FANTOM: Federated Adversarial Network for Training Multi Sequence Magnetic Resonance Imaging in Semantic Segmentation

Anupam Borthakur [1], Apoorva Srivastava [1], Avik Kar [2], Dipayan Dewan [1], Debdoot Sheet [1]

[1] Indian Institute of Technology, Kharagpur, Kharagpur India
[2] Indian Institute of Science, Bangalore, India

IEEE International Conference on Image Processing
27 - 30 October 2024 | Abu Dhabi, UAE

Introduction

Stroke is a leading cause of death worldwide

Ischemic Stroke emerging as its predominant form

MRI commonly used by clinicians to detect core and penumbra

DNN based medical image segmentation (**encoder & decoder architecture**)

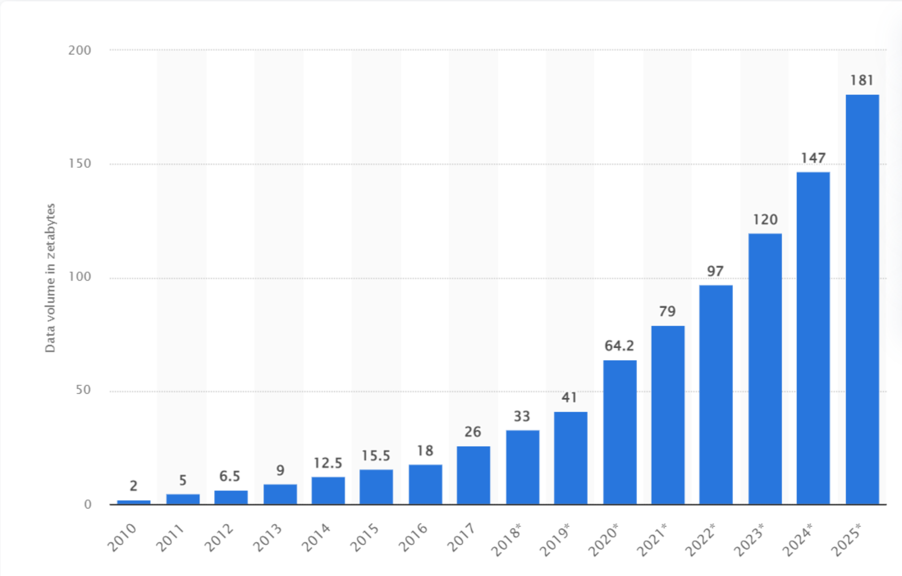
Challenge (a): complexity (b) small size dataset (c) non-IID

FANTOM comes to the rescue.

Federated **A**dversarial **N**etwork for **T**raining Multi-Sequence Magnetic Res**o**nance Imaging in Semantic Seg**m**entation

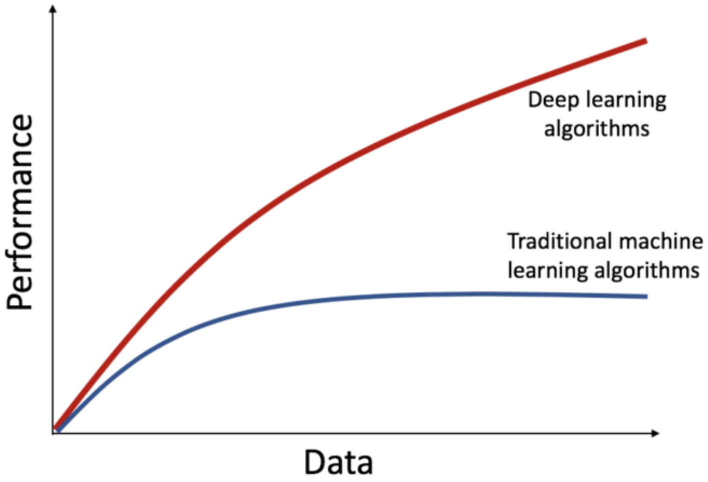
Hossein Abbasi, , et al., "Automatic brain ischemic stroke segmentation with deep learning: A review," Neuroscience Informatics, 2023.
Stefan Winzeck et al., "Isles 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral mri," Frontiers in neurology, 2018.
Jiaxu Miao , et al.,"Fedseg: Class-heterogeneous federated learning for se-mantic segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

Motivation for FL



Growth of Data Generation

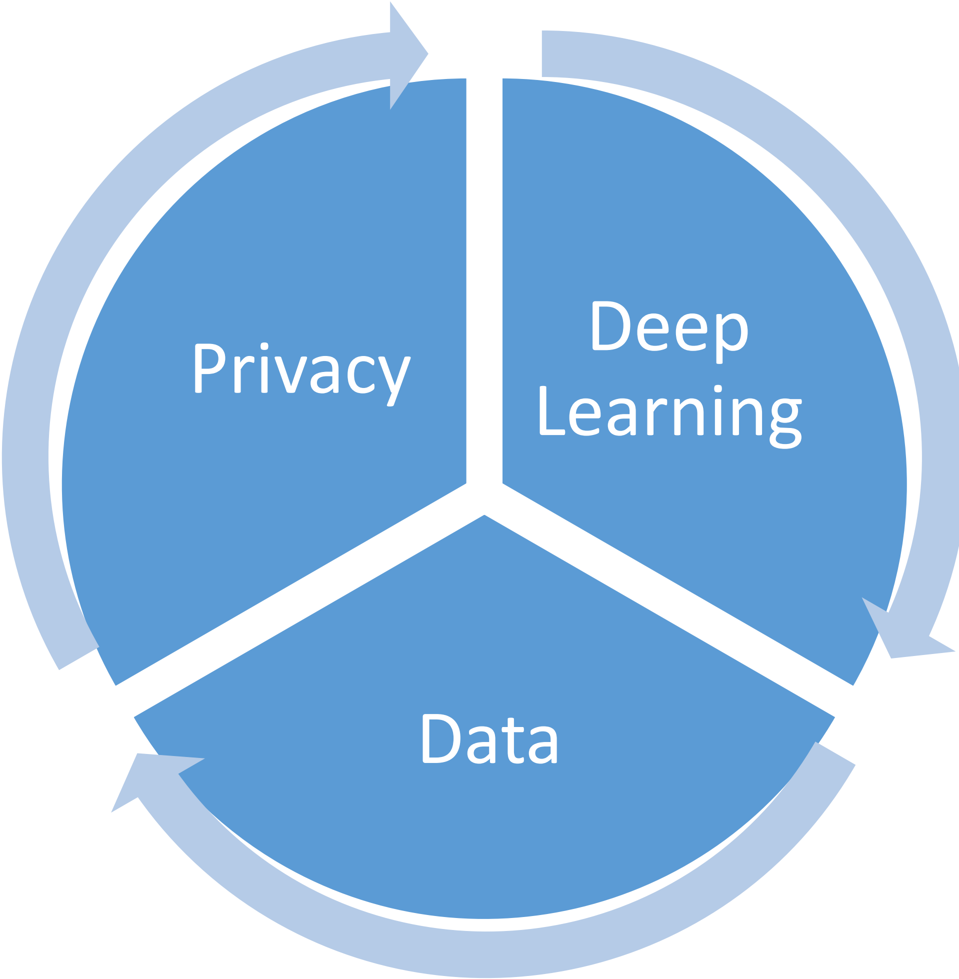
Privacy Concerns



Deep Learning

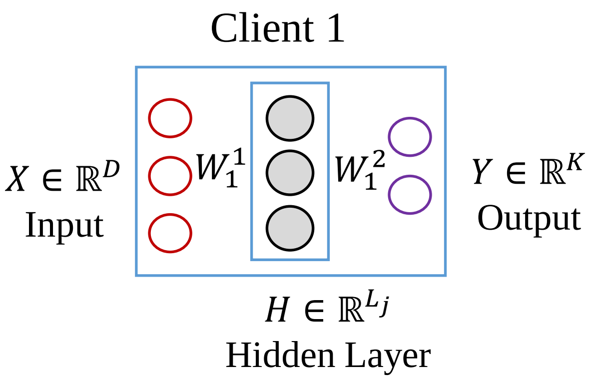
https://www.statista.com/statistics/871513/worldwide-data-created/
https://www.statista.com/chart/16400/internet-online-privacy/
https://abyssal.eu/were-data-hungry/

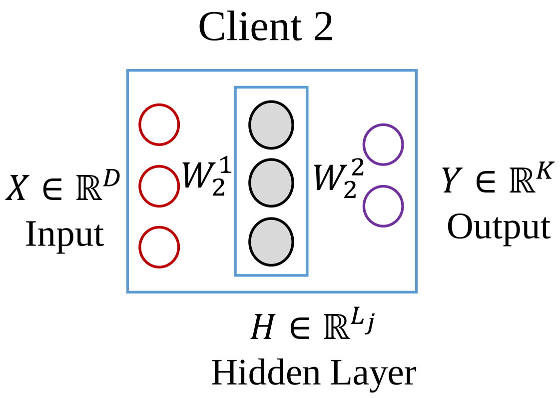
Motivation for FL



Jian Wang, , et al., "A review of deep learning on medical image analysis," Mobile Networks and Applications, 2021.
George J Annas, "Medical privacy and medical research: judging the new federal regulations," New England Journal of Medicine, 2012.

Weight Averaging methods





$$W_{new}^1 = \frac{W_1^1 + W_2^1}{2}$$

$$W_{new}^2 = \frac{W_1^2 + W_2^2}{2}$$

- Examples:
  - Federated Averaging (FedAvg) [1]
  - Federated Averaging with proximal term (FedProx) [2]

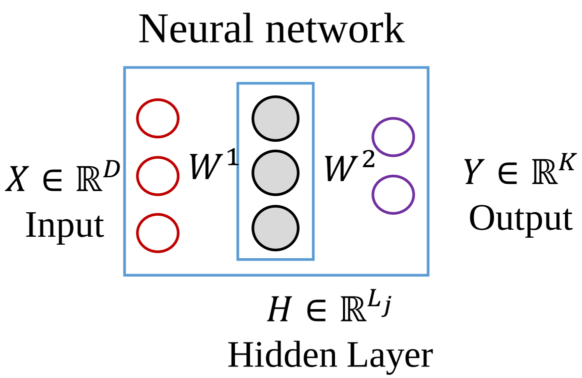
- These algorithms **don't converge** in non-IID cases

[1] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. (PMLR), 2017.
[2] Li, Tian, et al. "Federated optimization in heterogeneous networks." Proceedings of Machine learning and Systems 2 (MLSys), 2020.

Issue in Weight Averaging Methods

What can be the issue?

- Operations inside Neural networks are summations of products



$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad H = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

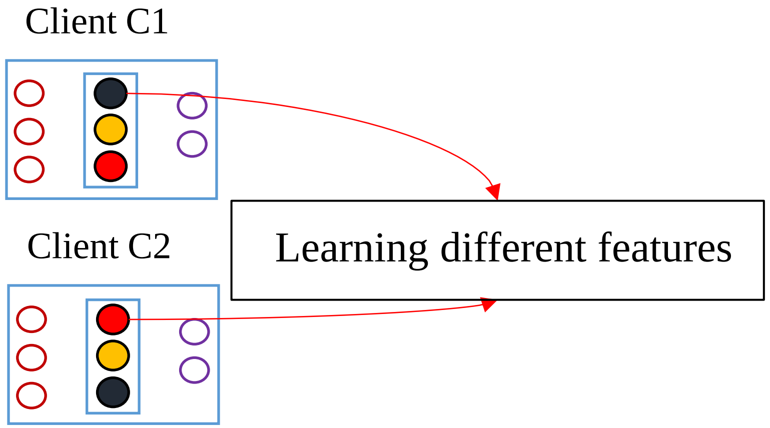
$$h_i = W_{i1}^1 \cdot x_1 + W_{i2}^1 \cdot x_2 + W_{i3}^1 \cdot x_3$$

(Ignore bias for simplicity)

$$y_j = W_{j1}^2 \cdot f(h_1) + W_{j2}^2 \cdot f(h_2) + W_{j3}^2 \cdot f(h_3)$$

- Summation is a permutation invariant operation

$$\left.\begin{aligned} y_1 &= W_{11}^2 \cdot f(h_1) + W_{12}^2 \cdot f(h_2) + W_{13}^2 \cdot f(h_3) \\ y_1 &= W_{13}^2 \cdot f(h_3) + W_{11}^2 \cdot f(h_1) + W_{12}^2 \cdot f(h_2) \end{aligned}\right\} \text{(same)}$$

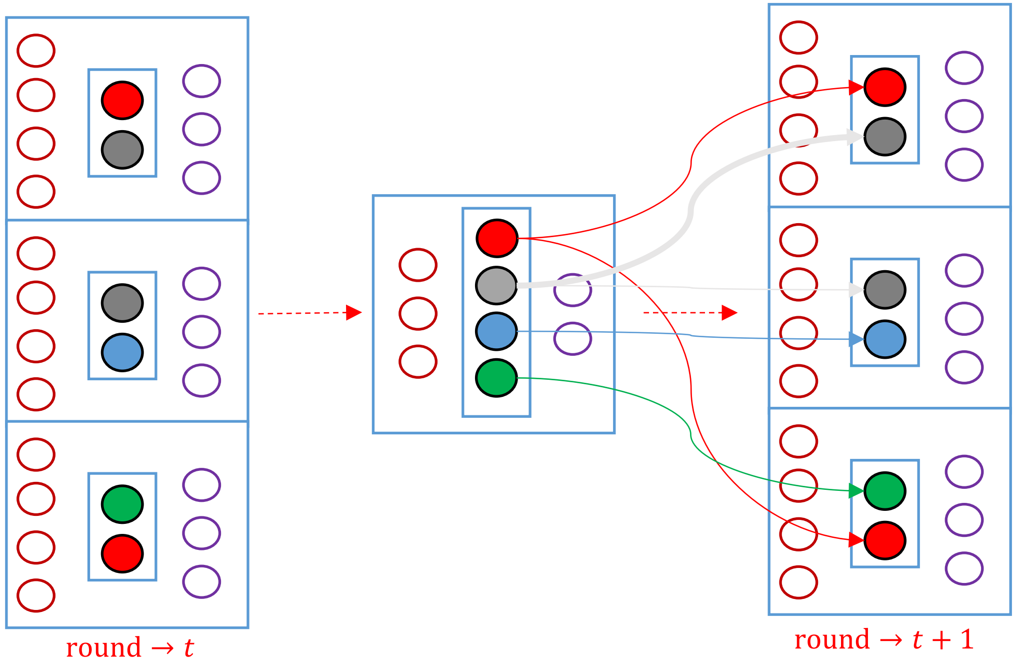


[Solution]:
To aggregate models neurons should be **properly matched** across all clients

Benjamin Bloem-Reddy, et al., "Probabilistic symmetries and invariant neural networks," Journal of Machine Learning Research (JMLR), 2020.

Multiple communication rounds

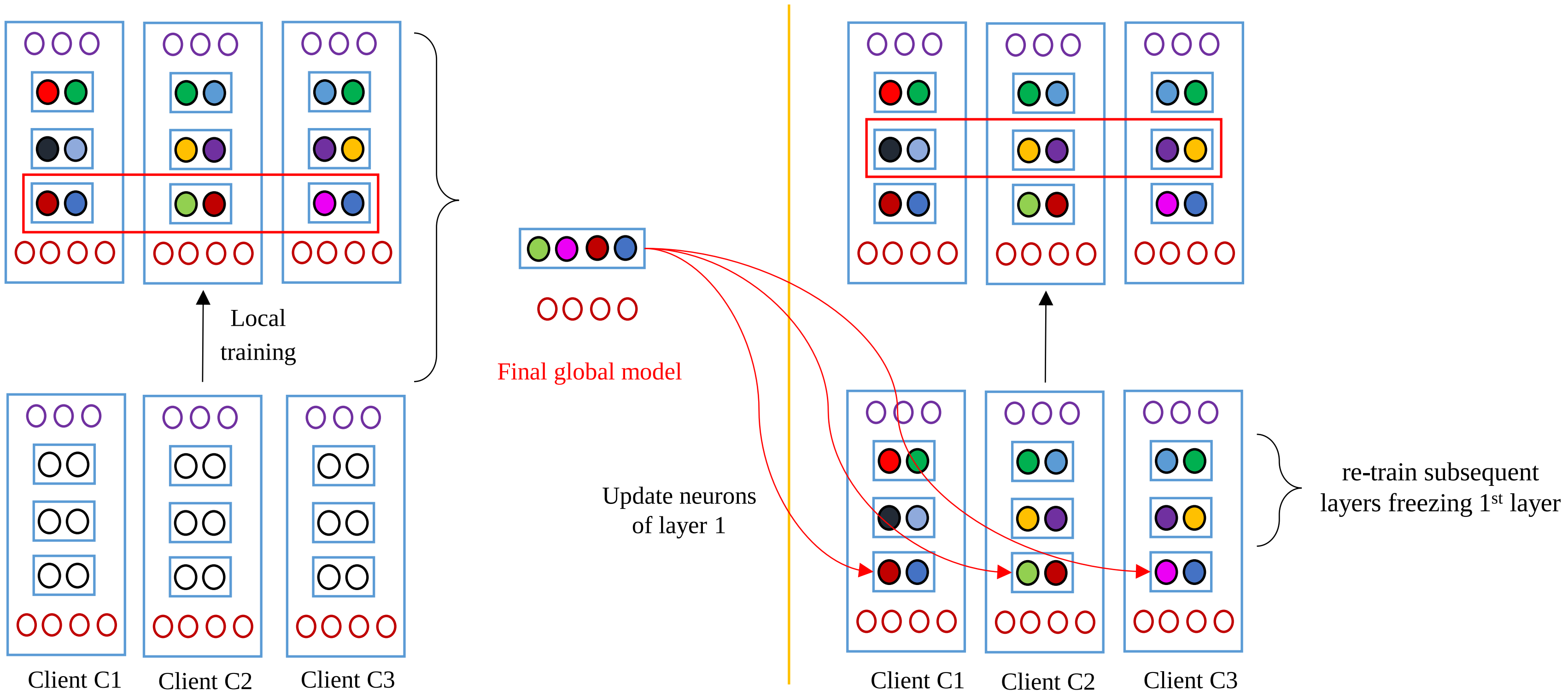
- In a FL scenario models are aggregated by a central server and then sent to local clients for **retraining**
- This continues for some communication rounds
- Since global models size is not fixed, only the matched neurons are set in local clients



Matching is based on the following

- Levy-Processes
- Beta-Process
- Bernoulli-Process

Federated Matched Averaging (FedMA)



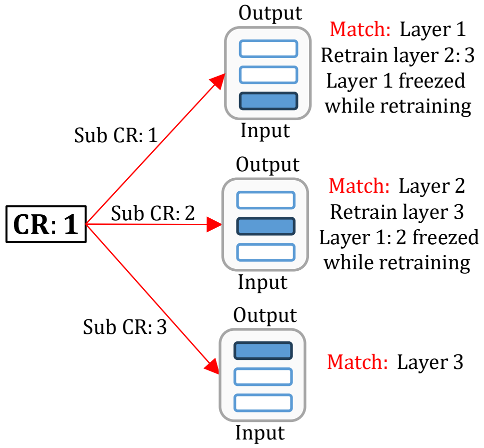
https://github.com/IBM/FedMA

Problem with FedMA

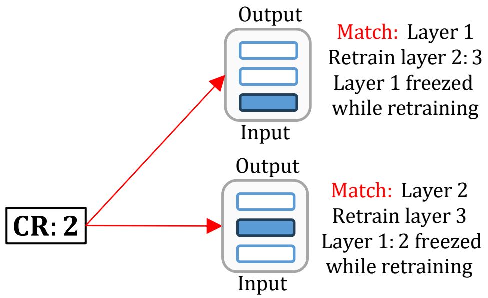
- 1 CR followed by one layer matching takes more CRs.
- A model with *N* layers required *N* rounds of communication → full model weights once
- Well trained model need not undergo matched average multiple times.
  - Local dataset not be changed through out training process

Propose: FedAvg with Initial matching → weights of all layers will be matched only in 1 CR

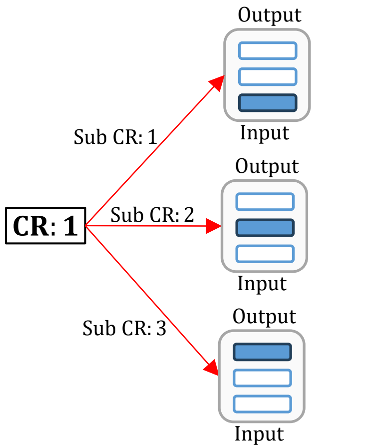
FedMA



Number of sub CRs: Num of learnable layers



FedAvg with Initial FedMA Matching



CR: 2 Federated Averaging

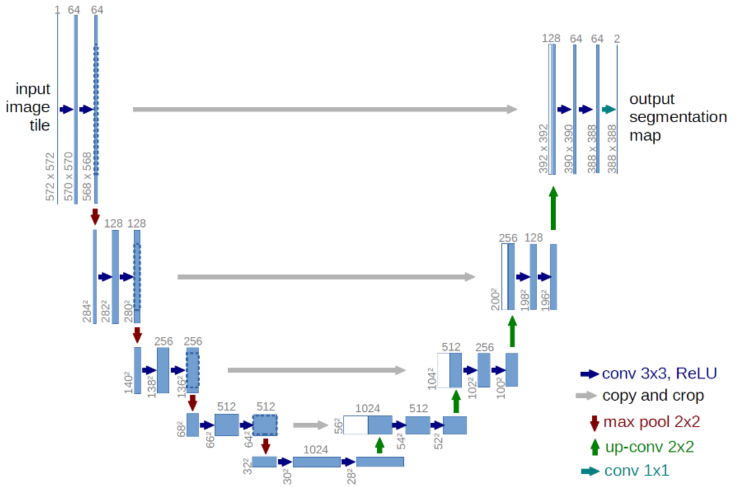
CR: 3 Federated Averaging

Neural Network with 3 hidden layers

CR Communication round

Medical image segmentation using DL

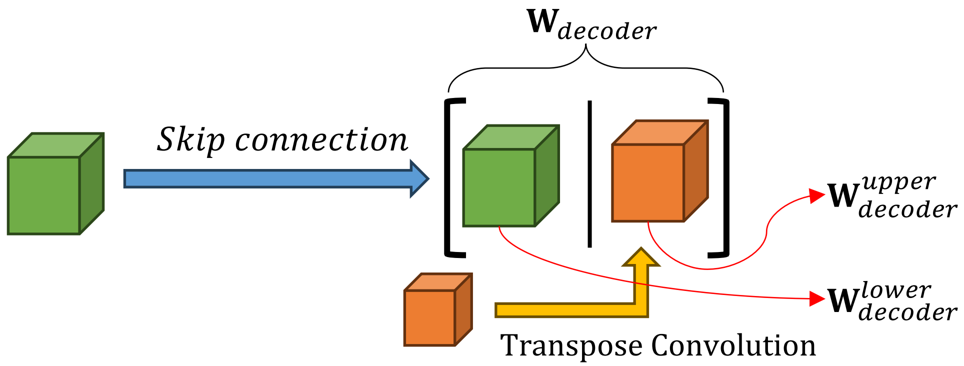
- Deep neural networks are very popular choice for medical image segmentation as they can learn very complex patterns
- Unet [1], SUMNet [2] are some of the popular networks for medical image segmentation
- They are encoder-decoder architectures which has feature concatenations that enhance the capabilities of these models



- Our proposed method needs to be modified to work with these type of architectures

More modifications needed in the proposed method

- Should be able to handle **feature concatenation [3]**

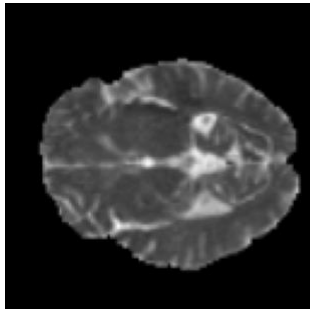


- Should be able to perform matching for **batch-normalization [4]**
- Should be able to **handle transpose convolution**

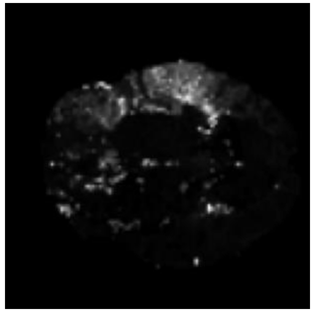
[1] Ronneberger, Olaf, , et al., "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.
[2] Nandamuri, Sumanth, et al., "Sumnet: Fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes." , in Proceedings International Symposium on Biomedical Imaging (ISBI), 2019.
[3] Kaiming He , et al., "Deep residual learning for image recognition," in Proceedings Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[4] Sergey Loffe , , et al., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proceedings International Conference on Machine Learning. (PMLR), 2015.

Experiments: Dataset

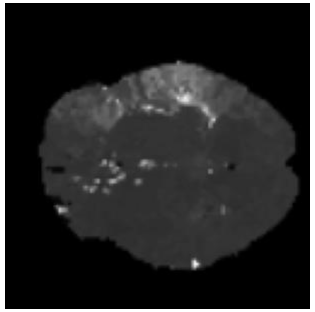
- We have performed experiments on Ischaemic Stroke Lesion Segmentation Challenge (ISLES)-2015 dataset
- It contains Magnetic Imaging Response (MRI) images
- Following are the channels which are present in the dataset
  - Diffusion Weighted Imaging (DWI)
  - Time to max (Tmax)
  - Time to peak (TTP)



DWI



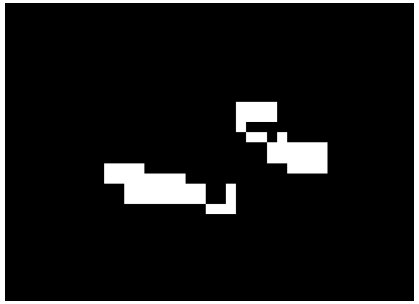
Tmax



TTP

- Following are the channels to be segmented
  - Penumbra
  - Core

Penumbra



Core

- There are total 30 volumes with an average of 70 slices per volume
- Size of each slice is 94x110 on an average

https://www.isles-challenge.org/ISLES2015/

Local and Global Training

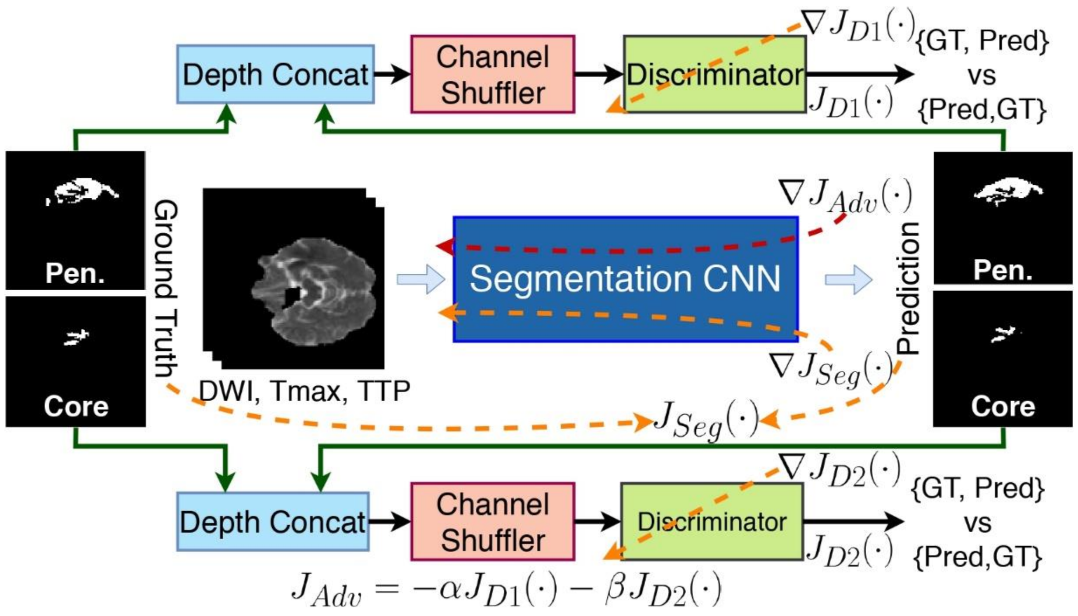


Fig: Overview of adversarial training

Algorithm 1 `FedAvg with Initial Matching.` K clients are indexed by k; B is the local minibatch size, E_t is the number of local epochs for t^{th} round and η is the learning rate. $E_0 > E_\tau$ where $\tau > 0$.

Server executes:
Initialize w_0;
for each round $t = 0, 1, 2...$ **do**
  for each client $k \in \mathbf{S}$ in parallel **do**
    $\mathrm{w}_{t+1}^k \leftarrow$ClientUpdate$(k, \mathrm{w}_t, E_t, t)$;
  end for
  if $t = 0$ **then**
    $\mathrm{w}_{t+1} \leftarrow$MatchedAverage$(\{\mathrm{w}_{t+1}^k\}_{k=1}^K)$;
  else
    $\mathrm{w}_{t+1} \leftarrow \frac{1}{K}\sum_{k=1}^K \mathrm{w}_{t+1}^k$;
  end if
end for

ClientUpdate(k, w_t, E, t)
$\mathcal{B} \leftarrow$(Split $\mathcal{P}_k$ into batches of size B);
for each local epoch i from 1 to E **do**
  for batch b in $\mathcal{B}$ **do**
    $\mathrm{w}_{t+1}^k \leftarrow$ModelUpdate$(\mathrm{w}_t)$;
  end for
end for

Rachana Sathish et al., "Adversarially trained convolutional neural networks for semantic segmentation of ischaemic stroke lesion using multisequence magnetic resonance imaging," in Proceedings International Conference of the Engineering in Medicine and Biology Society (EMBC), 2019.

Training Details

- Dataset: Ischaemic Stroke Lesion Segmentation Challenge (ISLES 2015)
  - Used – DWI, TTP and Tmax sequence
- Performance evaluated – 3 fold cross validation (6:2:2)
- Images resize – 128 x128
- 20 training subjects into 3 clients
- Segmentation model – SUMNet
- Initial epochs 230. Rest -200
- #CR: 20

Three experiments are carried under FL setup

1. FL Expt 1: Training without relativistic visual Turning test (rVTT)
2. FL Expt 2: rVTT discriminators are included in the FL framework
3. FL Expt 3: rVTT discriminators are excluded in the FL framework

Experimental Results

Method	Dice		Precision		Recall	
	Pen.	Core	Pen.	Core	Pen.	Core
CT	<u>0.7558</u> ± <u>0.01</u>	**0.7740 ± 0.06**	<u>0.7873</u> ± <u>0.03</u>	**0.7509 ± 0.07**	<u>0.7489</u> ± <u>0.03</u>	<u>0.7979</u> ± <u>0.05</u>
FL Exp.-1	0.7433 ± 0.02	0.7281 ± 0.06	0.7499 ± 0.01	0.6937 ± 0.17	0.7371 ± 0.02	0.8203 ± 0.07
FL Exp.-2	0.7507 ± 0.01	0.7380 ± 0.08	0.7735 ± 0.06	0.6987 ± 0.18	0.7345 ± 0.03	**0.8338 ± 0.08**
FL Exp.-3	**0.7713 ± 0.03**	<u>0.7720</u> ± <u>0.04</u>	**0.7875 ± 0.01**	<u>0.7448</u> ± <u>0.10</u>	**0.7581 ± 0.06**	0.8133 ± 0.03

Bold and <u>underline</u> specifies first and second best performance respectively.

Table: Evaluation results of the method with centralized (CT) setup and 3 different FL setups

Qualitative Results

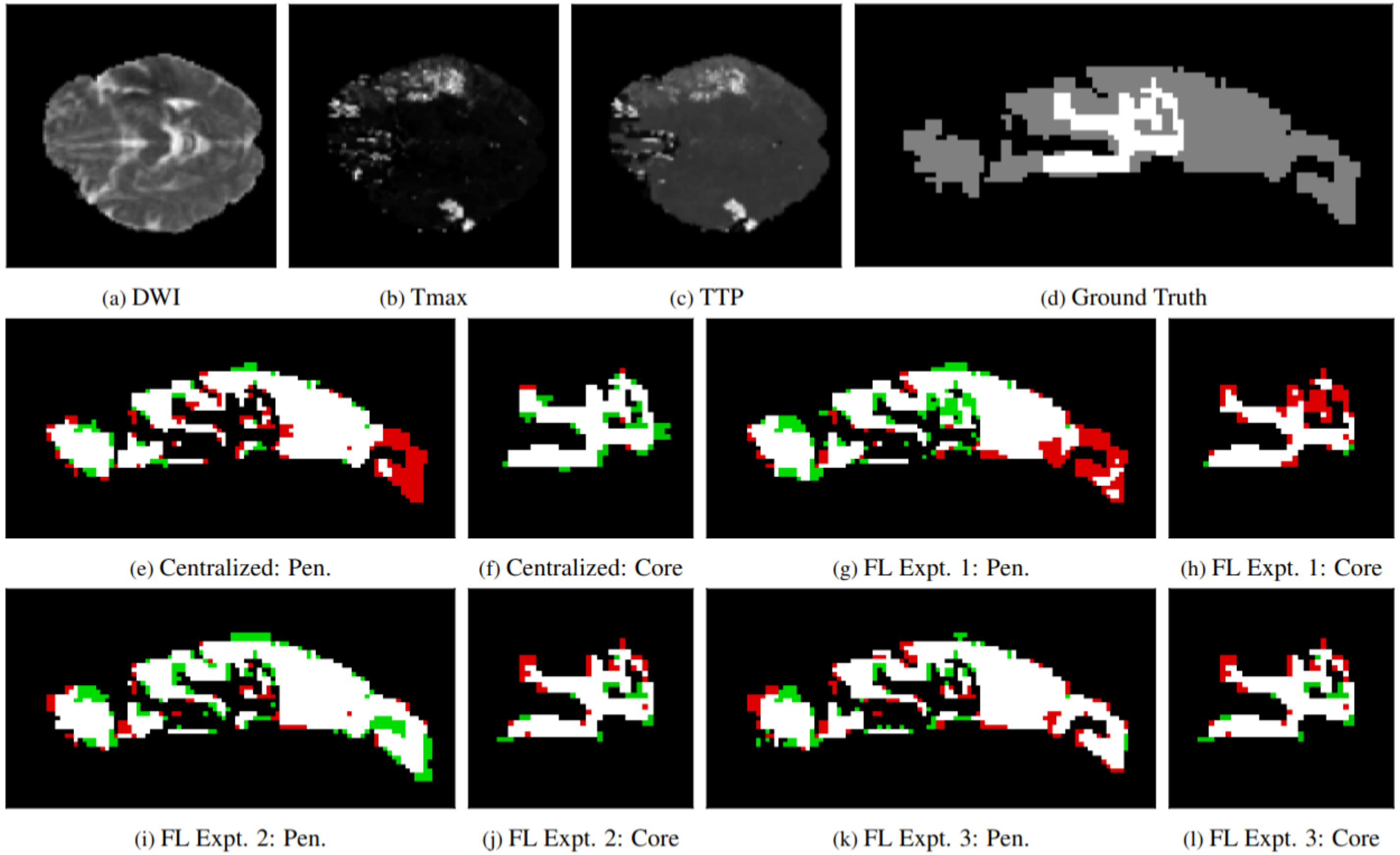


Fig: (a-c), Input Sequence GT, (d) gray for penumbra & white for core, (e-f) Centralized training (g-h) without rVTT, (i-j) rVTT included in FL setup and (k-l) rVTT excluded in FL setup

Communication Efficiency

- Let M is the total number of parameters
- Number of bits required are:
  - Vanilla Fed Avg: $2M$
  - FL Exp 2: $2(M + kN_D)$, where k discriminators are used each having N_D parameters
  - **FL Exp 3: $2M$**

Using discriminators locally in the FL framework not only gives better performance but also reduces a significant communication burden

Conclusion

- ✓ Proposed FANTOM to handle data and model specific challenge in distributed environment
- ✓ FANTOM gives the benefits of both kernel matching before aggregating along with FedAvg
- ✓ Handled kernel matching in CNNs
- ✓ Explored the effect of using adversarial mechanism in the FL framework
- ✓ Balance both the performance as well as communication burden