ACMS 60876 Report Draft

Title: Fine-tuning SAM2 for Automated Segmentation of Calcium Signaling in Simulated Non-Excitable Tissues

Authors: Yikang Gong

Claiming the use of AIGC: This report is hand-written, however, we used Grammarly to check the grammar and used Grammarly's Paraphraser for some sentences.

Abstract: Calcium ion(Ca^{2+}) patterns are pivotal in regulating cell development and regeneration. The current study approach is based on analyzing the light pattern of the cells by expressing the Genetically Encoded Calcium Indicators (GECIs) in cells, which produce fluorescence signals. However, earlier models are developed for excitable cells with clear contour and high intensity. To fill the gap of using GECIs to study the non-excitable cells, we implemented a transformer-based segmentation model, the Segment Anything Model 2 (SAM2), to accurately identify the activated cells in non-excitable epithelial tissue. We achieved this goal by simulating realistic GECI fluorescence image data based on a mathematical calcium model and attempting to adapt the model to real datasets.

1. Introduction

Calcium ions (Ca^{2+}) are ubiquitous secondary messengers that translate diverse extracellular stimuli into specific intracellular responses. They play pivotal roles in processes like cell proliferation, differentiation, migration, and apoptosis, which collectively drive tissue development and regeneration [1]. Understanding the intricate spatiotemporal dynamics of Ca^{2+} signals is crucial for deciphering cellular communication and function. While patch-clamping provides direct measurements in excitable cells, they are invasive and impractical for large cell populations or non-excitable cells (e.g., epithelial, endothelial) within tissues [2].
Genetically Encoded Calcium Indicators (GECIs) offer a non-invasive approach to studying the Ca^{2+} activity of non-excitable cells, converting intracellular Ca^{2+} fluctuations into fluorescent signals detectable by light microscopy [3]. This technology has unveiled complex Ca^{2+} patterns, including localized spikes and coordinated global fluttering or oscillations [4].
Analyzing datasets of images from non-excitable cells presents considerable challenges due to their innate characteristics. Epithelial and endothelial cells, the two major types of non-excitable cells, are defined by their cell-cell gap junctions and morphology. These junctions enable the transfer of calcium between cells, which, when visualized, typically results in clusters of fluorescing cells rather than the distinct, ring-like patterns seen in neuron cells. Moreover, non-excitable cells exhibit a self-illuminating property, meaning they can produce fluorescence signals independent of GECIs. This presents challenges for representation learning in models, especially when there's an intensity bias. Coupled with low-intensity and noisy images, traditional convolutional networks struggle to identify active, non-excitable cells [5] accurately.

Accurate segmentation of individual cells is a critical first step for extracting single-cell activity traces and classifying signaling patterns [7]. The transformer-based segmentation network, Segment Anything Model 2 (SAM2) [16], has the ability to automatically identify the region of interest via supervised fine-tuning on image and mask pairs. Here, we attempt to demonstrate a workflow that uses the calcium signaling model developed based on real biological data to generate realistic calcium patterns and coupled with synthetic imaging defects to mimic real samples and use them as supervised input to fine-tune the SAM2.
The workflow is as follows:

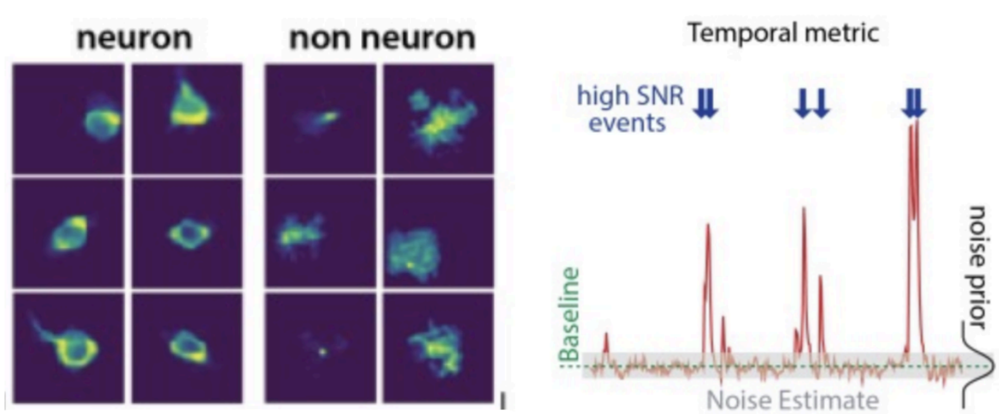
1. Generating realistic simulated GECI image data capturing known Ca^{2+} dynamics in non-excitable cell populations.
2. Fine-tuning the SAM2 model on this synthetic dataset to optimize cell detection.
3. Adapting the fine-tuned SAM2 model to real samples.

Success in this project will provide a powerful tool for automating the analysis of calcium signaling, enabling quantitative studies of intercellular communication networks and cellular behavior in engineered tissues and biological systems.

2. Related Work

CNN-based Segmentation and Activity identification: Tools like CalmAn [7], Suite2p [8], and NeuroSeg-II [9] primarily focus on neuronal two-phone imaging analysis. Two-phone microscopy provided earlier models a low noise and high intensity source images. While powerful for neuronal data, their performance can degrade when applied to non-excitable cells with different single cell fluorescence patterns, varied cell morphologies, denser packing, and subtler fluorescence changes.

To improve the classification performance, earlier models like CalmAn use the temporal data of the cell fluorescent brightness to exclude low SNR events. However, based on our model of non-excitable cells, such a filter also eliminates the information that induces the high SNR calcium events, leading the model to have bias due to biologically inappropriate data processing.



The temporal footprint of Ground Truth

Figure 1.The CalmAn's data processing methodology. CalmAn's training was based on the filtered data, which was based on the cells' temporal footprints, that only included the high Signal-to-noise ratio (SNR) events (adapted from CalmAn paper [7]).

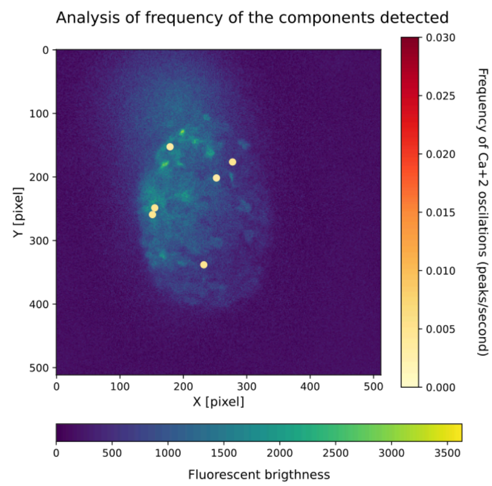


Figure 2. Real calcium images analyzed by CaImAn

When applying CalmAn to our in-house real calcium images of epithelial cells, the CaImAn fails to identify the majority of activated cells, which exhibit higher levels of green fluorescence. Additionally, it predicts false positives in areas with low fluorescence and noise (darker regions), and could not identify cells that are not rounded in shape. The strength of CalmAn lies in its ability to couple frequency with the brightness of the fluorescence, which reflects calcium activation. However, by excluding low SNR images during data processing, the model could not fully capture the entire calcium activation process, leading to a bias toward identifying only the cell shapes.

Pattern Classification: DISCo [10] explicitly addresses pattern classification in calcium imaging. It uses CNN-based segmentation combined with analyzing temporal correlations to distinguish different types of calcium activity (e.g., spikes vs. oscillations), providing insights into cellular states. However, its accuracy is dependent on the initial segmentation quality.

Foundation Models: The advent of foundation models like SAM [6] represents a paradigm shift. Pre-trained on vast datasets, they possess strong generalization capabilities. Fine-tuning these models on specific GECI images hold the potential to leverage their powerful learned representations while adapting them to the specific characteristics of the non-excitable cell images, potentially requiring less domain-specific annotated data than training a CNN from scratch.

3. Methodology

Our methodology comprises two main stages: generating realistic simulated training data and fine-tuning the SAM2 model for cell segmentation. The code implementation is available on GitHub [https://github.com/ygong2501/calcium].

3.1. In Silico Tissue Simulation

To generate training data with accurate ground truth, we employ a simulation approach based on the established methodology described by Soundarrajan [11]. We simulate a 2D monolayer of non-excitable cells with defined morphology and spatial arrangement to closely mimic real biological samples. The simulation incorporates a biophysical model of calcium dynamics, including intracellular Ca^{2+} release and intercellular communication via gap junctions. This model enables us to generate patterns such as spontaneous localized spikes and propagating waves or oscillations, consistent with experimental observations.

To realistically replicate experimental imaging defects, we introduce several types of imperfections into the simulation. Including:

1. Background fluorescence, which may appear as either a uniform glow like a faint light across the image or a spatially varying glow like a spotlight effect,
2. Spontaneous luminescence, representing random cell activity unrelated to true calcium signaling.
3. Optical distortions, such as radial distortion from lens curvature and chromatic aberration due to misalignment of color channels.
4. Multiple noise sources are incorporated, including Poisson noise (dependent on signal intensity), patterned readout noise, and random Gaussian noise.
5. Defocus and Kernel blur, which are applied either globally or locally to replicate image blurring caused by imperfect focal planes and stage motion.

Lastly, we generate masks based on each cell's simulated calcium activity. The aim is to rule out human bias during mask labeling. Conclusively, we generated calcium images with realistic visual representation and bias-free mask labeling.

3.2. SAM2 Model Fine-tuning

1. **Model Architecture:** We utilize the publicly available SAM2 published on Huggingface.
2. **Training Objective:** The model is trained to accurately predict instance segmentation masks—measured by Intersection over Union (IoU)—for all cells in each frame of the simulated video. The ground truth masks generated by the simulation serve as the training targets.
3. **Training:** The model is being fine-tuned using the PyTorch deep learning framework.

4. Evaluation

The SAM2 model was fine-tuned using our simulated GECI dataset. Its performance was then assessed on a test set composed of simulated images and real experimental GECI microscopy images.

The evaluation matrix is

1. Intersection over Union (IoU) / Jaccard Index: Measures the overlap between predicted and ground truth masks for each cell instance [13]. It is calculated as Area of Overlap / Area of Union. We will report the mean IoU (mIoU) across all detected cells.
2. Dice Coefficient (F1 Score for segmentation): Similar to IoU, it measures overlap, calculated as 2 * Area of Overlap / (Area of Prediction + Area of Ground Truth) [12]. It is often correlated with IoU but penalizes disagreements differently.
3. Pixel Accuracy: Measures the percentage of pixels correctly classified (as either belonging to a specific cell or the background) across the entire image [13].

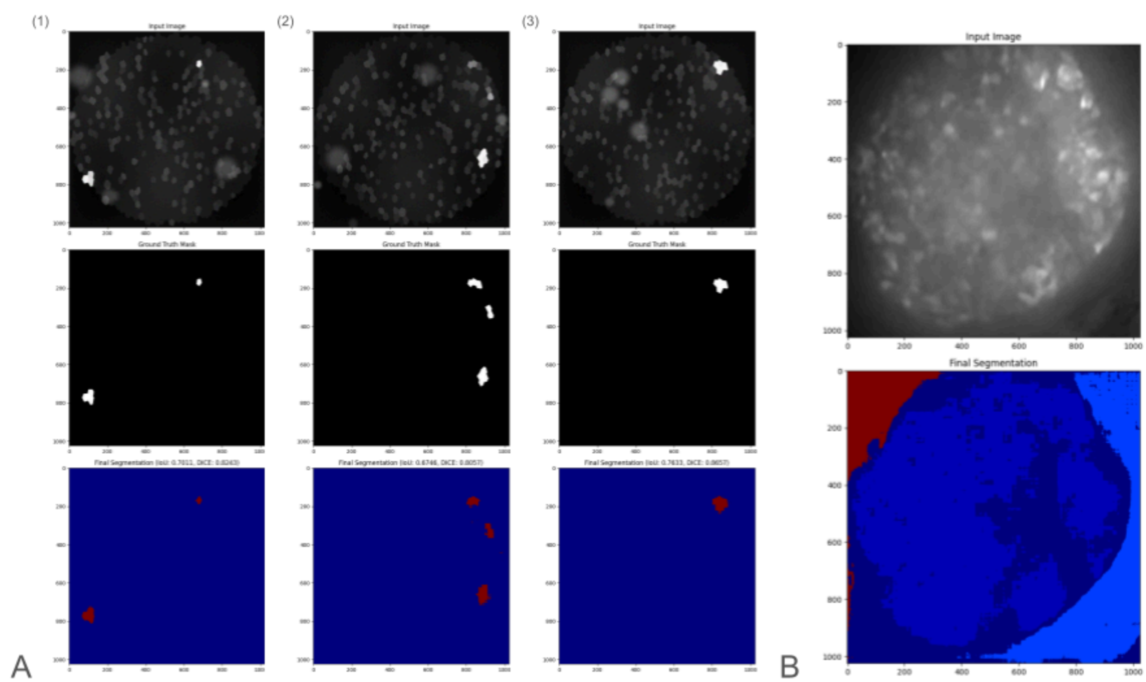
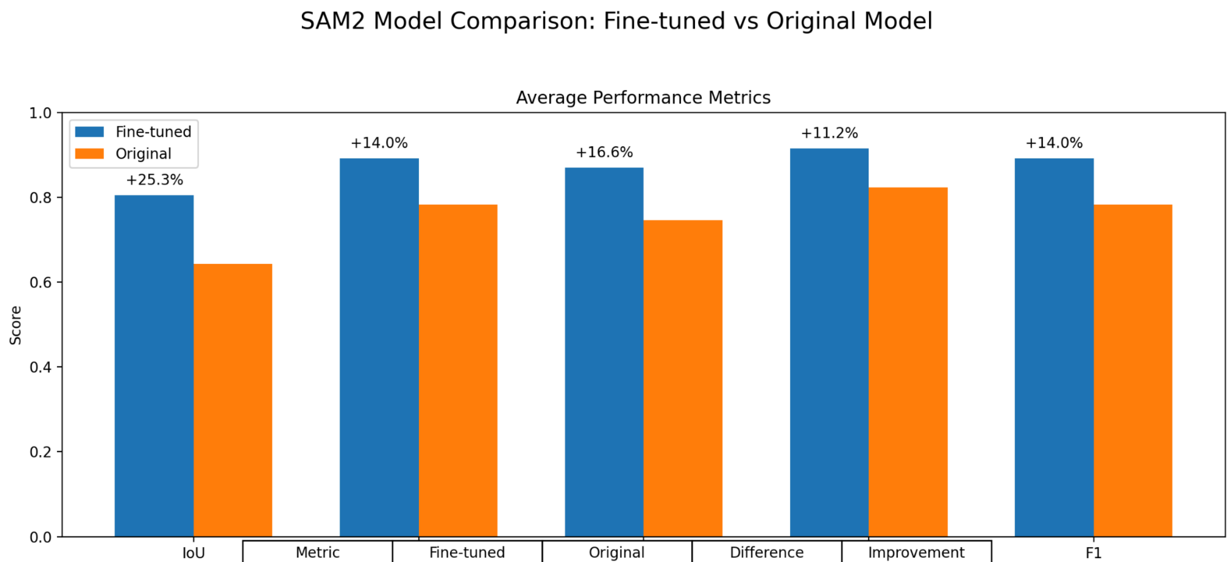


Figure 3. The visualization of fine tuned SAM2 on synthetic and real data. The darker the red color indicating higher certainty of the prediction. A. Test image 1 represents the single spot calcium activation; test image 2 represents the multispot calcium activation; test image 3 represents a challenging case when the activated cells are dimmer and are near defects. B. The inference results under the real sample. Different colors represent different classes.

Table 1. The evaluation metrics of fine tuned SAM2 compared to the original



Metric	Fine-tuned	Original	Difference	Improvement
IoU	0.8050	0.6427	+0.1623	+25.3%
Dice	0.8917	0.7823	+0.1093	+14.0%
Precision	0.8693	0.7453	+0.1240	+16.6%
Recall	0.9153	0.8233	+0.0920	+11.2%
F1	0.8917	0.7823	+0.1093	+14.0%

Note: The above metrics demonstrate how the fine-tuned SAM2 model outperforms the original model.
On average, fine-tuning improved IoU by 25.3% and Dice coefficient by 14.0%.

4.1. Performance on Simulated Data

On the simulated test set, the fine-tuned SAM2 model demonstrated that it had adapted to the characteristics of the simulated data. Qualitative inspection showed the model identifying cell-like structures and producing segmentation masks generally aligned with the simulated single cell boundaries. The model demonstrated an impressive ability to locate the cells despite the imaging defects on the data.

4.2. Performance on Real Experimental Data

When applied to real experimental GECI images, the fine-tuned model **failed to generalize**. The model demonstrated an inability to accurately detect individual cells, often failing to distinguish them from the background. In several cases, it incorrectly merged multiple distinct cells into a single, oversized segmentation mask. Additionally, the model predicts masks without any recognizable cellular structures present in the image.

5. Discussion

This project aimed to develop an automated cell segmentation tool for GECI calcium imaging by fine-tuning the SAM2 foundation model on simulated data. Nevertheless, our evaluation revealed a critical challenge: a significant **domain gap** between the simulated training and real world scenario.

Simulation vs. Real-World: While the fine-tuning process enabled the model to learn features present in our simulated dataset, these learned representations proved insufficient for handling the complexities of real microscopy images.

Diagnosing the Domain Gap: The discrepancy possibly originated from inherent differences between the simplified simulation and the actual imaging environment. We listed some of the potential cause:

1. Visual Defects: Differences in noise patterns (e.g., sensor noise, biological autofluorescence), illumination variations, contrast levels, and background textures.
2. Optical Variation: Microscope's blurring and light scattering within tissue.
3. Artifacts Disturbance: Debris and other random objects in real data not generated correctly.

Lack of Data Augmentation: To bridge this domain gap, enhancing the diversity and realism of the training data seen by the model during fine-tuning is essential. We propose focusing on **data augmentation** applied to the simulated data. By introducing variations that mimic those encountered in real images (e.g., realistic noise profiles, simulated blur, random contrast/brightness adjustments, elastic deformations for shape variability), we hypothesize the model can learn features that are more invariant to the shift between domains.

6. Conclusion and Future Work

In conclusion, our project demonstrated that while fine-tuning the SAM2 foundation model on simulated GECI data allowed it to adapt and perform segmentation tasks, it failed to generalize to real experimental images due to the domain gap between simulation and real world data. It highlights the difficulty of transferring models trained solely on simulations to the complexities of real-world biological microscopy.

Therefore, the future direction is to bridge the gap by building comprehensive data augmentation structure to avoid potential overfitting of models on simulation datasets.We plan to augment the simulated training data with more realistic variations mimicking experimental conditions, including adding diverse noise profiles, add random positioning of picture centers, adjusting contrast and brightness, and introducing elastic deformations to simulate morphological variability. The model will be iteratively re-trained with this augmented data and re-evaluated qualitatively on real images to assess generalization performance.

References

[1] Berridge, M. J., Bootman, M. D., & Roderick, H. L. (2003). Calcium signalling: dynamics, homeostasis and remodelling. *Nature reviews Molecular cell biology*, 4(7), 517-529.

[2] Stosiek, C., Garaschuk, O., Holthoff, K., & Konnerth, A. (2003). *In vivo* two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences*, 100(12), 7319-7324.

[3] Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., & Tsien, R. Y. (1997). Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature*, 388(6645), 882-7. doi:10.1038/42264

[4] Brunel, N., & Hakim, V. (1999). Fast Global Oscillations in Networks of Integrate-and-Fire Neurons with Low Firing Rates. *arXiv:cond-mat/9904278*.

[5] Park, J., et al. (2021). Cell Segmentation-Free Inference of Cell Types from in Situ Transcriptomics Data. *Nature Communications*, 12(1), 3545. doi:10.1038/S41467-021-23807-4 *(Note: While about transcriptomics, it highlights challenges with cell segmentation in complex tissues)*

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment Anything. *arXiv preprint arXiv:2304.02643*. [*Add Citation for SAM2 when available*]

[7] Giovannucci, A., et al. (2019). CalmAn: An Open Source Tool for Scalable Calcium Imaging Data Analysis. *eLife*, 8, e38173. doi:10.7554/eLife.38173

[8] Pachitariu, M., et al. (2017). Suite2p: Beyond 10,000 Neurons with Standard Two-Photon Microscopy. *bioRxiv*, 061507. doi:10.1101/061507

[9] Xu, Z., et al. (2023). NeuroSeg-II: A Deep Learning Approach for Generalized Neuron Segmentation in Two-Photon Ca^{2+} Imaging. *Frontiers in Cellular Neuroscience*, 17, 1127847. doi:10.3389/fncel.2023.1127847

[10] Kirschbaum, E., Bailoni, A., & Hamprecht, F. A. (2019). DISCo: Deep Learning, Instance Segmentation, and Correlations for Cell Segmentation in Calcium Imaging. *bioRxiv*, doi:10.1101/2019.08... *(Check for updated/published version)* [11] Soundarrajan, D. K., Huizar,

F. J., Paravitorghabeh, R., Robinett, T., & Zartman, J. J. (2021). From spikes to intercellular waves: Tuning intercellular calcium signaling dynamics modulates organ size control. *PLoS computational biology*, 17(11), e1009543.

[12] Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3), 297–302.

[13] Everingham, M., et al. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338.

[14] Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, 45(4), 427–437.

[15] Shou, Z., Wang, D., & Chang, S. F. (2016). Temporal Action Localization in Untrimmed Videos via Multi-Stage CNNs. *Proceedings of the IEEE International Conference on Computer Vision*, 267–275. *(Example for temporal evaluation metrics like tIoU)* [*Also add references cited in proposal for MOTA/MOTP, though perhaps less relevant if tracking isn't implemented yet:* Bernardin, K., & Stiefelhagen, R. (2008). Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP Journal on Image and Video Processing*, 2008, 246309.]

[16] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024, October 28). *Sam 2: Segment anything in images and videos*. arXiv.org. https://arxiv.org/abs/2408.00714