LEARNING CANONICAL COORDINATE SYSTEMS FOR DEFORMABLE ORGANS: A COMMON COORDINATE FRAMEWORK FOR THE MOUSE INTESTINE

Anonymous authors
Paper under double-blind review

ABSTRACT

Whole-organ tissue clearing and volumetric confocal microscopy have enabled cellular-resolution imaging of intact biological organs, generating multi-terabyte datasets that require new computational methods for standardized spatial analysis. Unlike the brain, the intestine exhibits substantial nonlinear deformation caused by tissue stretching, folding, and specimen preparation, preventing reliable anatomical correspondence across animals and limiting quantitative analysis of the enteric nervous system (ENS). Existing image registration approaches rely primarily on intensity matching or handcrafted landmarks and fail to preserve the hierarchical organization of intestinal tissue layers. Here, we present GutCCF, a learning- based Common Coordinate Framework (CCF) for the mouse intestine, formulating canonical coordinate learning as multi-task representation learning for deformable organs. We define the canonical space of the intestine as an intrinsic cylindrical parameterization $C = U \times V \times W$, where U, V, and W denote normalized longitudinal arc length, normalized circumferential position, and transmural depth (ordered tissue segment hierarchy), respectively. The CCF is a mapping $\Phi : \Omega \to C$ from a specimen-dependent image domain to a shared coordinate identity. Our multi-head 2.5D U-Net jointly predicts semantic tissue identity $s(x)$ via cross-entropy and a continuous depth field $d(x)$ via an RMSE objective. This depth-aware monotonic layer localization imposes an ordinal structure on the shared latent representation. Encoding each voxel's position requires confidence within the transmural hierarchy, yielding features that remain anatomically ordered under arbitrary tissue deformation and enabling segmentation and coordinate assignments to reinforce one another. The model is further refined through an iterative human-in-the-loop protocol that incorporates expert corrections into training to improve boundary localization and structural consistency. GutCCF is trained and evaluated using more than 150 high-resolution confocal image volumes of optically cleared intestines, including jejunum, ileum, and colon, from 71 transgenic mice containing seven genetically defined enteric neuron subtypes. Image datasets are at 0.863 μm isotropic resolution, spanning specimen lengths of 0.5–30 mm and between 200 and 700 optical sections. The training dataset contains over 2487 expertly annotated image-mask pairs with hierarchical tissue annotations including longitudinal muscle, circular muscle, submucosa, glands, villi, and out-of-tissue regions. The resulting coordinate system assigns every voxel to a continuous canonical anatomical space, enabling standardized localization of neuronal morphology, tissue structures, and multimodal biological measurements. Experimental results demonstrate that jointly learning anatomical depth and semantic segmentation significantly improves tissue boundary localization, spatial consistency, and cross-sample correspondence compared with segmentation-only baselines, achieving 82.48% mean Dice, 70.74% mIoU, 81.85% recall, and 83.28% overall precision. The proposed framework produces a high-resolution three-dimensional reference representation of the mouse intestine and enteric nervous system and provides a scalable foundation for multimodal spatial atlases, computational anatomy, and organ-scale digital twins.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Application of AI segmentation to neuroscience has been one of the biggest AI successes in biology. Most of the work is done on the Central Nervous System (CNS), while the largest nervous system in the periphery, the Enteric Nervous System (ENS), has not been addressed at comparable scale. The ENS, a neuronal network embedded within the gastrointestinal wall, is essential for regulating intestinal functions such as motility, secretion, and local sensory signaling. AI models have revolutionized the mapping of neural circuitry at high spatial resolution. However, mapping of the ENS is still in its infancy because tissue samples vary in thickness, curvature, staining, and imaging artifacts. Segmenting three-dimensional gut tissue samples from confocal microscopy is particularly difficult, and most segmentation is performed on 2D histology samples that do not capture the three-dimensional properties of the ENS.

To completely understand and map the ENS, we first need to create a standardized 3D twin that acts as a reference across samples and instruments. Recent advances in tissue clearing techniques (Richardson and Lichtman, 2015; Young et al., 2021) allow large-volume, high-resolution tissue imaging without distortions caused by dehydration or physical sectioning. Despite these advances, inherent limitations of light microscopy, including photon scattering and tissue penetration, persist. Furthermore, intestinal tissues are cut open and fixed flat before mounting on a slide for imaging. This makes larger-volume imaging possible but destroys the intrinsic 3D tubular structure and introduces variability through random tilting, waviness, and tissue folds. For tissue samples that need to be registered to a standard or compared with other samples, we propose a 3D canonical coordinate system for gut tissue. A reference atlas of the mouse brain is already available (Wang et al., 2020). Here, we present the Gut Common Coordinate Framework (GutCCF), a standardized digital twin that identifies tissue samples at voxel-wise precision for comparison across image datasets. At its core, GutCCF addresses a tissue registration problem in which deformation destroys direct anatomical correspondence, with tissue-layer segmentation serving as an intermediate step.

To solve deformable-object transformation and standardization, we propose a canonical coordinate system characterized by longitudinal, cross-sectional, and transmural variables in a cylindrical representation. The acquired image datasets contain limited fluorescence designed to highlight sparsely labeled neurons of the enteric nervous system. Therefore, the remaining tissue (more than 99% of all voxels) is often visible only through weak background autofluorescence comprising roughly the lowest 5% of potential intensity, making whole-tissue segmentation particularly challenging. A simple supervised algorithm does not capture the 3D context of tissue structure. To address this problem, we present a novel multi-head, multi-task 2.5D U-Net architecture for spatially aware gut tissue-layer segmentation. This depth-aware framework is designed as a foundational step toward a common coordinate framework for the mouse gut. Using deep learning models trained on expertly annotated confocal image volumes, we segment major gut tissue layers in a layer-aware and anatomically constrained manner.

Our contributions are as follows: (1) we introduce a high-resolution 3D GutCCF for analysis and research; (2) we provide a canonical coordinate representation for standardization of gut tissue; (3) we propose a depth-aware 2.5D multi-task, multi-head U-Net that jointly predicts semantic tissue identity and continuous transmural depth, explicitly modeling the ordered anatomical organization of intestinal layers from low-intensity image data; and (4) we perform comparative evaluation against segmentation-only and alternative baseline models to quantify the benefit of depth-aware supervision for tissue segmentation accuracy, boundary localization, and anatomical consistency.

2 RELATED WORK

CNS to ENS. The emergence of deep learning has substantially advanced mapping of neuronal circuits. The CNS has been extensively mapped in terms of morphology and omics by the Allen Brain Institute (Wang et al., 2020). Comparable efforts for the ENS have not been undertaken despite the fact that many digestive diseases, including irritable bowel syndrome and inflammatory bowel diseases, may involve ENS abnormalities (enteric neuropathies). The lack of standardized anatomical mapping has hindered mechanistic understanding of digestive disease pathophysiology.

Standardization and Segmentation. A standardized ENS atlas requires a segmentation pipeline that identifies biologically meaningful tissue layers despite specimen-dependent preparation and morphology. Classical methods based on thresholding, region growing, deformable contours, or atlas propagation were interpretable but often failed under low contrast, imaging artifacts, and anatomical variability, while requiring dataset-specific tuning. Deep learning networks are particularly useful for segmentation and annotation of tissue samples. Convolutional neural networks (CNNs), particularly U-Net and its variants, have become the standard architecture for biomedical image segmentation due to their ability to learn hierarchical image representations directly from annotated data (Krithika alias AnbuDevi and Suganthi, 2022). U-Net (Ronneberger et al., 2015) established the encoder-decoder architecture with skip connections that remains a dominant template for biomedical segmentation. Its volumetric extensions, 3D U-Net (Çiçek et al., 2016) and V-Net (Milletari et al., 2016), generalized the design to anisotropic 3D volumes, enabling direct exploitation of through-plane context in radiological and microscopic images. Attention U-Net (Oktay et al., 2018) introduced attention gates in skip connections to suppress artifacts introduced by tissue sample preparation and instrumental variation. More recently, nnU-Net demonstrated that self-configuring segmentation pipelines can achieve strong performance across diverse biomedical imaging tasks without extensive manual architecture design (Isensee et al., 2024). In standard histopathology, tissue segmentation has been successful in part because of the predominantly 2D characteristics of the data.

Registration. Classical deformable registration methods, including Demons (Thirion, 1998), B-spline free-form deformation (Rueckert et al., 1999), deformable registration on zebrafish brain (Fleishman et al., 2019), and SyN (Symmetric Normalization) in ANTs (Advanced Normalization Tools) (Avants et al., 2008), optimize image similarity and regularization iteratively but can require minutes to hours per volume pair. Learning-based methods such as VoxelMorph (Balakrishnan et al., 2019) instead predict dense deformation fields in a single forward pass, greatly accelerating registration.

Challenges. Despite recent advances, ENS models often generalize poorly because of substantial biological and imaging variability, while manual segmentation remains slow and difficult to scale. As a result, many ENS studies are limited to small regions or sample sizes, constraining population-level and disease-related analyses.

3 METHOD

3.1 MATERIALS

Data Preparation. Transgenic enteric neuron subtype-specific Cre or FlpO recombinase mice were sparsely labeled using low-dose systemic double inverse oriented (DIO) adeno-associated viruses (AAV) to deliver fluorescent proteins to a small proportion of enteric neurons. After fixation (or immunostaining, when performed), tissues were optically cleared using Ce3D immersion for 10 d and mounted on glass slides with #1 coverglass immersed in Ce3D solution. Each image dataset acquired on Nikon AXR or Olympus FV1000 laser-scanning confocal microscopes consisted of approximately 300–700 optical sections using > 1 Airy-unit pinhole values, producing the smallest cubic voxels supported by the detectors (0.83 μm for Nikon AXR and 1.0 μm for Olympus FV1000). Optical slices captured the full tissue thickness from serosa to mucosa. Tile-scanned multi-area acquisition with motorized stages (Prior) allowed capture of tissue areas with linear dimensions between 0.5 mm and 30 mm. Automated stitching created high-resolution large-area image datasets for subsequent analysis. An important characteristic of these data is that structure, texture, intensity variation, and autofluorescence differ substantially between viral labels and models (between tissue samples) and between filtered light channels (within tissue samples), so each tissue/channel emphasizes different textures.

A representative subset of volumetric image datasets (2487 single slices selected from the full dataset for variability and representation) was manually annotated by domain experts with experience in intestinal histology. Manual annotations were performed using in-house software, leveraging multi-planar visualization of confocal z-stacks to support interpretation of three-dimensional structures. Figure 1 shows representative 2D confocal slices and expert annotations.

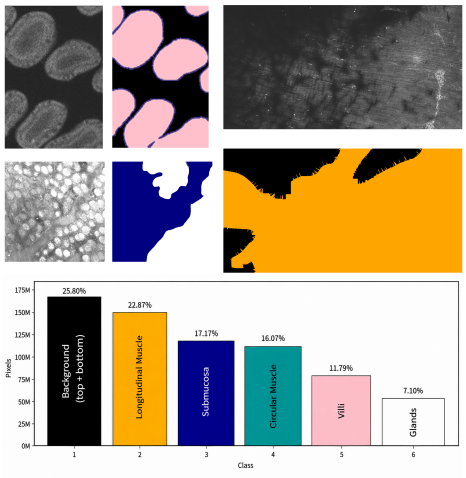

Figure 1: Single images from 3D image stacks of variable sizes obtained by cropping areas from stitched multi-area image datasets. Associated expert-annotated masks are shown alongside the images. Mask colors indicate tissue-region pixel identity, and the number and proportion of each identity are summarized in the bar chart.

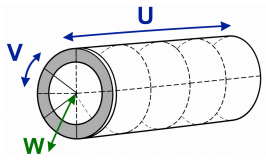

Figure 2: Canonical coordinate representation of a tissue sample. U denotes normalized longitudinal length along the intestinal centerline, V denotes periodic circumferential position within the local cross-sectional plane, and W denotes continuous transmural position defined by the ordered tissue-layer hierarchy. Together they represent the canonical coordinate representation used for deformable-organ transformation and standardized overlay.

3.2 Canonical Coordinate Formulation

Canonical coordinate mapping. GutCCF parameterizes the deformable intestine using three intrinsic coordinates (U, V, W) (Figure 2). U denotes relative longitudinal position along the intestinal centerline for cross-specimen correspondence; V denotes periodic circumferential position around the intestinal wall, represented by a normalized angular coordinate derived from the local cross-sectional angle in radians; and W denotes continuous transmural position (normalized depth) defined by the ordered tissue-layer hierarchy. The resulting mapping transforms specimen-specific voxel coordinates into a shared canonical space while remaining robust to bending, stretching, local compression, and variation in wall thickness. We represent each intestinal specimen as a deformable cylindrical domain $\Omega \subset \mathbb{R}^3$ and define the mapping

$$\Phi : \Omega \to C, \qquad \Phi(\mathbf{x}) = \big(U(\mathbf{x}), V(\mathbf{x}), W(\mathbf{x})\big), \tag{1}$$

where $C = U \times V \times W$ is the specimen-independent canonical coordinate space.

Longitudinal coordinate U. Let $\gamma : [0, L] \to \mathbb{R}^3$ denote the estimated intestinal centerline parameterized by arc length s, where L is the total centerline length. For a voxel $\mathbf{x} \in \Omega$, we determine its closest centerline location

$$s^*(\mathbf{x}) = \arg\min_{s \in [0,L]} \|\mathbf{x} - \gamma(s)\|_2^2, \tag{2}$$

and define the normalized longitudinal coordinate as

$$U(\mathbf{x}) = \frac{s^*(\mathbf{x})}{L}, \qquad U(\mathbf{x}) \in [0, 1]. \tag{3}$$

Thus, $U = 0$ and $U = 1$ represent the two longitudinal endpoints of a specimen, while equal increments in U correspond approximately to equal fractions of intestinal length rather than Euclidean distance. Physical distance in micrometers remains specimen specific and can be recovered as $s_{\mu\mathrm{m}} = UL_{\mu\mathrm{m}}$, where $L_{\mu\mathrm{m}}$ is the measured length of that specimen segment.

Circumferential coordinate V. At each centerline position, we define a local 2D cross-sectional plane perpendicular to the intestinal axis. For a voxel $\mathbf{x}$, its position within this plane is converted to an angle $\theta(\mathbf{x})$, which is normalized as

$$V(\mathbf{x}) = \frac{\theta(\mathbf{x}) + \pi}{2\pi}, \qquad V(\mathbf{x}) \in [0, 1). \tag{4}$$

Thus, V represents the voxel's circumferential position around the intestine. Because the coordinate is periodic, values near 0 and 1 correspond to neighboring locations on the circumference. This formulation remains valid even when the intestinal wall is locally compressed or noncircular.

Transmural coordinate W. Unlike a conventional radial coordinate, Euclidean distance from the centerline is not sufficient for intestinal tissue because tissue layers undergo nonuniform compression, folding, and stretching. We therefore define W from the ordered anatomical tissue hierarchy. Let $s(\mathbf{x})$ denote the class predicted for voxel $\mathbf{x}$ and let $d(\mathbf{x}) \in [0, 1]$ denote its normalized depth within that tissue layer. The canonical transmural coordinate is

$$W(\mathbf{x}) = b_{s(\mathbf{x})} + d(\mathbf{x}) \left[b_{s(\mathbf{x})+1} - b_{s(\mathbf{x})} \right], \tag{5}$$

where $0 = b_0 < b_1 < \cdots < b_K = 1$ represents the canonical boundaries of the ordered tissue layers. W describes how far a voxel lies through the intestinal wall according to anatomy rather than physical distance, allowing corresponding tissue locations to remain comparable even when tissue layers differ substantially in thickness or undergo nonlinear deformation during clearing and mounting.

3.3 Model Architecture

Multi-head U-Net for tissue and depth prediction. We formulate anatomical tissue reconstruction as a supervised multi-task learning problem using a shared U-Net encoder-decoder with two task-specific output heads (Figure 3). Intensity-normalized patches from 2D tissue slices are extracted for training with overlapping windows. Given an intestinal microscopy patch I, the network jointly predicts a tissue segmentation map $\widehat{S}$ and a continuous transmural depth map $\widehat{D}$. The segmentation head assigns each pixel or voxel $\mathbf{x}$ a probability of belonging to one of the annotated tissue classes,

$$p_\theta(s = k \mid I, \mathbf{x}), \tag{6}$$

including out-of-tissue, longitudinal muscle, circular muscle, submucosa, glands, and villi. The class with the highest probability is selected as the predicted tissue label. In parallel, the depth head predicts a normalized value $\widehat{D}(\mathbf{x}) \in [0, 1]$, indicating the voxel's relative position through the intestinal wall. Unlike semantic segmentation alone, the depth head explicitly encodes the ordinal organization of the intestinal wall. Both tasks share the same encoder-decoder representation. The encoder progressively transforms the input into multi-scale anatomical features, while skip connections preserve high-resolution boundary information. The decoder reconstructs spatially resolved feature maps, which are passed to two independent prediction heads.

This shared representation forces the network to learn features that are simultaneously discriminative for tissue identity and informative about relative anatomical position. The network is trained using the joint objective

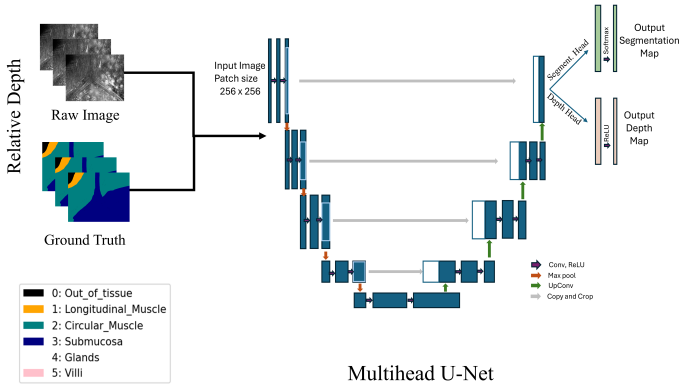

Figure 3: Multi-head U-Net for supervised anatomical representation learning. Raw confocal image patches and expert tissue annotations train a shared encoder-decoder network with two prediction heads. The segmentation head classifies intestinal tissue into six anatomical categories, while the regression head predicts a continuous transmural depth field. Domain experts inspect and correct anatomically inconsistent predictions in the HITL process.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{depth}}, \tag{7}$$

where the segmentation loss is cross-entropy,

$$\mathcal{L}_{\text{seg}} = -\frac{1}{N}\sum_{i=1}^{N} \log p_{i,y_i}, \tag{8}$$

and depth prediction is optimized using root-mean-square error,

$$\mathcal{L}_{\text{depth}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{d}_i - d_i\right)^2}. \tag{9}$$

2D predictions to 3D reconstruction. For volumetric confocal stacks, the network is applied consistently across the optical-section axis. Under a 2.5D formulation, the prediction for a slice can incorporate neighboring sections while producing segmentation and depth maps associated with the central section. The section-wise predictions are subsequently stacked to obtain volumetric fields. In this way, the model converts the raw confocal volume into a 3D semantic tissue representation and continuous anatomical-depth field. Rather than performing independent slice segmentation, joint depth supervision encourages the representation to preserve the ordering of intestinal layers across sections. This provides a stronger anatomical prior for reconstructing highly deformed tissue and supplies the transmural component W required by the GutCCF canonical coordinate system.

Human-in-the-loop refinement. To improve segmentation accuracy on anatomically ambiguous regions and rare tissue configurations, we introduce an iterative human-in-the-loop refinement stage. Starting from the initially supervised multi-head U-Net, the model is applied to previously unseen full-thickness confocal image stacks to generate a semantic tissue segmentation map and a continuous transmural depth map. Rather than manually annotating each image from scratch, domain experts inspect model predictions and correct anatomically inconsistent regions, including mislocalized tissue boundaries, missing structures, and incorrect layer assignments. Each refinement round incorporates expert knowledge obtained from errors made by the previous model.

End-to-end GutCCF pipeline. GutCCF integrates tissue segmentation, anatomical coordinate learning, human-in-the-loop refinement, and population-level atlas construction into a unified workflow. Raw 3D confocal volumes are first preprocessed and passed through the multi-head U-Net to

Table 1: Result of analysis on baseline models and the proposed model with anatomical assessment.

Model	Dice	Precision	Recall	RMSE
2D U-Net	0.6487	0.6956	0.6256	–
2.5D CNN (Multihead)				
Cross-entropy + RMSE	0.8248	0.8328	0.8185	0.2253
Uncertainty Loss	0.3390	–	–	0.2353
Focal Loss + RMSE	0.8016	0.8222	0.7848	0.0852
Consistency Loss	0.3413	–	–	0.2396
Weighted Loss + RMSE	0.6033	0.7258	0.6801	0.2089
SWINU-NetR (Encoder)	0.4726	–	–	0.2710
3D CNN	0.7277	–	–	–
Cellpose-SAM	0.5918	0.6182	0.5891	–
U-Net+GNN	0.3792	–	–	–

jointly predict semantic tissue labels and continuous transmural depth. These predictions provide the anatomical structure required to estimate the intrinsic longitudinal, circumferential, and transmural coordinates (U, V, W). The resulting specimen-specific coordinate maps are then aligned through coarse-to-fine registration into a shared canonical space. Expert-guided iterative refinement improves difficult tissue boundaries and structural consistency, while repeated mapping across animals enables construction of a population reference atlas. Once the CCF is established, neuronal morphology, vasculature, molecular measurements, and other multimodal signals can be projected into the same coordinate system, enabling standardized cross-sample and population-level spatial analysis.

Computation. We used an A100 GPU node at our in-house High-Performance Computing Cluster for model training and inference. HITL refinement was performed on a workstation with a 12th Gen Intel® Core i9-12900K CPU and NVIDIA GeForce RTX 3080 GPU.

4 EXPERIMENTS

Segmentation and depth prediction. For the supervised model, we benchmarked existing 2D and 3D segmentation models on our dataset (Table 1) and developed a multi-task U-Net model that jointly optimizes semantic segmentation and continuous transmural-depth estimation. Segmentation performance was evaluated using Dice similarity and precision, while depth prediction was assessed using root-mean-square error (RMSE). Training curves show stable convergence of both objectives over 300 epochs. The segmentation loss decreases consistently on both the training and validation sets, while validation Dice increases progressively and approaches the training performance, indicating limited train-validation divergence. After extensive analysis and experiments, our proposed multi-task model achieves a test accuracy of 81.28%, precision of 83.28%, recall of 81.85%, mDice of 82.48%, and mIoU of 70.74% across the annotated tissue classes (Figure 4). The auxiliary depth objective encourages the learned representation to preserve the ordinal organization of the intestinal wall. Whereas semantic predictions identify which tissue compartment contains a voxel, the predicted depth estimates where that voxel lies within the transmural anatomical hierarchy. This continuous representation is subsequently used to construct the W coordinate of GutCCF.

Table 1 summarizes evaluation across multiple existing models on the supervised, expert-annotated dataset. Domain experts additionally verified biological meaningfulness of inference outputs; therefore, model selection considered anatomical flow in addition to pixel-level performance.

Human-in-the-loop refinement. Human feedback qualitatively improved model performance by guiding the segmentation predictions toward more anatomically acceptable outputs over successive refinement epochs.

Overall output. An experiment was conducted on annotated enteric-neuron objects color-coded by tissue region and on automatically segmented color-coded objects. Manual annotators assigned neurites to specific regions using background autofluorescence, whereas GutCCF automatically assigned the same neurite features to tissue regions. Because expert annotations can differ across

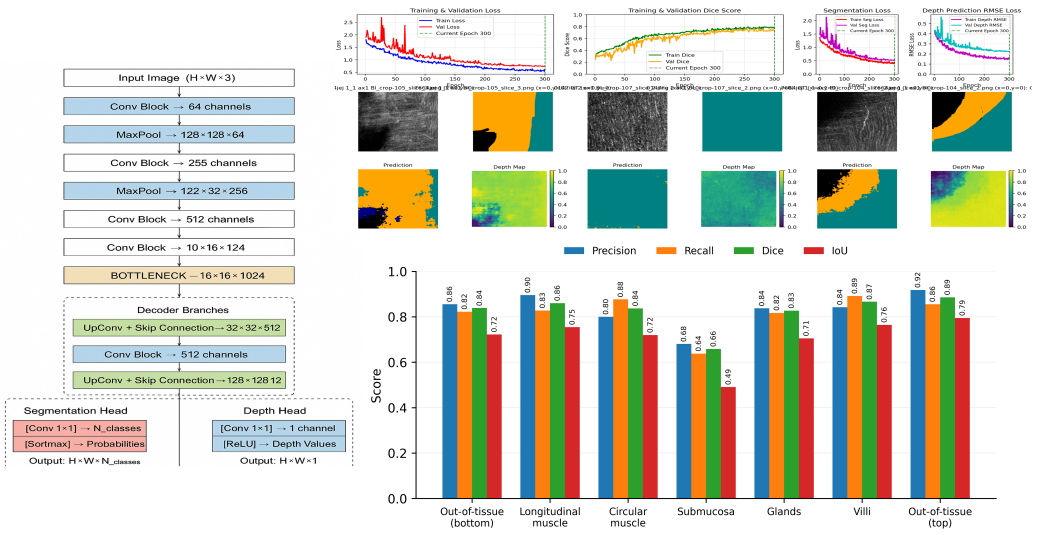

Figure 4: Quantitative and qualitative evaluation of the multi-head GutCCF model. Training and validation losses, Dice score, segmentation loss, and depth-prediction RMSE over 300 epochs are shown together with model performance metrics. The model converges on both tasks and produces spatially coherent tissue boundaries and smooth depth gradients across heterogeneous intestinal morphologies.

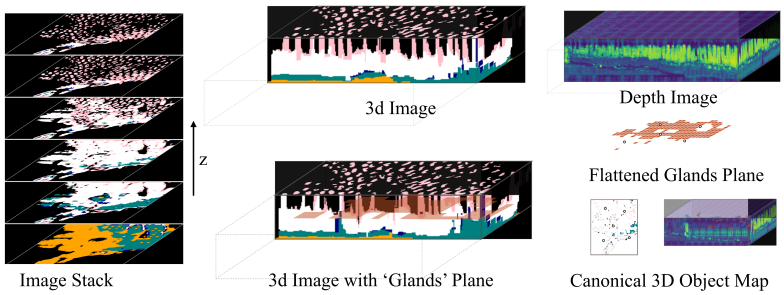

Figure 5: Canonical coordinate construction from a 3D image stack using (U, V, W). The 3D image is transformed into a volume with flattened layers representing a standardized tissue segment. U is measured along the oral-to-aboral direction, V along the circumferential direction, and the transmural position W is represented by the learned depth field.

tissue subregions, strict Dice/IoU may over-penalize plausible label variation. We therefore report part-composition similarity, yielding 0.513 ± 0.249 (mean $\pm$ SD, $n = 2$).

From the 2.5D multi-task model, the transmural depth map W is acquired together with segmentation. Coordinates U and V are obtained from tissue metadata and circumferential position, respectively. Using the depth information, a geometric transformation is applied for layer flattening of the 3D volume. Applying flattening and geometric transformation, we formulate a standardized 3D structure that provides the basis for GutCCF (Figure 5).

Because the final objective of GutCCF is not segmentation alone but canonical anatomical correspondence, we additionally evaluate whether each model preserves the expected ordering of intestinal tissue layers. For each cross-sectional profile, we measure layer-order violations and apply denoising along the predicted transmural direction.

Table 2: Ablation study with components removed from the basis model.

Model Component	Dice	RMSE
2D	0.4778	–
2.5D		
Fine Tuning (Encoder Frozen)	0.4205	0.2242
Retraining + CE loss	0.4845	0.1834
Retraining + Consistency Loss	0.3413	0.2396
Fine Tuning + Focal Loss	0.4100	0.2165
SwinU-NetR Encoder	0.4025	0.2410

5 Ablation Study

We conduct ablation experiments to quantify the contribution of the major components of GutCCF, with particular emphasis on depth-aware multi-task learning, 2.5D spatial context, and human-in-the-loop refinement. All variants use the same train/validation/test partitions and preprocessing pipeline to ensure that differences arise from the evaluated model components rather than data variation. Performance is measured using mean Dice and precision for semantic tissue segmentation and RMSE for continuous transmural depth prediction. Where applicable, we additionally evaluate anatomical consistency of the resulting canonical coordinate representation.

Table 2 shows the ablation results. This analysis tests whether improvements in conventional overlap metrics translate into more structurally valid representations for subsequent W-coordinate construction. The ablation study provides evidence for the contribution of individual components of the model, including the multi-head formulation, loss function, and depth branch. Taken together, these results support the central hypothesis of GutCCF: jointly encoding continuous transmural position and tissue identity yields representations better suited for anatomically consistent coordinate assignment in highly deformable intestinal tissue.

6 Discussion and Conclusion

GutCCF provides a solution for reproducible anatomical correspondence across highly deformable intestinal specimens. In contrast to rigid or approximately shape-preserving organs, the intestine exhibits substantial variability in length, curvature, folding, thickness, and preparation-induced distortion, making direct Cartesian registration unreliable. GutCCF instead defines anatomy in an intrinsic coordinate system, mapping each voxel to longitudinal, circumferential, and transmural coordinates (U, V, W). This representation separates anatomical identity from specimen-specific geometry and provides a common spatial reference for cross-animal comparison.

A central component of the framework is the joint learning of semantic tissue identity and continuous transmural depth. The 2.5D formulation incorporates local through-plane context while remaining computationally tractable for large confocal datasets, avoiding the memory requirements of fully volumetric 3D networks.

Several limitations remain beyond the scope of this work. Errors in segmentation, depth prediction, centerline estimation, or tissue-boundary localization can propagate into the final coordinate mapping. The current representation focuses on six major tissue compartments and does not capture finer-scale sublayers or region-specific microanatomy. Broader validation across strains, developmental stages, disease models, imaging protocols, and institutions will be required to establish robustness beyond the current dataset. Future work will therefore focus on uncertainty-aware coordinate estimation, stronger geometric consistency constraints, higher-resolution tissue hierarchies, and population-level statistical modeling of anatomical variability.

More broadly, GutCCF provides a foundation for integrating diverse biological measurements within a shared intestinal reference space. By providing consistent segmentations, geometric representations, and canonical coordinate outputs, GutCCF establishes a scalable framework for integrative enteric nervous system mapping and population-level gut atlas construction. This work further suggests a general strategy for coordinate-aware representation learning in deformable organs, with the long-term goal of supporting multimodal anatomical atlases and organ-scale digital twin models.

AI USE STATEMENT

Generative AI tools were used exclusively for copyediting to improve clarity, conciseness, and grammatical correctness. All scientific contributions, including research ideation, methodological design, experimental setup, data analysis, and interpretation of results, were conducted by the authors. The authors reviewed, edited, and assume full responsibility for the final content.

ETHICS STATEMENT

All animal procedures were reviewed and approved by the institution's Animal Use and Care Committee (IACUC) and were conducted in compliance with NIH Office of Laboratory Animal Welfare (OLAW) policies.

REPRODUCIBILITY STATEMENT

The main manuscript describes the data acquisition, annotation procedure, model architecture, training objectives, evaluation metrics, and ablation settings. Additional preprocessing and experimental details are provided in the appendix.

ACKNOWLEDGMENTS

This work was supported by NIH grant.

REFERENCES

Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432. Springer, 2016.

Greg M. Fleishman, Miaomiao Zhang, Nicholas J. Tustison, Ivan Espinosa-Medina, Yu Mu, Khaled Khairy, and Misha Ahrens. Deformable registration of whole brain zebrafish microscopy using an implementation of the flash algorithm within ANTs. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pages 213–217. IEEE, 2019.

Fabian Isensee, Paul F. Jäger, Peter M. Full, Ivo Wolf, Sandy Engelhardt, and Klaus H. Maier-Hein. nnU-Net revisited: A call for rigorous validation in 3D medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2024.

M. Krithika alias AnbuDevi and K. Suganthi. Review of semantic segmentation of medical images using modified architectures of UNET. *Diagnostics*, 12(12):3064, 2022.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

Douglas S. Richardson and Jeff W. Lichtman. Clarifying tissue clearing. *Cell*, 162(2):246–257, 2015.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

Daniel Rueckert, Luke I. Sonoda, Carmel Hayes, Derek L. G. Hill, Martin O. Leach, and David J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

Jean-Philippe Thirion. Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

Quanxin Wang, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, Maitham Naeemi, Benjamin Facer, Anh Ho, Tim Dolbeare, Brandon Blanchard, Nick Dee, Wayne Wakeman, Karla E. Hirokawa, Aaron Szafer, Susan M. Sunkin, Seung Wook Oh, Amy Bernard, John W. Phillips, Michael Hawrylycz, Christof Koch, Hongkui Zeng, Julie A. Harris, and Lydia Ng. The Allen mouse brain common coordinate framework: A 3D reference atlas. *Cell*, 181(4):936–953.e20, 2020.

David M. Young, Siavash Fazel Darbandi, Grace Schwartz, Zachary Bonzell, Deniz Yuruk, Mai Nojima, Laurent C. Gole, John L. R. Rubenstein, Weimiao Yu, and Stephan J. Sanders. Constructing and optimizing 3D atlases from 2D data with application to the developing mouse brain. *eLife*, 10:e61408, 2021.

A Supplementary Material

This appendix provides supplementary materials to support the main manuscript. We offer a more detailed interpretation of the multi-dimensional performance comparison and a comprehensive sensitivity analysis of the method's key hyperparameters.

A.1 Statement on LLM Usage

In accordance with conference guidelines, we disclose the use of a Large Language Model (LLM) during preparation of this manuscript. The LLM was employed exclusively for copyediting to improve clarity, conciseness, and grammatical correctness. All scientific contributions, including research ideation, methodological design, experimental setup, data analysis, and interpretation of results, were conducted entirely by the authors. The LLM did not generate scientific content or insights. The authors reviewed, edited, and assume full responsibility for the final content.

A.2 Data Preprocessing

A standard preprocessing pipeline was applied to all medical image data to ensure consistency across datasets and stable model training. This pipeline was uniformly applied to all methods under comparison to ensure a fair evaluation. For image resampling and spacing normalization, all images were resampled to a unified isotropic voxel spacing defined as the median voxel spacing of the dataset. Trilinear interpolation was used for intensity images, whereas nearest-neighbor interpolation was applied to segmentation masks to preserve discrete label integrity. During training, data augmentation was used to improve robustness and generalization, including rotation and geometric transformation, additive Gaussian noise, Gaussian blurring, and brightness adjustment.