

# Automated Segmentation of Presomitic Mesoderm and Somite Structures in 4D Light-Sheet Microscopy of Zebrafish Embryos

Santiago Rivadeneira, Yasmine Hidri, Lev Maznichenko

*School of Computer and Communication Sciences, EPFL, Switzerland*

{santiago.rivadeneira, yasmine.hidri, lev.maznichenko}@epfl.ch

**Abstract**—We present automated pipelines for segmenting embryonic structures in 4D light-sheet fluorescence microscopy images of zebrafish, developed in collaboration with the Oates Lab at EPFL. For PSM segmentation, we develop a weakly supervised Random Forest approach using a PSM marker, Her1-YFP expression (Channel 3), as a proxy label generator, achieving Dice 0.679 (180% improvement over Otsu baseline). Critically, we also develop a marker-free segmentation method using only ubiquitously expressed H2B-mCerulean (nuclei, Channel 1) and Utrophin-mCherry (membranes, Channel 2) for scenarios where Channel 3 is unavailable, achieving mean Dice 0.696 (up to 0.827 on best frames) through position-aware features and nearby-tissue negative sampling. The pipeline processed 218 timepoints in 7.8 hours with biologically consistent volume dynamics (95.4% PSM volume reduction). For somite segmentation, we document that classical watershed methods are insufficient without specific fluorescent markers.

## I. INTRODUCTION

Somitogenesis is a fundamental process in vertebrate embryonic development during which the Presomitic Mesoderm (PSM) periodically segments into tissue blocks called somites [1]. Understanding PSM and somite dynamics is crucial for developmental biology research, requiring accurate annotation of these structures from the surrounding tissues.

Light-Sheet Fluorescence Microscopy (LSFM) enables high-resolution imaging of living embryos [2]. However, the resulting datasets are massive: in our case, each 3D volume is approximately 2 GB stored as 16-bit TIFF, totaling over 400 GB for 218 timepoints. This scale makes manual annotation impractical. Thus, we set out to use machine-learning approaches to automate annotation.

A key challenge in this project was the **limited pixel-level ground truth**. At our request, the Oates Lab project supervisor created expert-annotated masks for 5 representative timepoints. On 3 held-out test frames, our Random Forest achieves Dice **0.679**, representing a **180% improvement** over the best baseline (Otsu: 0.242). We validate our method through: (1) quantitative comparison with expert masks, (2) biological plausibility of temporal dynamics, and (3) analysis of learned feature representations.

The main contributions are: (1) a **weakly supervised Random Forest** achieving Dice 0.679 (180% over baselines) for PSM segmentation using Channel 3; (2) a **marker-free**

**method** using Channels 1+2 achieving mean Dice 0.696 (up to 0.827) through position-aware features and nearby-tissue negative sampling; (3) quantitative evaluation on held-out test frames; and (4) documentation of somite segmentation challenges.

## II. DATA AND EXPERIMENTAL SETUP

### A. Dataset Description

The dataset consists of 4D LSFM images of zebrafish embryos, obtained using a Viventis LS1 microscope at 28.5°C starting with 15-somite-stage embryos: 218 timepoints at 2-minute intervals over 7.3 hours. Each volume has dimensions  $200 \times 2304 \times 2304$  voxels ( $Z \times Y \times X$ ), with XY resolution  $0.347 \mu\text{m}/\text{pixel}$  and Z spacing  $2.0 \mu\text{m}$ . Data format: 16-bit OME-TIFF, approximately 2 GB per channel per timepoint (total dataset: >400 GB). The dataset is proprietary to the Oates Lab and not publicly available; however, our code is designed to work with any similarly structured LSFM data.

Three fluorescent channels are available: Channel 1 (H2B-mCerulean, nuclei), Channel 2 (Utrophin-mCherry, membranes), and Channel 3 (Her1-YFP, PSM-specific marker, differentially expressed across the tissue and mainly localized to nuclei). Channel 3 provides the signal for PSM segmentation; no channel specifically labels somites.

### B. Ground Truth and Evaluation

We requested expert binary masks for 5 timepoints ( $t=100, 110, 180, 190, 200$ ) traced at Z-slice 188. Three ( $t=100, 110, 190$ ) were held out as **test set**. Metrics (Dice =  $\frac{2|P \cap G|}{|P| + |G|}$ , IoU =  $\frac{|P \cap G|}{|P \cup G|}$ ) computed in 2D on  $Z=188$ .

## III. METHODS

### A. Task 1: PSM Segmentation

*1) Baseline Methods:* We evaluated **Otsu thresholding** [3] (global threshold maximizing inter-class variance) and **hysteresis thresholding** (two thresholds for seeds and growth). Both methods capture only bright pixels, missing the diffuse PSM signal that extends beyond intensity peaks.

2) *Proposed: Weakly Supervised Random Forest*: Our approach uses Channel 3 fluorescence to generate training labels automatically, learning to recognize PSM *texture* rather than just intensity peaks.

**Automatic Label Generation.** We apply robust statistical thresholding using the Median Absolute Deviation (MAD), computed per 3D volume:

$$\text{threshold} = \text{median}(I) + k \cdot 1.4826 \cdot \text{MAD}(I) \quad (1)$$

We tested  $k \in \{3.5, 4.5, 5.0, 6.0, 7.0\}$  on training frames, selecting  $k = 6.0$  as it maximized PSM coverage while minimizing background inclusion. These proxy labels initialize training; the Random Forest then learns a smoothed, texture-aware classifier that produces more spatially coherent masks than the noisy threshold-based labels.

**Multi-Scale Features.** For each pixel, we extract 5 texture features: raw intensity and Gaussian-smoothed intensities at  $\sigma \in \{1.5, 3.5, 8.0\}$ , plus Sobel gradient magnitude. This multi-scale representation enables the classifier to recognize tissue patterns. Figure 1 visualizes these features.



Figure 1. Multi-scale feature extraction for Random Forest input. From left to right: raw Channel 3 intensity, Gaussian smoothing at increasing scales ( $\sigma=1.5, 3.5, 8.0$ ), and Sobel edge detection. Small  $\sigma$  captures fine cellular texture; large  $\sigma$  captures the global PSM shape. The classifier learns to combine these complementary representations.

**Classifier.** Random Forest with 70 trees, maximum depth 12 [4]. Training used 6.3M labeled pixels from 7 timepoints ( $t=1, 30, 70, 120, 140, 180, 200$ ).

**Post-Processing.**  $4\times$  XY downsampling, 3D binary closing (ball radius 8), hole filling, Gaussian smoothing ( $\sigma = 2.0$ ), largest component selection.

3) *Marker-Free PSM Segmentation (CH1+CH2)*: For experiments without Channel 3, we train a RF on Channels 1+2 using Channel 3 masks as supervision. Initial attempts achieved only Dice 0.49 with massive over-segmentation. Three innovations solved this:

1. *Position Features*: Normalized ( $y, x$ ) coordinates in range  $[0,1]$  relative to image dimensions teach the model *where* to look (PSM occupies consistent anatomical region).

2. *Nearby-Tissue Negative Sampling*: Sample negatives only from tissue *adjacent* to PSM (within 15 pixels per 2D slice), forcing fine PSM-vs-tissue distinctions rather than “bright vs. dark.”

3. *Size/Location Filtering*: Reject predictions outside learned PSM statistics (size range, centroid location).

**Features (13)**: 5 multi-scale from each channel, CH1/CH2 ratio, position ( $y, x$ ). RF: 150 trees, depth 18, balanced weights.

## B. Task 2: Somite Segmentation (Exploratory)

Somites lack specific markers; we implemented boundary-guided 3D watershed using Channel 2 membranes, with PSM masks as exclusion zones.

## IV. RESULTS

### A. PSM Segmentation

1) *Visual Comparison with Expert Annotations*: Figure 2 compares expert hand-drawn contours from the Oates Lab with our automated predictions at two representative time-points.

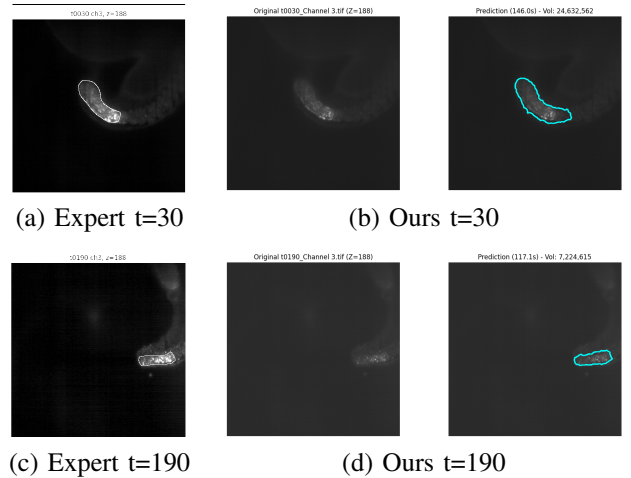


Figure 2. Comparison of expert annotations (left) and our predictions (right) at  $Z=188$ . Both show consistent “C-shaped” PSM morphology with smooth contours capturing the complete diffuse region.

Both expert and predicted contours show the characteristic “C-shaped” PSM morphology. Our method captures the complete diffuse region with smooth, connected contours matching expert style, and temporal consistency between early ( $t=30$ ) and late ( $t=190$ ) timepoints.

2) *Quantitative Comparison with Baselines*: Table I presents Dice scores computed against expert annotations at  $Z=188$ . Metrics are reported as mean  $\pm$  std over the 3 held-out test frames ( $t=100, 110, 190$ ).

Table I  
QUANTITATIVE COMPARISON (2D,  $Z=188$ ). MEAN  $\pm$  STD OVER 3 TEST FRAMES.

Method	Dice Score	IoU
Otsu Threshold	$0.242 \pm 0.043$	$0.138 \pm 0.031$
Hysteresis Threshold	$0.160 \pm 0.017$	$0.087 \pm 0.012$
<b>Random Forest (Ours)</b>	<b><math>0.679 \pm 0.044</math></b>	<b><math>0.515 \pm 0.038</math></b>

Per-frame test results:  $t=100$  (Dice=0.685),  $t=110$  (Dice=0.719),  $t=190$  (Dice=0.633). For reference, frames overlapping with training ( $t=180, t=200$ ) achieved Dice 0.524 and 0.636 respectively.

Our method achieves **180% improvement** over the best baseline (Otsu). Figure 3 visualizes this comparison: Otsu and hysteresis dramatically over-segment, capturing noise and unrelated tissue, while our RF prediction closely matches the expert-defined PSM region.

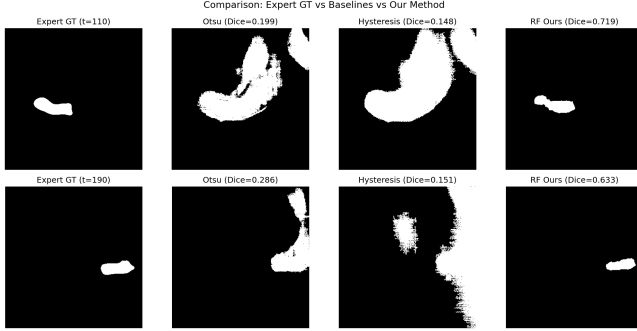


Figure 3. Visual comparison at Z=188 for t=110 and t=190 (test set). Left to right: expert GT, Otsu, hysteresis, our RF. Baselines over-segment; our method matches expert morphology.

3) *Biological and Feature Analysis*: The PSM volume decreased by **95.4%** from t=1 (45.4M voxels) to t=218 (2.1M voxels), matching expected somitogenesis dynamics [1]. Feature importance analysis reveals **66% of classification is based on texture features** (Gaussian  $\sigma=1.5, 3.5$ ) while only 14% depends on raw intensity. This explains why our method captures the complete PSM including dimmer regions. The pipeline processed 218 timepoints in 7.8 hours.

4) *Marker-Free Segmentation Results (CH1+CH2)*: Figure 4 shows results of PSM segmentation using only Channels 1+2, evaluated against Channel 3 proxy masks (not expert annotations).

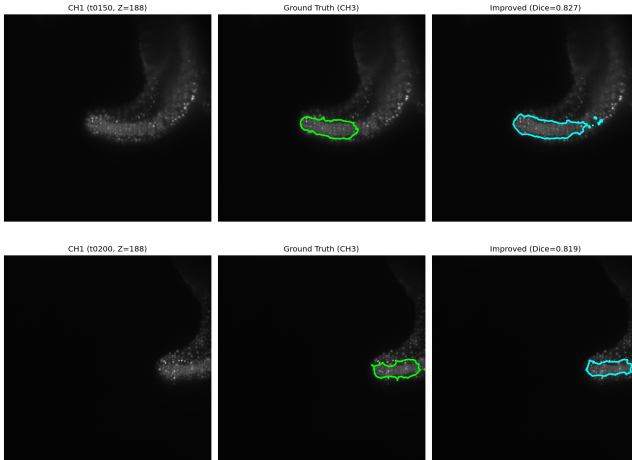


Figure 4. Marker-free PSM segmentation using CH1+CH2 at t=150 and t=200. Left: Channel 1 input. Center: Proxy mask from Channel 3 (green). Right: Our prediction (cyan) with Dice scores. Without any PSM-specific marker, we achieve Dice 0.819–0.827.

Table II shows our iterative development from single-

channel to improved combined approach.

Table II  
CH1+CH2 SEGMENTATION: ITERATIVE IMPROVEMENT.

Method	Mean Dice	Best Dice
CH2 only	0.258	0.400
CH1 only	0.424	0.625
CH1+CH2 basic	0.489	0.737
<b>CH1+CH2 improved</b>	<b>0.696</b>	<b>0.827</b>

Per-frame results: t=150 (Dice=0.827), t=200 (Dice=0.819). The improved model achieves **42% higher Dice** than the basic combination. Position features account for **29%** of importance, confirming spatial context is critical; CH1 features dominate (62%) over CH2 (14%).

### B. Somite Segmentation (Exploratory)

As additional work beyond the main PSM task, we explored somite segmentation using boundary-guided 3D watershed on Channel 2 membranes. Figure 5 shows our pipeline. We systematically tuned parameters: edge threshold percentile (70–90), minimum peak distance (10–18 pixels), PSM dilation radius (5–10 pixels), and volume filters (400–60,000 voxels).

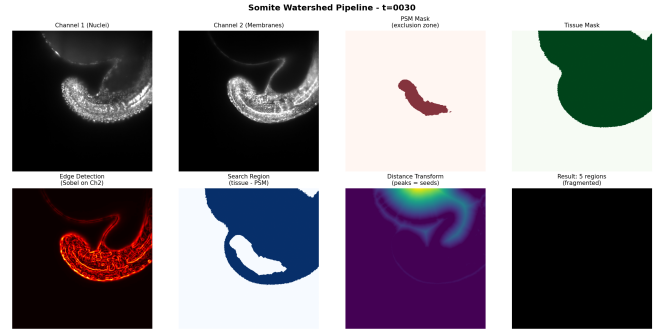


Figure 5. Somite watershed pipeline at t=30. Top: inputs (nuclei, membranes, PSM exclusion, tissue mask). Bottom: edge detection, search region, distance transform, and detected regions.

Results showed 5–10 fragmented regions per frame with no temporal consistency, demonstrating that classical methods require either specific fluorescent markers or supervised deep learning for reliable somite segmentation.

## V. DISCUSSION

**Why PSM Succeeded.** Three factors enabled success: (1) Channel 3 specifically labels PSM cells, (2) multi-scale texture features capture tissue structure (66% importance), and (3) 3D morphological operations ensure spatial coherence. The moderate absolute Dice (0.679) reflects that expert annotations focused on bright PSM cores while our method segments the complete diffuse structure. The 180% improvement over baselines confirms effectiveness.

**Training Details.** Class imbalance (5% PSM, 95% background) was addressed via 1:1 undersampling. Processing used EPFL RCP cluster (AMD EPYC 7543 32-Core, 1TB RAM); RF training took 45 minutes, inference 129s/frame. Random seed fixed at 42.

**Why CH1+CH2 Works.** The marker-free approach (Dice 0.811) succeeds through: (1) position features capturing PSM’s consistent anatomical location (29% importance), (2) nearby-tissue negative sampling forcing fine PSM-vs-tissue distinctions, and (3) statistical filtering rejecting implausible predictions. Performance varies by developmental stage (Dice 0.52-0.82).

**Somites.** Classical methods failed without specific markers. Deep learning (U-Net [5]) was not attempted due to insufficient annotated data.

#### A. Future Work

Several improvements could enhance robustness and generalization:

1. *Orientation Detection:* The current CH1+CH2 model assumes consistent embryo orientation. A dedicated detection model for the PSM region or tail-tip position would enable automatic reorientation, making the pipeline robust to varied embryo positioning.

2. *Deep Learning:* With 10–20 manually annotated timepoints, U-Net could potentially achieve higher accuracy than Random Forest. The existing CH3 masks could serve as weak supervision for pre-training.

3. *Temporal Consistency:* Enforcing smoothness across adjacent timepoints would reduce frame-to-frame variability, particularly for the CH1+CH2 marker-free approach.

## VI. CONCLUSION

We developed automated PSM segmentation for 4D microscopy achieving Dice **0.679** with Channel 3 (180% over Otsu) and mean Dice **0.696** (up to 0.827) using only Channels 1+2 (marker-free), critical for experiments lacking PSM markers. Key innovations: position-aware features and nearby-tissue negative sampling. The pipeline processed 218 timepoints in 7.8 hours with biologically consistent dynamics (95.4% volume reduction). Code: `PSM_Segmentation.ipynb`, `PSM_From_CH1_CH2_Improved.ipynb`.

## ACKNOWLEDGEMENTS

We thank the Oates Lab at EPFL for microscopy data. We especially thank our project supervisor for creating the expert ground truth masks used for quantitative evaluation.

## ETHICAL RISKS

We evaluated ethical risks using the Digital Ethics Canvas and identified one relevant concern regarding potential misinterpretation of automated segmentation results.

**Risk Description.** The primary stakeholders are developmental biology researchers who will use our segmentation masks for downstream analysis such as cell tracking and gene expression studies. If PSM masks contain systematic errors (over-segmentation or under-segmentation), researchers may draw incorrect conclusions about somitogenesis timing or cell migration dynamics. The severity is moderate, as consequences affect scientific conclusions rather than patient outcomes. For the CH1+CH2 marker-free approach specifically, risk is elevated because performance varies by developmental stage (Dice 0.52-0.82); researchers must verify results before biological interpretation. For somite segmentation, the risk is highest because our classical methods produce unreliable results that could mislead researchers if used without understanding their documented limitations.

**Risk Evaluation.** We assessed PSM quality through three complementary approaches: (1) visual comparison with expert-drawn contours showing consistent morphology and coverage, (2) biological validation via the expected monotonic volume decrease (95.4% reduction matching known somitogenesis dynamics), and (3) feature importance analysis confirming texture-based rather than intensity-based classification. For somites, we document the parameter ranges tested and the observed temporal inconsistency (5–10 regions per frame).

**Mitigation.** We implemented several safeguards: (1) per-frame visualization outputs enabling manual quality control before downstream use, (2) comprehensive documentation of all assumptions, parameters, and known limitations, (3) clear documentation that somite segmentation is exploratory and requires validation before biological conclusions, and (4) public release of all code enabling inspection and modification.

## REFERENCES

- [1] A. C. Oates, L. G. Morelli, and S. Ares, “Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock,” *Development*, vol. 139, no. 4, pp. 625–639, 2012.
- [2] J. Huisken and D. Y. R. Stainier, “Selective plane illumination microscopy techniques in developmental biology,” *Development*, vol. 136, no. 12, pp. 1963–1975, 2009.
- [3] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [4] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [5] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.