Two-Stage Learning Approach for Glioblastoma Cell Segmentation

Yusuf ATA

Content

Introduction & Motivation

- **Context**
  - Glioblastoma is one of the most aggressive brain tumors
  - Understanding cell behavior is crucial for treatment and research
- **Importance of Segmentation**
  - Helps track cell division, shape changes, and motion
  - Enables analysis of tumor growth and drug response
- **Challenges**
  - Images often suffer from low contrast and noise
  - Manual labeling is expensive and time-consuming
  - Need for automatic and reliable segmentation methods
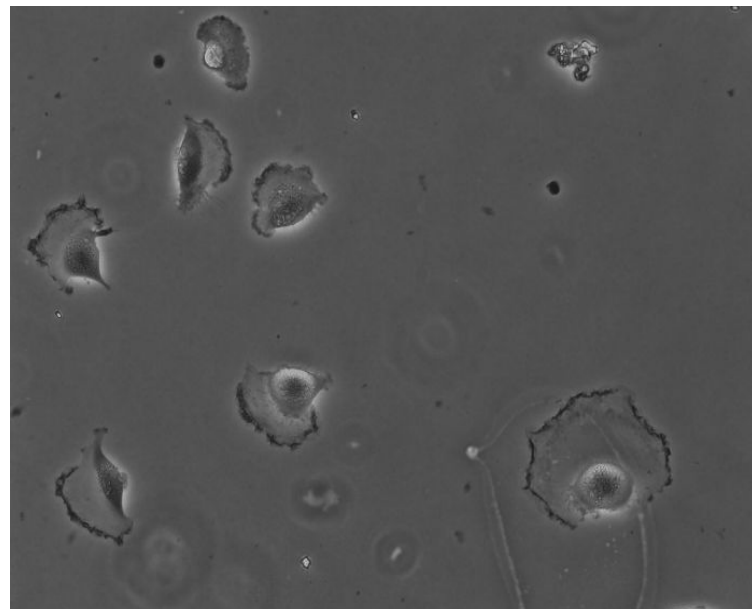
Glioblastoma-astrocytoma U373 cells image example

Problem Definition & Hypothesis

- **Annotation Problem**
  - Gold Truth (GT): highly accurate, but limited and costly
  - Silver Truth (ST): abundant, automatically generated, but noisy
- **Core Question**
  - How to train a robust model using many noisy (ST) and few accurate (GT) masks?

- **Hypothesis**
  - Two-stage learning can first teach general structure with ST
  - Then correct errors and refine with GT for high accuracy
  - Reduces dependency on expert labels while maintaining performance

Original Image   Reference Mask (ST)   Reference Mask (GT)
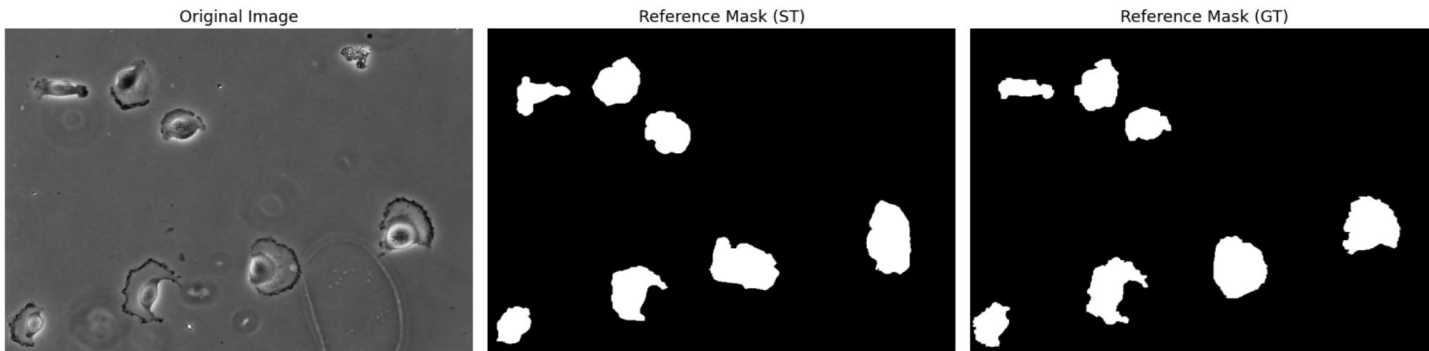
ST Mask vs GT Mask

Literature Review

Model Foundations

- U-Net: widely used in medical imaging; encoder-decoder with skip connections
- nnU-Net: adaptive configuration for any medical dataset

Labeling Strategies

- Semi-supervised learning: combines labeled and unlabeled data
- Multi-stage pipelines: pretraining and refinement stages

Identified Gap

- Limited studies on using both ST and GT in a structured two-phase pipeline for glioblastoma segmentation
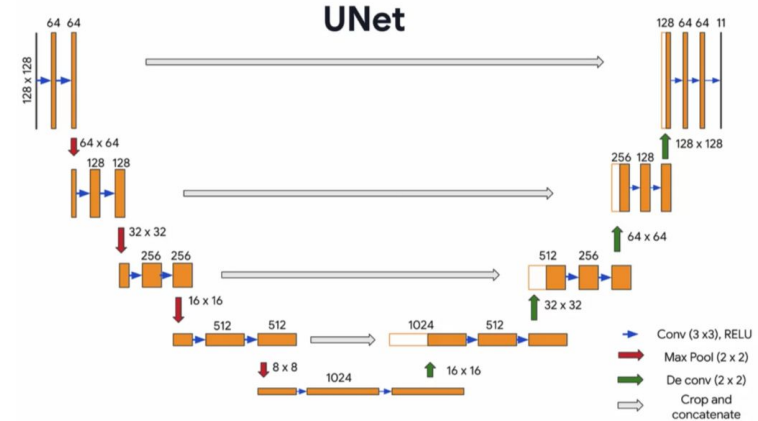
U-Net Architecture Example

Dataset & Preprocessing

Dataset: PhC-C2DH-U373

- 230 grayscale phase-contrast microscopy images
- ST masks: auto-generated for all frames
- GT masks: manually annotated, limited subset

Preprocessing Techniques

- CLAHE for contrast enhancement
- Data augmentations: flip, rotate, elastic transformation
- Padding for U-Net input size constraints

Data Management

- Custom data loader (CTCDataManager) organizes image-mask pairs
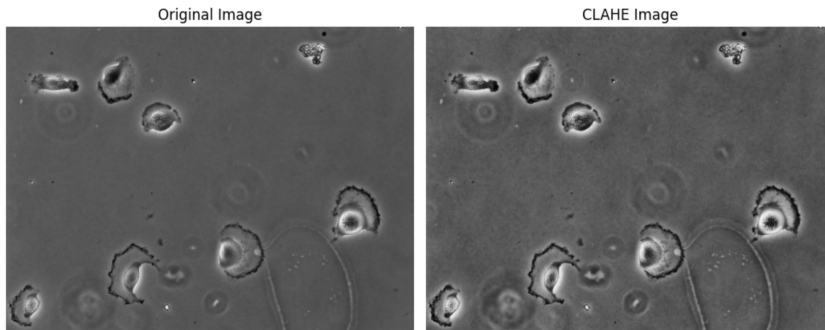
Original image vs CLAHE applied image

Two-Stage Learning Pipeline
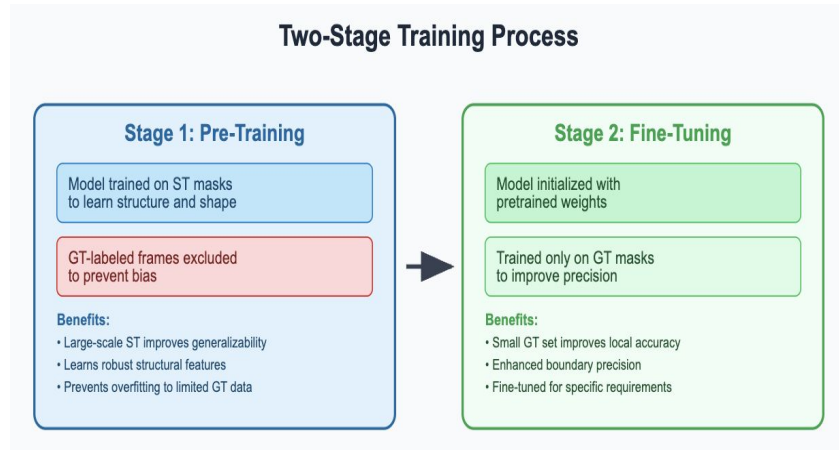
Stage 1: Pre-Training

- Model trained on ST masks to learn structure and shape
- GT-labeled frames excluded to prevent bias

Stage 2: Fine-Tuning

- Model initialized with pretrained weights
- Trained only on GT masks to improve precision

Benefits

- Large-scale ST improves generalizability
- Small GT set improves local accuracy and boundary precision

Two-Stage Training Process Flowchart

Model Architecture

Architecture Overview

- **Framework:** MONAI (Medical Open Network for AI)
- **Purpose:** Segmentation of Glioblastoma (U373) Cells from PCM Images
- **Input:** CLAHE-enhanced Grayscale PCM Images
- **Output:** Binary Mask of Cells (Probability Map)

Why U-Net & MONAI?

- High performance on biomedical segmentation
- Efficiently learns from small datasets
- Skip connections preserve spatial resolution
- MONAI's optimized medical imaging modules

Encoder	Bottleneck	Decoder
4 Levels: 64→128→256→512	1024 Channels	**4 Levels:** 512→256→128→64
Strides (2,2,2,2)	Feature extraction	Restores spatial details

Detailed Structure
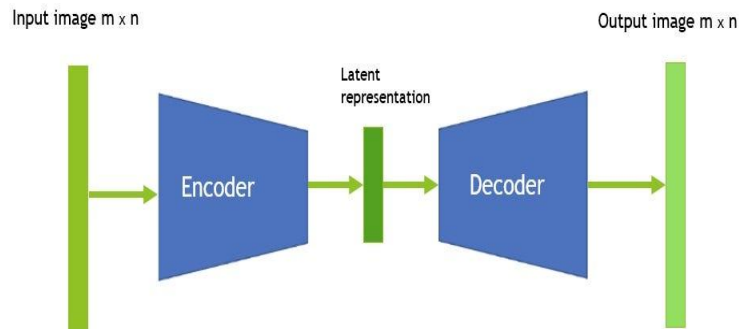
Basic U-Net Diagram

Evaluation Metrics
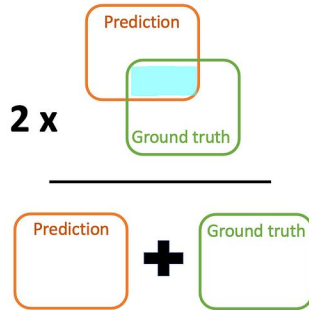
Quantitative Metric

- **Dice Coefficient**
  - Measures overlap between prediction and ground truth

Qualitative Evaluation

- Visual inspection by overlaying masks on raw images
- Helps interpret false positives, under-segmentation, boundary issues

Why Dice?

- More informative than accuracy for imbalanced data (small object regions)

$$Dice = \frac{2 \: X \: Area \: of \: overlap}{Total \: area} =$$

Experimental Setup

Environment

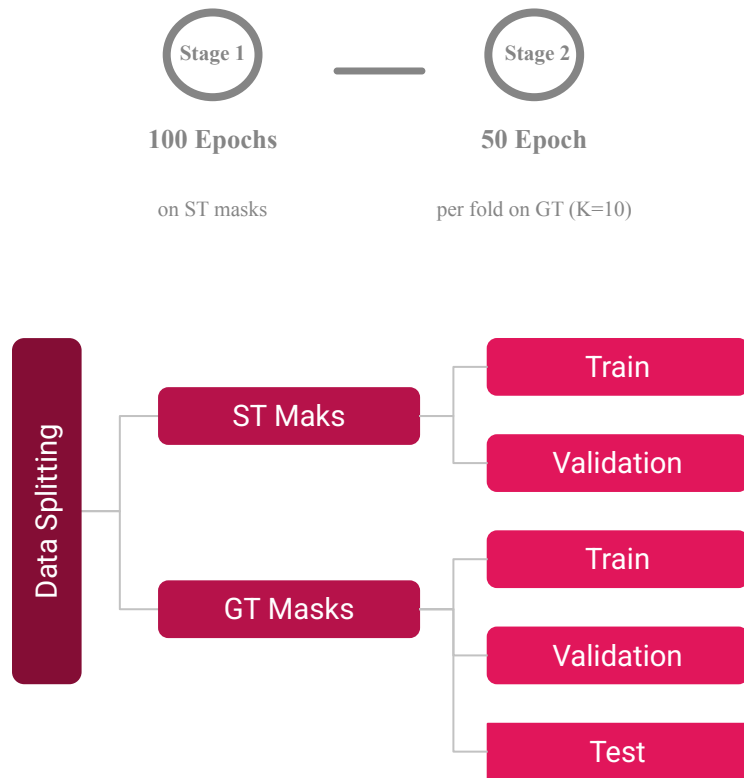- Google Colab Pro with CUDA-enabled GPU

Training Details

- Stage 1: 100 epochs on ST masks
- Stage 2: 50 epochs per fold on GT (K = 10)

Data Splitting

- ST-only images for pretraining
- GT-only images for fine-tuning and testing
- GT set split into train/val/test with no overlap

Hyperparameters

- ST: LR = 1e-4, threshold = 0.6
- GT: LR = 1e-5, threshold = 0.8

Stage 1 — Stage 2

100 Epochs **50 Epoch**

on ST masks per fold on GT (K=10)

Data Splitting

ST Maks
- Train
- Validation

GT Masks
- Train
- Validation
- Test

Model Comparisons & K-Fold Validation

Tested Variants

- MONAI U-Net vs Pure U-Net
- With/without augmentation
- CLAHE on/off
- Loss functions: BCEDiceLoss vs BCEWithLogitsLoss

Best Setup

- MONAI U-Net + CLAHE + augmentation + BCEDiceLoss

K-Fold Results (GT)

- K=10 performed mean Dice of 0.9492 ± 0.0321
- Consistent generalization across folds

Mode	Loss Function	Augmentation	Preprocessing	Val Score (Pre)	Val Score (Fine)	Threshold
MONAI U-Net (K-Fold)	**BCEDiceLoss (0.3)**	**Yes**	**CLAHE**	**0.9685**	**0.9469**	**0.60 (Pre) 0.8 (Fine)**
MONAI U-Net	BCEDiceLoss (0.3)	Yes	CLAHE	0.9770	0.9273	0.7
MONAI U-Net	BCEDiceLoss (0.3)	No	CLAHE	0.9631	0.9172	0.7
MONAI U-Net	BCEDiceLoss (0.3)	No	No	0.9689	0.9207	0.7
Pure U-Net	BCEDiceLoss (0.3)	No	No	0.9798	0.9375	0.5
Pure U-Net	BCEWithLogitsLoss	Yes	No	0.9290	0.8947	0.5

Ensemble Learning & Results

Why Ensemble?

- Models from K-folds have slightly different strengths
- Reduces overfitting from individual folds

Ensemble Method

- Pixel labeled as "cell" if at least eight of ten models agreed.

Quantitative Results

- Pre-trained (ST only): 0.9030 ± 0.0980
- Only-GT (single stage on GT masks): 0.8317 ± 0.1204
- Mixed (single stage on ST + GT masks): 0.9063 ± 0.1081
- Fine-tuned (GT): 0.9128 ± 0.0951
- Ensemble of 10 folds (majority vote): 0.9202 ± 0.0792

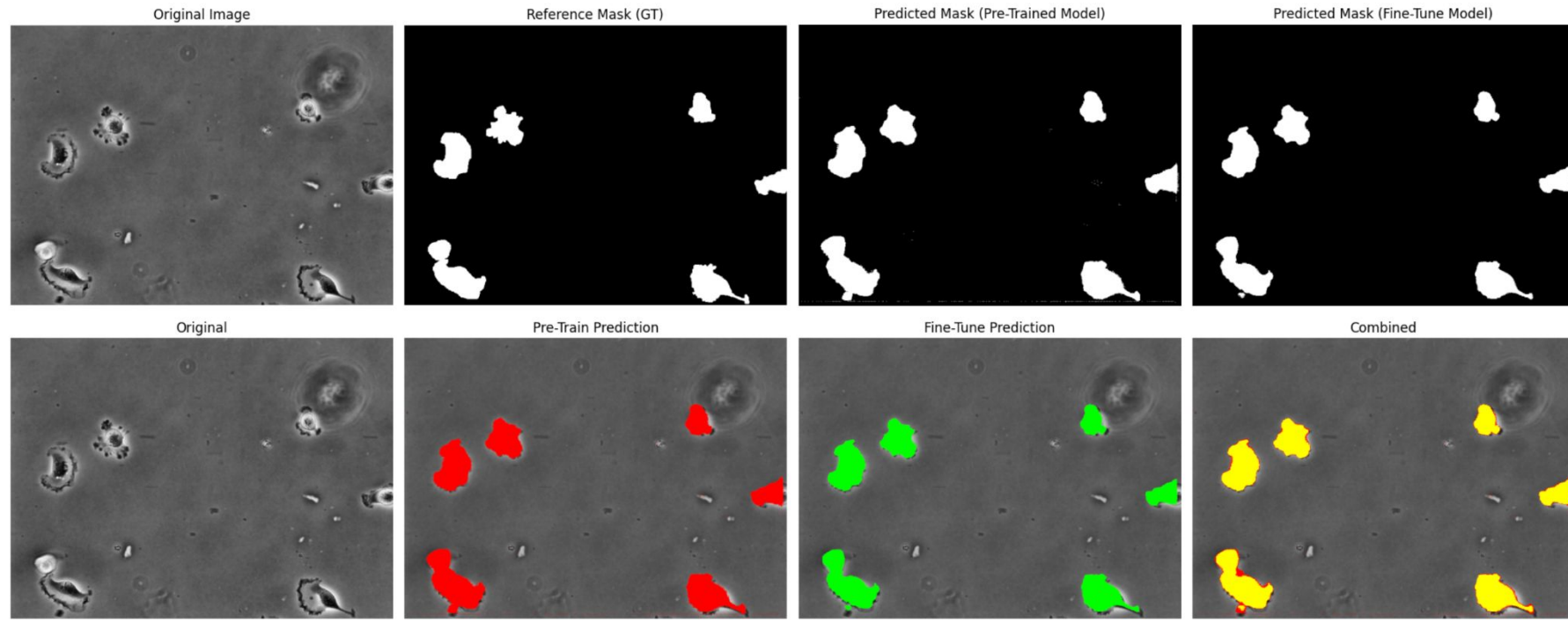**Significance**

- Wilcoxon test: $p = 0.015625$ (improvement is statistically significant)
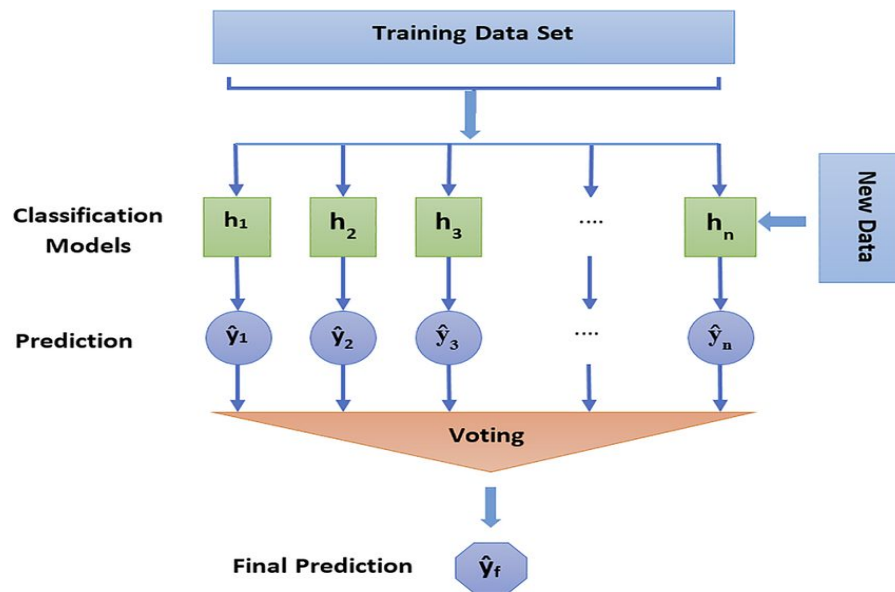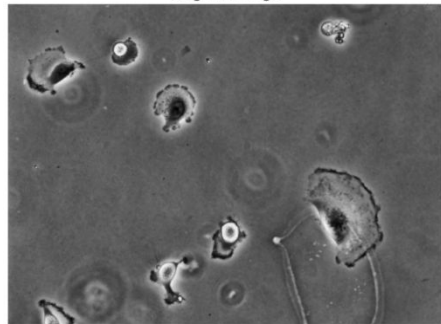

Majority Voting Example

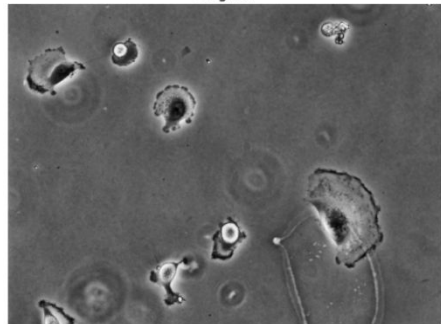# Results

Original Image | Reference Mask (GT) | Predicted Mask (Pre-Trained Model) | Predicted Mask (Fine-Tune Model)

Original | Pre-Train Prediction | Fine-Tune Prediction | Combined

Results

| Original Image | Reference Mask (GT) | Predicted Mask (Pre-Trained Model) | Predicted Mask (Fine-Tune Model) |

| Original | Pre-Train Prediction | Fine-Tune Prediction | Combined |

Results

Original Image | Reference Mask (GT) | Predicted Mask (Pre-Trained Model) | Predicted Mask (Fine-Tune Model)
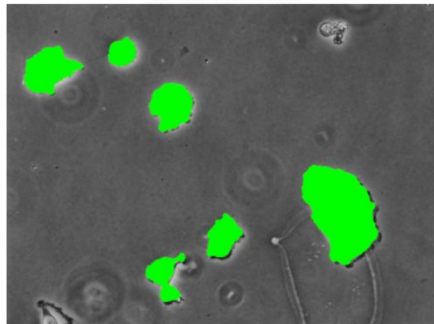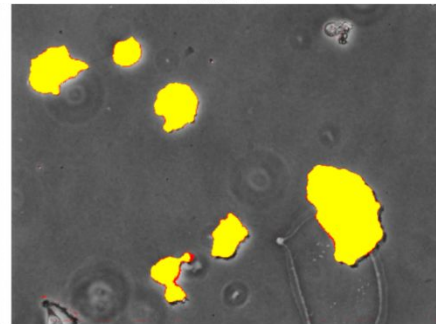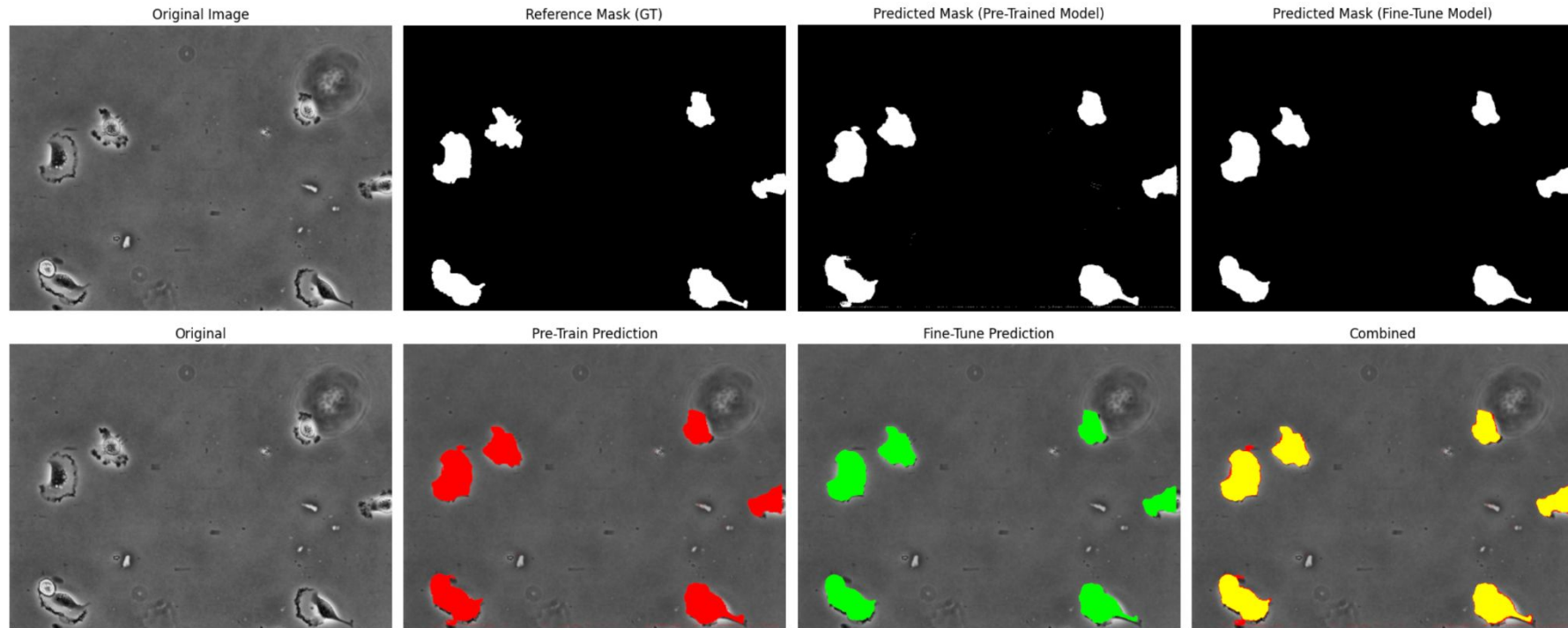
Original | Pre-Train Prediction | Fine-Tune Prediction | Combined
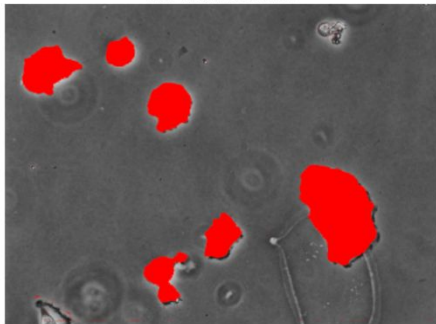
Discussion & Conclusion & Future Work

Summary

- Two-stage training approach is effective and efficient
- Combines strengths of both ST (quantity) and GT (quality)
- Ensemble learning increases stability and accuracy

Final Dice Score

- 0.9202 on test set with GT masks

Future Directions

- Adaptive ensembling strategies such as weighted voting or stacking based on pre-model confidence could push the ensemble's performance beyond simple majority voting
- Larger and more diverse datasets
- Post-processing (morphological ops, watershed) to refine masks