SSNet: A Hybrid CNN–Transformer Architecture for Organ Segmentation

TL;DR: We evaluate a hybrid semantic segmentation architecture (SSNet) for liver and lung segmentation using multiple MiT transformer backbones. For liver segmentation, three variants (MiTB0, MiTB3, MiTB5) are compared, where MiTB3 achieves the best overall performance with a Dice score of **96.03%** and an HD95 value of **7.49 mm**, outperforming several recent state-of-the-art methods on the MSD liver dataset. For lung segmentation, MiTB3 demonstrates competitive performance, achieving a Dice score of **96.03%** and an HD95 value of **3.65 mm**. While slightly behind the top-performing methods in terms of Dice, SSNet shows strong boundary accuracy and robust generalization on a comparatively smaller dataset. Overall, the results highlight the effectiveness of mid-scale transformer backbones for accurate and efficient organ segmentation.

1 Experimental Setup

Table 1: Experimental Configuration 1

Parameter	**Value**
Dataset	MSD Task 03 Liver
Classes	2 (Organ and Background)
Train / Val / Test volumes	86 / 18 / 19
Train / Val/ Test slices	∼13k/∼3k/∼3k
Image Size	256×256 grayscale
Hardware	Local GPU (CUDA)
Batch Size	4/8
Fair Comparison	Different variants of MiT transformer(MiTB0, MiTB3, MiTB5)

Table 2: Experimental Configuration 2

Parameter	**Value**
Dataset	Covid19 CT Lung and Infection segmentation dataset
Classes	2 (Organ and Background)
Train / Val / Test volumes	16 / 1 / 3
Train / Val/ Test slices	1763/411/642
Image Size	256×256 grayscale
Hardware	Local GPU (CUDA)
Batch Size	8

2 Training Behavior

Table 3: Training Comparison for Liver Segmentation

Metric	MiTB0	MiTB3	MiTB5
GFlOPs	3.73	13	20
No. of parameters(Millions)	20	96	133.99
Dice Score Coefficient ↑	95.94	**96.03**	95.13
HD95(mm) ↓	7.54	**7.49**	10.50
Precision	97.69	**97.83**	97.08
Recall	96.44	**96.49**	96.26
mIoU	94.39	**94.49**	93.65

Table 4: Training metrics for Lung Segmentation

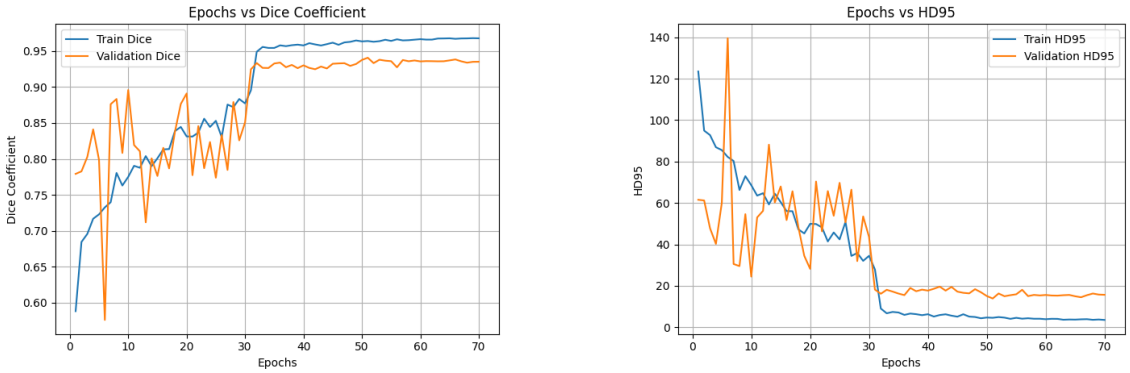
Metric	MiTB3
GFlOPs	13
No. of parameters(Millions)	96
Dice Score Coefficient ↑	96.0
HD95(mm) ↓	3.65
Precision	96.88
Recall	95.35
mIoU	92.67

3 Training Performance Analysis

This section presents the quantitative training performance of the proposed segmentation model for liver and lung organs using standard evaluation metrics.

3.1 Liver Segmentation Performance

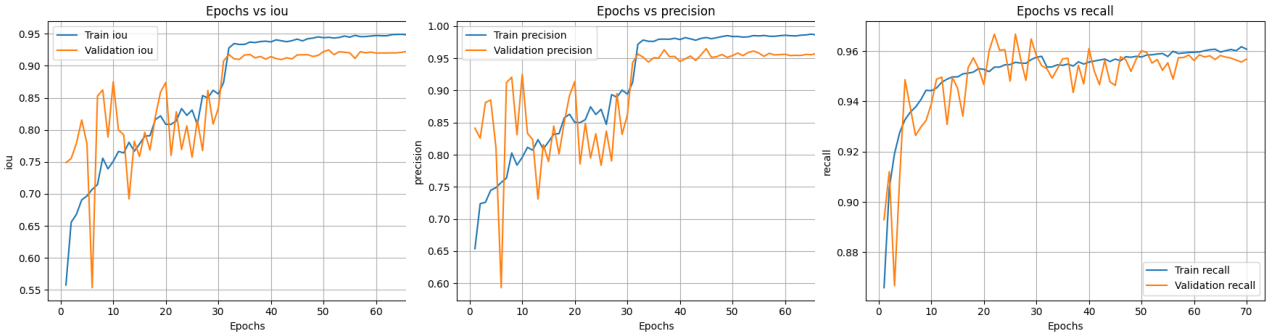
3.1.1 Overlap and Boundary Accuracy



(a) Dice Coefficient (b) HD95

Figure 1: Training performance of liver segmentation in terms of (a) Dice coefficient and (b) 95th percentile Hausdorff Distance (HD95).

3.1.2 Region-Based Evaluation Metrics



(a) IoU (b) Precision (c) Recall

Figure 2: Region-based training metrics for liver segmentation: (a) Intersection over Union (IoU), (b) Precision, and (c) Recall.

3.2 Lung Segmentation Performance

3.2.1 Overlap Accuracy

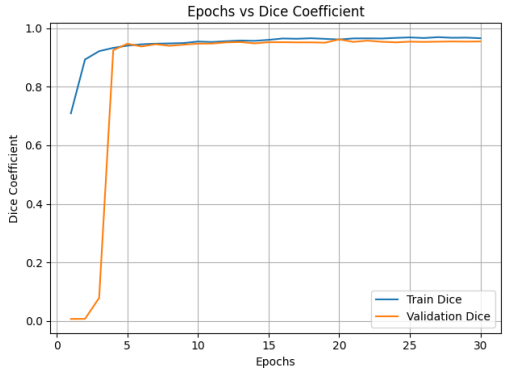
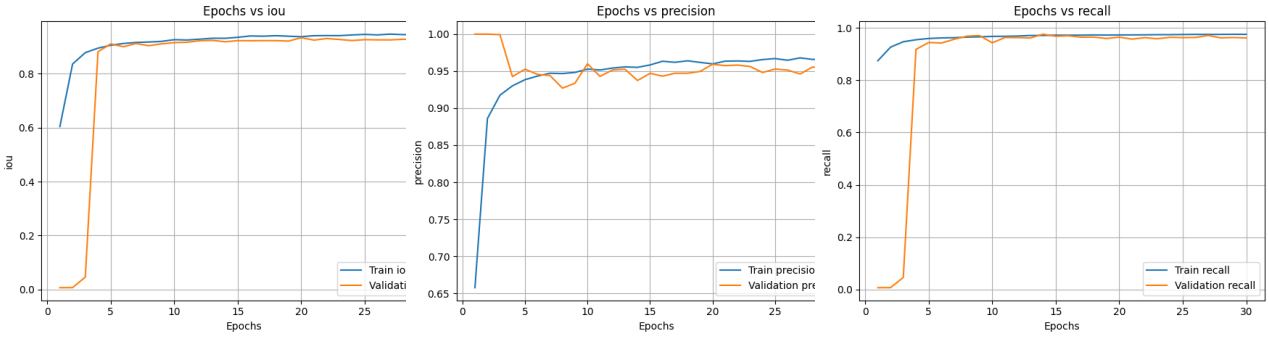


Figure 3: Training Dice coefficient curve for lung segmentation.

3.2.2 Region-Based Evaluation Metrics



(a) IoU (b) Precision (c) Recall

Figure 4: Region-based training metrics for lung segmentation: (a) Intersection over Union (IoU), (b) Precision, and (c) Recall.

4 Comparison with Existing Methods

This section presents a quantitative comparison between the proposed SSNet architecture and previously published state-of-the-art methods evaluated on the same datasets. The comparison is performed using standard overlap and boundary accuracy metrics reported in the respective papers.

Table 5: Performance Comparison with Existing Methods on MSD Liver Dataset

Model	**Dice (%) ↑**	**HD95 (mm) ↓**
FF Swin-Unet	94.42	15.94
STD-Net	91.9	8.13
vMixer	94.89	10.26
UNetFormer	95.73	7.68
nnUnet	95.75	7.94
TransUnet	92.66	-
SwinUnet	94.17	-
FSS ULivR	94.78	-
Autoregressive sequence model	95.9	-
Universal model	95.4	-
Swin UnetR	95.35	-
DiNTs	95.35	-
EffiDec 3D	93.68	-
VNet with attention gate	95.54	-
SSNet (Ours)	**96.03**	**7.49**

Table 6: Performance Comparison with Existing Methods on COVID-19 Lung Dataset

Model	**Dice (%) ↑**	**HD95 (mm) ↓**
Unet	89	-
UNet++	98.3	-
MultiResUnet	89.88	-
QAPNet	81.63	-
DMDF Net	98.66	-
3D-CUNet	96.15	-
3D CovidSegNet	textbf98.7	-
3D-Unet	95.6	-
3D-VGGUnet	96.24	-
CHSNet	96.3	-
DANEt with ECA-Attention	97.14	-
nnUnet	87.90	-
LungQuant2	96.01	-
SSNet (Ours)	96.03	**3.65**

5 Results Snapshot

5.1 Qualitative Segmentation Results

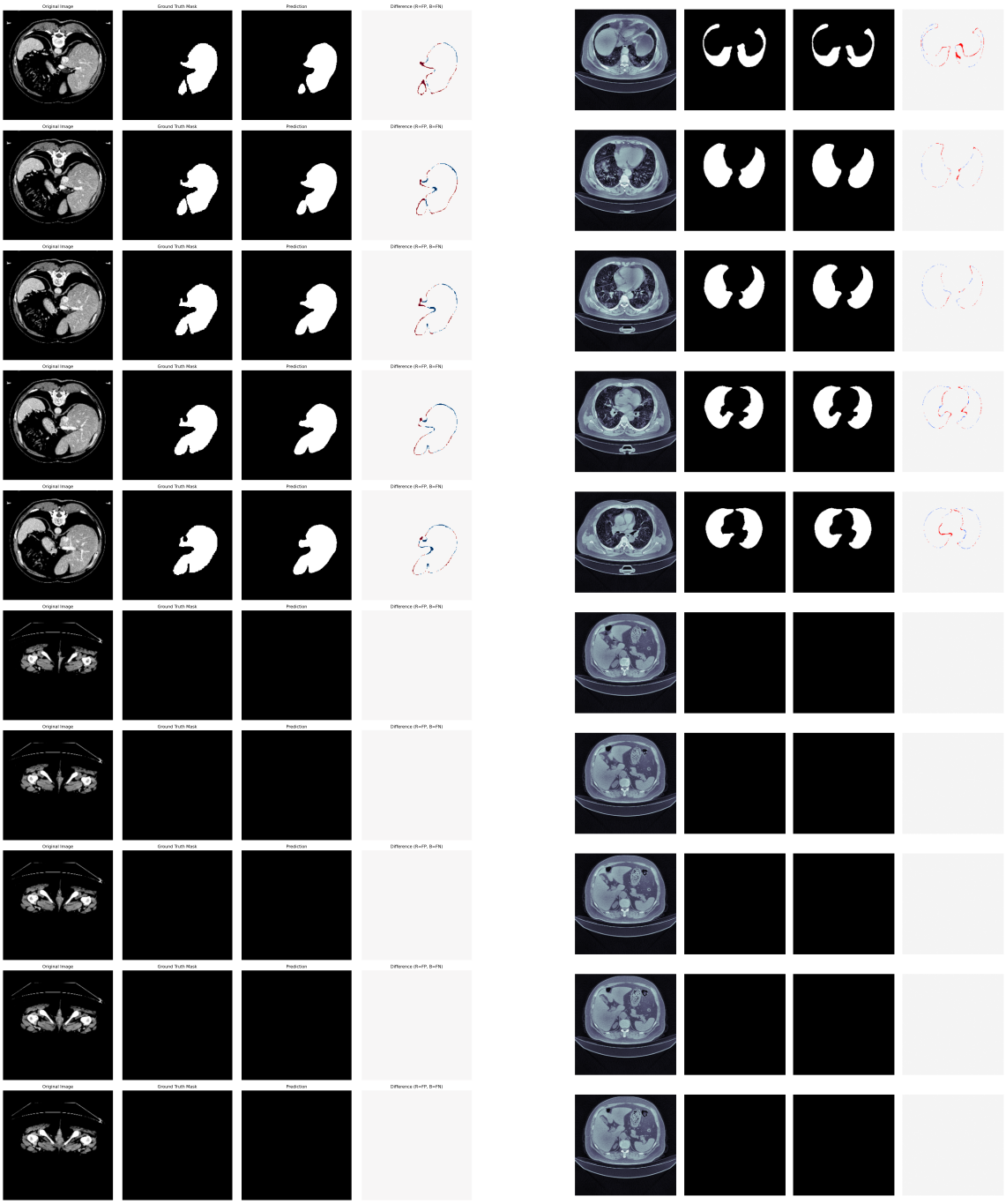


Figure 5: Qualitative segmentation results for liver (left) and lung (right). For each organ, the columns correspond to: (a) input CT image, (b) ground truth segmentation mask, (c) model prediction, and (d) difference map illustrating discrepancies between prediction and ground truth.

6 Conclusion

In this work, we presented an extensive evaluation of SSNet, a hybrid CNN–Transformer semantic segmentation architecture, for liver and lung segmentation tasks on CT imaging datasets. Multiple variants of the MiT transformer backbone were investigated to analyze the trade-off between model complexity and segmentation accuracy.

For liver segmentation, a comparative study was conducted using MiTB0, MiTB3, and MiTB5 backbones. Among these, MiTB3 achieved the best overall performance, attaining a Dice coefficient of **96.03%** and the lowest HD95 value of **7.49 mm**, indicating improved boundary delineation. As shown in Table 5, SSNet outperforms several recent state-of-the-art approaches, including UNetFormer, nnU-Net, FF Swin-Unet, and vMixer, in terms of both overlap accuracy and boundary precision. These results demonstrate that the proposed hybrid architecture provides a favorable balance between representational capacity and computational efficiency for liver organ segmentation.

For lung segmentation, SSNet with the MiTB3 backbone demonstrated competitive performance across overlap-based and region-based metrics, achieving a Dice score of **96.03%** and an HD95 of **3.65 mm**. While the Dice score is slightly lower than the best-performing methods such as DANet with ECA-Attention and CHSNet, as reported in Table 6, SSNet remains highly competitive and offers improved boundary accuracy, as reflected by the reported HD95 value, which is not available for most of the compared methods. Furthermore, considering the comparatively smaller training dataset used for lung segmentation, the obtained results indicate strong generalization capability of the proposed model under limited data conditions.

Overall, the experimental results indicate that mid-scale transformer backbones strike an effective balance between accuracy and efficiency for organ segmentation tasks. SSNet establishes new state-of-the-art performance on the MSD liver dataset and delivers competitive results on the COVID-19 lung dataset, highlighting its robustness across different organ structures and dataset scales. Future work will explore multi-organ segmentation, robustness across larger and more diverse datasets, and integration of advanced regularization strategies to further enhance generalization performance, particularly for small-scale and heterogeneous clinical datasets.