Privacy-Preserving Collaboration for Multi-Organ Segmentation via Federated Learning from Sites with Partial Labels

Adway Kanhere*, Pranav Kulkarni, Paul H. Yi, Vishwa S. Parekh

University of Maryland School of Medicine, Baltimore, MD

*correspondence: akanhere@som.umaryland.edu adwaykanhere @AdwayKanhere

CVPR JUNE 17-21, 2024 SEATTLE, WA

Summary

- Federated learning can **aggregate knowledge from distributed medical imaging datasets** with **partial annotations** into a single 'global' model **without the need to share data**.
- We aim to show **training one unified segmentation model** is possible
- This is achieved by combining and sharing knowledge from different segmentation tasks working on different anatomical structures and fields of view
- SegViz will enable different groups around the world curating single-task datasets to collaborate and train large-scale clinically translational segmentation models.

Motivation

- Manual annotation of 3D medical images is time-consuming and costly, leading to datasets focusing on single organs.
- Specialized models lack generalizability, limiting clinical utility. There is a need for a collaborative approach to develop multi-organ segmentation models.
- We show that SegViz aims to enable privacy-preserving collaboration among sites with partial labels, enhancing AI-assisted medical image analysis.

Data

- Used 4 heterogenous public datasets from Medical Segmentation Decathlon challenge and BTCV as an independent test set
- The training data varied in dimensions and voxel spacing, resulting in heterogeneous, partially labeled data.

Set	Annotated Organ	Dataset	Modality	Imaging Protocol	Median Shape	Median Spacing (mm)	Sample Size	Intensity Range
Training + Validation	Liver	MSD	CT	Portal Venous Phase	432 x 512 x 512	1 x 0.77 x 0.77	131	[-200 , 200]
Training + Validation	Spleen	MSD	CT	Portal Venous Phase	90 x 512 x 512	5 x 0.79 x 0.79	41	[-57 , 164]
Training + Validation	Kidneys	KiTs-2019	CT	Preop. Late Arterial Phase	107 x 512 x 512	3 x 0.78 x 0.78	206	[-79 , 304]
Training + Validation	Pancreas	MSD	CT	Portal Venous Phase	93 x 512 x 512	2.5 x 0.8 x 0.8	282	[-87 , 199]
Testing	Liver, Spleen, Kidneys, Pancreas	BTCV	CT	Portal Venous Phase	128 x 512 x 512	3 x 0.76 x 0.76	30	[-175, 250]

Methods

- SegViz uses an FL setup with a 3D U-Net architecture. The model has a representation block for shared features and a task block for specific features.
- Selective weight synchronization aggregates shared representation weights while maintaining task-specific knowledge into a meta-model.
- The local model was trained with the Adam optimizer and data augmentation to handle distribution shifts.
- Evaluated SegViz across different FL setups, comparing it to conventional FL, single-organ models, and centralized aggregation models. Applied a privacy-preserving strategy by synchronizing only partial networks, thus enhancing data security.
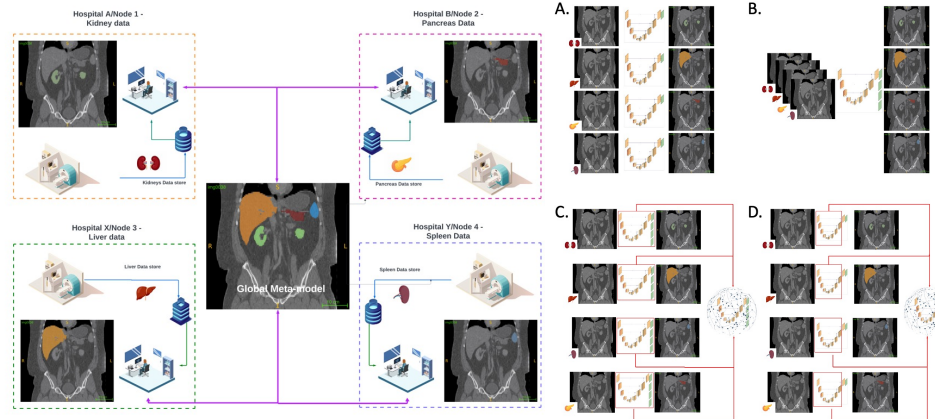

Figure 1: Illustration of SegViz. Suppose there are four research groups, each working on segmenting a single organ using abdominal CT scans. Using our proposed method, they can collaboratively train a multi-organ segmentation model capable of segmenting all four organs.

Figure 2: (A) Baseline: Independent models trained on each dataset's labels.(B) Central aggregation: Single model trained on combined datasets(C) Conventional federated learning: Tasks pre-defined, full weight aggregation. (D) SegViz: Selective aggregation preserving task specificity, no prerequisite task knowledge required

Results

Experimental Setup	Models	Organ Segmented			
		Liver	Spleen	Pancreas	Kidneys
Single-Dataset Models	Liver	0.88 ± 0.15	-	-	-
	Spleen	-	0.79 ± 0.17	-	-
	Pancreas	-	-	0.46 ± 0.20	-
	Kidneys	-	-	-	0.64 ± 0.21
2-node	Central Aggregation	0.00 ± 0.00	0.65 ± 0.14	-	-
	Conventional FL	0.89 ± 0.06	0.84 ± 0.09	-	-
	SegViz	**0.90 ± 0.04**	**0.84 ± 0.12**	-	-
3-node	Central Aggregation	0.00 ± 0.00	0.61 ± 0.18	0.00 ± 0.00	-
	Conventional FL	0.75 ± 0.29	0.71 ± 0.17	0.45 ± 0.17	-
	SegViz	**0.91 ± 0.06**	**0.81 ± 0.17**	**0.55 ± 0.19**	-
4-node	Central Aggregation	0.65 ± 0.14	0.00 ± 0.00	**0.55 ± 0.18**	0.68 ± 0.21
	Conventional FL	0.91 ± 0.03	0.46 ± 0.15	0.53 ± 0.18	0.68 ± 0.12
	SegViz	**0.93 ± 0.02**	**0.78 ± 0.14**	0.40 ± 0.20	**0.78 ± 0.12**

CT Image	Ground Truth	Baseline model	Central Agg	Vanilla FL	SegViz
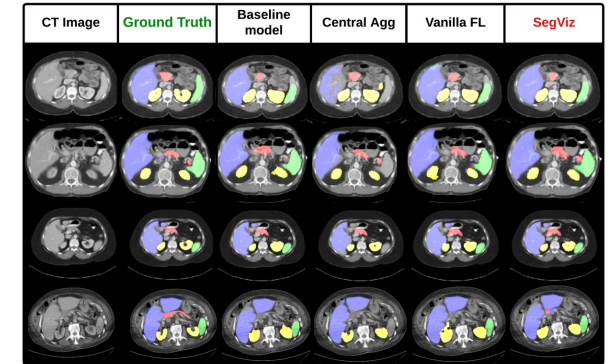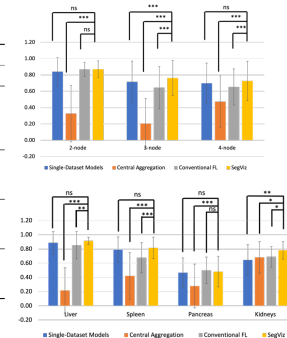
- SegViz outperformed conventional FL, single-organ models, and a centralized model.
- Achieved higher Dice similarity scores for liver, spleen, pancreas, and kidney segmentation across different FL setups. In a 4-node setup, SegViz scored 0.93 for liver segmentation, outperforming all other methods.
- Demonstrated effective knowledge aggregation from heterogeneous datasets with high segmentation accuracy.

Limitations & Extensions

- SegViz is trained and tested only on CT, and requires the same UNet backbone at each client.
- Future work will extend SegViz to more modalities, accommodate new tasks, and test with overlapping labels at clients.

Live Demo with Flower.ai