Supplementary Information (SI) for Digital Discovery.

MultiTaskDeltaNet: Change Detection-based Image Segmentation for operando ETEM with Application to Carbon Gasification Kinetics

Yushuo Niu, Tianyu Li, Yuanyuan Zhu, Qian Yang
University of Connecticut, Storrs

Supplementary Information

A. Minimal dataset size for effective training.

To evaluate the minimum dataset size required for effective training, we progressively reduced the size of each ID's training set by 30%, 40%, and 50%, while maintaining the overall variation across the dataset. As noted in the manuscript, the performance differences between models with and without weight initialization are marginal, and the results obtained using different prediction fusion strategies are also comparable. For simplicity and clarity, all experiments in this section are therefore conducted using the MTDN model without weight initialization and employing backward fusion.

The results in Table S1 demonstrate that, compared to a U-Net baseline, our MTDN model can still achieve similar segmentation performance on the challenging A_2 prediction even after reducing the training dataset by 50%. This indicates the robustness and data efficiency of the proposed model under limited training conditions.

The detailed quantitative results are presented below:

Model	F1 Score (Dice Score) ↑					
	A_1 Prediction			A_2 Prediction		
	Filament ID 6	Filament ID 7	Test Total	Filament ID 6	Filament ID 7	Test Total
U-Net	0.90293	0.97206	0.94102	0.76477	0.75970	0.76276
MTDN	0.91678	0.96940	0.94588	0.83800	0.82295	0.83117
MTDN (reduce by 30%)	0.89622	0.96154	0.93280	0.81474	0.80283	0.80924
MTDN (reduce by 40%)	0.88273	0.95773	0.92590	0.79450	0.80006	0.79737
MTDN (reduce by 50%)	0.86052	0.95655	0.91680	0.76554	0.75001	0.75852

Table S1. Performance comparison of MTDN under different training data reduction ratios. Results are reported for A_1 and A_2 predictions in terms of F1 Score (Dice Score) ↑ on test datasets Filament ID 6 and Filament ID 7, as well as the overall test performance. The U-Net model is included as the baseline.

B. Leave-one-filament-diameter-out cross-validation

To verify robustness of our model across filament diameters, we conducted leave-one-filament-diameter-out (LOFDO) cross-validation experiments. As noted in the main manuscript, the performance differences between models with and without weight initialization are marginal, and the results obtained using different prediction fusion strategies are also comparable. Therefore, for simplicity and clarity, all experiments in this section were performed using the MTDN model without weight initialization and employing backward fusion.

The LOFDO cross-validation results are presented below in Table S2. When comparing the LOFDO cross-validation results with those in the original experiments (Table 3 in the main manuscript), both demonstrate consistently similar performance, further validating the robustness and generalization of the proposed MTDN model across filament diameters.

Test Diameter	F1 Score (Dice Score) ↑		IoU (Intersection over Union) ↑	
	A_1 Prediction	A_2 Prediction	A_1 Prediction	A_2 Prediction
24 nm	0.94553	0.84934	0.89986	0.76262
37 nm	0.95304	0.91050	0.91276	0.84613
14 nm	0.91436	0.86272	0.85238	0.78317
34 nm	0.96938	0.79580	0.94189	0.70606
Mean ± SD	0.9456 ± 0.0265	0.8546 ± 0.0471	0.9017 ± 0.0372	0.7745 ± 0.0578

Table S2. Performance comparison for LOFDO cross-validation results using MTDN. Each test filament diameter dataset is used as the test set in turn, and the mean ± standard deviation (SD) of the overall performance is also reported. Results are presented for both A_1 and A_2 predictions in terms of F1 Score (Dice Score)↑ and IoU (Intersection over Union)↑.

C. Shuffled data partitioning

We conducted additional experiments with a modified partitioning strategy where the dataset was shuffled such that the training, validation, and test sets partially overlap in filament diameters. Specifically, the new training set includes Filament IDs 3, 4, and 6 (diameters: 24, 37, and 14 nm), the validation set includes Filament IDs 1 and 5 (24 and 14 nm), and the test set includes Filament IDs 2 and 7 (24 and 34 nm).

As noted in the main manuscript, the performance differences between models with and without weight initialization are marginal, and different prediction fusion strategies yield comparable results. Therefore, for simplicity, all experiments in this section were conducted using the MTDN model without weight initialization and employing backward fusion.

The results in Table S3 below show that MTDN continues to outperform U-Net under the modified data partition. This further validates the robustness and generalization capability of the proposed MTDN model across filament diameters.

Model	F1 Score (Dice Score)					
	A_1 Prediction			A_2 Prediction		
	Filament	Filament	Test total	Filament	Filament	Test

	ID 2	ID 7		ID 2	ID 7	total
U-Net	0.97666	0.96952	0.97573	0.92841	0.76484	0.91010
MTDN	0.98242	0.97156	0.98098	0.94178	0.79863	0.92611

Table S3. Performance comparison of U-Net and MTDN under the modified data partition with overlapping filament diameters.

D. Change detection performance comparison

All performance comparisons in the main manuscript are based on segmentation results. For completeness, the change-detection results obtained by predicting each frame independently with U-Net and then computing their difference are presented below in Table S4. As shown, MTDN consistently outperforms U-Net in the change detection setting as well.

Model	F1 Score (Dice Score) ↑					
	A_1 Prediction			A_2 Prediction		
	Filament ID 6	Filament ID 7	Test total	Filament ID 6	Filament ID 7	Test total
U-Net	0.83636	0.92992	0.87092	0.60858	0.56297	0.59365
MTDN_no_int	0.87146	0.92209	0.88811	0.73837	0.68089	0.71548
MTDN_init1	0.87562	0.91269	0.88647	0.75603	0.67999	0.72816
MTDN_init2	0.87141	0.91665	0.88545	0.77349	0.67351	0.73636

Table S4. Comparison of change detection results between U-Net and different variants of MTDN.

E. Computational cost comparison

Table S5 below compares the total trainable parameters and training time between U-Net and MTDN. Both models were trained with the same early stopping strategy on a single Nvidia GeForce A5000 GPU, ensuring comparable training budgets.

The parameter counts and execution time for U-net and MTDN are below:

Model	**Total Params (M)**	**Execution Time**
U-Net	7.76	1.46h
MTDN	7.88	7.32h

Table S5. Comparison of trainable parameters and training time between U-Net and MTDN. All models were trained using an early stopping strategy on single Nvidia GeForce A5000 GPU.