

***CrossSpec*: Cross-Species Data Integration for Enhanced Multi-Organ Tissue Segmentation**

Fardeen Bablu¹

¹Department of Computer Science, Vanderbilt University, Nashville, TN, USA ,
`fardeen.e.bablu@vanderbilt.edu`

June 9, 2025

Abstract

****Work in progress and outdated abstract for now**** This report investigates the use of cross-species data integration to enhance layer segmentation for kidney pathology. Building on prior work [2], we explore incorporating both homologous and analogous structures using public datasets such as NuInsSeg [1]. This approach aims to improve model generalization and segmentation performance, especially in scenarios with limited clinical data.

Keywords: Layer Segmentation, Kidney Pathology, Cross-Species Data Integration

1. Introduction

Robust segmentation of histological images is critical in kidney pathology research. Cross-species data integration through using both homologous and analogous structures offers a promising avenue to enhance segmentation models. Homologous structures are anatomical features inherited from a common ancestor. While they may serve different functions, they share underlying structural similarities. In contrast, analogous structures perform similar functions but evolved independently and do not share a common evolutionary origin. Building on prior work [2], which focused on homologous structures, this study extends the exploration to include analogous structures as well. We hypothesize that combining both forms of biological similarity can further improve segmentation outcomes. We utilize public datasets such as NuInsSeg [1] to implement this approach. Our goal is to enhance model generalization and performance, particularly in settings where annotated clinical data is scarce.

2. Method

2.1 Cross-Species Training Framework

We implemented a multi-modal training framework supporting four distinct approaches: *separate*, *homologous*, *analogous*, and *combined* training. Following Zhu et al. [2], we employ hybrid loss

functions combining Cross Entropy Loss (CE) and Dice Loss for independent training:

$$L_s = \lambda_1 \cdot CE(y, y') + \lambda_2 \cdot Dice(y, y') \quad (1)$$

For cross-species joint training, we replace traditional Cross Entropy Loss with Focal Loss (FL) to address class imbalance, incorporating specific weights w_i for each category:

$$L_t = \lambda_3 \cdot w_i \cdot FL(y, y') + \lambda_4 \cdot Dice(y, y') \quad (2)$$

2.2 Dataset and Class Imbalance Handling

Our analysis utilized the NuInsSeg dataset [1], containing histopathological images from both human and mouse organs. The dataset exhibited significant class imbalance across organ-species combinations:

- **Kidney:** 11 human + 40 mouse images (51 total)
- **Liver:** 40 human + 36 mouse images (76 total)
- **Spleen:** 34 human + 7 mouse images (41 total)

To address this imbalance, we implemented weighted loss functions as described by Zhu et al. [2], which included ‘a specific weight w_i for each category i within the Focal Loss framework’, ultimately improving the model’s ability to handle imbalanced data by prioritizing the ‘rare’ categories.

Our class weights were calculated based on the inverse relationship between sample counts:

- **Kidney:** $w_{human} = 2.5$, $w_{mouse} = 1.0$ (compensating for 3.6× mouse majority)
- **Liver:** $w_{human} = 1.0$, $w_{mouse} = 1.1$ (roughly balanced)
- **Spleen:** $w_{human} = 1.0$, $w_{mouse} = 4.8$ (compensating for 4.8× human majority)

This weighting strategy prevents model bias toward majority classes and ensures balanced learning across species, particularly critical for mouse spleen data with only 7 samples compared to 34 human spleen samples.

2.3 Model Architecture

We evaluated three segmentation architectures: U-Net, PSPNet, and DeepLabv3+, using simplified implementations with ResNet50 backbones for feature extraction. Models were trained with 1024×1024 input resolution to preserve fine-grained structural details essential for accurate layer segmentation.

References

- [1] Amirreza Mahbod, Christine Polak, Katharina Feldmann, Rumsha Khan, Katharina Gelles, Georg Dorffner, Ramona Woitek, Sepideh Hatamikia, and Isabella Ellinger. Nuinsseg: A fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. *Scientific Data*, 11(1):295, March 2024. doi:10.1038/s41597-024-03117-2.
- [2] Junchao Zhu, Mengmeng Yin, Ruining Deng, Yitian Long, Yu Wang, Yaohong Wang, Shilin Zhao, Haichun Yang, and Yuankai Huo. Cross-species data integration for enhanced layer segmentation in kidney pathology, August 2025. URL: <https://arxiv.org/abs/2408.09278>, arXiv:2408.09278.