Enhancing Cardiac Cell Networks Segmentation via Hybrid Supervised and Zero-Shot Strategies*

Sarah Costa[1] [0009-0004-3209-6227], Hassan Eshkiki[1] [0000-0001-7795-453X], Fabio Caraffini[1] [0000-0001-9199-7368], and Christopher H. George[2] [0000-0001-9852-1135]

[1] Department of Computer Science, Swansea University, Swansea SA1 8EN, UK
`{2038107,h.g.eshkiki,fabio.caraffini}@swansea.ac.uk`
[2] Swansea University Medical School, Faculty of Medicine, Health and Life Sciences, Swansea SA2 8PP, UK
`christopher.george@swansea.ac.uk`

Abstract. Cardiac cell network segmentation is uniquely challenging because cardiomyocytes, unlike other cell types, form morphologically complex multicellular structures, causing generalist models like Cellpose to oversegment and perform inaccurately. We use our unique live cell imaging dataset of self-organised HL-1 networks to propose and assess various algorithmic configurations based on combinations of the Cellpose model and the Segment Anything Model, equipped with multiple pre- and post-processing routines. Our results demonstrate the advantages of integrating equalisation-based pre-processing with median filtering, fine-tuning Cellpose and incorporating our post-processing routine into the segmentation pipeline, achieving up to 85% accuracy, 96% recall, 91% DICE and 88% precision, while mitigating oversegmentation.

Keywords: Cell Segmentation · Cellpose · Segment Anything Model · Cardiomyocyte Networks

1 Introduction

Modern advances in biomedical imaging have vastly increased the complexity of datasets, creating a need for novel computer vision (CV) methods to automate quantitative data extraction. Contemporary artificial intelligence (AI) paradigms are crucial for complex CV tasks like image segmentation, or dividing images into segments based on colour, intensity, and texture [28, 17]. Segmentation is widely applied in biomedicine for tumour detection in MRI/CT scans [18], diagnosing pathological specimens [1], and microscopy cell counting [11].

Cell segmentation spans traditional thresholding and edge detection to promising machine learning (ML) algorithms [4, 27, 9, 25], usually performing binary

---

* Supported by the Morgan Advanced Studies Institute, Wales, UK; the National Cardiovascular Research Network (funded by Health and Care Research Wales); the British Heart Foundation.

classification to differentiate discrete individual cells (ICs) from the image background [28]. Performances excel on high-quality images with clear cell boundaries but struggle with heterogeneous lighting, zoom, and focus, or when cells grow in dense clusters with unclear edges [8]. This is especially pertinent for engineered monolayers of HL-1 cardiomyocytes (CM), used as cardiovascular research tools to study CM connectivity in the heart [7]. HL-1 cells self-organise into complex networks of ICs and larger multicellular aggregates (CAs) formed by the fusion of ICs [4]. The irregular morphology of CAs and the absence of internal boundaries affect segmentation by conventional ML algorithms, causing image fragmentation [4, 25]. Currently, no model accurately segments HL-1 networks.

This study combines supervised and zero-shot learning (ZSL) to improve ICs and CAs segmentation in HL-1 networks. Our objectives are: 1) reducing image variability with pre-processing, 2) enhancing cell detection, and 3) addressing CAs oversegmentation. The article is structured as follows: Section 2 reviews the background and current research; Section 3 outlines our methodology and proposed algorithms, evaluated in Section 4; Section 5 offers recommendations for using these algorithms beyond this study and concludes the article.

2 Background

Cell segmentation uses semantic and instance methods for identifying, counting, and measuring ICs [25, 26, 11]. Semantic segmentation marks cells as foreground and background with binary labels, creating a mask for comparison with the ground-truth (GT) [22, 28]. Instance segmentation assigns unique labels to individual objects and assesses for their overlap (i.e., concordance) with each GT mask [6, 28]. By distinguishing and outlining objects, instance segmentation provides both pixel and object-level detection, exceeding semantic methods [6].

Numerous algorithms have been developed using ML and deep learning (DL) architectures, integrating semantic and instance strategies for segmentation [25, 26], classification [24, 9], and live-cell tracking [27]. Specialist models (e.g., [27]) excel in specific segmentation tasks on tailored datasets, performing well on known imaging types but lacking generalisation [27, 28]. Hence, research has focused on creating generalist models [25, 26, 9, 24] that can recognise diverse cell types across various microscopy modalities, such as brightfield and fluorescent, ensuring consistent performance under different imaging conditions [28].

Cellpose [25] is a commonly used generalist DL algorithm for cell segmentation based on the U-shaped encoder-decoder architecture of U-Net to generate segmented masks [23]. Cellpose performs segmentation in a supervised fashion, as it requires manual annotations for model training. Despite its promise to generalise well for all types of cells, its high sensitivity to round cellular shapes and nuclei hinders the accurate detection of irregularly shaped CAs lacking internal boundaries, as observed in HL-1 networks, which are often oversegmented (see Section 1) [4]. Oversegmentation causes the unnecessary fragmentation of the GT annotations, requiring the manual and laborious post-processing of the predicted masks to correct errors.

Segment Anything Model (SAM) is a ZSL algorithm designed to generalise the segmentation of any type of boundary-defined object [16]. As a ZSL method, SAM automatically performs segmentation without requiring manual annotations for training. SAM has gained significant popularity within the imaging community, as one of the most widely used algorithms due to its adaptability, ease to use, and generalisable properties, which make it suitable for application in any field. Since its vanilla pipeline was trained on large natural imaging datasets, the model displays marked sensitivity for boundaries, demonstrating more versatility than Cellpose in the recognition of irregularly shaped objects [16]. These properties make SAM a suitable candidate model to enhance Cellpose's performance on HL-1 networks, as evaluated in the next sections of this paper.

3 Methodology

Our proposed pipeline advances the methodology used in [4] by evaluating novel image pre-processing techniques (described in Section 3.3) and three novel hybrid algorithms for HL-1 networks segmentation (presented in Section 3.6). Segmentation is followed by post-processing (explained in Section 3.5) and the evaluation of the segmentation results obtained from each model (see Section 3.7).

3.1 HL-1 live cell imaging

We employ a unique live cell imaging dataset from confocal laser scanning microscopy experiments in self-organised networks of HL-1 CM. A phenotypic description of HL-1 cells has been done previously [4, 7, 12, 13, 5]. Cell edges were visualised in HL-1 CM loaded with the fluorescent Ca_{2+} reporter Fluo3 [14]. Fluorescently stained HL-1 networks were imaged using a confocal microscope for 50s at 512×512 pixel resolution [4, 13].

3.2 Image dataset preparation, annotation and splitting

We extracted 150 frames from each HL-1 recording, then stacked, averaged and normalised them to a pixel intensity range of $0-255$. This produced a composite greyscale image (see Fig. 1) for annotation using OpenCV [3], as in [4].

Two researchers with experience of HL-1 CM network formation independently used QuPath, a free histopathology software [2], to label and verify ICs and CAs in HL-1 networks, forming our GT annotations. Following the exclusion of poor-quality images (scarce resolution, low brightness, excessive blur, extreme cellular density, etc.), we worked with 77 composite greyscale images and respective annotated GT masks.

Due to the heterogeneous subsets of images captured at different confocal microscope zoom settings, we initially arranged the 77 images into two groups based on their levels of digital magnification and subsequently randomly allocated images within each group to the same training and testing sets using an

80/20 partition ratio. This ensures a balanced distribution of images with higher and lower zoom levels for training and testing, preventing model under- or overfitting. The resulting training set includes 61 images, with 9 at higher zoom (see Fig. 1a), while the testing set includes 16 images, with 3 of them at at higher zoom.

3.3 Exploratory data analysis for images pre-processing

We performed an exploratory data analysis (ExDA) on a random subset of 6 training images to evaluate pre-processing methods for improved image quality and segmentation performance. We focused on optimising: 1) contrast and brightness enhancement, 2) background noise removal, and 3) preservation of image sharpness to identify cell boundaries, which are crucial for classifying CM networks into ICs and CAs (as described in Section 1).

From [20], we evaluated global histogram equalisation (HE), local contrast limited adaptive histogram equalisation (CLAHE) and two noise filtering techniques, namely gaussian filtering (GF) and median filtering (MF), on the 6 composite greyscale images. Furthermore, we tested all four combinations obtainable by cascading the two above-mentioned equalisation methods and the two noise-filtering techniques. Processed images were then ulteriorly normalised $(0-255)$ to maintain consistent intensity ranges across the dataset after pre-processing. Results of this analysis are graphically reported in Fig. 1d.

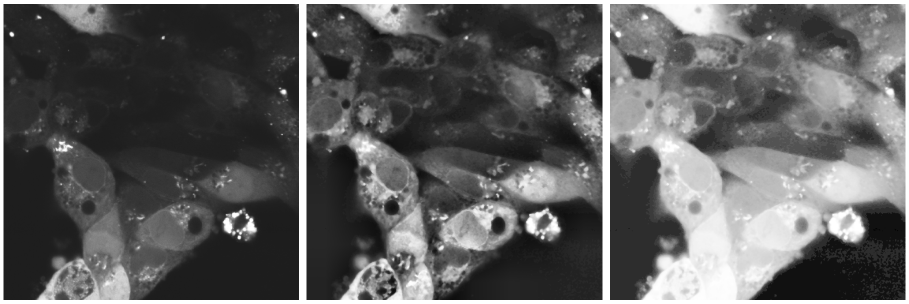
Fig. 1d shows comparable semantic segmentation performance across all pre-processing methods for the evaluated metrics. CLAHE + MF (Fig. 1b) slightly outperforms for Cellpose, while HE + MF (Fig. 1c) excels for SAM. Training and testing sets for evaluation of our proposed approaches were pre-processed accordingly.

3.4 Baseline segmentation algorithms

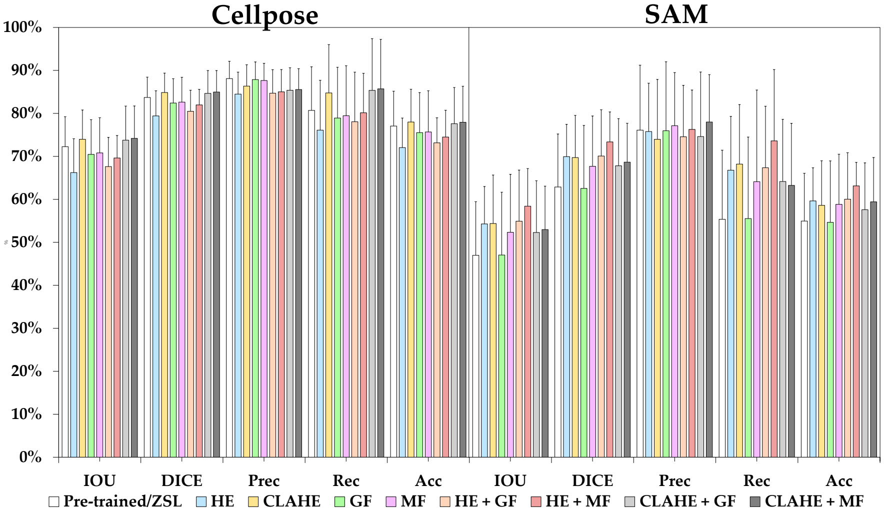
We propose hybrid approaches based on the vanilla segmentation algorithms `SAM` [16] with ZSL, i.e., using its pre-trained ViT-H model [10] as image encoder and its built-in automatic masks generation (AMG) function, and the pre-trained 'Cyto2' version of Cellpose [25], as in [4], referred to as $\texttt{CellPose}_1$. We include these algorithms in our comparative analysis.

Note that $\texttt{CellPose}_1$ can take parameters to specify: 1) the expected cell diameter (d), 2) the flow threshold (t_f), and 3) the cell probability threshold (t_p), which can substantially affect the model segmentation results in terms of detected cell numbers. Empirically, we determined $d = 50$ pixels, $t_f = 0.8$ and $t_p = -5$ to be satisfactory on the HL-1 image dataset.

Extending on [4], we follow the same approach by performing a full fine-tuning of $\texttt{CellPose}_1$ with default hyper-parameters (Table 1) - but this time with CLAHE + MF training images and new GTs annotated by experts. We refer to this version as $\texttt{CellPose}_2$. Note that in $\texttt{CellPose}_2$, d is directly estimated by the algorithm while t_f and t_p are set as for $\texttt{CellPose}_1$.



(a) Greyscale (b) CLAHE + MF (c) HE + MF



(d) ExDA

Fig. 1. ExDA of image pre-processing strategies. (a) Example of a composite greyscale image processed using CLAHE (b) or HE (c) followed by MF. (d) Quantification of Cellpose and SAM's semantic segmentation performance on 6 randomly selected HL-1 networks. Data are reported as mean ± stdev values.

Learning rate	**weight decay**	**Batch Size**	**Epochs**	**Optimiser**	**Loss fucntion**	**Augmentation**
0.1	0.0001	8	100	Adam	Dice Loss	False

Table 1. Hyper-parameter settings for fine-tuning $\texttt{CellPose}_2$

3.5 Post-processing routine

As evident from Fig. 2a, `SAM` struggles with cell segmentation tasks in the HL-1 dataset, despite using context-aware features to generalise to new data [16]. Its boundary sensitivity causes misidentification of cellular nuclei and organelles as separate entities, often leading to incorrect overlaps, image fragmentation and reduced model accuracy. To enhance the quality of predicted results, we designed a post-processing routine consisting of four sequential steps.

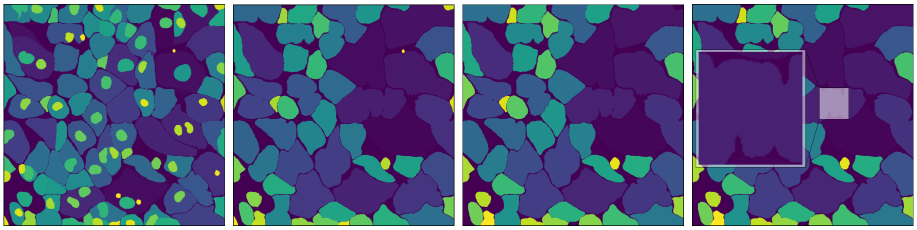
1. Removing predicted objects larger than $250,000$ pixels (i.e., 95% of the image area), to correct errors where the entire background was segmented due to poorly visible boundaries caused by low brightness.
2. Removal of masks significantly overlapping with larger segments due to unwanted detection of intracellular structures (Fig. 2b).
3. Removing small masks (SM) below the presumed pixel cell size limit. In light of $\texttt{CellPose}_\texttt{i}$ ($i \in \{1, 2\}$) single-cell segmentation capabilities, we used the two configurations as the references for a frequency distribution analysis of the predicted HL-1 cell areas. Considering positive skewness, lower-size outliers were identified using Tukey's rule to have an area inferior to 372 pixels for $\texttt{CellPose}_1$ and 424 pixels for $\texttt{CellPose}_2$. Therefore, we distinguish between $\texttt{SAM}_1$ as the configuration using the $\texttt{CellPose}_1$ threshold (Fig. 2c) and $\texttt{SAM}_2$ as the one using the $\texttt{CellPose}_2$ threshold for post-processing.
4. Filling small holes within cells generated from the previous steps using morphological operations of erosion and dilation (Fig. 2d).

For consistency, we also tested this routine in the $\texttt{CellPose}_\texttt{i}$ algorithms by skipping the first two steps that are unnecessary in this case. To indicate the application of post-processing, we use the superscript 'p', obtaining the $\texttt{SAM}^p_\texttt{i}$ and $\texttt{CellPose}^p_\texttt{i}$ approaches.

3.6 Hybrid algorithmic approaches

We present multiple new segmentation pipelines that enhance semantic and instance segmentation by combining results from the baseline algorithms.

Algorithm 1 shows the working mechanism of the $\texttt{CellPose}^p_\texttt{i}$-$\texttt{SAM}^p_\texttt{i}$ approach, where ($i \in \{1, 2\}$) and is fixed across both configurations (i.e., either $i = 1$ or $i = 2$, but never a mix of both). In these algorithms, the baseline models are followed by post-processing and their outputs are selected so that additional masks from $\texttt{SAM}^p_\texttt{i}$ that do not significantly intersect with any segment from $\texttt{CellPose}^p_\texttt{i}$ are added to those obtained from $\texttt{CellPose}^p_\texttt{i}$. Note that the threshold $\theta = 91$ in line 5 is set empirically to ensure minimal intersection between any $\texttt{SAM}^p_\texttt{i}$ and $\texttt{CellPose}^p_\texttt{i}$ prediction before inclusion in the combined output.



(a) `SAM` (b) Remove overlap (c) Remove SM (d) Fill holes

Fig. 2. `SAM` results before post-processing (a) and at various steps of the post-processing routine to generate the $\mathtt{SAM}_{\mathtt{i}}^{p}$ configuration.

Similarly, $\mathtt{SAM}_{\mathtt{i}}^{p}$-$\mathtt{CellPose}_{\mathtt{i}}^{p}$ refers to the approach obtained by adding masks predicted by $\mathtt{CellPose}_{\mathtt{i}}^{p}$ to those from $\mathtt{SAM}_{\mathtt{i}}^{p}$, following a dual logic with respect to Algorithm 1.

Algorithm 1 $\mathtt{CellPose}_{\mathtt{i}}$-$\mathtt{SAM}_{\mathtt{i}}$

```
CS←C←$\mathtt{CellPose}_{\mathtt{i}}^{p}$    ▷ Returns $N^c$ tuples $\langle m_n, l_n \rangle$ of masks $m_n$ and labels $l_n$
S←$\mathtt{SAM}_{\mathtt{i}}^{p}$    ▷ Returns $N^s$ tuples $\langle m_n, l_n \rangle$ of masks $m_n$ and labels $l_n$
for each mask $m_i \in$ S do
    for each mask $m_j \in$ C do
        if $|m_i \cap m_j| \leq \theta$ then    ▷ Intersection threshold $\theta = 91$ pixels
            Add $\langle m_i, l_i \rangle$ to CS
        end if
    end for
end for
Return CS    ▷ Combined predictions
```

We propose Algorithm 2 to reduce $\mathtt{CellPose}_{\mathtt{i}}$ oversegmentation [4]. We first use $\mathtt{SAM}_{\mathtt{i}}^{p}$, which does not exhibit this artefact, to find oversegmented regions in $\mathtt{CellPose}_{\mathtt{i}}^{p}$ and merge them into a single mask. To find oversegmentation, we calculate the percentage of overlap between each $\mathtt{CellPose}_{\mathtt{i}}^{p}$ and $\mathtt{SAM}_{\mathtt{i}}^{p}$ mask. An overlap of 40% or more indicates oversegmentation, while lower values prompt the original $\mathtt{CellPose}_{\mathtt{i}}^{p}$ mask to be included in the result (Algorithm 2, line 6). If two or more distant $\mathtt{CellPose}_{\mathtt{i}}^{p}$ masks overlap with the same $\mathtt{SAM}_{\mathtt{i}}^{p}$ prediction, they are instead replaced with the original $\mathtt{SAM}_{\mathtt{i}}^{p}$ mask (not shown in Algorithm 2). We refer to the outcome of this algorithm as the $\mathtt{CellPose}_{\mathtt{i}}^{p}$ RO configuration.

3.7 Evaluation metrics

We quantify the model's segmentation performance using already established pixel-level evaluation metrics for semantic segmentation, namely IoU score, DICE (F-1) score, precision, recall and accuracy [22, 4].

Algorithm 2 Resolve Oversegmentation

```
C ← $\texttt{CellPose}_{\texttt{i}}^{p}$    ▷ Return $N^c$ tuples $\langle m_n, l_n \rangle$ of masks $m_n$ and labels $l_n$
S ← $\texttt{SAM}$    ▷ Return $N^s$ tuples $\langle m_n, l_n \rangle$ of masks $m_n$ and labels $l_n$
RO ← $\emptyset$    ▷ Used to store resolved prediction
for each mask $m_i \in$ S do
    for each mask $m_j \in$ C do
        if $\frac{|m_i \cap m_j|}{|m_j|} \cdot 100 \geq \theta$ then    ▷ $\theta = 40\%$ common pixels
            Add $\langle m_j, l_i \rangle$ to RO
        end if
    end for
end for
Add remaining $\langle m_n, l_n \rangle \in$ C to RO    ▷ Add all non-oversegmented masks
Return RO    ▷ Resolved (non-oversegmented) predictions
```

To quantify oversegmentation, we focus on instance segmentation where a match (i.e., IoU ≥ 0.5) is sought between each GT and predicted mask [15]. Hence, when IoU ≥ 0.5, we record the predicted mask as true positive (TP); when $0 \leq$ IoU < 0.5, we record the predicted mask as false positive (FP); The difference between the GT masks and the TP gives false negatives (FN). From this classification, we calculate the object-wise IoU, DICE, precision, and recall metrics. We also report the 'average IoU' score computed across each predicted segment and GT for each image.

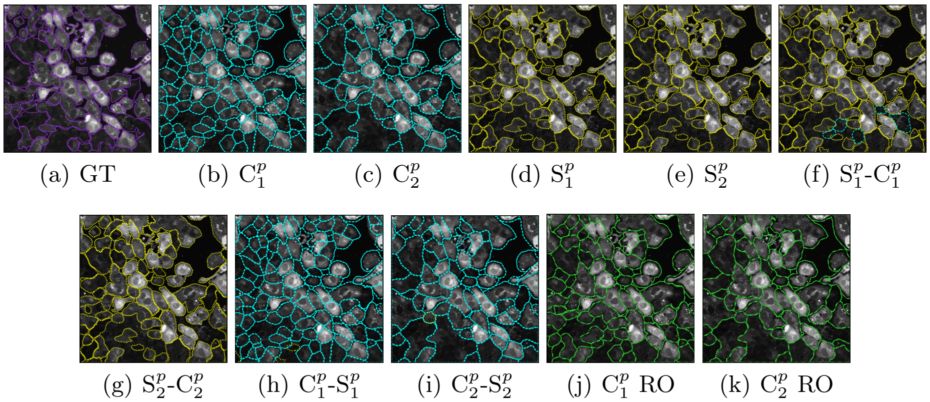
Undersegmentation is calculated using an adapted version of the undersegmentation error (USE) formula from [19]. For each predicted mask, we find the GT mask with the highest IoU, add the overflowing pixel count to the total USE, and divide by the total GT pixels to obtain a percentage.

4 Results and discussion

We evaluated a total of 13 segmentation methods, including the baseline algorithms outlined in Section 3.4, whose results for SAM are shown in Fig. 2a, their versions using the post-processing routines defined in Section 3.5, and their combined variants described in Section 3.6. For consistency, all segmentation results shown here are superimposed on the composite greyscale images, despite being obtained after pre-processing (see Section 1).

From Figs. 3 and 4, it is immediate to notice that the $\texttt{CellPose}_{\texttt{1}}^{p}$ configuration performs better on smaller cells in less magnified (Fig. 3b) than more magnified images (Fig. 4b). Generally, $\texttt{CellPose}_{\texttt{1}}^{p}$ results in more fragmented segmentation than the GT annotations and the other methods. This result aligns with [4], which identified $\texttt{CellPose}_{\texttt{1}}$'s over-segmentation, a tendency which we also note in $\texttt{CellPose}_{\texttt{1}}^{p}$-$\texttt{SAM}_{\texttt{1}}^{p}$ (Fig. 3h and Fig. 4h), where image fragmentation seems to be further aggravated by the addition of small, minimally overlapping masks from the $\texttt{SAM}_{\texttt{1}}^{p}$ prediction to the $\texttt{CellPose}_{\texttt{1}}^{p}$ result.

$\texttt{SAM}_{\texttt{i}}^{p}$ (Figs. 3d and 4d) recognises fewer cells than $\texttt{CellPose}_{\texttt{1}}^{p}$, but its segmentation results are less fragmented and more accurate in recognising CAs,

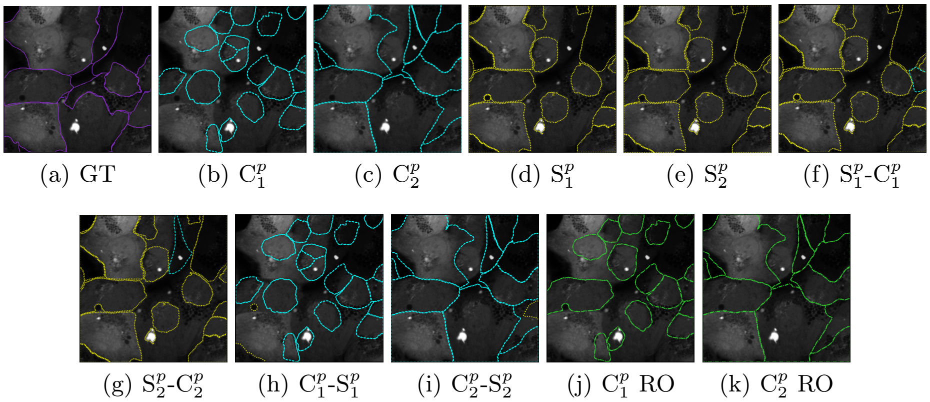

(a) GT (b) C^p_1 (c) C^p_2 (d) S^p_1 (e) S^p_2 (f) S^p_1-C^p_1

(g) S^p_2-C^p_2 (h) C^p_1-S^p_1 (i) C^p_2-S^p_2 (j) C^p_1 RO (k) C^p_2 RO

Fig. 3. Example of segmentation results on the training set

particularly in brightly lit areas with clear cell boundaries. However, it struggles with distinguishing between foreground and background, often misclassifying these areas (see Fig. 4d). We hypothesise that the enhancement of background pixel darkness due to HE median pre-processing (Fig. 1c) may introduce noisy artefacts, leading to segmentation errors.

The `CellPose`$^p_1$ RO pipeline addresses the limitations of the baseline algorithm by leveraging the accurate segmentation of `CellPose`$^p_1$ single cells and the reduced oversegmentation of `SAM`$^p_1$, yielding better results (Fig. 3j and Fig. 4j).


(a) GT (b) C^p_1 (c) C^p_2 (d) S^p_1 (e) S^p_2 (f) S^p_1-C^p_1

(g) S^p_2-C^p_2 (h) C^p_1-S^p_1 (i) C^p_2-S^p_2 (j) C^p_1 RO (k) C^p_2 RO

Fig. 4. Example of segmentation results on the testing set

The fine-tuned `CellPose`$^p_2$ algorithm (Fig. 3c and Fig. 4c) returns less fragmented segmentation results, but appears to be affected by the heterogeneous intensity of the Fluo3 reporter across the cellular network. Consequently, objects with similar Fluo3-loading intensities are often grouped under a single mask, causing cell undersegmentation in some regions. Undersegmentation is also observed in $\texttt{CellPose}_1^p$ RO (Fig. 3j and Fig. 4j), though less so than the $\texttt{CellPose}_2^p$ and $\texttt{CellPose}_2^p$ RO algorithms (Fig. 3k and Fig. 4k).

4.1 Semantic segmentation results

Semantic evaluation metrics (see Section 3.7) are presented in Tables 2 (training) and 3 (testing) as mean (%) ± stdev. Mean values are calculated by averaging segmentation masks per image and then across the dataset. The best metric (highest mean, lowest stdev) is shown in boldface.

Our hybrid approaches generally outperform baseline algorithms in pixel-wise segmentation across most metrics. The $\texttt{CellPose}_2^p\texttt{-SAM}_2^p$ configuration excelled in semantic evaluation on the training set across all metrics except precision, where $\texttt{CellPose}_1^p$ RO outperformed all other models.

IoU scores from $\texttt{CellPose}_2^p\texttt{-SAM}_2^p$ in training (82.11%) and $\texttt{CellPose}_2^p$ RO in testing (83.91%), along with recall values of 95.91% (training) and 96.21% (testing) from $\texttt{CellPose}_2^p\texttt{-SAM}_2^p$, surpassed in training ($\texttt{CellPose}_2^p\texttt{-SAM}_2^p$), improved the best pipeline in [4] by more than 10%. However, none of the hybrid approaches surpassed the precision and accuracy in [4], likely due to `SAM`'s tendency to misclassify background as foreground (see Section 4). Classification errors can introduce FP pixels into the $\texttt{CellPose}_1^p$ result, reducing precision and accuracy due to their sensitivity to wrong segment predictions.

Overall, lower stdev values for all metrics in both baseline and novel models (Tables 2 and 3) indicate the effectiveness of our training-testing split (Section 3.2) and model-specific pre-processing (Section 3.3) in standardizing segmentation performance across diverse imaging datasets like HL-1.

	IoU score(%)	**DICE(%)**	**Precision(%)**	**Recall(%)**	**Accuracy(%)**
C$_1$	79.26 ± 5.96	88.30 ± 3.92	85.99 ± 5.22	91.67 ± 8.39	80.94 ± 5.35
C$_2$	81.83 ± 5.58	89.90 ± 3.39	85.20 ± 5.05	95.52 ± 5.01	83.22 ± 4.64
C$_1^p$	79.14 ± 5.92	88.23 ± 3.90	85.98 ± 5.23	91.52 ± 8.36	80.83 ± 5.33
C$_2^p$	81.78 ± 5.56	89.88 ± 3.38	85.23 ± 5.04	95.43 ± 5.03	83.19 ± 4.61
S	72.40 ± 10.15	83.54 ± 7.74	84.97 ± 7.44	83.26 ± 11.84	74.74 ± 9.15
S$_1^p$	71.57 ± 10.41	82.94 ± 7.98	85.46 ± 7.49	81.77 ± 12.28	74.11 ± 9.47
S$_2^p$	71.38 ± 10.47	82.81 ± 8.02	85.57 ± 7.50	81.46 ± 12.38	73.97 ± 9.55
S$_1^p$-C$_1^p$	73.62 ± 8.37	84.52 ± 5.97	85.70 ± 6.86	84.34 ± 9.75	75.92 ± 7.63
S$_2^p$-C$_2^p$	71.99 ± 9.52	83.33 ± 7.00	85.61 ± 7.06	82.34 ± 11.36	74.46 ± 8.81
C$_1^p$-S$_1^p$	79.37 ± 5.67	88.38 ± 3.71	85.82 ± 5.23	91.95 ± 7.87	81.01 ± 5.06
C$_2^p$-S$_2^p$	**82.11 ± 5.47**	**90.07 ± 3.32**	85.18 ± 5.07	**95.91 ± 4.65**	**83.47 ± 4.52**
C$_1^p$ RO	79.19 ± 5.73	88.27 ± 3.74	**86.02 ± 5.28**	91.53 ± 8.14	80.87 ± 5.16
C$_2^p$ RO	81.76 ± 5.55	89.86 ± 3.37	85.25 ± 5.04	95.36 ± 5.01	83.17 ± 4.61

Table 2. Semantic segmentation results over the training set

	IoU score(%)	**DICE(%)**	**Precision(%)**	**Recall(%)**	**Accuracy(%)**
C_1	78.52 ± 10.94	87.45 ± 8.50	87.72 ± 5.01	89.25 ± 13.64	79.74 ± 10.38
C_2	83.84 ± 4.83	91.14 ± 2.86	87.02 ± 4.99	96.08 ± 4.82	84.67 ± 4.31
C_1^p	78.42 ± 10.91	87.39 ± 8.48	87.74 ± 5.01	89.10 ± 13.61	79.65 ± 10.36
C_2^p	83.76 ± 4.75	91.09 ± 2.81	87.06 ± 4.98	95.94 ± 4.80	84.60 ± 4.25
S	75.03 ± 7.73	85.50 ± 5.26	87.09 ± 5.42	84.54 ± 8.51	76.62 ± 7.01
S_1^p	74.23 ± 7.92	84.96 ± 5.42	87.58 ± 5.49	83.13 ± 8.88	76.00 ± 7.21
S_2^p	74.06 ± 8.00	84.85 ± 5.46	87.71 ± 5.51	82.81 ± 8.96	75.88 ± 7.30
S_1^p-C_1^p	75.32 ± 7.52	85.70 ± 5.05	87.39 ± 5.82	84.60 ± 7.78	76.92 ± 6.92
S_2^p-C_2^p	74.54 ± 7.64	85.19 ± 5.21	87.62 ± 5.61	83.45 ± 8.30	76.28 ± 7.00
C_1^p-S_1^p	78.66 ± 10.65	87.57 ± 8.19	87.72 ± 4.95	89.38 ± 13.30	79.87 ± 10.14
C_2^p-S_2^p	83.91 ± 4.80	91.18 ± 2.85	86.98 ± 4.98	$\mathbf{96.21 \pm 4.65}$	84.73 ± 4.29
C_1^p RO	78.56 ± 10.37	87.53 ± 7.92	$\mathbf{87.76 \pm 5.06}$	89.23 ± 13.08	79.78 ± 9.83
C_2^p RO	$\mathbf{83.91 \pm 4.54}$	$\mathbf{91.19 \pm 2.68}$	87.22 ± 4.98	95.93 ± 4.45	$\mathbf{84.75 \pm 4.13}$

Table 3. Semantic segmentation metrics over the testing set

4.2 Instance segmentation results

In terms of instance segmentation, results in Tables 4 and 5 show a trend similar to that for semantic segmentation, with all approaches originating from $\texttt{CellPose}_2$ generally outperforming the counterparts employing $\texttt{CellPose}_1$. Therefore, fine-tuning is also successful in this context, improving recall by 7% to 9% and precision by 18% to 20% in both training (Table 4) and test sets (Table 5).

Coherently with visual inspection, $\texttt{CellPose}_2^p$ improves instance segmentation metrics, reducing fragmentation (see Fig. 3 and 4) and addressing $\texttt{CellPose}_1$'s oversegmentation issues [4]. $\texttt{CellPose}_2^p$ achieved the highest average IoU, precision, DICE and IoU score among $\texttt{CellPose}_2$ algorithms in the training set and excelled in precision, DICE and IoU in the test set. This underscores our post-processing pipeline's effectiveness in enhancing segmentation quality by reducing false positive segments (see Section 3.7).

It should be noted that $\texttt{CellPose}_2^p$-$\texttt{SAM}_2^p$ performs similarly (slightly worse) than $\texttt{CellPose}_2$ and outperforms the baseline model in terms of recall on the training set (Table 4). Arguably, the two configurations are comparable for object-level detection.

Finally, we stress that the USE values in Tables 4 and 5, with the worst values highlighted in red, show undersegmentation issues, especially with $\texttt{CellPose}_2^p$ RO, followed by all other $\texttt{CellPose}_2^p$ configurations. We attribute this issue to arise during the fine-tuning process in $\texttt{CellPose}_1$, where the algorithm can learn to differentiate single cells from multicellular aggregates based on variations in Fluo3-loading intensities rather than actual cell boundaries, resulting in incorrect segmentation of distinct cellular entities (see Figs. 3c and 4c).

5 Conclusions and recommendations

Our findings highlight the benefit of combining the proposed equalisation-based pre-processing (HE and CLAHE) with noise filtering (median filtering) in images

	Average IoU(%)	**Precision(%)**	**Recall(%)**	**DICE(%)**	**IoU score(%)**	**USE(%)**
C_1	33.42 ± 8.88	29.50 ± 12.34	57.29 ± 14.15	38.41 ± 13.46	24.62 ± 10.28	21.66 ± 6.32
C_2	46.54 ± 9.65	49.46 ± 14.52	64.44 ± 11.39	55.39 ± 13.13	39.42 ± 12.39	26.56 ± 6.47
C_1^p	33.86 ± 9.02	29.91 ± 12.52	57.12 ± 14.03	38.70 ± 13.52	24.86 ± 10.36	21.64 ± 6.32
C_2^p	$\mathbf{48.28 \pm 9.10}$	$\mathbf{51.36 \pm 14.39}$	64.31 ± 11.31	$\mathbf{56.56 \pm 12.86}$	$\mathbf{40.53 \pm 12.36}$	26.53 ± 6.45
S	19.12 ± 6.13	16.05 ± 6.91	61.99 ± 14.07	25.21 ± 9.61	14.77 ± 6.44	16.86 ± 9.01
S_1^p	35.03 ± 9.97	32.74 ± 12.29	49.55 ± 11.41	38.87 ± 11.96	24.80 ± 9.12	24.48 ± 14.56
S_2^p	35.93 ± 10.06	33.63 ± 12.46	49.27 ± 11.28	39.41 ± 11.92	25.22 ± 9.14	24.35 ± 4.57
S_1^p-C_1^p	34.32 ± 10.06	31.75 ± 12.54	50.12 ± 11.45	38.25 ± 12.25	24.35 ± 9.30	24.87 ± 14.55
S_2^p-C_2^p	35.82 ± 9.97	33.45 ± 12.37	49.72 ± 10.99	39.43 ± 11.84	25.22 ± 9.10	24.61 ± 14.62
C_1^p-S_1^p	33.24 ± 9.17	29.39 ± 12.58	57.30 ± 13.90	38.26 ± 13.62	24.53 ± 10.41	21.73 ± 6.29
C_2^p-S_2^p	47.52 ± 9.33	50.48 ± 14.52	$\mathbf{64.68 \pm 11.41}$	56.14 ± 13.03	40.15 ± 12.43	26.65 ± 6.48
C_1^p RO	38.12 ± 9.42	33.60 ± 13.20	44.26 ± 11.63	37.56 ± 12.08	23.79 ± 9.05	29.47 ± 11.62
C_2^p RO	46.67 ± 9.34	47.07 ± 13.87	49.96 ± 11.44	47.92 ± 11.89	32.28 ± 9.93	$\mathbf{32.33 \pm 9.87}$

Table 4. Instance segmentation and USE error results over the training set

	Average IoU(%)	**Precision(%)**	**Recall(%)**	**DICE(%)**	**IoU score(%)**	**USE(%)**
C_1	31.02 ± 9.75	26.03 ± 12.07	53.10 ± 14.36	34.37 ± 13.88	21.59 ± 10.09	19.66 ± 5.94
C_2	44.00 ± 9.66	44.41 ± 14.74	$\mathbf{62.20 \pm 12.64}$	51.42 ± 14.37	35.84 ± 12.76	24.73 ± 5.75
C_1^p	31.55 ± 9.63	26.48 ± 12.10	52.99 ± 14.32	34.75 ± 13.81	21.87 ± 10.10	19.63 ± 5.93
C_2^p	46.02 ± 9.94	$\mathbf{46.53 \pm 15.95}$	61.53 ± 12.60	$\mathbf{52.45 \pm 14.69}$	$\mathbf{36.85 \pm 13.13}$	24.69 ± 5.76
S	18.23 ± 6.04	14.58 ± 5.99	60.26 ± 11.65	23.22 ± 8.63	13.40 ± 5.50	14.24 ± 5.54
S_1^p	35.86 ± 8.80	32.51 ± 10.66	51.98 ± 10.00	39.42 ± 10.74	25.11 ± 8.37	20.79 ± 9.03
S_2^p	36.96 ± 8.94	33.60 ± 10.95	51.87 ± 9.99	40.21 ± 10.86	25.75 ± 8.60	20.60 ± 9.00
S_1^p-C_1^p	34.93 ± 9.20	31.36 ± 10.77	52.01 ± 10.11	38.45 ± 10.92	24.37 ± 8.39	21.22 ± 9.45
S_2^p-C_2^p	36.73 ± 9.11	33.10 ± 10.99	51.87 ± 9.99	39.81 ± 10.87	25.43 ± 8.55	20.84 ± 9.04
C_1^p-S_1^p	31.53 ± 9.48	26.64 ± 11.81	53.80 ± 13.08	35.06 ± 13.36	22.05 ± 9.84	19.66 ± 5.87
C_2^p-S_2^p	45.61 ± 10.12	45.89 ± 15.66	61.53 ± 12.60	52.06 ± 14.76	36.49 ± 13.08	24.71 ± 5.76
C_1^p RO	35.62 ± 9.60	30.02 ± 11.70	42.30 ± 9.74	34.34 ± 10.57	21.24 ± 7.94	26.74 ± 9.30
C_2^p RO	$\mathbf{46.06 \pm 8.54}$	43.66 ± 13.46	49.22 ± 8.65	45.76 ± 10.89	30.33 ± 9.37	$\mathbf{29.28 \pm 7.41}$

Table 5. Instance segmentation metrics and USE error in the testing set

before segmentation, and the key roles of fine-tuning `CellPose` and adding our post-processing routine to both `CellPose` and `SAM` to segment HL-1 cells.

Furthermore, we demonstrate the use of popular baseline segmentation algorithms in hybrid pipelines by jointly using the proposed `CellPose` and `SAM` configurations for multicellular (HL-1) layers, with $\texttt{CellPose}_2^p$-$\texttt{SAM}_2^p$ resulting as the recommended approach for both semantic segmentation and instances despite having slightly lower performance metric values than $\texttt{CellPose}_2^p$, which, however, displays a higher tendency to undersegmentation.

The identified USE will be subject to future studies for its mitigation by improving the $\texttt{CellPose}_1$ fine-tuning process, e.g., by using a human-in-the-loop approach, as in [21], for manual correction. We will also focus on increasing the dataset size to be able to fully retrain Cellpose rather than fine-tuning it.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Banerji, S., Mitra, S.: Deep learning in histopathology: A review. WIREs Data Mining and Knowledge Discovery **12**(1), e1439 (2022). https://doi.org/https://doi.org/10.1002/widm.1439, https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1439
2. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al.: Qupath: Open source software for digital pathology image analysis. Scientific reports **7**(1), 1–7 (2017)
3. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
4. Caraffini, F., Eshkiki, H., Mohammadpour, M., Sullo, N., George, C.H.: Towards improving single-cell segmentation in heterogeneous configurations of cardiomyocyte networks. In: Xie, X., Styles, I., Powathil, G., Ceccarelli, M. (eds.) Artificial Intelligence in Healthcare. pp. 104–117. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-67285-9_8
5. CH, G.: Arrhythmogenic mutation-linked defects in ryanodine receptor autoregulation reveal a novel mechanism of ca^{2+} release channel dysfunction. Circ Res **98**, 88–97 (2006)
6. Chen, L., Wu, Y., Stegmaier, J., Merhof, D.: Sortedap: Rethinking evaluation metrics for instance segmentation (2023), https://arxiv.org/abs/2309.04887
7. Claycomb, W.C., Lanson Jr, N.A., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., Izzo Jr, N.J.: Hl-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proceedings of the National Academy of Sciences **95**(6), 2979–2984 (1998)
8. Cunha, I., Latron, E., Bauer, S., Sage, D., Griffié, J.: Machine learning in microscopy–insights, opportunities and challenges. Journal of cell science **137**(20) (2024)
9. Dao, D., Fraser, A.N., Hung, J., Ljosa, V., Singh, S., Carpenter, A.E.: Cellprofiler analyst: interactive data exploration, analysis and classification of large biological image sets. Bioinformatics **32**(20), 3210–3212 (2016)
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
11. Emre Dedeagac, C., Koyuncu, C.F., Adams, M.M., Edemen, C., Ugurdag, B.C., Ilgim Ardic-Avci, N., Fatih Ugurdag, H.: A guided-ensembling approach for cell counting in fluorescence microscopy images. IEEE Access **12**, 195552–195560 (2024). https://doi.org/10.1109/ACCESS.2024.3517641
12. George, C.H., Higgs, G.V., Lai, F.A.: Ryanodine receptor mutations associated with stress-induced ventricular tachycardia mediate increased calcium release in stimulated cardiomyocytes. Circulation research **93**(6), 531–540 (2003)
13. George, C.H., Rogers, S.A., Bertrand, B.M., Tunwell, R.E., Thomas, N.L., Steele, D.S., Cox, E.V., Pepper, C., Hazeel, C.J., Claycomb, W.C., et al.: Alternative splicing of ryanodine receptors modulates cardiomyocyte ca^{2+} signaling and susceptibility to apoptosis. Circulation research **100**(6), 874–883 (2007)
14. Kao, J., Harootunian, A., Tsien, R.: Photochemically generated cytosolic calcium pulses and their detection by fluo-3. The Journal of biological chemistry **264**, 8179–84 (06 1989). https://doi.org/10.1016/S0021-9258(18)83166-0

15. Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P.: Panoptic segmentation (2019), https://arxiv.org/abs/1801.00868
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything (2023), https://arxiv.org/abs/2304.02643
17. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., Van Valen, D.: Deep learning for cellular image analysis. Nature methods **16**(12), 1233–1246 (2019)
18. Murmu, A., Kumar, P.: Deep learning model-based segmentation of medical diseases from mri and ct images. In: TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON). pp. 608–613 (2021). https://doi.org/10.1109/TENCON54134.2021.9707278
19. Neubert, P., Protzel, P.: Superpixel benchmark and comparison. In: Proc. Forum Bildverarbeitung. vol. 6, pp. 1–12 (2012)
20. Ningsih, D.R., et al.: Improving retinal image quality using the contrast stretching, histogram equalization, and clahe methods with median filters. International Journal of Image, Graphics and Signal Processing **14**(2), 30 (2020)
21. Pachitariu, M., Stringer, C.: Cellpose 2.0: how to train your own model. Nature methods **19**(12), 1634–1641 (2022)
22. Rainio, O., Teuho, J., Klén, R.: Evaluation metrics and statistical tests for machine learning. Scientific Reports **14**(1), 6086 (2024)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
24. Stoltzfus, C.R., Filipek, J., Gern, B.H., Olin, B.E., Leal, J.M., Wu, Y., Lyons-Cohen, M.R., Huang, J.Y., Paz-Stoltzfus, C.L., Plumlee, C.R., Pöschinger, T., Urdahl, K.B., Perro, M., Gerner, M.Y.: Cytomap: A spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. Cell Reports **31**(3), 107523 (2020). https://doi.org/https://doi.org/10.1016/j.celrep.2020.107523, https://www.sciencedirect.com/science/article/pii/S221112472030423X
25. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature methods **18**(1), 100–106 (2021)
26. Weigert, M., Schmidt, U., Haase, R., Sugawara, K., Myers, G.: Star-convex polyhedra for 3d object detection and segmentation in microscopy. In: The IEEE Winter Conference on Applications of Computer Vision (WACV) (March 2020). https://doi.org/10.1109/WACV45572.2020.9093435
27. Wen, C., Miura, T., Voleti, V., Yamaguchi, K., Tsutsumi, M., Yamamoto, K., Otomo, K., Fujie, Y., Teramoto, T., Ishihara, T., Aoki, K., Nemoto, T., Hillman, E.M., Kimura, K.D.: 3deecelltracker, a deep learning-based pipeline for segmenting and tracking cells in 3d time lapse images. eLife **10**, e59187 (mar 2021). https://doi.org/10.7554/eLife.59187, https://doi.org/10.7554/eLife.59187
28. Winfree, S.: User-accessible machine learning approaches for cell segmentation and analysis in tissue. Frontiers in Physiology **13** (2022). https://doi.org/10.3389/fphys.2022.833333, https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.833333