Deep Learning-Based Multi-Organelle Segmentation in Human Liver Electron Microscopy Images

KAUST Academy Summer Research Program 2025

Student: Atheer Alwuthaynani

SUDS Summer Scholar, Data Science Institute, University of Toronto

Project Mentor: Prof. Gary D. Bader

Co-supervisor: Cheng Xing

Bader Lab, Department of Molecular Genetics, University of Toronto

Project Duration: June 1, 2025 - August 16, 2025

August 18, 2025

Contents

1 Project Information **3**

2 Project Objectives and Goals **3**

2.1 Primary Objectives 3

2.2 Specific Goals 3

3 Methodology and Approach **4**

3.1 Data Preparation and Annotation 4

3.2 Deep Learning Framework 4

3.3 Evaluation Protocol 5

3.4 Annotation Quality and Challenges 5

4 Current Progress and Milestones Achieved **7**

4.1 Initial Phase (June 2025): Foundation and Single Organelle Work 7

4.2 Intermediate Phase (July 2025): Model Development and Evaluation 8

4.3 Advanced Phase (July-August 2025): Multi-Organelle Integration 8

4.4 Error Analysis and Pattern Recognition 9

4.5 Technical Achievements and Computational Challenges 9

5 Results and Findings **10**

5.1 Quantitative Performance Results 10

5.2 Key Findings 14

6 Challenges and Obstacles Encountered **14**

6.1 Annotation Complexity and Biological Challenges 14

6.2 Technical and Methodological Obstacles 15

6.3 Data Quality and Consistency Issues 15

7 Next Steps and Future Work **16**

7.1 Immediate Next Steps 16

7.2 Long-term Objectives 16

8 Resources and Tools Used 16

8.1 Software and Frameworks 16

8.2 Hardware 17

8.3 Data Sources 17

8.4 Source Code and Reproducibility 17

9 Project Impact and Contributions 17

9.1 Scientific Contributions 17

9.2 Technical Contributions 18

10 Acknowledgments 18

11 Conclusion 19

1 Project Information

Project Title	Deep Learning-Based Multi-Organelle Segmentation in Human Liver Electron Microscopy Images
Student	Atheer Alwuthaynani
Institution	SUDS Summer Scholar, Data Science Institute, University of Toronto
Project Mentor	Prof. Gary D. Bader
Co-supervisor	Cheng Xing
Lab	Bader Lab, Department of Molecular Genetics, University of Toronto
Duration	June 1, 2025 - August 16, 2025

2 Project Objectives and Goals

2.1 Primary Objectives

1. **Generate high-quality labeled datasets** for training AI models in electron microscopy (EM) image segmentation
2. **Develop an automated multi-organelle segmentation pipeline** using deep learning techniques
3. **Evaluate model performance** on human liver tissue samples across five organelle types
4. **Establish a scalable workflow** for EM image analysis that can accelerate biological research

2.2 Specific Goals

- Train a 2D nnU-Net model for multi-class segmentation of liver organelles

- Achieve accurate identification of mitochondria, nucleus, endoplasmic reticulum (ER), lipid droplets, and cell boundaries
- Quantitatively evaluate segmentation performance using standard metrics (Dice coefficient, IoU, F1 score)
- Create reproducible analysis pipeline for large-scale EM studies

3 Methodology and Approach

3.1 Data Preparation and Annotation

- **Manual Annotation:** Used VAST software for expert-guided annotation of cellular organelles
- **Ground Truth Creation:** Single high-quality training image with comprehensive multi-organelle labels
- **Data Format:** Converted TIF/PNG images to NIfTI format for nnU-Net compatibility
- **Multi-class Mask Generation:** Combined individual organelle masks into unified segmentation labels

3.2 Deep Learning Framework

- **Model Architecture:** 2D nnU-Net (self-configuring biomedical segmentation network)
- **Training Strategy:** Single image training with data augmentation
- **Class Labels:**
  - Background (0)
  - Mitochondria (1)
  - Nucleus (2)
  - Endoplasmic Reticulum (3)

- Lipid Droplets (4)
- Cell Boundaries (5)

3.3 Evaluation Protocol

- **Test Set:** 3 manually annotated images not seen during training
- **Metrics:** Dice Similarity Coefficient, Intersection over Union (IoU), Precision, Recall, F1 Score, Specificity, Accuracy
- **Error Analysis:** Pixel-wise confusion analysis between organelle classes

3.4 Annotation Quality and Challenges

The manual annotation process revealed important insights about organelle distribution and annotation complexity in EM images.

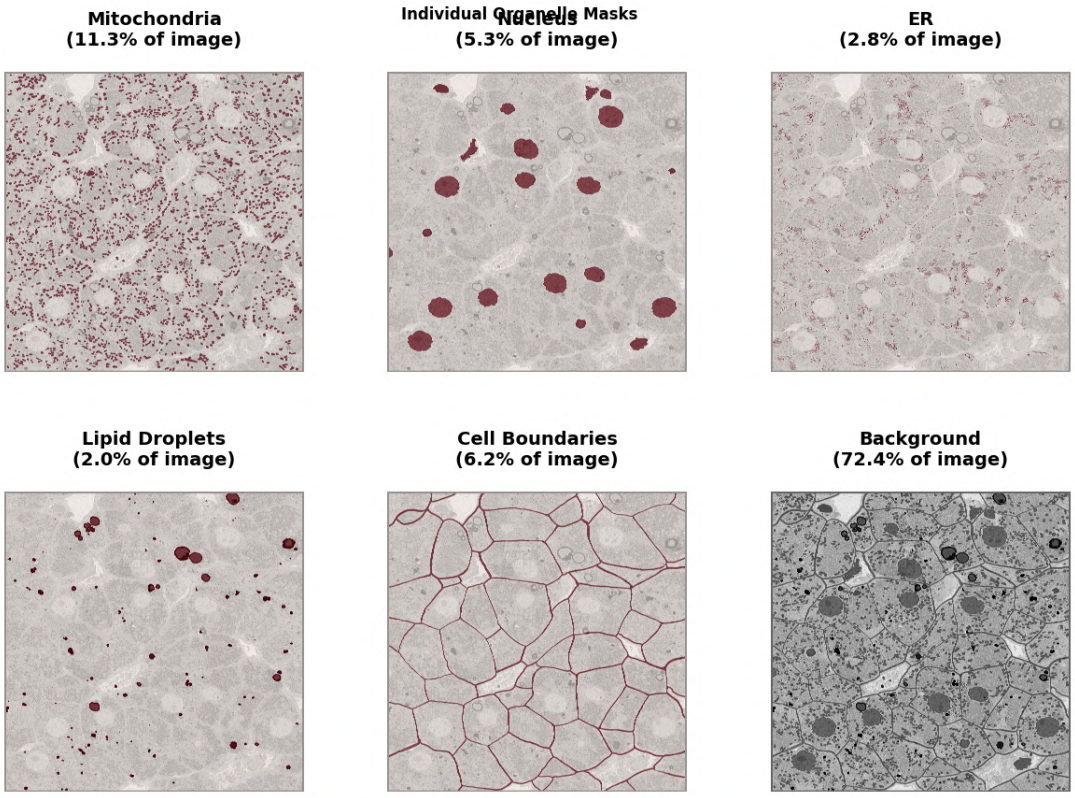


Figure 2: Individual Organelle Masks: Detailed breakdown of manual annotations showing each organelle type separately overlaid on the original EM image. Percentages indicate the relative area coverage of each organelle class in the training image, demonstrating the class distribution used for model training.

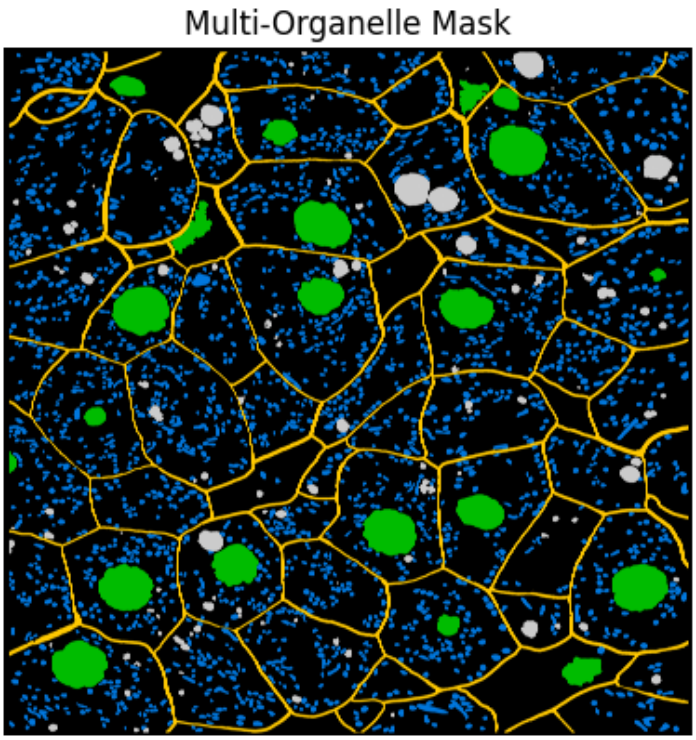


Figure 3: Multi-Organelle Combined Mask: Unified segmentation mask showing all five organelle types in a single image. Color coding: blue (mitochondria), white (nucleus), green (lipid droplets), with ER and cell boundaries also visible. This combined mask serves as ground truth for multi-class model training.

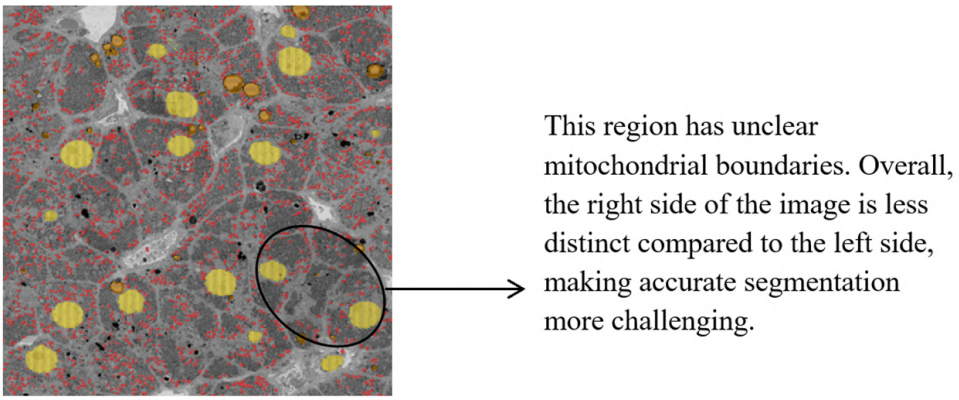


Figure 4: Annotation Challenges Example: Illustration of segmentation difficulties encountered during manual annotation. The circled region highlights unclear mitochondrial boundaries where structures become less distinct, demonstrating the inherent challenges in creating precise ground truth labels for complex cellular regions.

As shown in the individual organelle masks, the training image contained varying proportions of different organelles, with background comprising the majority (72.4%) and mitochondria being the most abundant organelle type (11.3%). This class imbalance influenced

our evaluation strategy and model performance expectations. The annotation challenges highlighted in Figure 4 demonstrate the complexity of creating precise ground truth labels, particularly in regions where organelle boundaries become ambiguous.

4 Current Progress and Milestones Achieved

4.1 Initial Phase (June 2025): Foundation and Single Organelle Work

The project began with comprehensive literature review and domain knowledge acquisition, focusing on understanding cellular organelle morphology and annotation techniques. Using the VAST annotation tool, manual labeling commenced with a single EM slice from human liver tissue with 8 nm isotropic pixel size, initially targeting nucleus and mitochondria structures.

Key insights were gained from analyzing "Regulation of liver subcellular architecture controls metabolic homeostasis" (Parlakgül et al., 2022), which provided understanding of how cellular organelles like the endoplasmic reticulum and mitochondria undergo structural remodeling under different metabolic conditions. This biological knowledge proved essential for accurate annotation of organelle structures.

The first major milestone was completing mitochondria labeling, though this process revealed significant annotation challenges. Unclear boundaries in certain image regions made precise segmentation difficult, particularly when structures were tightly packed or partially overlapping. These challenges were systematically documented and addressed through iterative refinement protocols, with reference to established protocols from the OpenOrganelle database for validation.

Initial nnU-Net model training was successfully implemented following the official pipeline, with TIF and PNG files converted to NIfTI format for compatibility. The model architecture utilized 2D U-Net with early stopping mechanisms to prevent overfitting during the 1-hour training sessions on Google Colab's A100 GPU.

4.2 Intermediate Phase (July 2025): Model Development and Evaluation

Following initial model training, quantitative evaluation revealed promising results for single organelle segmentation. The mitochondria-only model achieved a Dice coefficient of 96.29% and Intersection over Union of 92.84% when evaluated on the training image, demonstrating the feasibility of the approach despite limited training data.

However, this evaluation highlighted a critical methodological concern: testing on the same image used for training introduced potential overfitting bias. This observation led to the development of a more rigorous evaluation protocol requiring independent test images with manual ground truth annotations.

The abstract development process during this phase involved multiple iterations, with particular attention to clarifying the biological focus on human liver tissue and emphasizing the gap in available open source datasets for human studies compared to animal models.

4.3 Advanced Phase (July-August 2025): Multi-Organelle Integration

The project evolved significantly with expansion to multi-organelle segmentation, incorporating mitochondria, nucleus, endoplasmic reticulum, lipid droplets, and cell boundaries. This required substantial technical development, including mask format standardization and multi-class integration protocols.

A critical breakthrough was the creation of proper test sets using images 120, 164, and 208 from a stack of 129 sliding sections. These images were carefully selected as representatives from three equal divisions of the dataset and manually annotated for all five organelle types, ensuring completely independent evaluation data.

The multi-organelle model revealed distinct performance patterns across organelle types. Lipid droplets emerged as the best-performing class with a Dice coefficient of 0.806 ± 0.045, benefiting from their simple spherical morphology and high contrast against cytoplasmic background. Nucleus segmentation achieved solid performance (0.699 ± 0.049) due to distinct boundary features.

In contrast, endoplasmic reticulum segmentation presented the greatest challenge, achiev-

ing only 0.182 ± 0.025 Dice coefficient. Detailed analysis revealed that while the model showed high precision (0.775), recall was extremely low (0.104), indicating conservative predictions that accurately identified ER structures but missed a significant portion of actual ER networks.

4.4 Error Analysis and Pattern Recognition

Comprehensive pixel-wise error analysis across the three test images revealed systematic misclassification patterns. The most significant finding was frequent confusion between mitochondria and nucleus, with up to 190,171 pixels misclassified in a single test image. This suggests the model relies heavily on shared intensity characteristics between these organelles, particularly in regions where mitochondria cluster around the nuclear envelope.

Additional confusion patterns emerged between mitochondria and cell boundaries, likely in regions where organelles overlap with membrane structures or where annotation completeness varies near slice borders.

4.5 Technical Achievements and Computational Challenges

The project successfully implemented a complete pipeline from VAST annotations to nnU-Net training and evaluation. Technical accomplishments included format conversion protocols, multi-class mask generation, comprehensive evaluation metrics implementation, and visualization frameworks for result analysis.

Computational challenges were addressed through Google Colab Pro utilization with A100 GPU access, though extended training requirements for multi-organelle models highlighted the need for more robust computational resources. Initial attempts to access KAUST's Ibex computational cluster were made to address these limitations.

3D visualization capabilities were developed using ITK-SNAP with proper voxel spacing configuration (8×8×75 nm), though technical challenges remained in optimal visualization techniques for complex organelle morphology across multiple slices.

5 Results and Findings

5.1 Quantitative Performance Results

Table 1: Model Performance Across Different Organelles

Organelle	Dice	IoU	Precision	Recall	F1 Score
Lipid Droplets	0.806 ± 0.045	0.678 ± 0.063	0.863 ± 0.078	0.758 ± 0.051	0.807 ± 0.053
Nucleus	0.699 ± 0.049	0.539 ± 0.068	0.599 ± 0.078	0.835 ± 0.059	0.699 ± 0.049
Mitochondria	0.616 ± 0.045	0.447 ± 0.057	0.521 ± 0.063	0.748 ± 0.073	0.616 ± 0.045
Cell Boundaries	0.473 ± 0.037	0.311 ± 0.046	0.362 ± 0.063	0.673 ± 0.142	0.473 ± 0.037
ER	0.182 ± 0.025	0.100 ± 0.015	0.775 ± 0.162	0.104 ± 0.015	0.182 ± 0.025

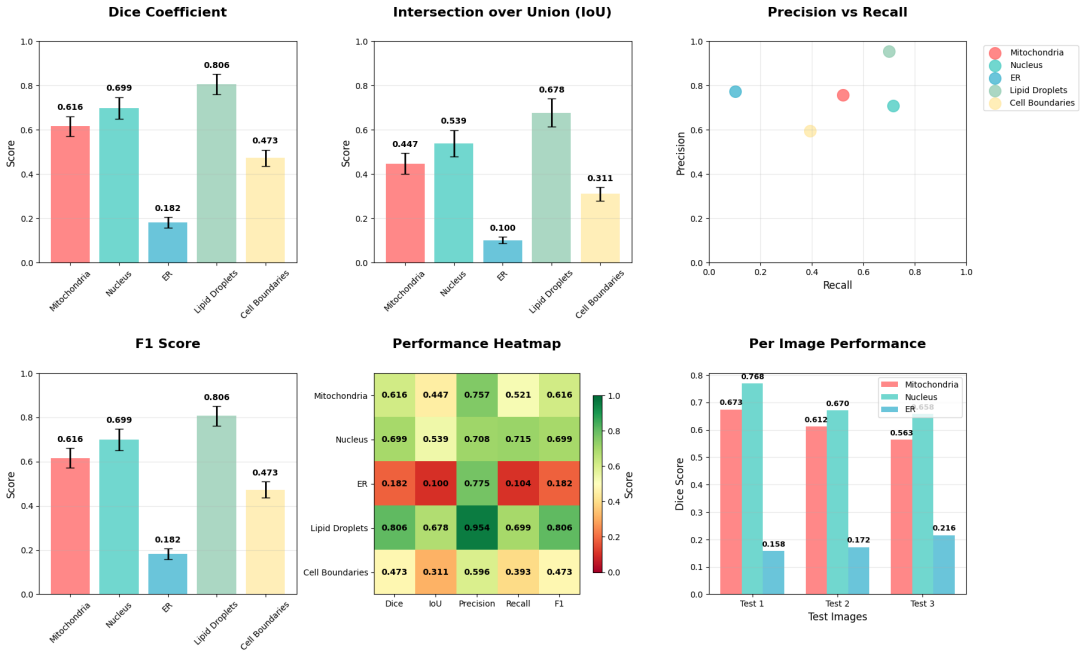


Figure 5: Comprehensive Performance Analysis showing Dice coefficients, IoU scores, precision vs recall, F1 scores, performance heatmap, and per-image comparison across all organelle types.

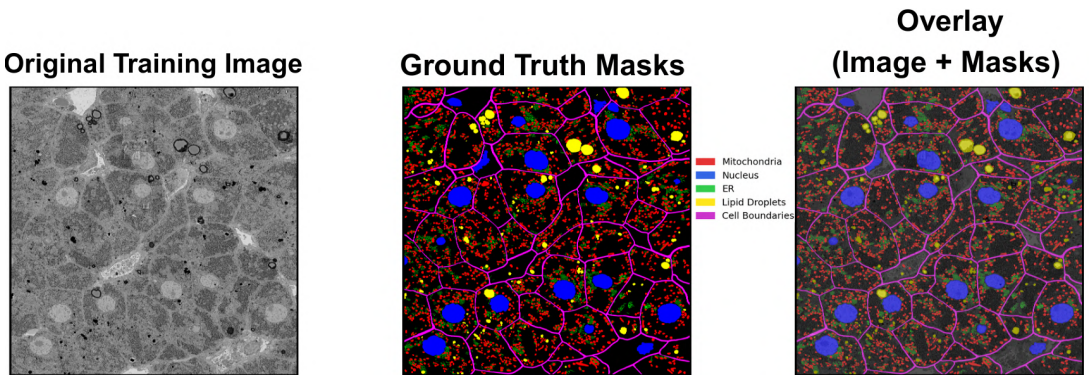


Figure 6: Training Image Analysis: Original EM image, ground truth annotations, and overlay showing the single manually annotated image used for model training.

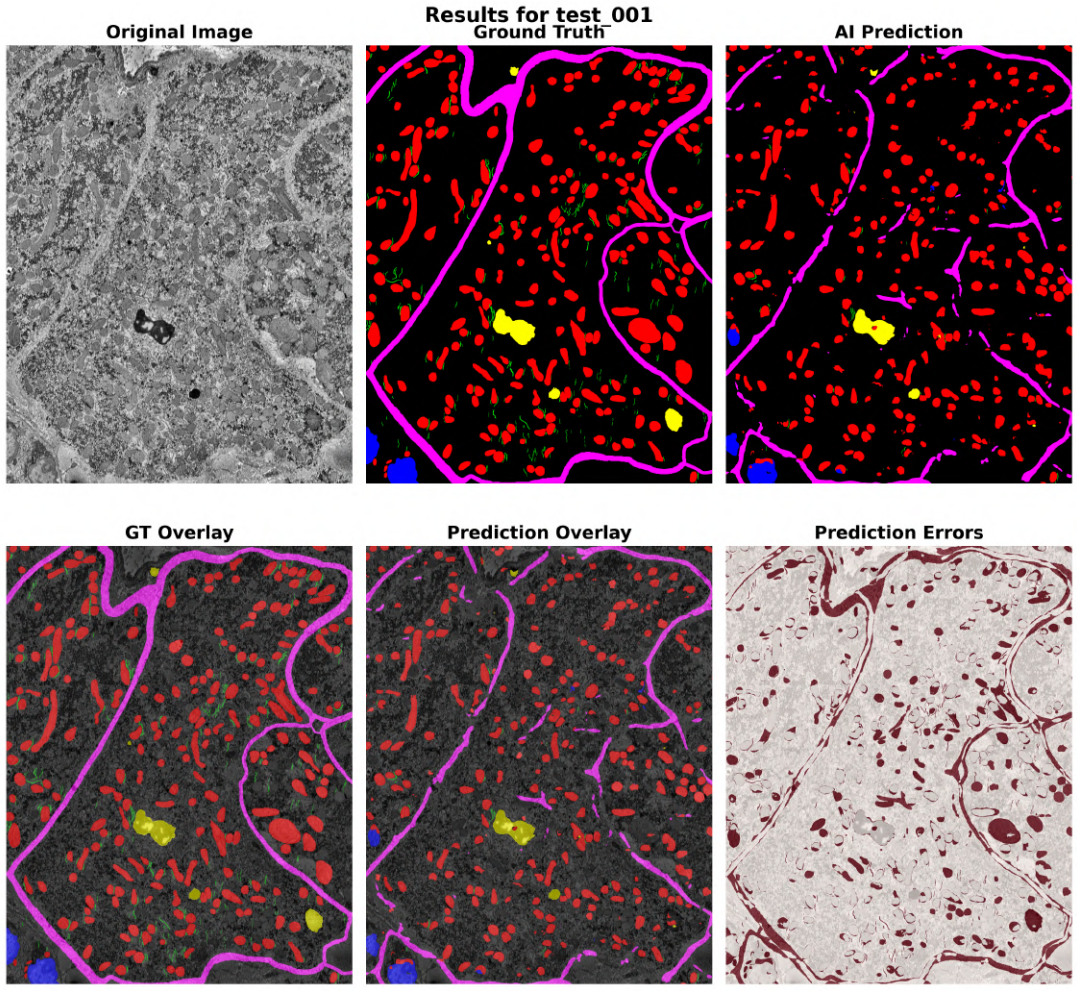


Figure 7: Test Results - Image 1: Model performance evaluation on first test image with error analysis visualization.

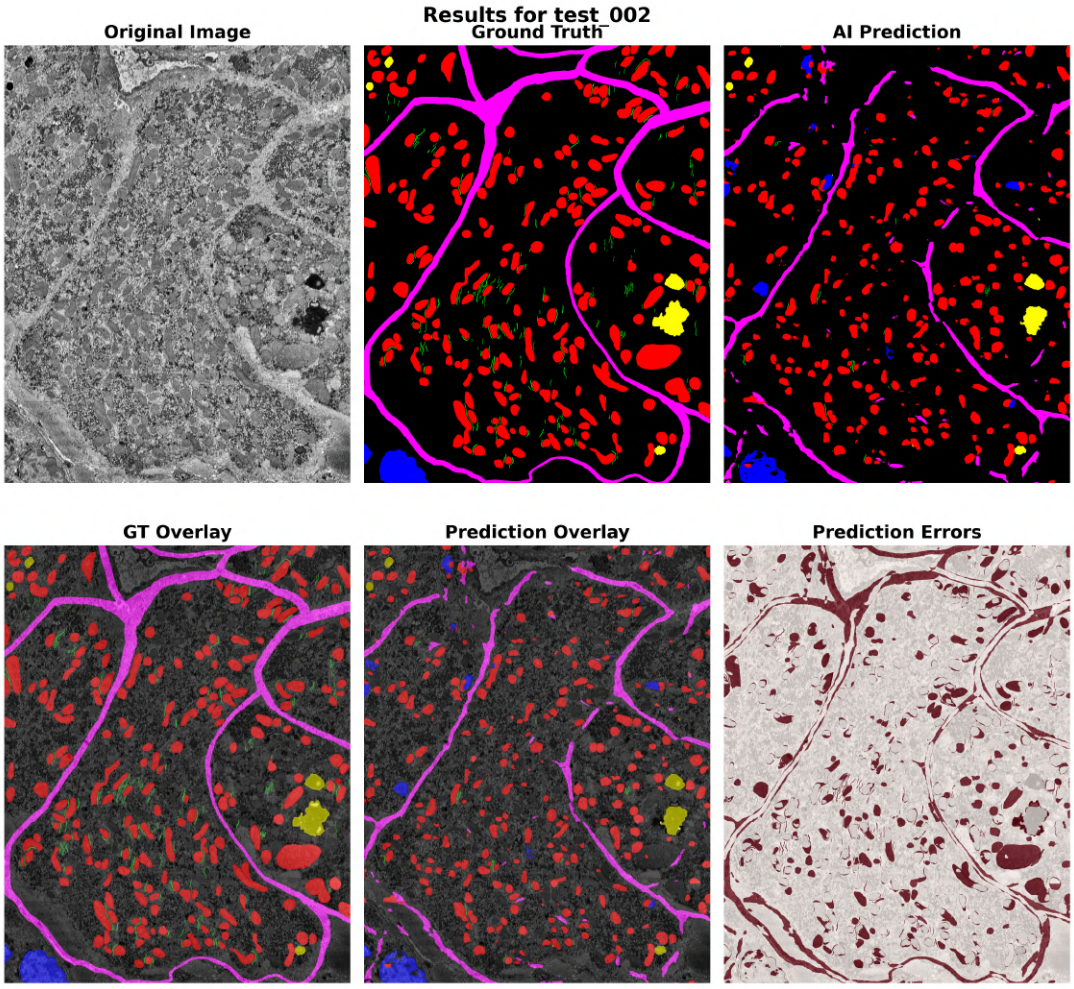


Figure 8: Test Results - Image 2: Model performance evaluation on second test image with error analysis visualization.

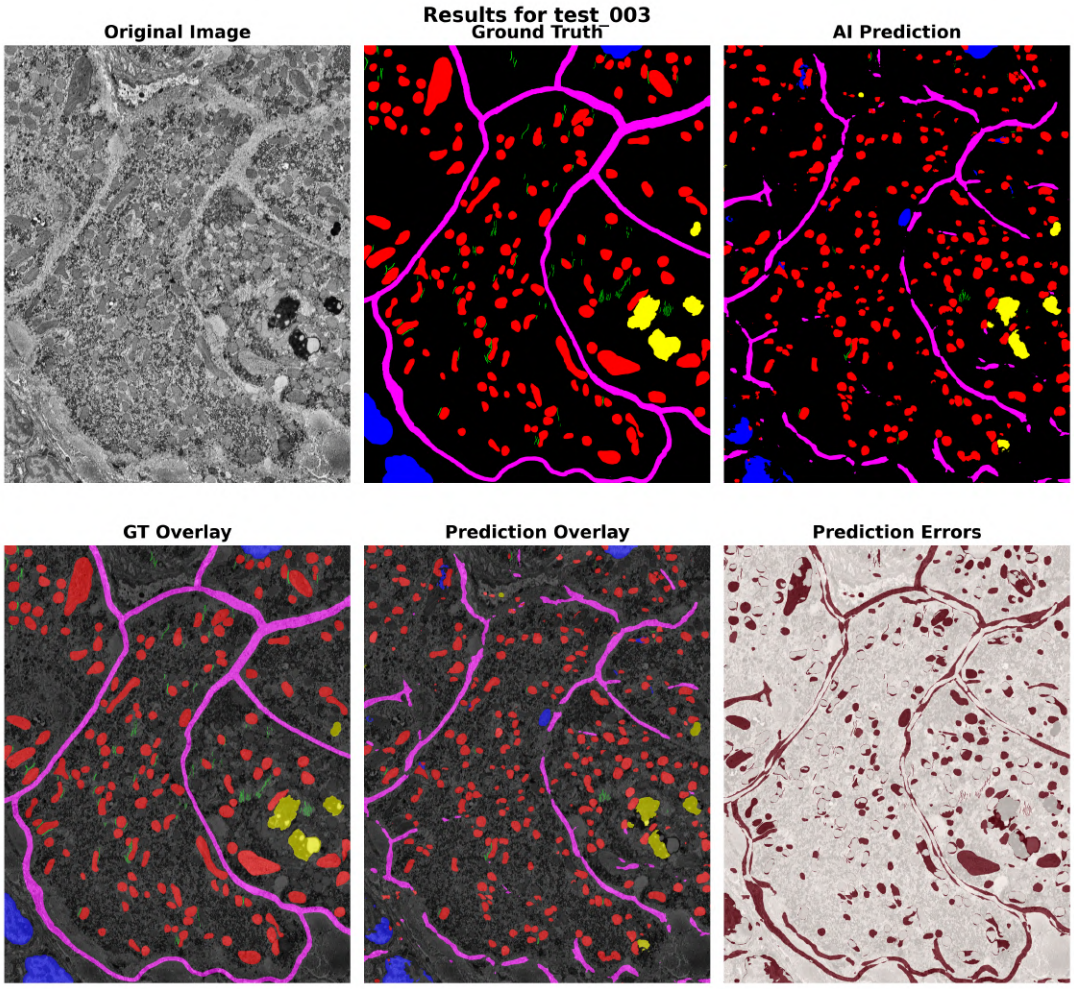


Figure 9: Test Results - Image 3: Model performance evaluation on third test image with error analysis visualization.

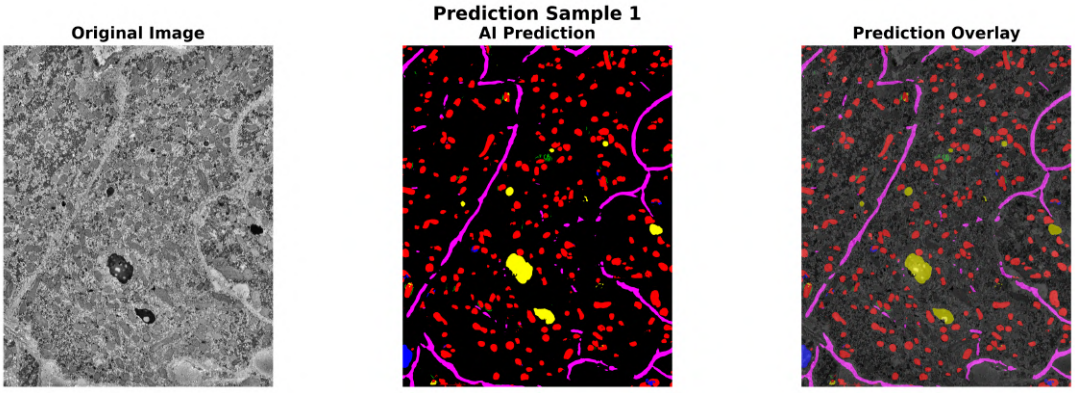


Figure 10: Prediction Sample 1: AI model segmentation results on unseen EM images demonstrating automated organelle identification capabilities.

5.2 Key Findings

1. **Morphology Impact:** Organelles with distinct shapes (lipid droplets, nuclei) achieved highest accuracy
2. **Complex Structures:** ER segmentation was most challenging due to intricate tubular network
3. **Precision vs Recall Trade-off:** ER showed high precision but low recall, indicating conservative predictions
4. **Single Image Training:** Demonstrated feasibility of training with limited high-quality annotated data

6 Challenges and Obstacles Encountered

6.1 Annotation Complexity and Biological Challenges

The manual annotation process revealed significant biological complexity that directly impacted segmentation quality. Organelle boundary ambiguity emerged as a primary challenge, particularly in regions where structures were tightly packed or partially overlapping. This was especially problematic for mitochondria, where determining whether closely adjacent structures represented single or multiple organelles proved difficult without complete 3D contextual information.

The endoplasmic reticulum presented unique annotation difficulties due to its intricate tubular network architecture. The interconnected sheets and tubules often appeared similar to background noise, particularly when approaching the resolution limits of electron microscopy. This complexity was reflected in the model's conservative prediction behavior, achieving high precision but low recall for ER structures.

Lipid droplets, while generally spherical, exhibited boundary challenges where edges faded into surrounding cytoplasm, requiring careful annotation decisions about precise boundary locations. Cell boundaries proved particularly challenging near slice borders where anno-

tation completeness varied due to incomplete structural context.

6.2 Technical and Methodological Obstacles

VAST software limitations created significant workflow complications, particularly the inability to export multi-organelle masks directly in proper format. This required developing custom pipelines to export each organelle separately and programmatically combine them into unified multi-class masks, adding complexity to the annotation workflow.

Computational resource constraints emerged as training progressed from single to multi-organelle models. Google Colab's session limitations required implementing timeout mechanisms and careful session management, while the increased complexity of multi-class models significantly extended training duration even with A100 GPU acceleration.

The evaluation methodology itself presented challenges, particularly the initial oversight of testing on training data, which risked overfitting bias. This led to the development of more rigorous evaluation protocols requiring completely independent test sets with manual ground truth annotations.

6.3 Data Quality and Consistency Issues

Ensuring annotation consistency across different cellular regions proved challenging, particularly when organelle morphology varied due to local cellular conditions or imaging artifacts. The class imbalance inherent in cellular imagery, with background comprising 72.4% of pixels while some organelles represented less than 3%, required careful consideration in evaluation metric interpretation.

Resolution limitations of electron microscopy created fundamental challenges for certain structures, particularly the endoplasmic reticulum where fine tubular networks approached the imaging resolution limit. This technical limitation directly impacted both manual annotation accuracy and automated segmentation performance.

7 Next Steps and Future Work

7.1 Immediate Next Steps

1. **Dataset Expansion:** Annotate additional training images to improve generalization
2. **Human-in-the-Loop Refinement:** Implement iterative annotation correction workflow
3. **ER Segmentation Optimization:** Develop specialized approaches for complex tubular structures
4. **3D Volume Analysis:** Complete 3D reconstruction and morphological analysis

7.2 Long-term Objectives

1. **Cross-validation Study:** Implement proper train/validation/test splits with larger dataset
2. **Multi-scale Analysis:** Investigate different resolution approaches for various organelles
3. **Comparative Model Study:** Evaluate alternative architectures (U-Net variants, Vision Transformers)
4. **Clinical Translation:** Extend methodology to disease state analysis and metabolic studies

8 Resources and Tools Used

8.1 Software and Frameworks

- **VAST:** Volume annotation tool for EM data
- **nnU-Net v2:** Self-configuring deep learning framework
- **ITK-SNAP:** 3D volume visualization and analysis
- **Google Colab:** Cloud computing platform with GPU acceleration

- **Python Libraries:** SimpleITK, PIL, NumPy, Matplotlib, Seaborn

8.2 Hardware

- **Google Colab Pro:** A100 GPU for model training
- **KAUST Ibex Cluster:** (Attempted access for extended training)

8.3 Data Sources

- **Human Liver EM Images:** High-resolution tissue samples
- **OpenOrganelle Database:** Reference annotations for validation
- **Expert Annotations:** Manual ground truth labels from domain specialists

8.4 Source Code and Reproducibility

- **Project Repository:** All source code, notebooks, and documentation are available at: https://github.com/atheerwuthaynani/liver-organelle-segmentation
- **Google Colab Notebooks:** Interactive notebooks for data processing, model training, and evaluation
- **Data Processing Pipeline:** Complete workflow from VAST annotations to nnU-Net training
- **Evaluation Scripts:** Comprehensive metrics calculation and visualization tools

9 Project Impact and Contributions

9.1 Scientific Contributions

1. **Methodology Development:** Established single-image training protocol for EM segmentation

2. **Performance Benchmarking:** Comprehensive evaluation across multiple organelle types
3. **Workflow Documentation:** Reproducible pipeline from annotation to evaluation
4. **Error Analysis Framework:** Systematic approach to understanding segmentation failures

9.2 Technical Contributions

1. **Multi-class Integration:** Combined individual organelle masks into unified segmentation
2. **Evaluation Pipeline:** Comprehensive metrics calculation and visualization
3. **Format Conversion Tools:** EM image to nnU-Net compatible data processing
4. **Quality Assessment:** Error pattern analysis for model improvement

10 Acknowledgments

We gratefully acknowledge the contributions of:

- **Data Generation Team:** Ali Darbandi, Ben Mulcahy, Ian D. McGilvray, Mei Zhen, Ronald Xie, and Sonya MacParland
- **Funding Support:** KAUST Academy and the Data Sciences Institute at the University of Toronto
- **Technical Support:** Bader Lab members and SUDS program coordinators
- **Computing Resources:** Google Colab and University of Toronto computational facilities

11 Conclusion

This project successfully demonstrates that high-quality manual annotations from a small dataset can effectively train deep learning models for EM organelle segmentation. The nnU-Net framework proved adaptable across different organelle types, offering a promising approach for automated cellular analysis.

Key achievements include:

- Establishment of a scalable annotation and training pipeline
- Quantitative evaluation revealing organelle-specific performance patterns
- Identification of morphological factors affecting segmentation accuracy
- Foundation for future human-in-the-loop refinement approaches

The work provides a solid foundation for large-scale studies of hepatocyte organelle architecture and metabolic regulation that would be prohibitively time-consuming with manual annotation alone.

Future efforts will focus on expanding the annotated dataset, improving accuracy for complex structures like the ER, and extending validation to 3D structural analysis.

Submitted by: Atheer Alwuthaynani

Date: August 18, 2025

KAUST Academy Summer Research Program 2025