

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.rivadeneiraquintero, 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 ($\text{Dice} = \frac{2|P \cap G|}{|P| + |G|}$, $\text{IoU} = \frac{|P \cap G|}{|P \cup G|}$) computed in 2D on Z=188.

III. METHODS

A. Why Random Forest Instead of Deep Learning

We chose Random Forest over U-Net/CNNs for three reasons: (1) **Limited annotations**: only 5 expert-annotated frames exist, insufficient for deep learning which typically requires hundreds of annotated volumes; (2) **Interpretability**: RF provides feature importance analysis revealing what the model learns (texture vs intensity); (3) **No GPU required**: RF runs on CPU, critical for the Oates Lab's infrastructure. With 10–20 annotated timepoints, U-Net could be explored.

B. Task 1: PSM Segmentation with Channel 3

1) Baseline Methods: We evaluated **Otsu thresholding** [3] (threshold $t^* = \arg \max_t \sigma_B^2(t)$ maximizing between-class variance) and **hysteresis thresholding** (percentile-based thresholds $t_{low} = P_{70}$, $t_{high} = P_{90}$). Both capture only bright pixels, missing diffuse PSM signal.

2) Step 1: Automatic Proxy Label Generation: We generate training labels from Channel 3 using robust MAD thresholding, computed per 3D volume V :

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

where $\text{MAD}(V) = \text{median}(|V - \text{median}(V)|)$ and factor 1.4826 normalizes to Gaussian σ . We tested $k \in \{3.5, 4.5, 5.0, 6.0, 7.0\}$; $k = 6.0$ maximized PSM coverage while excluding background (validated visually on $t=1, 30, 70$).

3) *Step 2: Multi-Scale Feature Extraction*: For each pixel at position (y, x) in slice z , we compute 5 features from intensity I :

$$f_1 = I(y, x) \quad (\text{raw intensity}) \quad (2)$$

$$f_2 = (G_{\sigma=1.5} * I)(y, x) \quad (\text{fine texture}) \quad (3)$$

$$f_3 = (G_{\sigma=3.5} * I)(y, x) \quad (\text{medium texture}) \quad (4)$$

$$f_4 = (G_{\sigma=8.0} * I)(y, x) \quad (\text{coarse context}) \quad (5)$$

$$f_5 = |\nabla I|(y, x) \quad (\text{Sobel edges}) \quad (6)$$

where G_σ is Gaussian kernel with standard deviation σ . Scale selection: $\sigma = 1.5$ captures cell-level patterns (~ 5 pixels $\approx 1.7\mu\text{m}$), $\sigma = 3.5$ captures tissue texture, $\sigma = 8.0$ captures regional PSM shape. Figure 1 visualizes these features.

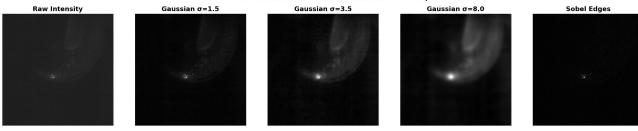


Figure 1: Multi-scale features: raw intensity, Gaussian at $\sigma=1.5, 3.5, 8.0$, and Sobel edges. Small σ captures cellular texture; large σ captures PSM shape.

4) *Step 3: Random Forest Training: Hyperparameters*: 70 trees (sufficient for convergence, validated via OOB error), max depth 12 (prevents overfitting to noisy proxy labels), `class_weight='balanced'` (handles 5% PSM / 95% background imbalance). Training: 6.3M pixels from 7 frames ($t=1, 30, 70, 120, 140, 180, 200$), spanning early-to-late developmental stages. Seed=42 for reproducibility.

5) *Step 4: Post-Processing Pipeline*:

- 1) **Downsampling**: $4 \times XY$ reduction ($2304 \rightarrow 576$ pixels)
- 2) **RF prediction**: classify each pixel in downsampled space
- 3) **3D closing**: ball structuring element $r = 8$ voxels
- 4) **Hole filling**: `binary_fill_holes`
- 5) **Upsampling**: nearest-neighbor interpolation to original size
- 6) **Smoothing**: Gaussian $\sigma = 2.0$ then threshold > 0.5
- 7) **Component selection**: keep largest connected region

This downsampling reduces processing time from ~ 20 min to ~ 2 min per frame ($16\times$ speedup) with minimal quality loss, as PSM boundaries are smooth at the mesoscale.

C. Task 1b: Marker-Free Segmentation (CH1+CH2)

For experiments without Channel 3, we train on CH1 (nuclei) + CH2 (membranes) using CH3 masks as supervision. Initial attempts (Dice 0.49) failed due to over-segmentation: the model learned “tissue vs background” instead of “PSM vs other tissue.”

1) *Innovation 1: Extended Feature Set (13 features)*:

$$f_{1-5} : \text{CH1 multi-scale (intensity, } G_{1.5}, G_{3.5}, G_{8.0}, \text{Sobel}) \quad (7)$$

$$f_{6-10} : \text{CH2 multi-scale (same 5 features)} \quad (8)$$

$$f_{11} = G_{2.0} * \frac{I_{\text{CH1}}}{I_{\text{CH2}} + \epsilon} \quad (\text{ratio, } \epsilon = 10^{-6}) \quad (9)$$

$$f_{12} = y/H \in [0, 1] \quad (\text{normalized Y position}) \quad (10)$$

$$f_{13} = x/W \in [0, 1] \quad (\text{normalized X position}) \quad (11)$$

Position features encode anatomical prior: PSM is always in posterior tail region ($y \approx 0.6-0.8$, $x \approx 0.4-0.6$).

2) *Innovation 2: Nearby-Tissue Negative Sampling*: Instead of sampling negatives from entire background, we sample only from

tissue adjacent to PSM:

$$\text{Positives} : \text{erode}(M_{\text{PSM}}, r = 2) \quad (12)$$

$$\text{Negatives} : [\text{dilate}(M_{\text{PSM}}, r = 15) - \text{dilate}(M_{\text{PSM}}, r = 3)] \cap T \quad (13)$$

where T is tissue mask ($I > P_{20}$). This forces the classifier to distinguish PSM from adjacent non-PSM tissue rather than “bright vs dark.”

3) *Innovation 3: Size/Location Filtering*: From training masks, we compute PSM statistics: size range $[s_{\min}, s_{\max}]$, centroid distribution $(\mu_y, \sigma_y), (\mu_x, \sigma_x)$. At inference, we reject components with area outside $[0.3 \cdot s_{\min}, 3 \cdot s_{\max}]$ or centroid beyond 3σ from mean.

RF parameters: 150 trees (more than CH3 model due to harder task), depth 18, balanced weights.

D. Task 2: Somite Segmentation (Exploratory)

Somites lack specific markers; we implemented boundary-guided 3D watershed using Channel 2 membrane edges with PSM masks as exclusion zones. Parameters tested: edge threshold percentile (70–90), min peak distance (10–18px), volume filters (400–60k voxels).

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 timepoints.

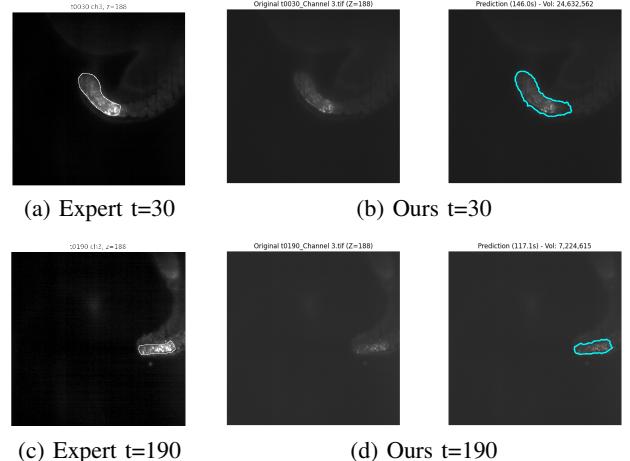


Figure 2: Comparison of expert annotations (left) and our predictions (right) at $Z=188$. Both show consistent “C-shaped” PSM morphology with smooth contours capturing the complete diffuse region.

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 ($t=100, 110, 190$).

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.

Table I: Quantitative comparison (2D, Z=188). Mean \pm std over 3 test frames.

Method	Dice Score	IoU
Otsu Threshold	0.242 ± 0.043	0.138 ± 0.031
Hysteresis Threshold	0.160 ± 0.017	0.087 ± 0.012
Random Forest (Ours)	0.679 ± 0.044	0.515 ± 0.038

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.

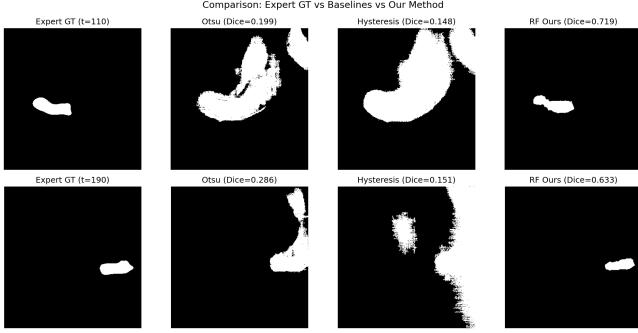


Figure 3: Visual comparison at Z=188 for t=110 and t=190 (test set). Left to right: expert GT, Otsu, hysteresis, our RF. Baselines over-segment; our method matches expert morphology.

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).

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
CH1+CH2 improved	0.696	0.827

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

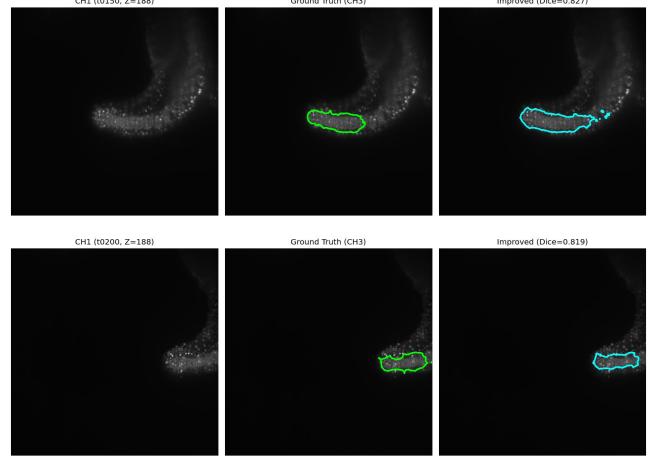


Figure 4: Marker-free PSM segmentation using CH1+CH2 at t=150 and t=200. Left: Channel 1 input. Center: Proxy mask from Channel 3 (green). Right: Our prediction (cyan) with Dice scores. Without any PSM-specific marker, we achieve Dice 0.819–0.827.

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).

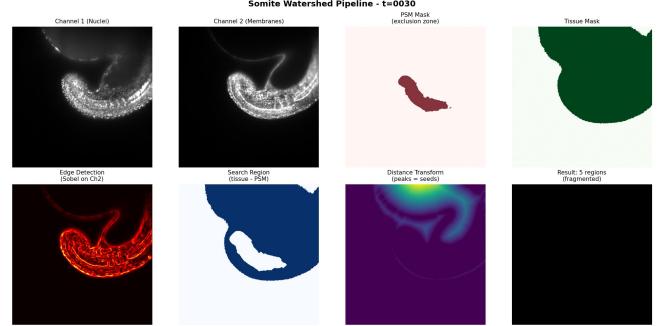


Figure 5: Somite watershed pipeline at t=30. Top: inputs (nuclei, membranes, PSM exclusion, tissue mask). Bottom: edge detection, search region, distance transform, and detected regions.

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 Multi-Scale Features Work. Feature importance analysis reveals texture features (Gaussian $\sigma = 1.5, 3.5$) contribute 66% of classification power, while raw intensity contributes only 14%. This explains why our method captures diffuse PSM regions that intensity-based thresholds miss. The moderate absolute Dice (0.679) reflects annotation style differences: experts traced bright cores while our method segments complete diffuse structure.

Hyperparameter Justification. (1) Tree count 70: OOB error plateaus at ~ 50 trees; 70 provides margin. (2) Depth 12: deeper trees overfit noisy proxy labels. (3) $k = 6.0$ for MAD: validated on training frames; lower k over-segments, higher k misses dim PSM. (4) Gaussian σ values: $\sigma = 1.5$ matches cell diameter ($\sim 5\text{px}$), $\sigma = 8.0$ matches PSM width ($\sim 25\text{px}$).

CH1+CH2 Success Factors. Position features contribute 29% importance, confirming spatial prior is critical. Nearby-tissue sampling increased Dice from 0.49 to 0.70 by forcing PSM-vs-tissue distinctions. Size/location filtering removes anatomically implausible predictions. Performance varies by stage (Dice 0.52–0.82) because PSM morphology changes during somitogenesis.

Computational Details. EPFL RCP cluster (AMD EPYC 7543 32-Core, 1TB RAM). RF training: 45 min (CH3 model), 60 min (CH1+CH2 model). Inference: 129s/frame. No GPU required. Seed=42.

Somite Limitations. Classical watershed failed (5–10 fragments/frame) because somite boundaries lack clear membrane signal enhancement. U-Net [4] requires >50 annotated volumes; we have 0 for somites.

Limitations. The CH1+CH2 model assumes consistent embryo orientation (tail in lower-left quadrant). If embryos are flipped, the model must be retrained with appropriately oriented training data. Additionally, 3D annotation for deep learning would require labeling each Z-slice separately (~4,000 images per timepoint), making it impractical with current resources.

Future Work. (1) *Orientation detection:* A dedicated model to detect 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 achieve higher accuracy; 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.

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] 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.