DELTASEGMENT: MULTIFIDELITY MEDICAL IMAGE SEGMENTATION VIA CHANGE DETECTION

Yushuo Niu[*], *Hailey Reed*[†], *Qian Yang*[‡]

School of Computing, University of Connecticut, USA

ABSTRACT

We propose **DeltaSegment**, a lightweight, data-efficient framework for medical image segmentation that leverages multi-fidelity modeling and change detection. Differences between low-fidelity and high-fidelity images enable learning from fewer labeled samples. For each original, high-fidelity input image, we generate a coarse-grained, color-quantized version as a low-fidelity input. These paired inputs are then processed by a Semi-Siamese neural network trained end-to-end to predict segmentation maps, encouraging focus on salient morphological differences and boundary details critical for accurate segmentation. DeltaSegment achieves relative improvements of 1.3% in Dice and 2.3% in IoU on GlaS, 1.5% in Dice and 2.4% in IoU on MoNuSeg, and 1.3% in Dice and 1.7% in IoU on TNBC, demonstrating a simple yet effective approach that harnesses multi-fidelity change detection as a paradigm for small-data medical segmentation tasks.

***Index Terms*—** image segmentation, change detection, small data, multifidelity, Siamese neural network

1. INTRODUCTION

Medical image segmentation is crucial for clinical applications, enabling precise localization of anatomical and pathological structures. Deep learning approaches such as U-Net [1] and transformer-based models like ViT [2] achieve state-of-the-art results but are constrained by scarce, highly specialized datasets that require costly expert annotation. Thus, there is a need for segmentation methods that perform well with limited labeled data in the medical domain.

Existing solutions to data scarcity include few-shot and semi-supervised learning [3, 4], data augmentation and synthetic generation [5, 6], and transfer learning from pretrained models. Foundation models such as SAM [7] and MedSAM [8] mitigate the data scarcity issue through extensive pre-training but remain limited by domain variability, small fine-tuning sets, and high computational cost [9]. These challenges motivate lightweight, data-efficient alternatives.

We propose a multi-fidelity segmentation framework built on a Semi-Siamese neural network to address both data scarcity and the multi-scale feature representation. Siamese networks, originally developed for few-shot learning in small data contexts [3], change detection tasks using paired images from different time points or imaging modalities [10, 11, 12, 13, 14], and small-data medical imaging tasks [15, 16, 17, 18], are well suited for small-data domains. Traditional CNNs struggle with global context, while transformers face difficulties with fine-grained localization. Recent work fuses multi-scale representations [19] or combine CNN and transformer architectures [20, 21]. In contrast, our method leverages multi-fidelity *inputs*: each high-fidelity image is paired with a low-fidelity version generated via color quantization, enabling more precise segmentation by highlighting salient differences revealed through multi-fidelity comparisons. This approach offers a practical, data-efficient strategy for medical image segmentation without additional data collection.

2. METHODS

2.1. Multifidelity Modeling via Change Detection

Multi-fidelity modeling leverages relationships between different fidelity levels of data to improve learning efficiency. Traditional multi-fidelity approaches combine many low-fidelity samples with a smaller number of high-fidelity ones. In contrast, we auto-generate low-fidelity inputs from high-fidelity images via color quantization. The key idea is that differences between low- and high-fidelity representations better highlight structural information, such as boundaries, than high-fidelity representations alone. The model takes each high–low-fidelity image pair as input, learns feature representations for both in the encoder-decoder branches, computes their difference, and uses this difference to predict the corresponding segmentation mask, allowing the model to exploit multi-fidelity cues without any additional data collection (Fig. 2).

Fig. 1 shows the difference between high- and low-fidelity MoNuSeg [22] inputs after computing the color difference and applying min-max normalization. The process enhances

[*]ORCID: 0000-0001-8077-1005; [†]ORCID: 0009-0002-1506-4739; [‡]ORCID: 0000-0001-5519-1092, corresponding author: qyang@uconn.edu

This research was supported by funding from the National Science Foundation under Grant No. DMR-2102406.

visual features and structural boundaries, enabling DeltaSegment's precise mask predictions.

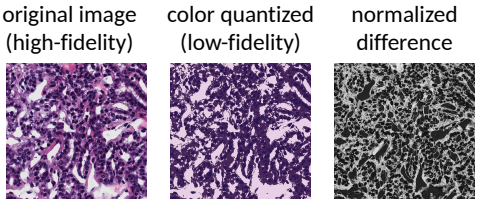

Fig. 1. Difference between high- and low-fidelity MoNuSeg inputs after color quantization and normalization, revealing enhanced structural boundaries for segmentation.

2.2. Low Fidelity Images via Color Quantization

To construct low-fidelity inputs, we apply color quantization to reduce each image to k colors while retaining the overall structural information. The value of k is selected using the elbow method on the k-means distortion curve. We evaluate four representative approaches: Centroid (mean), Mode (most frequent), Medoid (nearest), and Random. The Random method samples k pixel values at random to form a representative color palette, providing a simple non-clustering baseline. For the other three approaches, we first cluster the pixel colors in the original high-fidelity image using k-means clustering. The Centroid approach then replaces each pixel by its corresponding cluster mean, which is the arithmetic average of the pixel values within that cluster. This tends to produce smooth but sometimes over-blurred color transitions. The Medoid approach instead assigns each pixel to the original pixel color closest to the corresponding cluster mean, preserving realistic tones and mitigating extreme color outliers. Finally, the Mode approach identifies, for each cluster, the most frequently occurring original pixel value and assigns it as the representative color. This method preserves one of the exact color values that actually exists in the input image, ensuring high color fidelity such that the resulting low-fidelity image maintains the same visual palette as the source. Since it selects a single discrete color for each cluster, the output may exhibit abrupt color transitions or block-like artifacts in regions where colors change gradually. These artifacts, while less visually smooth, are highly correlated with structural boundaries and therefore preserve the high- frequency details and contrast necessary for accurate change-based segmentation. As a result, Mode quantization produces faithful and discriminative low-fidelity representations and is adopted as the default setting in this study, with all four methods compared in the ablation study.

2.3. Model Architecture

DeltaSegment adopts a Semi-Siamese neural network built upon a U-Net backbone [14] for multi-fidelity segmentation. As illustrated in Fig. 2, the network takes as input a high–low-fidelity image pair generated via color quantization. Each input passes through an independent encoder to extract domain-specific features, with symmetric skip connections to the shared decoder to preserve spatial alignment and enable multi-scale feature fusion. The Euclidean distance between the feature maps produced by both encoder branches highlights structural and textural discrepancies that are particularly informative for boundary delineation and region separation (Fig. 1). This difference map is then passed through a series of convolutional layers to generate the final segmentation output. In our ablation, this Euclidean distance-based operation outperforms a concatenation-based fusion approach. The network is trained end-to-end using focal loss, which downweights well-classified pixels and focuses learning on difficult regions such as blurred boundaries or overlapping nuclei.

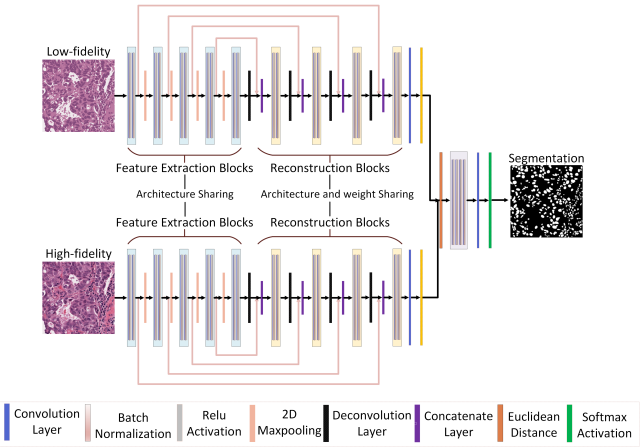

Fig. 2. Architecture of our multifidelity DeltaSegment model.

3. EXPERIMENTS AND RESULTS

3.1. Implementation details

Datasets. To evaluate the performance of our method, we conducted experiments on three publicly available and widely used biomedical image segmentation datasets. These benchmarks are well-known for their small dataset sizes, representative of the challenges posed by limited labeled data in medical imaging: MoNuSeg (tissue nuclei segmentation) [22], GlaS (gland segmentation) [23, 24], and TNBC (breast cancer nuclei segmentation) [25]. Each dataset contains high-resolution microscopy images with expert-annotated masks, offering diverse cross-organ and cross-scanner variations for small-data evaluation.

Training details. Experiments are conducted on an NVIDIA GeForce A5000 GPU (24 GB VRAM) using Python 3.12, PyTorch 2.0, and CUDA 12.4. All models (U-Net, UCTransNet, TSCA-Net and ours) share identical training settings. Following transformer-based baselines [26, 21], all images are resized to 224×224 for a fair comparison

across all models, particularly those based on Vision Transformers. To ensure statistically reliable evaluation on small datasets, we perform three rounds of 5-fold cross-validation and report the mean and standard deviation of the results. We evaluate segmentation quality using standard metrics, Dice and Intersection over Union (IoU) scores, which quantify the similarity between predicted segmentations and ground-truth masks [27]. Early stopping is applied to prevent overfitting.

Comparison with state-of-the-art. Given the rapid development of medical image segmentation techniques, we compare our proposed model against several representative and widely adopted architectures, including U-Net [1], TransUNet [28], Swin-UNet [29], UCTransNet [26], and TSCA-Net [21]. For fair comparison, we follow the same training settings as described in the UCTransNet and TSCA-Net papers for our model. The reported results for U-Net, TransUNet, Swin-UNet, and UCTransNet on the MoNuSeg and GlaS datasets are reproduced from UCTransNet [26], while the TNBC results for these baselines are taken from TSCA-Net [21]. We evaluate all models under the same image resolution, data preprocessing, and evaluation metrics to ensure consistency in our comparative analysis.

3.2. Results

To ensure a fair comparison across all models, we evaluate the performance of our method under a uniform 224×224 input resolution, as summarized in Table 1. All reported percentage improvements refer to relative gains over the strongest baseline for each dataset: UCTransNet for GlaS and TSCA-Net for MoNuSeg and TNBC. On the GlaS dataset, our model achieves an improvement of 1.3% in Dice and 2.3% in IoU. On the MoNuSeg dataset, it further achieves gains of 1.5% in Dice and 2.4% in IoU. Lastly, on the TNBC dataset, it achieves an improvement of 1.3% in Dice and 1.7% in IoU. Additionally, as shown in Table 4, cropping 128×128 patches yields further improvements of 2.4% in Dice and 4.0% in IoU on the MoNuSeg dataset, indicating that retaining higher-resolution local detail benefits CNN-based models.

As shown in Fig. 3, qualitative comparisons on the GlaS dataset demonstrate the superiority of our method, as it captures more distinct boundaries between segmented features than other methods.

3.3. Ablation Study

To further validate the effectiveness of each component in our proposed framework, we conduct a series of ablation experiments focusing on three key aspects: (1) feature fusion strategy, (2) low-fidelity image construction, and (3) input image preprocessing. All experiments are performed under identical training and evaluation settings to ensure fair comparison.

Feature Fusion Strategy. A key design choice in our framework is the use of a distance-based operation to compare features between low- and high-fidelity branches. We evaluate this design against a concatenation-based fusion scheme, where feature maps from both branches are concatenated before being passed to the last few convolutional layers for final segmentation. As summarized in Table 2, the Euclidean distance (Ours (dis)) consistently outperforms concatenation (Ours (con)) across all datasets.

Low-Fidelity Image Construction. From Table 3, we observe that the choice of color quantization strategy significantly influences how structural information is preserved. Among all strategies, Mode produces the sharpest transitions aligned with structural boundaries—these high-frequency artifacts, though visually less smooth, are highly informative for change-based segmentation. Quantitatively, as shown in Table 3, the Mode-based method achieves the highest Dice and IoU scores on both GlaS and MoNuSeg datasets, confirming that frequency-based color preservation provides more discriminative and structurally faithful low-fidelity representations for multifidelity learning.

Input Image Preprocessing. All images are resized to 224×224 for fair comparison with transformer-based baselines, though this resolution is not optimal for our CNN-based

Table 1. Comparison of segmentation performance on the GLaS, MoNuSeg, and TNBC datasets.

Dataset	Method	Dice ↑	IoU ↑
GlaS	U-Net	0.8545±0.0130	0.7478±0.0170
	TransUNet	0.8840±0.0070	0.8040±0.0100
	Swin-UNet	0.8958±0.0060	0.8206±0.0070
	UCTransNet	0.9018±0.0070	0.8296±0.0110
	TSCA-Net	0.8867±0.0050	0.8072±0.0076
	Ours (Mode)	**0.9139±0.0728**	**0.8488±0.1094**
MoNuSeg	U-Net	0.7645±0.0260	0.6286±0.0300
	TransUNet	0.7853±0.0110	0.6505±0.0130
	Swin-UNet	0.7769±0.0090	0.6377±0.0120
	UCTransNet	0.7908±0.0070	0.6550±0.0090
	TSCA-Net	0.8023±0.0069	0.6713±0.0095
	Ours (Mode)	**0.8142±0.0196**	**0.6871±0.0280**
TNBC	U-Net	0.8136±0.0078	0.6892±0.0112
	TransUNet	0.7939±0.0107	0.6624±0.0119
	Swin-UNet	0.7589±0.0250	0.6167±0.0320
	UCTransNet	0.8108±0.0107	0.6860±0.0127
	TSCA-Net	0.8190±0.0101	0.6975±0.0123
	Ours (Mode)	**0.8295±0.0221**	**0.7092±0.0327**

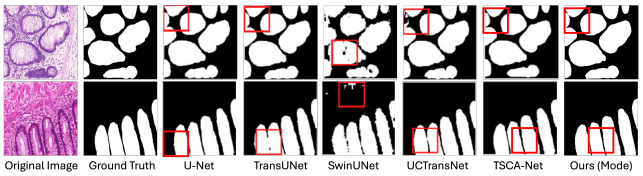

Fig. 3. Qualitative comparison of segmentation results across different models on the GlaS dataset. The predictions from other models are reproduced from TSCA-Net [21]. The red boxes highlight representative regions used for comparison across methods. As shown, our model produces more accurate and sharp segmentation results.

Table 2. Comparison of models using **difference layer** or **concatenate layer**.

Dataset	Method	Dice ↑	IoU ↑
GlaS	**Ours (dis)**	**0.9139±0.0728**	**0.8488±0.1094**
	Ours (con)	0.9088±0.0794	0.8414±0.1172
MoNuSeg	**Ours (dis)**	**0.8142±0.0196**	**0.6871±0.0280**
	Ours (con)	0.8059±0.0245	0.6757±0.0347
TNBC	**Ours (dis)**	**0.8295±0.0221**	**0.7092±0.0327**
	Ours (con)	0.8149±0.0067	0.6877±0.0095

Table 3. Comparison of segmentation performance under different color quantization strategies.

Dataset	Method	Dice ↑	IoU ↑
GlaS	**Ours (Mode)**	**0.9139±0.0728**	**0.8488±0.1094**
	Ours (Medoid)	0.9095±0.0694	0.8407±0.1051
	Ours (Centroid)	0.9127±0.0660	0.8455±0.1021
	Ours (Random)	0.9012±0.0604	0.8252±0.0933
MoNuSeg	**Ours (Mode)**	**0.8142±0.0196**	**0.6871±0.0280**
	Ours (Medoid)	0.8131±0.0221	0.6857±0.0314
	Ours (Centroid)	0.8103±0.0237	0.6817±0.0335
	Ours (Random)	0.8088±0.0216	0.6795±0.0306

architecture. To assess its impact, we compare uniform resizing against cropping, where multiple 128×128 patches are extracted from each image. As shown in Table 4, cropping yields higher segmentation performance than uniform resizing, reported in Tables 1, particularly on the MoNuSeg dataset. This gain arises from the model's ability to retain fine structural details that are partially lost during global resizing. Our fully convolutional architecture supports variable input sizes without architectural modification, offering flexibility for cropping-based inference and scalable deployment.

Table 4. Comparison of segmentation performance under **cropping vs. uniform resizing** on the MoNuSeg dataset.

Method	Dice ↑	IoU ↑
Ours (Mode_crop)	**0.8219±0.0237**	**0.6983±0.0341**
Ours (Mode_resize)	0.8142±0.0196	0.6871±0.0280

4. CONCLUSION

Our experiments demonstrate that DeltaSegment effectively reframes medical image segmentation as a multi-fidelity change detection problem, providing clear advantages in delineating complex tissue boundaries and nuclear structures. By automatically generating low-fidelity inputs through color quantization, DeltaSegment eliminates the need for additional data collection while leveraging cross-fidelity differences to capture subtle boundaries and improve segmentation consistency. The method is most effective for medical images with a limited range of dominant colors, such as histopathology images. Datasets with high color complexity or minimal color variation (e.g., grayscale modalities) may require modified coarse-graining strategies. While we explored multiple quantization methods, other forms of low-fidelity generation—such as frequency filtering, patch masking, or learned degradation models—could further expand the applicability of this approach. The cropping analysis in our ablation study also suggests practical value for clinical deployment, enabling efficient patch-based processing of large whole-slide images, improving accuracy and enhancing scalability for real-time or resource-constrained environments. Overall, DeltaSegment presents a strategy for segmentation when labeled data are limited, combining robustness, data efficiency, and adaptability across diverse medical imaging domains.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by the MoNuSeg[22], GlaS[23, 24], and TNBC[25] datasets. Ethical approval was not required as confirmed by the licenses attached with the open access data.

6. REFERENCES

[1] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. 2015, vol. 9351 of *LNCS*, pp. 234–241, Springer.

[2] Alexey Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[3] Gregory R. Koch, "Siamese neural networks for one-shot image recognition," in *ICML Deep Learning Workshop*, 2015.

[4] Aneesh Rangnekar, Christopher Kanan, and Matthew Hoffman, "Semantic Segmentation with Active Semi-Supervised Learning," in *2023 IEEE/CVF WACV*, Waikoloa, HI, USA, Jan. 2023, pp. 5955–5966, IEEE.

[5] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le, "AutoAugment: Learning Augmentation Strategies From Data," in *2019 IEEE/CVF CVPR*, Long Beach, CA, USA, June 2019, pp. 113–123, IEEE.

[6] Mahmoud Ibrahim et al., "Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges," *Computers in Biology and Medicine*, vol. 189, pp. 109834, May 2025.

[7] Alexander Kirillov et al., "Segment Anything," Apr. 2023, arXiv:2304.02643 [cs].

[8] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, pp. 654, Jan. 2024, Publisher: Nature Publishing Group.

[9] Maciej A. Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang, "Segment anything model for medical image analysis: An experimental study," *Medical Image Analysis*, vol. 89, pp. 102918, Oct. 2023.

[10] Enqiang Guo, Xinsha Fu, J. Zhu, Min Deng, Y. Liu, Qing Zhu, and Haifeng Li, "Learning to measure change: Fully convolutional siamese metric networks for scene change detection," *arXiv:1810.09111*, 2018.

[11] Hongruixuan Chen, Chen Wu, B. Du, Liangpei Zhang, and L. Wang, "Change detection in multisource vhr images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, pp. 2848–2864, 2020.

[12] Rodrigo Caye Daudt, B. L. Saux, and Alexandre Boulch, "Fully convolutional siamese networks for change detection," *2018 25th IEEE ICIP*, pp. 4063–4067, 2018.

[13] Sayan Banerjee, Avik Hati, Subhasis Chaudhuri, and Rajbabu Velmurugan, "CoSegNet: Image Co-segmentation using a Conditional Siamese Convolutional Network," pp. 673–679, 2019.

[14] Yushuo Niu, Ethan Chadwick, Anson W. K. Ma, and Qian Yang, "Semi-Siamese Network for Robust Change Detection Across Different Domains with Applications to 3D Printing," in *Computer Vision Systems*, Henrik I. Christensen, Peter Corke, Renaud Detry, Jean-Baptiste Weibel, and Markus Vincze, Eds., vol. 14253, pp. 183–196. Springer Nature Switzerland, Cham, 2023, Series Title: Lecture Notes in Computer Science.

[15] Yu-An Chung and Wei-Hung Weng, "Learning Deep Representations of Medical Images using Siamese CNNs with Application to Content-Based Image Retrieval," Dec. 2017, arXiv, arXiv:1711.08490 [cs].

[16] Matthew D. Li et al., "Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging," *npj Digital Medicine*, vol. 3, no. 1, pp. 48, Mar. 2020, Publisher: Nature Publishing Group.

[17] Zihao Huang et al., "Application of attention-based Siamese composite neural network in medical image recognition," Mar. 2024, arXiv:2304.09783 [eess] version: 3.

[18] Brandon Mac, Alan R Moody, and April Khademi, "Siamese Content Loss Networks for Highly Imbalanced Medical Image Segmentation," *MIDL*, 2020.

[19] Yongqi Yuan and Yong Cheng, "Medical image segmentation with UNet-based multi-scale context fusion," *Scientific Reports*, vol. 14, no. 1, pp. 15687, Oct. 2024, Publisher: Nature Publishing Group.

[20] Ning Zhang et al., "CT-Net: Asymmetric compound branch Transformer for medical image segmentation," *Neural Networks*, vol. 170, pp. 298–311, Feb. 2024.

[21] Yinghua Fu, Junfeng Liu, and Jun Shi, "TSCA-Net: Transformer based spatial-channel attention segmentation network for medical images," *Computers in Biology and Medicine*, vol. 170, pp. 107938, Mar. 2024.

[22] Neeraj Kumar et al., "A Multi-Organ Nucleus Segmentation Challenge," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, May 2020.

[23] Korsuk Sirinukunwattana, David R. J. Snead, and Nasir M. Rajpoot, "A Stochastic Polygons Model for Glandular Structures in Colon Histology Images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2366–2378, Nov. 2015.

[24] Korsuk Sirinukunwattana et al., "Gland segmentation in colon histology images: The glas challenge contest," *Medical Image Analysis*, vol. 35, pp. 489–502, Jan. 2017.

[25] Tahir Mahmood et al., "Accurate segmentation of nuclear regions with multi-organ histopathology images using artificial intelligence for cancer diagnosis in personalized medicine," *Journal of Personalized Medicine*, vol. 11, no. 6, pp. 515, 2021.

[26] Haonan Wang, Peng Cao, Jiaqi Wang, and Osmar R Zaiane, "Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer," in *Proc. AAAI*, 2022, vol. 36, pp. 2441–2449.

[27] Varduhi Yeghiazaryan and Irina Voiculescu, "Family of boundary overlap metrics for the evaluation of medical image segmentation," *Journal of Medical Imaging*, vol. 5, no. 01, pp. 1, Feb. 2018.

[28] Jieneng Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[29] Hu Cao et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Proc. ECCV*. Springer, 2022, pp. 205–218.