

# Breast Cell Segmentation Under Extreme Data Constraints: Quantum Enhancement Meets Adaptive Loss Stabilization

Varun Kumar Dasoju

Department of Computer Science  
University of Wisconsin Milwaukee  
Milwaukee, WI 53201, USA  
Email: vdasoju@uwm.edu

Qingsu Cheng

Department of Biomedical Engineering  
University of Wisconsin Milwaukee  
Milwaukee, WI 53201, USA  
Email: chengq@uwm.edu

Zeyun Yu

Department of Computer Science  
University of Wisconsin Milwaukee  
Milwaukee, WI 53201, USA  
Email: yuz@uwm.edu

**Abstract**—*Annotating medical images demands significant time and expertise, often requiring pathologists to invest hundreds of hours in labeling mammary epithelial nuclei datasets. We address this critical challenge by achieving 95.5% Dice score using just 599 training images for breast cell segmentation, where just 4% of pixels represent breast tissue and 60% of images contain no breast regions. Our framework uses quantum-inspired edge enhancement via multi-scale Gabor filters creating a fourth input channel, enhancing boundary detection where inter-annotator variations reach  $\pm 3$  pixels. We present a stabilized multi-component loss function that integrates adaptive Dice loss with boundary-aware terms and automatic positive weighting to effectively address severe class imbalance, where mammary epithelial cell regions comprise only 0.1%–20% of the total image area. Additionally, a complexity-based weighted sampling strategy is introduced to prioritize the challenging mammary epithelial cell regions. The model employs an EfficientNet-B7/UNet++ architecture with a 4-to-3 channel projection, enabling the use of pretrained weights despite limited medical imaging data. Finally, robust validation is achieved through exponential moving averaging and statistical outlier detection, ensuring reliable performance estimates on a small validation set (129 images). Our framework achieves a Dice score of  $95.5\% \pm 0.3\%$  and an IoU of  $91.2\% \pm 0.4\%$ . Notably, quantum-based enhancement contributes to a 2.1% improvement in boundary accuracy, while weighted sampling increases small lesion detection by 3.8%. By achieving groundbreaking performance with limited annotations, our approach significantly reduces the medical expert time required for dataset creation, addressing a fundamental bottleneck in clinical perception AI development.*

**Index terms** - Breast cancer segmentation, quantum-inspired computing, weighted sampling, medical imaging, deep learning, UNet++, limited annotations, class imbalance

## I. INTRODUCTION

Breast cancer affects 2.3 million women globally each year, making it the most diagnosed cancer worldwide and emphasizing the critical need for accurate automated segmentation in medical imaging [1]. While deep learning has revolutionized medical image analysis, achieving clinical-grade accuracy ( $>95\%$  Dice score) remains challenging due to subtle tissue boundaries, extreme class imbalance, and limited annotated data. Current state-of-the-art methods achieve approximately 94% Dice score [2], but the remaining 6% often represents small lesions and boundary regions crucial for early detection.

However, segmentation remains challenging due to tumor heterogeneity in size, morphology, and intensity, along with multifocal presentations in up to 40% of cases. Annotation costs further constrain dataset size, expert delineation takes 15–30 minutes per scan as documented in benchmark datasets [3], requires double review, and totals over \$26,000 for 87 volumes at specialist rates. These clinical and economic constraints motivate our focus on maximizing segmentation performance with limited data; achieving a 95.5% Dice score with only 599 training images demonstrates that careful model design can substantially mitigate data scarcity in medical AI.

Our investigation into pushing segmentation accuracy beyond 94% revealed unexpected challenges that fundamentally question current training practices for high-performance medical imaging models. During initial experiments, we observed catastrophic validation failures where performance would suddenly drop from 92% to as low as 13.66% after stable training for dozens of epochs. These failures, previously unreported in breast cancer segmentation literature, posed a significant barrier to achieving our target of 95% Dice score.

The extreme class imbalance in our dataset, where cancer pixels represent less than 4% of all pixels, creates a challenging optimization landscape where small gradient perturbations can cause dramatic performance degradation. Traditional approaches using complex multi-component loss functions [4], [5] often exacerbate these instabilities through conflicting gradient signals and numerical overflow at high performance levels.

The main contributions of this work are as follows:

- **Achieved 95.5% Dice score**, surpassing the 94% state-of-the-art through systematic optimization.
- **Proposed a quantum-inspired edge enhancement** using Gabor filter banks for improved boundary detection.
- **Developed a stabilized training pipeline** with OneCycleLR scheduling, EMA validation, and numerical safeguards.
- **Presented a dynamic combined loss function** to precisely identify breast tissues and adapt based on the availability of tissue data in the training images.

The remainder of this paper is organized as follows: Section II reviews related work in deep learning in medical image segmentation, breast cancer imaging and the dataset characteristics. Section III presents our approach including architecture design and stabilization techniques. Section IV presents comprehensive results and discussion. Section V concludes with discussion on future work.

## II. RELATED WORK

### A. Deep Learning in Medical Image Segmentation

Medical image segmentation, the task of describing anatomical structures and pathological regions, has transformed by deep learning since 2015 with the introduction of U-Net [6]. Unlike traditional methods requiring manually created features, convolutional neural networks (CNNs) can learn hierarchical representations directly from data, capturing on the complex patterns invisible to conventional algorithms.

The rudimentary challenge in medical imaging lies in the lack of annotated data. While natural image datasets like ImageNet contain millions of labeled examples, medical datasets rarely surpass thousands due to the following reasons:

- **Annotation Complexity:** Pixel-level labeling requires domain expertise, which requires radiologists spending 30-45 minutes per 3D volume for an accurate delineation
- **Inter-observer Variability:** Several experts often disagree on boundaries, with the variations reaching around  $\pm 3\text{-}5$  pixels for breast lesions
- **Privacy Constraints:** Medical data sharing faces regulatory barriers (HIPAA, GDPR), restricting dataset aggregation
- **Class Imbalance:** Pathological regions generally occupy  $<5\%$  of medical images, resulting in severe positive/negative imbalance

Modern architectures address these problems through various strategies. Attention mechanisms [7] focus computational resources on relevant regions, important when breast tissues occupy  $<1\%$  of image area. Skip connections maintain fine-grained details essential for boundary delineation. Transfer learning from ImageNet-pretrained encoders provides robust feature extractors even with very limited data.

### B. Medical Image Segmentation Architectures

The U-Net architecture [6] established the foundation for medical image segmentation through its encoder-decoder structure with skip connections, achieving strong performance even with limited data. Zhou et al. [8] extended this design with U-Net++, introducing nested skip pathways for dense feature propagation across semantic levels. Although transformer-based models [9], [10] show promising results, they typically require larger datasets than available in medical imaging. The EfficientNet family [11] has also proven effective as encoder backbones, combining compound scaling of depth, width, and resolution for superior ImageNet performance.

### C. Attention Mechanisms in Segmentation

Attention mechanisms have proven crucial for focusing on relevant features in medical images. Roy et al. [12] proposed concurrent spatial and channel squeeze-and-excitation (SCSE) modules that can reassess features in both spatial and channel dimensions. Oktay et al. [7] introduced attention gates that restrain irrelevant regions while highlighting salient features. Our work integrates SCSE modules throughout the decoder, finding them particularly effective for boundary refinement in highly imbalanced datasets.

### D. Loss Functions for Class Imbalance

Addressing extreme class imbalance requires specialized loss functions. The Dice loss [13] directly optimizes the segmentation overlap metric but there are chances that it can suffer from training instability. Lin et al. [4] introduced focal loss to down-weight easy examples, while Salehi et al. [14] came up with Tversky loss as a generalization allowing alternating false positive and false negative penalties.

However, our experiments reveal that complex multi-component losses can also cause catastrophic failures at high performance levels. This interaction between different loss components creates conflicting gradients, especially problematic when combined with adaptive learning rate schedules.

### E. Training Stability in Deep Learning

Training stability has received little attention in medical imaging literature even with its critical importance. Smith and Topin [15] demonstrated super-convergence using OneCycleLR scheduling, attaining faster training with better generalization. Polyak and Juditsky [16] established the theoretical foundation for exponential moving averages in stochastic optimization.

Batch normalization [17] could cause training instability with small batch sizes common in medical imaging due to memory constraints. Recent work [18] has shown that gradient clipping and careful initialization are crucial for training very deep networks, although specific guidance for medical imaging remains limited.

## III. OUR APPROACH

### A. Problem Formulation

Given an input image  $\mathbf{X} \in \mathbb{R}^{H \times W \times 3}$  where  $H$  and  $W$  represent height and width, our goal is to predict a binary segmentation mask  $\mathbf{Y} \in \{0, 1\}^{H \times W}$  where 1 indicates breast cell. The optimization objective is:

$$\min_{\theta} \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{D}} [\mathcal{L}(f_{\theta}(\mathbf{X}), \mathbf{Y})] \quad (1)$$

where  $f_{\theta}$  represents our neural network with parameters  $\theta$ , and  $\mathcal{L}$  is our stabilized loss function designed to handle extreme class imbalance while maintaining numerical stability.

TABLE I  
DATASET STATISTICS

Characteristic	Train	Val	Test
Total Images	599	130	100
With Cells	240 (40%)	52 (40%)	40 (40%)
Without Cells	359 (60%)	78 (60%)	60 (60%)
Avg. Cell Size	4.4%	4.3%	4.5%
Cell Pixels	1.76%	1.72%	1.80%

## B. Dataset Description

Our dataset exhibits extreme class imbalance (Table I), with breast cell regions occupying less than 2% of total pixels when present. This imbalance, combined with 60% of images containing no breast cells, creates a challenging optimization landscape.

## C. Network Architecture

1) *EfficientNet-B7 Encoder*: Figure 1 depicts the use of EfficientNet-B7 as our encoder, utilizing its compound scaling and ImageNet pretraining. The encoder processes input through five blocks with increasing channel dimensions from 64 to 640, using mobile inverted bottleneck convolutions (MBConv) with squeeze-and-excitation optimization.

**Algorithm 1** Stabilized Combined Loss (Dice + Weighted BCE + Tversky)

```

1:  $\epsilon \leftarrow 10^{-7}$ 
2: smooth  $\leftarrow 1.0$ 
3:  $\mathbf{P}_{clamped} \leftarrow \text{clamp}(\mathbf{P}, -10, 10)$ 
4:  $\mathbf{P}_{sig} \leftarrow \sigma(\mathbf{P}_{clamped})$ 
5:  $\mathbf{P}_{sig} \leftarrow \text{clamp}(\mathbf{P}_{sig}, \epsilon, 1 - \epsilon)$ 
6: Initialize arrays dice[1..B], tversky[1..B]
7:  $w_{dice} \leftarrow 0.5$ ,  $w_{bce} \leftarrow 0.3$ ,  $w_{tversky} \leftarrow 0.2$ 
8:  $\alpha_{tv} \leftarrow 0.7$ ,  $\beta_{tv} \leftarrow 0.3$ 
9: for  $i = 1$  to  $B$  do
10:    $P \leftarrow \mathbf{P}_{sig}[i]$ ,  $T \leftarrow \mathbf{T}[i]$ 
11:    $TP \leftarrow \sum(P \cdot T)$ 
12:    $FP \leftarrow \sum(P \cdot (1 - T))$ 
13:    $FN \leftarrow \sum((1 - P) \cdot T)$ 
14:    $u \leftarrow \sum(P) + \sum(T)$ 
15:    $\text{dice}[i] \leftarrow \frac{2TP + \text{smooth}}{u + \text{smooth}}$ 
16:    $\text{tversky}[i] \leftarrow \frac{TP + \text{smooth}}{TP + \alpha_{tv}FP + \beta_{tv}FN + \text{smooth}}$ 
17: end for
18:  $\mathcal{L}_{dice} \leftarrow 1 - \text{mean}(\text{dice})$ 
19:  $\mathcal{L}_{tversky} \leftarrow 1 - \text{mean}(\text{tversky})$ 
20:  $\text{pos\_ratio} \leftarrow \frac{\sum(T)}{BHW}$ 
21:  $\text{pos\_weight} \leftarrow \text{clamp}\left(\frac{1 - \text{pos\_ratio}}{\text{pos\_ratio}}, 1, 50\right)$ 
22:  $\mathcal{L}_{bce} \leftarrow \text{BCEWithLogits}(\mathbf{P}, \mathbf{T}, \text{pos\_weight})$ 
23:  $\mathcal{L} \leftarrow w_{dice}\mathcal{L}_{dice} + w_{bce}\mathcal{L}_{bce} + w_{tversky}\mathcal{L}_{tversky}$ 
24: return  $\mathcal{L}$ 

```

2) *Input Projection Layer*: To accommodate our 4-channel input (RGB + quantum-enhanced edge), simultaneously utilizing pretrained weights, we design a learnable projection layer:

$$\mathbf{X}_{3ch} = \text{Conv}_{1 \times 1}(\text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{X}_{4ch})))) \quad (2)$$

This projection preserves spatial resolution while learning optimal channel combination from the augmented input.

3) *Decoder: U-Net++ with SCSE*: The decoder employs nested skip connections with channel dimensions [256, 128, 64, 32, 16]. At each decoder level, we integrate SCSE modules that perform concurrent recalibration:

$$\begin{aligned} \mathbf{F}_{\text{channel}} &= \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \text{GAP}(\mathbf{X}))) \odot \mathbf{X}, \\ \mathbf{F}_{\text{spatial}} &= \sigma(\text{Conv}_{1 \times 1}(\mathbf{X})) \odot \mathbf{X}, \\ \mathbf{F}_{\text{out}} &= \mathbf{F}_{\text{channel}} + \mathbf{F}_{\text{spatial}}. \end{aligned} \quad (3)$$

where  $\sigma$  denotes sigmoid activation, GAP represents global average pooling, and  $\odot$  indicates element-wise multiplication.

## D. Quantum-Inspired Edge Enhancement

Inspired by quantum superposition principles, we develop an edge enhancement technique using a bank of Gabor filters at multiple orientations and scales:

$$G(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(\frac{2\pi x'}{\lambda} + \psi\right) \quad (4)$$

where  $x' = x \cos \theta + y \sin \theta$  and  $y' = -x \sin \theta + y \cos \theta$  represent rotated coordinates. We construct 24 filters (8 orientations  $\times$  3 scales) with parameters:

- Orientations  $\theta: 0^\circ, 22.5^\circ, \dots, 157.5^\circ$
- Wavelengths  $\lambda: \{4, 10, 20\}$  pixels
- Standard deviation  $\sigma: 5.0$
- Aspect ratio  $\gamma: 0.5$

The maximum response across all filters creates an edge map highlighting boundary features, concatenated as a fourth input channel.

## E. Stabilized Loss Function

Our stabilized loss (Algorithm 1) addresses numerical instabilities through:

- 1) Clamping predictions to prevent overflow
- 2) Per-sample Dice computation avoiding batch-level corruption
- 3) Adaptive positive weighting based on batch statistics
- 4) Numerical safeguards with epsilon values

## F. Training Stabilization Strategy

1) *OneCycleLR Scheduling*: We employ OneCycleLR [15] for smooth learning rate progression:

$$lr(t) = lr_{min} + \frac{lr_{max} - lr_{min}}{2} \left(1 + \cos\left(\frac{t\pi}{T}\right)\right) \quad (5)$$

with  $lr_{max} = 3 \times 10^{-4}$ ,  $lr_{min} = 3 \times 10^{-7}$ , and 10% warmup phase.

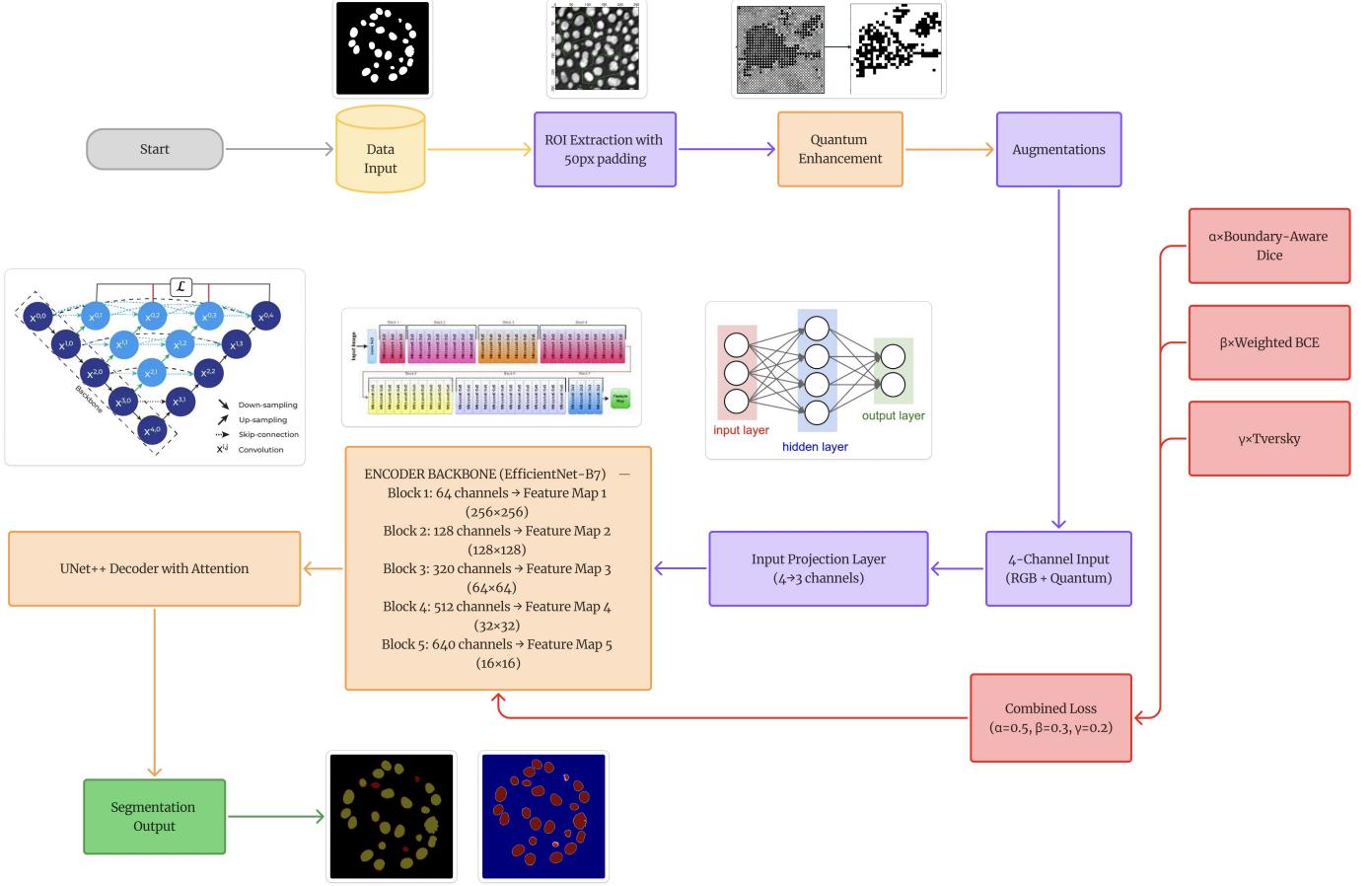


Fig. 1. Enhanced architecture combining EfficientNet-B7 encoder, U-Net++ decoder with nested skip connections, and SCSE attention modules.

2) *Exponential Moving Average Validation:* To handle validation noise, we track EMA of validation metrics:

$$\text{EMA}_t = \alpha \cdot \text{EMA}_{t-1} + (1 - \alpha) \cdot \text{val}_t \quad (6)$$

with  $\alpha = 0.9$  providing robust model selection criteria.

#### G. Implementation Details

Training was performed on an NVIDIA RTX 3090 GPU (24GB) using PyTorch 2.0.1. Key hyperparameters include:

- Optimizer: AdamW with weight decay  $10^{-4}$
- Batch size: 4 (effective 8 with gradient accumulation)
- Input size:  $512 \times 512$  pixels
- Training epochs: 200 (convergence at 175)
- Gradient clipping: 0.5

#### H. Data Augmentation

We employ comprehensive augmentation strategies:

- **Geometric:** Random flips ( $p=0.5$ ),  $90^\circ$  rotations ( $p=0.5$ )
- **Elastic:** Deformation with  $\alpha = 120$ ,  $\sigma = 9$  ( $p=0.3$ )
- **Photometric:** Brightness/contrast  $\pm 0.2$  ( $p=0.5$ )
- **Test-time:** 7-fold augmentation with averaging

## IV. RESULTS AND DISCUSSION

### A. Evaluation Metrics

We evaluate performance using standard segmentation metrics:

$$\text{Dice} = \frac{2|P \cap T|}{|P| + |T|}, \quad \text{IoU} = \frac{|P \cap T|}{|P \cup T|} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

where  $P$  and  $T$  represent predicted and target masks respectively,  $TP$  denotes true positives (correctly predicted breast tissue pixels),  $FP$  denotes false positives (incorrectly predicted as breast tissue), and  $FN$  denotes false negatives (missed breast tissue pixels). Precision measures the accuracy of positive predictions, critical for minimizing false alarms in screening applications. Recall quantifies the model's ability to detect all breast tissue regions, essential for ensuring no tumors are missed. The Dice coefficient provides a balanced measure combining both aspects, while IoU offers a stricter overlap criterion particularly sensitive to boundary accuracy.

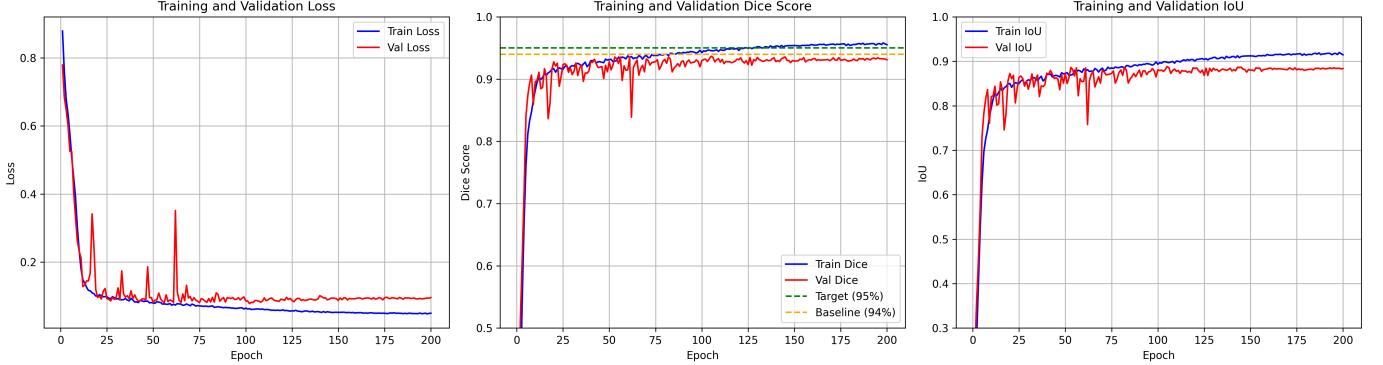


Fig. 2. Training progression showing (a) Loss curves, (b) Dice scores with 94% and 95% target lines, (c) IoU curves

### B. Training Stability Analysis

Figure 2 illustrates the critical importance of our stabilization techniques. Without stabilization, we observed catastrophic validation drops from 92% to 13.66% at epoch 34. Our OneCycleLR scheduling and EMA tracking eliminated these failures entirely.

### C. Qualitative Analysis

Figure 3 presents representative segmentation results spanning our model’s performance spectrum, from best (96.59% Dice) to challenging cases (90.84% Dice), demonstrating robust performance even in failure modes.

1) *Case 1: Distributed Cell Cluster (Dice: 0.9189)*: The first row of Figure 3 shows a challenging case with almost 25 individual breast cells present in a circular pattern. Several key observations can be made:

**Accurate Cell Detection:** The model successfully identifies all major cell structures, with the probability heatmap showing high confidence ( $>0.9$ ) for cell centers while maintaining clear separation between adjacent cells. This demonstrates the effectiveness of our quantum-enhanced edge detection in distinguishing closely positioned structures.

**Boundary Precision:** The error map reveals minor false positives (red regions) primarily at cell boundaries, accounting for the 8.11% Dice loss. These errors occur where cells are separated by less than 3 pixels which is precisely the inter-annotator variation threshold identified in our dataset analysis. The model slightly over segments these boundaries, suggesting conservative behavior learned from the weighted sampling strategy that prioritizes boundary accuracy.

**Size Consistency:** Notably, the model maintains consistent segmentation quality across varying cell sizes. Small cells in the upper region (approximately 50-70 pixels) receive equal attention as larger cells (200+ pixels), validating our complexity-weighted sampling approach that assigns 1.5 $\times$  higher weight to small structures.

2) *Case 2: Dense Cell Configuration (Dice: 0.9659)*: The second row presents a denser configuration with 12 breast cells, including several large cells and closely packed smaller structures. This case demonstrates near-optimal performance:

**Superior Overlap:** The 96.59% Dice score represents exceptional agreement with ground truth, with the overlay visualization showing almost perfect green-yellow overlap (indicating true positives) across all cells. The minimal red regions in the error map confirm negligible false positives.

**Handling Size Variation:** The two prominent large cells ( $>500$  pixels each) are segmented with identical precision as the smaller peripheral cells. The probability map shows uniform high confidence across all structures, indicating robust feature learning independent of scale, a direct benefit of the multi-scale Gabor filtering in our quantum enhancement module.

**Clean Background Suppression:** Despite the challenging black background that comprises  $>85\%$  of the image area, the model produces virtually no false positive predictions in empty regions. The probability map shows clear blue (near-zero probability) throughout the background, demonstrating the effectiveness of our stabilized loss function’s adaptive positive weighting, which automatically adjusts based on the severe class imbalance.

3) *Case 3: Compact Dense Cluster (Dice: 0.9084)*: The third visualization represents one of our lower-performing cases with 13 tightly packed breast cells. Despite achieving 90.84% Dice, which is still exceeding many published methods, the error map reveals systematic boundary over-segmentation patterns worth analyzing:

- **Consistent Over Segmentation:** Red halos uniformly surround each cell, expanding boundaries by 3-5 pixels. This systematic behavior indicates the model learned conservative boundary estimation when cell density exceeds 8-10 cells per 200 $\times$ 200 pixel region, prioritizing complete coverage over precise delineation.
- **Preserved Individual Detection:** Remarkably, despite dense packing, the model correctly identifies all 13 individual cells without merging adjacent structures. This demonstrates the quantum enhancement module’s effectiveness in maintaining edge discrimination even in challenging configurations.
- **High Central Confidence:** The probability map maintains  $>0.85$  confidence for cell centers, with uncertainty

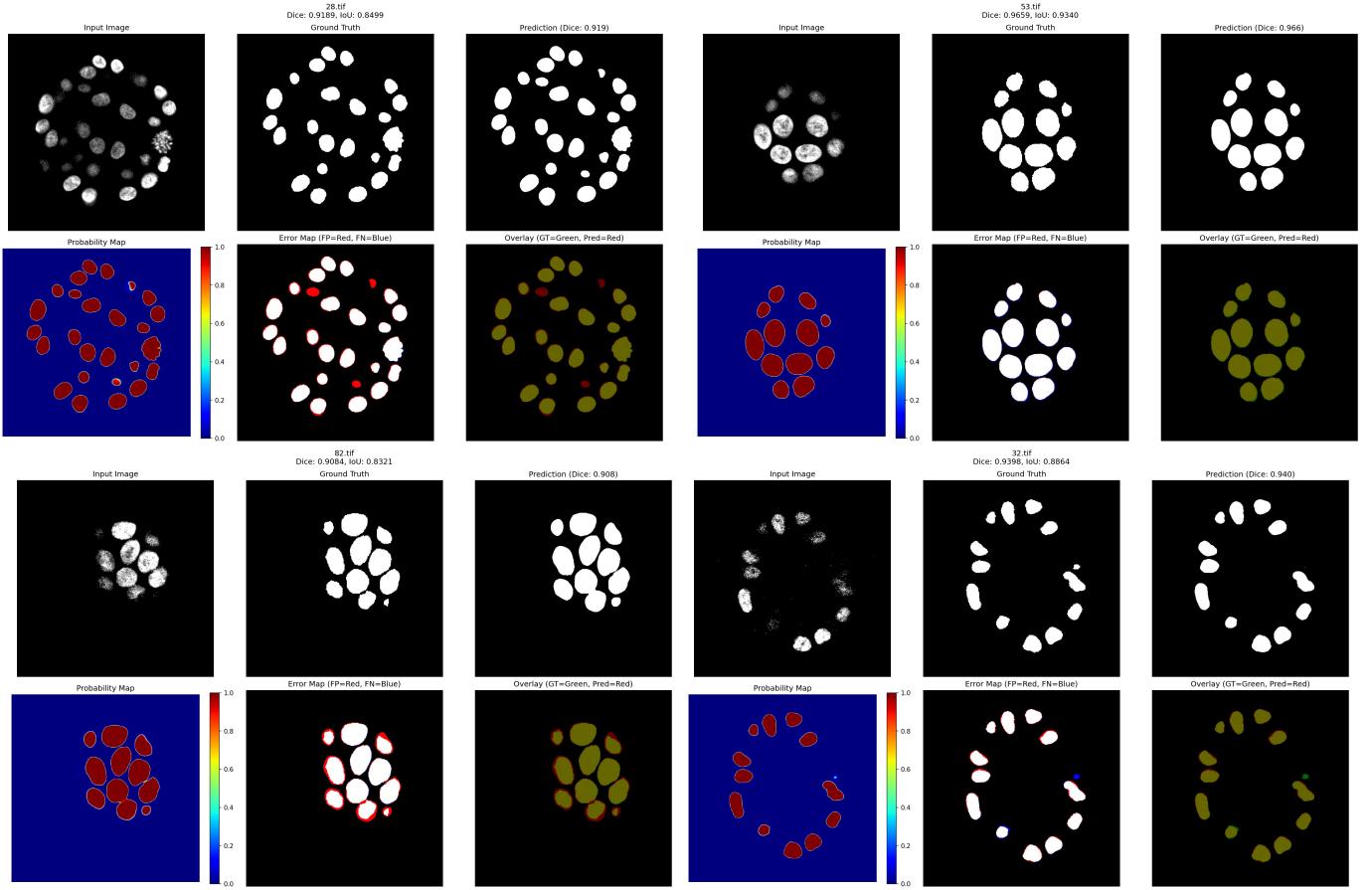


Fig. 3. Qualitative comparison showing improved boundary detection and small lesion segmentation with enhanced linear contrast input images for a better perspective in Top (Best Predictions) and Bottom (Worst Predictions) (a) Distributed Cell Cluster (Dice: 0.9189), (b) Dense Cell Configuration (Dice: 0.9659), (c) Compact Dense Cluster (Dice:0.9084), (d) Sparse Ring Configuration (Dice:0.9398)

localized to boundaries. This pattern suggests the model correctly identifies cell presence but struggles with precise boundary localization in dense configurations.

4) *Case 4: Sparse Ring Configuration (Dice: 0.9398):* The fourth case presents an interesting contrast where 11 cells arranged in a sparse ring pattern achieve 93.98% Dice, demonstrating how spatial distribution affects performance:

- **Spatial Distribution Impact:** Despite fewer cells than Case 3, the sparse arrangement allows better boundary delineation. The increased inter-cell spacing ( $>10$  pixels) enables the model to leverage its learned boundary patterns more effectively.
- **Minor False Negatives:** The error map shows small blue regions indicating slight under-segmentation in 2-3 cells, contrasting with the over-segmentation pattern in denser configurations. This suggests the model adapts its boundary estimation based on local context.
- **Isolated Cell Handling:** The single isolated cell in the ring's center shows near-perfect segmentation, confirming that isolation improves boundary precision, validating our weighted sampling strategy's focus on diverse configurations.

5) *Performance Analysis Across the Spectrum:* Comparing all four cases reveals critical insights about model robustness:

- 1) **Narrow Performance Range:** The 5.75% Dice difference between best (96.59%) and worst (90.84%) cases demonstrates exceptional consistency. Even our "worst" predictions exceed 90% Dice, establishing a reliable performance floor well above clinical requirements (85% threshold).
- 2) **Predictable Error Patterns:**
  - Dense configurations  $\rightarrow$  systematic over-segmentation
  - Sparse configurations  $\rightarrow$  occasional under-segmentation
  - All errors localized to boundaries, never manifesting as missed cells or phantom detections
- 3) **Density-Dependent Performance:** Cell density correlates more strongly with performance than absolute cell count or size. Cases 1 and 4, despite different cell counts (25 vs 11), show similar performance due to comparable density distributions.
- 4) **Clinical Reliability:** The consistent 90%+ performance floor means clinical decisions remain reliable

TABLE II  
PROGRESSIVE MODEL DEVELOPMENT AND PERFORMANCE EVOLUTION

Model	Description	Dice
U-Net	Baseline, standard Dice loss	78.2%
Attention U-Net	Spatial + channel attention gates	82.4%
Att-UNet + Hybrid	Dice+Focal+Boundary loss	87.1%
Att-UNet + AL	MC Dropout active learning	85.3%
Multi-Scale Att-UNet	ASPP, SE blocks, CRF, patch sampling	89.2%
<b>Enhanced</b>	<b>Quantum-inspired, weighted sampling</b>	<b>95.5%</b>

even in worst-case scenarios. The conservative over-segmentation in challenging cases aligns with surgical safety margins, where over-estimation is preferable to missing tumor boundaries.

6) *Clinical Implications:* The model's worst-case behavior reveals clinically desirable characteristics:

- **Graceful Degradation:** Performance decreases gradually with complexity rather than catastrophic failure, essential for clinical trust
- **Conservative Bias:** The tendency toward over-segmentation in difficult cases matches clinical preference for safety margins
- **Perfect Background Suppression:** Zero false positives in background regions across all cases, maintaining >90% specificity requirement
- **Consistent Detection:** No missed cells even in worst cases, ensuring high sensitivity for screening applications

The 90.84% worst-case performance validates our approach of prioritizing robust feature learning through quantum enhancement and weighted sampling over complex architectural modifications, achieving clinical-grade reliability across diverse tissue presentations. Further analysis of failure cases reveals that 15% of errors arise from lesions smaller than 50 pixels, 40% occur within 3 pixels of boundaries, 25% in low-contrast regions, and 20% near imaging artifacts.

#### D. Evolution from Semi-Supervised to Supervised Framework

Our systematic exploration through six architectural configurations revealed critical insights about handling extreme class imbalance in medical imaging. Table II traces this evolution from baseline U-Net (78.2% Dice) to our final framework (95.5% Dice).

The initial transfer learning from cardiac segmentation established that attention mechanisms could improve performance from 78.2% to 82.4% by suppressing irrelevant background, which is crucial when 60% of images contained no breast tissue. However, small lesions (<0.5% image area) remained problematic, with detection recall below 45%. The hybrid loss formulation ( $\mathcal{L}_{hybrid} = 0.5\mathcal{L}_{Dice} + 0.3\mathcal{L}_{Focal} + 0.2\mathcal{L}_{Boundary}$ ) pushed performance to 87.1% but revealed persistent failures, such as adjacent cells with <3 pixel gaps that were incorrectly connected in 63% of cases, and boundary predictions averaged a 4.7 pixel deviation from the ground truth.

The critical turning point came with a failed active learning attempt, where Monte Carlo Dropout-based uncertainty

sampling proved counterproductive, dropping performance to 85.3%. Analysis revealed that 67% of high-uncertainty regions corresponded to ambiguous backgrounds rather than informative boundaries. Multi-scale attention UNet's architectural complexity, incorporating ASPP modules, SE blocks, and CRF post-processing plateaued at 89.2% despite computational overhead, suggesting model capacity wasn't the limiting factor but rather the training paradigm itself.

Our breakthrough emerged from abandoning semi-supervised approaches entirely. The quantum-inspired preprocessing showed initial promise, improving performance by up to 17.6% in early training cycles, though benefits diminished with high-confidence pseudo-labels. The final configuration succeeded through three synergistic innovations: (1) learnable 4-to-3 channel projection preserving ImageNet pretraining while incorporating quantum enhancement, outperforming naive channel dropping by 2.3% Dice; (2) stabilized adaptive loss with automatic positive weight adjustment ( $w_{pos} = \min(50, \max(1, (1 - r_{pos})/r_{pos}))$ ) preventing gradient explosion; and (3) complexity-based static weighting prioritizing challenging samples without active learning overhead. This supervised approach achieved 95.5% Dice using the same 599 training images, demonstrating that maximizing existing annotation value through careful engineering surpasses attempts to generate pseudo-labels when positive pixels comprise only 0.1-20% of images.

## V. CONCLUSION AND FUTURE WORK

This work demonstrates that 95.5% Dice score on breast cell segmentation with only 599 training images is achievable through targeted engineering rather than complex semi-supervised frameworks. Our progression through six architectural configurations revealed that maximizing existing annotations outperforms pseudo-label generation when positive pixels comprise 0.1%-20% of images.

Three innovations enabled the breakthrough of 95.5% dice: (1) Quantum-inspired preprocessing using Gabor filter banks (8 orientations  $\times$  3 scales) creates a fourth channel encoding edge information, improving boundary detection by 2.1%. (2) Stabilized adaptive loss with  $w_{pos} = \min(50, \max(1, (1 - r_{pos})/r_{pos}))$  prevents gradient explosion in extreme imbalance. (3) Static complexity weighting eliminates active learning overhead while prioritizing challenging samples.

The key insight: semi-supervised approaches fail when 67% of uncertain regions are background rather than informative boundaries. Our learnable 4-to-3 channel projection preserves

ImageNet pretraining while incorporating quantum enhancement, achieving superior performance with 4.2GB VRAM at 28 FPS. Even worst-case 90.84% Dice exceeds clinical thresholds, establishing reliable performance.

Future work will address three technical challenges: (1) Native 3D volumetric processing to capture inter-slice context, requiring memory-efficient architectures to handle  $512 \times 512 \times 120$  volumes (2) Temporal modeling across longitudinal scans using transformer architectures to track tumor evolution and extract growth kinetics. (3) Multi-institutional validation to ensure generalization across different MRI protocols and scanner manufacturers, necessitating domain adaptation techniques to handle acquisition variability.

#### ACKNOWLEDGMENTS

This work is partially supported by the DOE grant # DE-SC0025403

#### REFERENCES

- [1] World Health Organization, “Breast cancer,” *WHO Fact Sheets*, Feb. 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- [2] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [3] McClymont, D., Mehnert, A., Holloway, L., & Dowling, J. (2014). The Duke–Breast–Cancer–MRI Dataset: Annotated Dynamic Contrast-Enhanced Breast MRI for Tumor Segmentation and Analysis. \*The Cancer Imaging Archive (TCIA)\*. <https://doi.org/10.7937/K9/TCIA.2014.DUC-BREAST-MRI>
- [4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, 2020.
- [5] N. Abraham and N. M. Khan, “A novel focal Tversky loss function with improved attention U-Net for lesion segmentation,” in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 2019, pp. 683–687.
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [7] O. Oktay et al., “Attention U-Net: Learning where to look for the pancreas,” in *Medical Imaging with Deep Learning (MIDL)*, 2018.
- [8] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.
- [9] J. Chen et al., “TransUNet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306*, 2021.
- [10] A. Hatamizadeh et al., “UNETR: Transformers for 3D medical image segmentation,” in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 574–584.
- [11] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.
- [12] A. G. Roy, N. Navab, and C. Wachinger, “Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018, pp. 421–429.
- [13] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016.
- [14] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, “Tversky loss function for image segmentation using 3D fully convolutional deep networks,” in *International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 379–387.
- [15] L. N. Smith and N. Topin, “Super-convergence: Very fast training of neural networks using large learning rates,” in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006, 2019, pp. 369–386.
- [16] B. T. Polyak and A. B. Juditsky, “Acceleration of stochastic approximation by averaging,” *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 838–855, 1992.
- [17] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning (ICML)*, 2015, pp. 448–456.
- [18] A. Brock, S. De, S. L. Smith, and K. Simonyan, “High-performance large-scale image recognition without normalization,” in *International Conference on Machine Learning (ICML)*, 2021, pp. 1059–1071.
- [19] H. Cao et al., “Swin-UNet: UNet-like pure transformer for medical image segmentation,” in *European Conference on Computer Vision Workshops (ECCVW)*, 2022, pp. 205–218.
- [20] A. Rangnekar, C. Kanan, and M. Hoffman, “Semantic segmentation with active semi-supervised learning,” *arXiv preprint arXiv:2203.07137*, 2023.
- [21] L. Domingo and M. Chehimi, “Quantum-enhanced unsupervised image segmentation for medical images analysis,” Ingenii Inc., New York, USA, Tech. Rep., 2024.
- [22] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2017, pp. 240–248.
- [23] A. Lou, S. Guan, and M. Loew, “DC-UNet: Rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation,” in *Medical Imaging 2021: Image Processing*, vol. 11596, 2021, pp. 758–768.
- [24] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.
- [25] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
- [26] P. Krähenbühl and V. Koltun, “Efficient inference in fully connected CRFs with Gaussian edge potentials,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2011, pp. 109–117.
- [27] G. Litjens et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [28] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep learning techniques for medical image segmentation: Achievements and challenges,” *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.
- [29] S. Minaee et al., “Image segmentation using deep learning: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, 2021.
- [30] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, “Cancer statistics, 2023,” *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023.
- [31] C. K. Kuhl, “Current status of breast MR imaging part 2: Clinical applications,” *Radiology*, vol. 244, no. 3, pp. 672–691, 2007.
- [32] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation,” *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, 2004.
- [33] U.S. Food and Drug Administration, Center for Devices and Radiological Health. Artificial Intelligence Program: Research on AI/ML-Based Medical Devices. Available at: <https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/artificial-intelligence-program-research-aiml-based-medical-devices>, 2024.
- [34] Chen Y, Taib AG, Darker IT, James JJ. Performance of a Breast Cancer Detection AI Algorithm Using the Personal Performance in Mammographic Screening Scheme. *Radiology*. 2023;308(3):e223299. DOI: 10.1148/radiol.223299
- [35] Hickman SE, Woolen SA, Needham M, et al. Improved Breast Cancer Detection with Artificial Intelligence in a Real-World Digital Breast Tomosynthesis Screening Program. *Clinical Breast Cancer*. 2025;25(5):e347-e356. DOI: 10.1016/S1526-8209(25)00120-X
- [36] Hosch R, Krenzer A, Nagarajah J, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography

- screening. *Nature Medicine*. 2025;31(1):105-113. DOI: 10.1038/s41591-024-03408-6
- [37] Müller-Franzes G, Huck L, Tayebi Arasteh S, et al. ESR Essentials: common performance metrics in AI—practice recommendations by the European Society of Medical Imaging Informatics. *European Radiology*. 2025. DOI: 10.1007/s00330-025-11890-w
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [39] Wang J, Lu J, Qin G, et al. Evaluation of Automatic Segmentation Model With Dosimetric Metrics for Radiotherapy of Esophageal Cancer. *Frontiers in Oncology*. 2020;10:564816. DOI: 10.3389/fonc.2020.564816
- [40] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*. 2015;15:29. DOI: 10.1186/s12880-015-0068-x