAF on the kibble X-ray tomography dataset. The qualitative results are presented in Supplementary 4, Fig. **??**, while the quantitative results are displayed in Supplementary 4, Fig. **??**.

5 Software utilized

All code was implemented in Python (3.10) using Pytorch (2.0) as the base deep learning framework. We also used several Python packages for data analysis and results visualization, including torchvision (0.15.2), numpy (1.24.3), scikit-image (0.20.0), scipy (1.10.1), matplotlib (3.7.1), and opencv-python (4.8.0).

6 Report summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

7 Data availability

The training and validating datasets used in this study are available in the public domain and can be downloaded. We confirmed that all the image datasets in this study are publicly accessible and permitted for research purposes. Source data are provided in this paper.

8 Code availability

The training script, inference script, and trained model have been publicly available at XXX.

9 Acknowledgment

The authors would like to thanks the support from NSF-CAREER-2239810, NSF-CRII-2222739, and Colgate-Palmolive Experiential Learning Program.

References

[1] M. Monti et al. Influence of dietary fiber on macrostructure and processing traits of extruded dog foods. Animal Feed Science and Technology **220**, 93–102 (2016). URL https://www.sciencedirect.com/science/article/pii/S0377840116303625.

[2] Ye, L. et al. Using confocal microscopy to estimate the distribution of natural antioxidants in poultry meal and extruded kibbles. European Journal of Lipid Science and Technology **121**, 1800374 (2019).

[3] Alvarenga, I. C., Aldrich, C. G. & Shi, Y.-C. Factors affecting digestibility of starches and their implications on adult dog health. Animal Feed Science and Technology **282**, 115134 (2021).

[4] Samant, S. S., Crandall, P. G., Jarma Arroyo, S. E. & Seo, H.-S. Dry pet food flavor enhancers and their impact on palatability: a review. Foods **10**, 2599 (2021).

[5] Le Guillas, G., Vanacker, P., Salles, C. & Labouré, H. Insights to study, understand and manage extruded dry pet food palatability. Animals **14**, 1095 (2024).

[6] Hołda, K. & Głogowski, R. Selected quality properties of lipid fraction and oxidative stability of dry dog foods under typical storage conditions. Journal of Thermal Analysis and Calorimetry **126**, 91–96 (2016).

[7] Majid Naderi. in *Chapter fourteen - surface area: Brunauer–emmett–teller (bet)* (ed.Steve Tarleton) Progress in Filtration and Separation 585–608 (Academic Press, Oxford, 2015). URL https://www.sciencedirect.com/science/article/pii/B9780123847461000148.

[8] Tan, J., Zhang, H. & Gao, X. Sem image processing for food structure analysis. Journal of Texture Studies **28**, 657–672 (1997).

[9] Sharma, V. & Bhardwaj, A. Scanning electron microscopy (sem) in food quality evaluation 743–761 (2019).

[10] Ong, L., Dagastine, R. R., Kentish, S. E. & Gras, S. L. Microstructure of milk gel and cheese curd observed using cryo scanning electron microscopy and confocal microscopy. LWT-Food Science and Technology **44**, 1291–1302 (2011).

[11] Holt, M., Harder, R., Winarski, R. & Rose, V. Nanoscale hard x-ray microscopy methods for materials studies. Annual Review of Materials Research **43**, 183–211 (2013).

[12] Ou, X. et al. Recent development in x-ray imaging technology: Future and challenges. Research (2021).

[13] Ciaramitaro, V. et al. A new methodological approach to correlate protective and microscopic properties by soft x-ray microscopy and solid state nmr spectroscopy: The case of cusa's stone. Applied Sciences **11**, 5767 (2021).

[14] Mansikkala, T. et al. Lignans in knotwood of norway spruce: Localisation with soft x-ray microscopy and scanning transmission electron microscopy with energy dispersive x-ray spectroscopy. Molecules **25**, 2997 (2020).

[15] Nowak-Stępniowska, A. et al. Nanometer-resolution imaging of living cells using soft x-ray contact microscopy. Applied Sciences **12**, 7030 (2022).

[16] Weinhardt, V. et al. Imaging cell morphology and physiology using x-rays. Biochemical Society Transactions **47**, 489–508 (2019).

[17] Wu, Z. et al. Principles of transmission x-ray microscopy and its applications in battery study. Advanced X-ray Imaging of Electrochemical Energy Materials and Devices 65–90 (2021).

[18] Li, Q. et al. Microstructural study of hydration of c3s in the presence of calcium nitrate using scanning transmission x-ray microscopy (stxm).

Journal of Nanomaterials **2020**, 1–9 (2020).

[19] Wu, H., Liu, Q. & Liu, X. A review on deep learning approaches to image classification and object segmentation. Computers, Materials & Continua **60** (2019).

[20] Zhou, L., Zhang, C., Liu, F., Qiu, Z. & He, Y. Application of deep learning in food: a review. Comprehensive reviews in food science and food safety **18**, 1793–1811 (2019).

[21] Zhu, L., Spachos, P., Pensini, E. & Plataniotis, K. N. Deep learning and machine vision for food processing: A survey. Current Research in Food Science **4**, 233–249 (2021).

[22] Jeong, N., Gan, Y. & Kong, L. Emerging non-invasive microwave and millimeter-wave imaging technologies for food inspection. Critical Reviews in Food Science and Nutrition 1–12 (2024).

[23] Ma, P. et al. Deep learning accurately predicts food categories and nutrients based on ingredient statements. Food Chemistry **391**, 133243 (2022).

[24] Ji, H. et al. Recent advances and application of machine learning in food flavor prediction and regulation. Trends in Food Science & Technology (2023).

[25] Thermo Fisher Scientific. Amira-Avizo Software. Thermo Fisher Scientific, Waltham, MA, USA (2022). URL https://www.thermofisher.com/amira-avizo. Version 2022.1.

[26] Object Research Systems (ORS) Inc. Dragonfly. Object Research Systems (ORS) Inc., Montreal, Canada (2023). URL https://www.theobjects.com/dragonfly. Version 2023.1.

[27] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015 234–241 (2015).

[28] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. DLMIA in conjunction with MICCAI 3–11 (2018).

[29] Huang, H. et al. Unet 3+: A full-scale connected unet for medical image segmentation. International conference on acoustics, speech and signal processing (ICASSP) 1055–1059 (2020).

[30] Cao, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. European conference on computer vision (ECCV) 205–218 (2022).

[31] Oktay, O. et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018).

[32] Wang, L. et al. Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS Journal of Photogrammetry and Remote Sensing **190**, 196–214 (2022).

[33] Li, X. et al. Multi-scale reconstruction of undersampled spectral-spatial OCT data for coronary imaging using deep learning. IEEE Transactions on Biomedical Engineering **69**, 3667–3677 (2022).

[34] Tian, Y., Ye, Q. & Doermann, D. Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 (2025).

[35] Zhang, Y. et al. Image super-resolution using very deep residual channel attention networks. Proceedings of the European conference on computer vision (ECCV) 286–301 (2018).

[36] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).