AI-enabled live-dead cell viability classification and motion forecasting

Anzhe Cheng[1], Chenzhong Yin[1], Michael A.S. Lamba[2], Mathieu Sertorio[2], Jorge Maldonado[3], Alexandre R. Sathler[4], Yu Chang[5], Catalin Chiritescu[6], Catherine A. Best-Popescu[7], Dan Ionascu[2], Nicholas Kotov[8], Shahin Nazarian[1], and Paul Bogdan[1,*]

[1]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90007, USA.
[2]Department of Radiation Oncology, College of Medicine, University of Cincinnati, Cincinnati, OH 45267, USA.
[3]Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA
[4]Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720 USA.
[5]Department of Statistics, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[6]Phi Optics, Inc., Champaign, IL 61820, USA.
[7]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[8]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, United States
[*]Correspondence and requests for materials should be addressed to P.B. (email: pbogdan@usc.edu)

ABSTRACT

Distinguishing live from dead cells is crucial in a wide variety of research fields, including regenerative medicine, toxicology, pharmacology, and cellular product manufacturing, because it allows researchers to evaluate the efficacy and toxicity of molecules, materials, and therapies and ensures the quality of manufactured cell products. The cost of failure or uncertainty for live-dead analysis can be particularly high for therapeutic compound screening and evaluating medical therapies. Here, we present a novel deep learning framework that integrates a self-attention UNet for segmentation and a transformer network for dynamic tracking of cell movements. Our proposed model achieves state-of-the-art performance, with a high intersection-over-union (IoU) score of **96%** and an area-under-curve (AUC) score of **99%** for cell segmentation and over 65% IoU of full image cell motion forecasting, highlighting its ability to predict cell dynamics accurately. The self-attention mechanism significantly enhances the model's ability to differentiate live and dead cells, even in densely packed or morphologically diverse environments. Additionally, the transformer network effectively captures temporal dependencies, enabling precise predictions of future cell movements. This integrated framework demonstrates robust performance across diverse datasets, consistently outperforming existing methods. By offering high-accuracy segmentation and predictive modeling, our approach provides a transformative tool for advancing cellular analysis in research and clinical applications, including cell therapeutics, cancer diagnostics, drug development, and regenerative medicine.

Supplementary information

Code and Data Availability

Our data and code to run the experiments can be found at https://github.com/Belis0811/Unet.

More Experiments and Dicussion

Model Comparison and Metrics

Loss Functions. The model was trained using a combination of cross-entropy loss and Jaccard loss. Cross-entropy loss measures the difference between the predicted probability distribution and the true distribution. Jaccard loss penalizes mismatches between predicted and true masks, often used in segmentation tasks.

Evaluation Metrics. All models(self-attention Unet, SAM2, and E-UNet) were evaluated on the same test set with identical criteria, allowing for a fair comparison. The primary metrics include:

- **Intersection over Union (IoU)**:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}},$$

where TP, FP, and FN stand for true positive, false positive, and false negative, respectively. IoU measures how much the predicted region overlaps with the ground truth.

- **Area Under the Curve (AUC)**: This is the area under the receiver operating characteristic (ROC) curve. It captures how well the model distinguishes between positive and negative examples over a range of classification thresholds.

- **F1-score**: The harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

  It emphasizes scenarios where both precision and recall are important.

- **Recall**:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

  This indicates the proportion of true positives that the model successfully identifies.

- **Precision**:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

  This indicates the proportion of predicted positives that are correctly predicted.

All models were tested under identical conditions, using the same dataset and the same evaluation procedures. This enables a consistent comparison of their performance across these metrics.

Learning curve of traditional Unet

Figure S1 (a) and (b) display the learning curves for IoU and Cross Entropy loss for training the traditional Unet model with our dataset. The increasing IoU values and decreasing loss in these figures indicate that the model's performance improved steadily during training. After convergence, the model achieved an average IoU score of 0.842 and a loss of 0.00567.

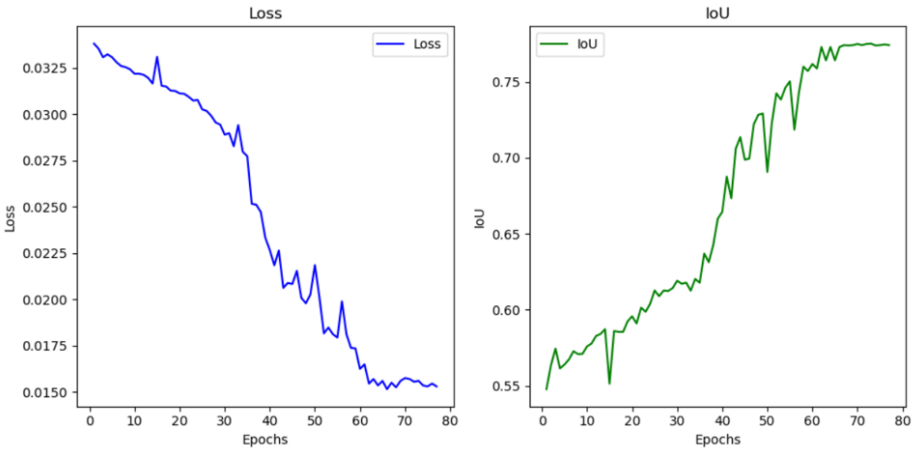

Figure S1. Performance of the baseline UNet model in unclassified cell segmentation, showing IoU and loss metrics.

Learning curve of Self-Attention Unet

We initialized our model using weights pre-trained on the traditional U-Net architecture and further refined it using the same dataset. The progression of training and validation losses is depicted in Figure S2. The training loss decreased to 1.1×10^{-4}, while the validation loss remained slightly higher, indicating a consistent but slightly conservative generalization.

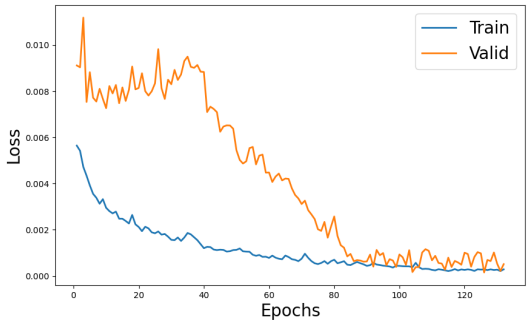

Figure S2. Learning curve for Self-Attention Unet on 10×10 magnification cell images

Confusion matrix on 40× magnification cell images

We also applied our trained model to 562 SLIM images not used in training. We compared the dominant semantic label between the predicted cell nuclei and the ground truth for individual nuclei. The result is shown in Table S1.

		Ground truth	
		Live ($n = 101718$)	Dead ($n = 99083$)
Cells	Live	99.2%	0.8%
	Dead	0.8%	99.2%
Evaluation	Precision	96.4%	97.5%
	Recall	96.6%	97.9%
	F1 Score	96.5%	97.7%

Table S1. Pixel-wise model prediction and evaluation metrics for live and dead CHO cell images with 40× magnification. Confusion matrix entries are normalized relative to the number of cells in each class.

Through Table S1, we could easily distinguish live and dead cells with an error rate of less than 1%. The model achieved high precision, recall, and F1 scores for both live and dead cells, with all metrics exceeding 96%. These results highlight the model's reliability in accurately segmenting and classifying live and dead CHO cells from 40× magnification images, ensuring robust performance in challenging microscopy data.

Velocity Correlations and Displacement Moments of model

We validated the performance of our model by comparing the spatial and temporal velocity correlations, as well as the displacement moments, between the predicted results and the ground truth data. The results are shown in Figure S3; the spatial velocity correlation of our model's prediction exhibits strong agreement with the ground truth across spatial, with minor deviations observed at larger distance scales (Figure S3a), validating the model's ability to capture the spatial dependency of cell movements. The trends of temporal velocity correlations (Figure S3b) observed demonstrate high consistency between different temporal scales, capturing the temporal coherence of cell movements. For the scaling behaviors of displacement moments(Figure S3c), the predicted moments align well with the ground truth, capturing both linear and nonlinear scaling behaviors. These results validate the model's ability to predict cell movement dynamics with high accuracy.

Model Complexity Comparison

To assess the efficiency and performance of our model, we compared it against Unet and a transformer-based segmentation model (ViT-B-16). The results, summarized in Table S2, indicate that despite having 66.34 million trainable parameters, our model maintains a significantly lower computational burden than ViT-B-16 (88.57M) while outperforming U-Net (31.55M) in accuracy. Additionally, our model exhibits a 40.5% reduction in execution time compared to ViT-B-16, requiring 75.43 ms per training versus 135.60 ms for the transformer-based model.

Memory consumption is a critical factor in real-world deployment. While ViT-B-16 requires 12.05 GB of RAM, our model reduces this demand to 10.85 GB, improving efficiency without sacrificing performance. Compared to U-Net, which has the lowest memory footprint (3.59 GB) and execution time (30.36 ms), our model provides a balance between computational efficiency and accuracy.

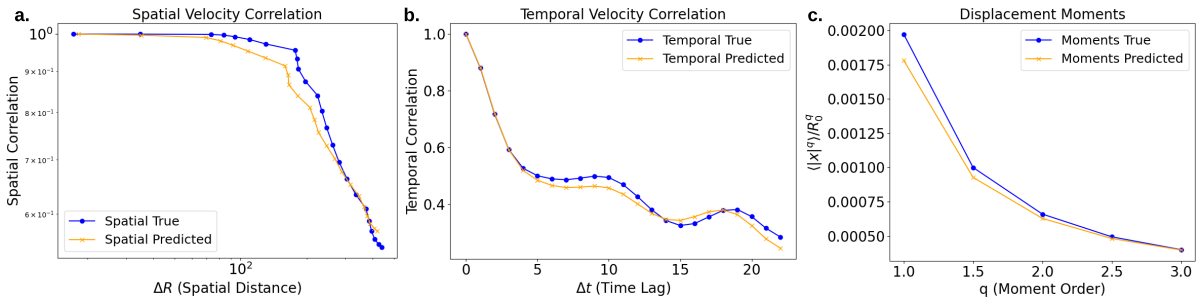

Figure S3. Validation of predicted spatial and temporal velocity correlations and displacement moments. a. Spatial velocity correlation as a function of spatial distance (ΔR) for both the predicted and ground truth data. **b.** Temporal velocity correlation as a function of temporal lag (Δt) comparing predicted and ground truth data. **c.** Displacement moments normalized by the maximum displacement (R_0) for moment orders $q = 1.0, 1.5, 2.0, 2.5, 3.0$.

	Unet	ViT-B-16	Our model
RAM usage (GB) ↓	3.59	12.05	**10.85**
Execution time (ms) ↓	30.36	135.60	**75.43**
Trainable parameters(M) ↓	31.55	88.57	**66.34**
Accuracy(%) ↑	84.2	84.7	**85.8**

Table S2. Comparison of segmentation models in terms of computational efficiency and accuracy. The table presents a comparative analysis of U-Net, ViT-B-16(Transformer), and our proposed model in terms of RAM usage, execution time, number of trainable parameters, and accuracy. All computational metrics are evaluated based on one batch of training.