Enhancing Communication Efficiency and Robustness in Split-Federated Learning with Rate-Distortion inspired Compression

Chamani Shiranthika, Hadi Hadizadeh, Parvaneh Saeedi, Ivan V. Baji'c

SIMON FRASER UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

- SplitFed learning [1] → Federated Learning + Split Learning
- Communication challenges → High latency, bandwidth constraints, Synchronization overhead
- **SplitFedZip** → Employs rate-distortion inspired compression
- **SplitFedZip** → Preserves performance, reduced data transfer, enhanced communication efficiency & robustness

Contributions

- Factorized Prior (AE) [2] and Cheng_2020 [3] codecs
- FG - Both features and gradients compression, F - Features compression
- First rate-distortion inspired compression approach for SplitFed
- Medical image segmentation on blastocysts and skin lesions

SplitFedZip methodology

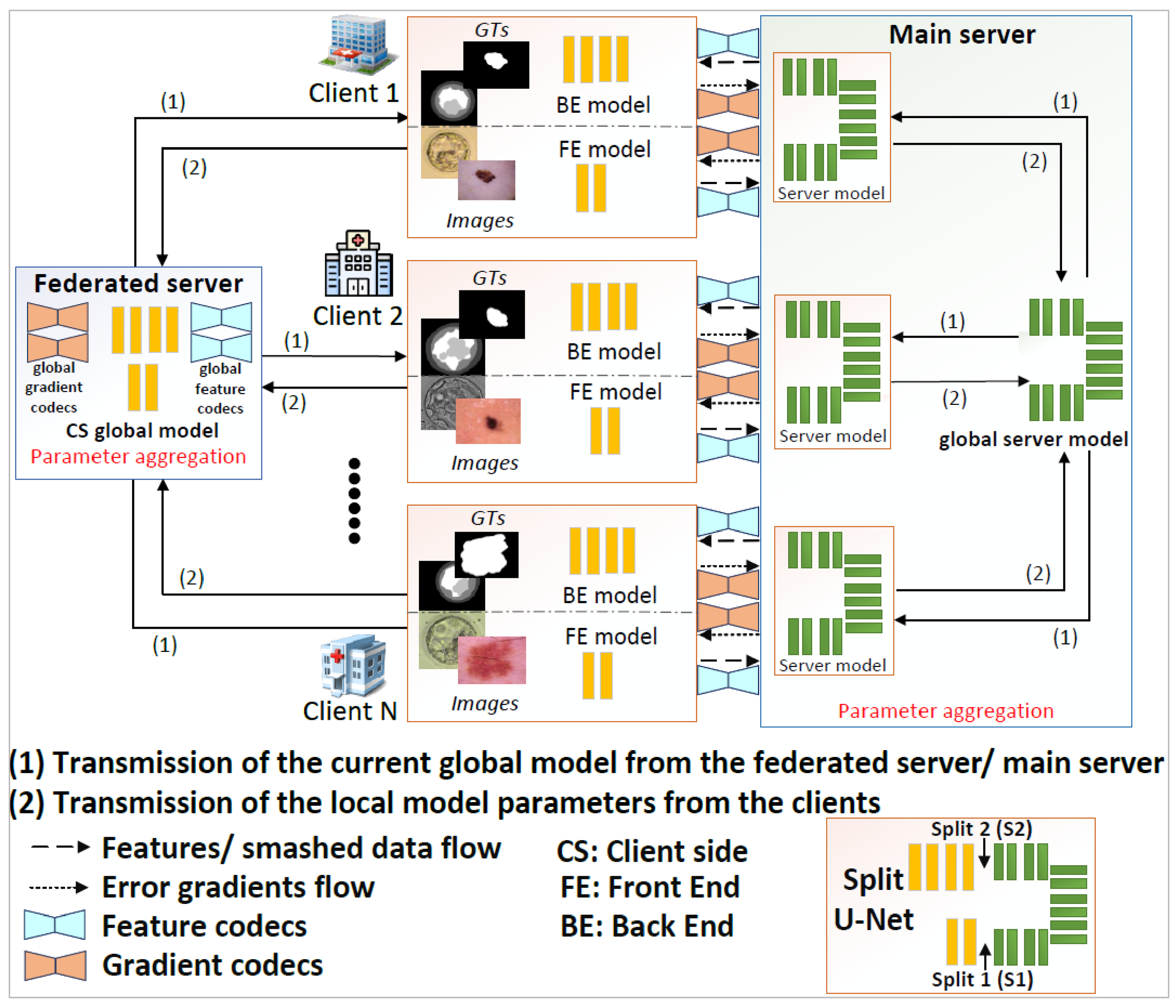


Figure 1: SplitFedZip network.

Each client's loss function:

$$L = L_r + \lambda \cdot \{L_{Dice} + L_{mse}\}$$

Loss function during FG:

$$L = \sum_{i=1,2} \left(L_r^{Si,F} + L_r^{Si,G}\right) + \lambda \cdot \left(\sum_{i=1,2} \left(L_{mse}^{Si,F} + L_{mse}^{Si,G}\right) + L_{Dice} \right)$$

Loss function during F:

$$L = \sum_{i=1,2} L_r^{Si,F} + \lambda \cdot \left(\sum_{i=1,2} L_{mse}^{Si,F} + L_{Dice} \right)$$

Loss function during no compression:

$$L_{Dice}$$

R-A curves

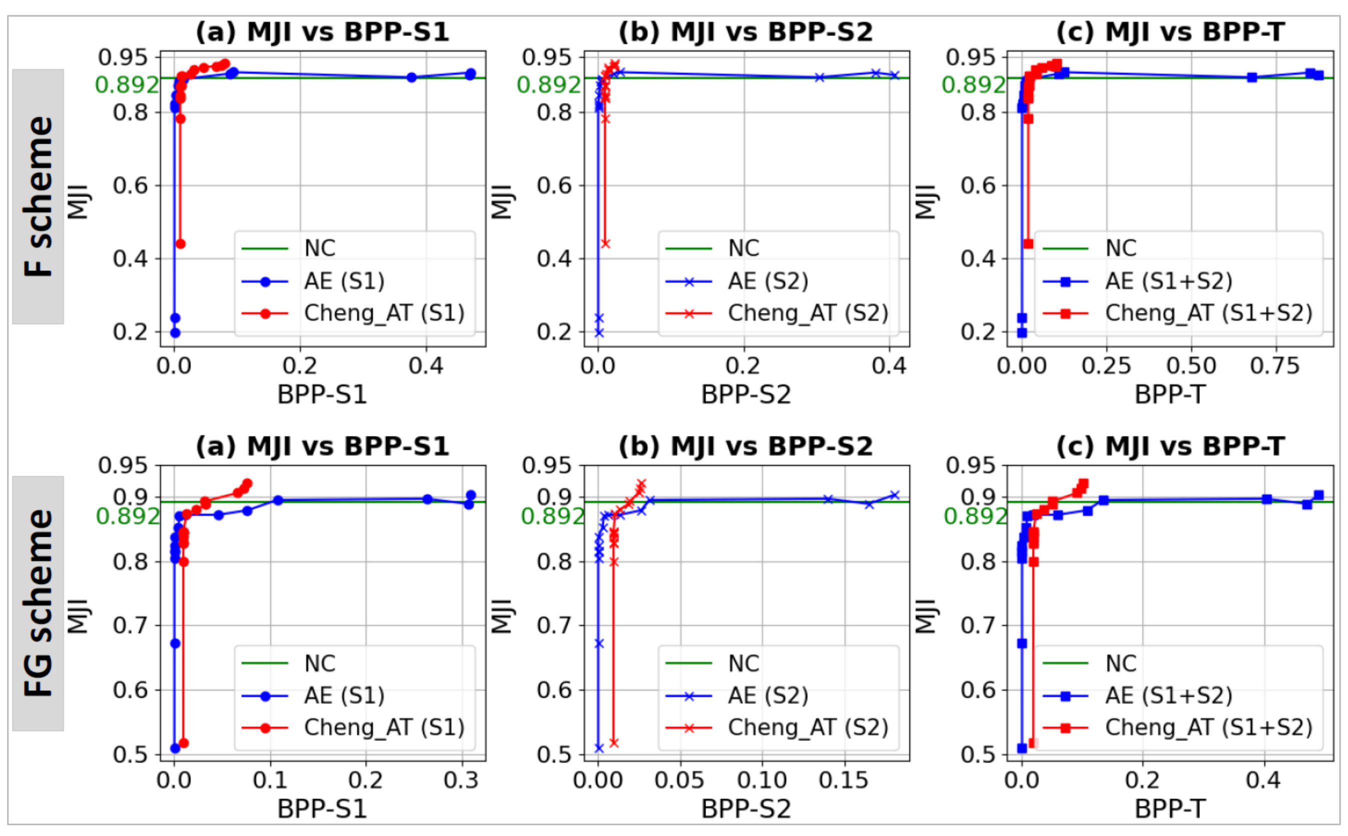


Figure 2: R-A curve for the HAM10K dataset.

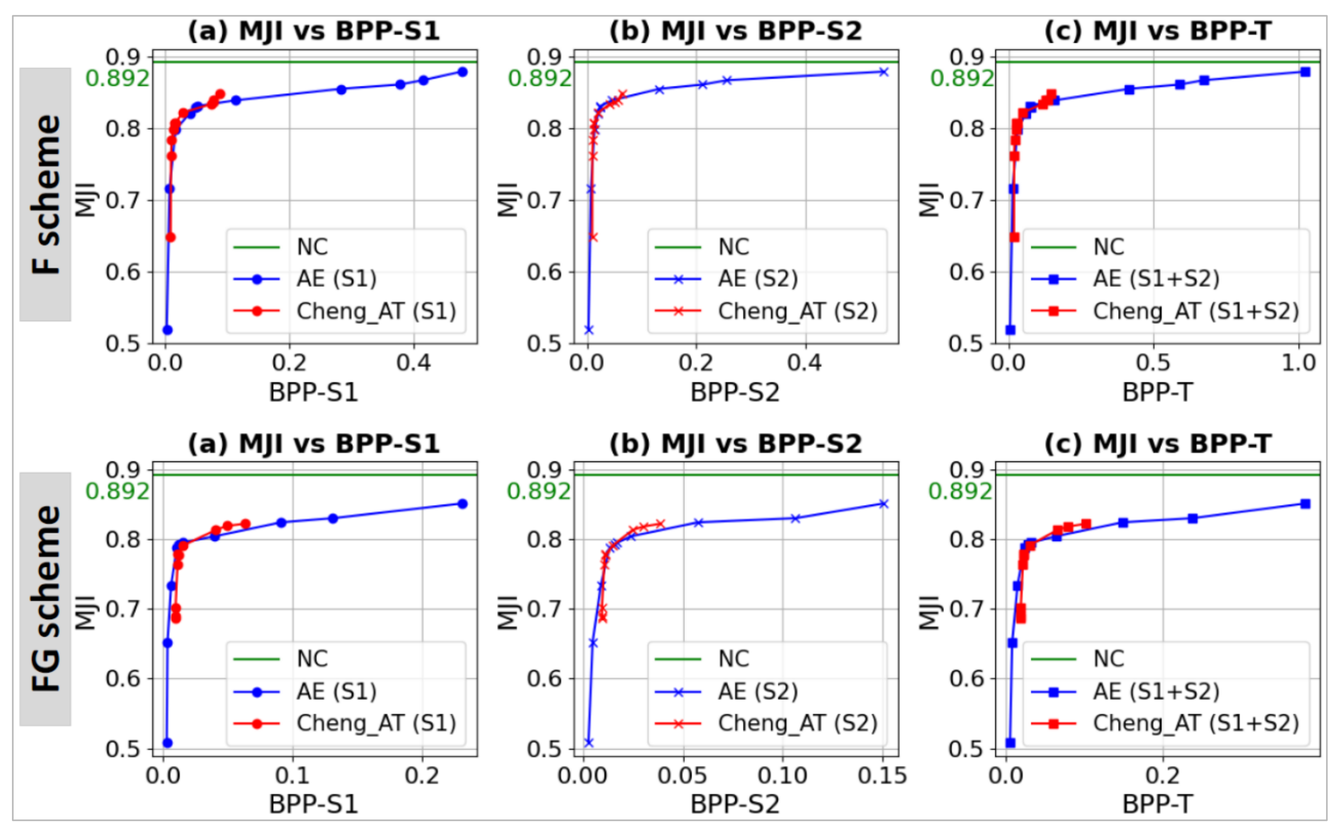


Figure 3: R-A curve for the Blastocyst dataset.

BPP =

Total bits in the compressed bitstream at S1 or S2 / input image resolution

CR =

(Original feature size at S1 or S2) / (BPP-T * input image resolution)

BPP analysis

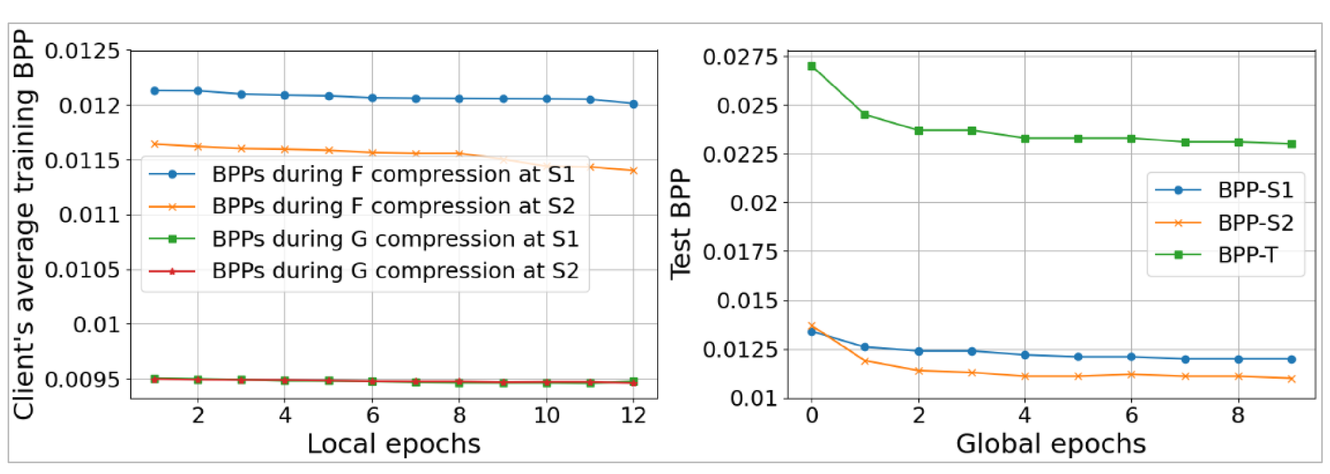


Figure 4: BPP analysis during SplitFed training for the blastocyst dataset..

CR analysis

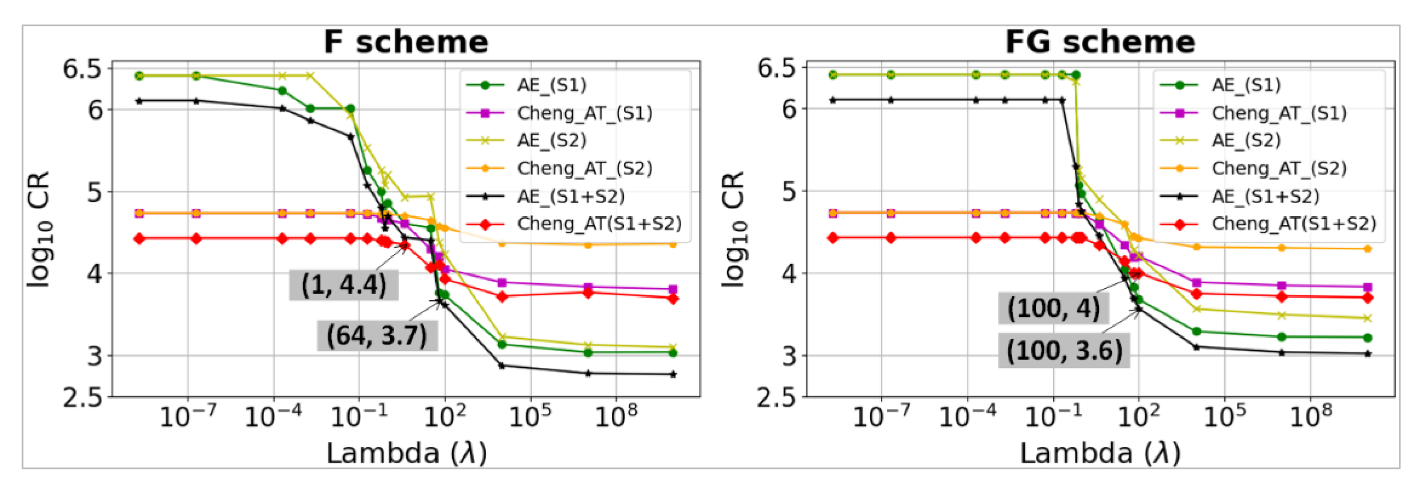


Figure 5: CR vs. λ for the HAM10K dataset.

Bjøntegaard Delta (BD) analysis

Table 1: BD-MJI and BD-BPP values.

Split point	F scheme		FG scheme	
	BD-MJI	BD-BPP	BD-MJI	BD-BPP
Blastocyst dataset (corresponds to Fig. 2)				
S1	0.004	–33.681	0.005	–99.195
S2	0.007	–2.418	0.007	–14.124
S1+S2	0.001	–26.634	0.004	–9.924
HAM10K dataset (corresponds to Fig. 3)				
S1	0.001	–62.019	0.010	–14.190
S2	0.001	–25.089	0.007	–26.663
S1+S2	0.002	–71.620	0.009	–25.007
WAvg based on testing dataset sizes of the two datasets				
S1+S2: WAvg	0.002	**–68.704**	0.009	**–24.029**

SOTA comparison

- Compared to AE in [4] → SplitFedZip offers better compression accuracy

M1 (pretrained AE used as non-trainable layers in the split network) → Two-stage training

M2 (AE trained for a few global epochs, then frozen) → Two-phase training

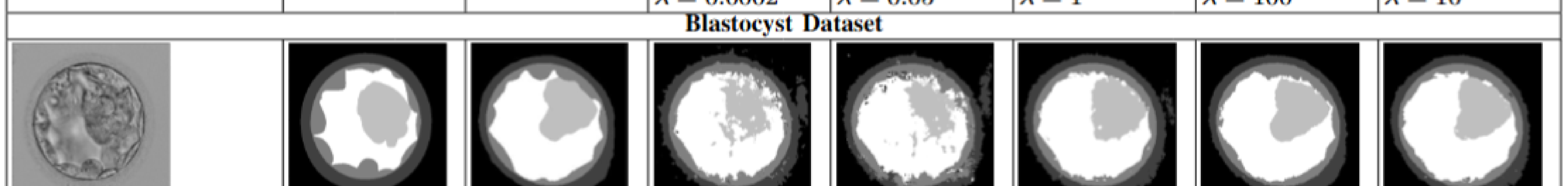
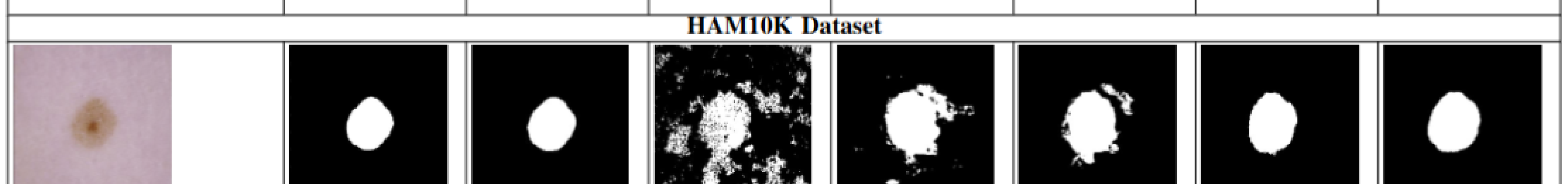
Table 2: MJI and DT: AE in [4].

gthres	Blastocyst Dataset				HAM10K Dataset			
	M1		M2		M1		M2	
	MJI	DT	MJI	DT	MJI	DT	MJI	DT
NC	0.892	40.6	0.892	40.6	0.892	514	0.892	4.0
inf	0.797	0.32	0.475	0.32	0.839	4.0	0.802	4.0
5	0.798	0.32	0.548	0.32	0.836	4.0	0.808	4.0
3	0.799	0.32	0.717	0.32	0.835	4.0	0.865	4.0
1.5	0.800	0.32	0.789	0.32	0.873	4.0	0.860	4.0
1.0	0.799	0.32	0.787	0.32	**0.876**	4.0	0.860	4.0
0.8	0.800	0.32	0.796	0.39	0.874	4.0	0.860	4.0
0.5	**0.801**	0.32	0.790	0.63	0.838	4.0	0.876	4.0
0.25	0.799	0.32	**0.817**	0.63	0.841	4.0	0.873	4.0
0.2	0.799	0.32	0.807	0.63	0.868	4.0	0.869	4.1
0.15	0.798	0.32	0.800	0.63	0.869	4.0	0.876	4.1
0.10	0.798	0.41	0.812	0.63	0.844	4.0	0.879	4.3
0.08	0.800	0.52	0.808	0.63	0.823	4.1	0.870	4.4
0.05	0.795	0.63	0.816	0.63	0.828	4.2	0.883	5.0
0	0.797	0.63	0.812	0.63	0.857	8.0	**0.886**	8.0

Table 3: MJI and DT: SplitFedZip's AE.

λ	Blastocyst Dataset				HAM10K Dataset			
	Two-stage		Two-phase		Two-stage		Two-phase	
	MJI	DT	MJI	DT	MJI	DT	MJI	DT
NC	NA	NA	NA	NA	NA	NA	NA	NA
10^{10}	0.878	1.04	0.845	0.07	0.894	0.8	**0.907**	0.6
100	0.869	0.5	**0.847**	0.04	**0.898**	0.3	0.899	0.12
64	**0.884**	0.4	0.846	0.03	0.896	0.27	0.892	0.09
32	0.877	0.4	0.823	0.03	0.893	0.06	0.882	0.04
16	0.869	0.3	0.822	0.02	0.891	0.06	0.894	0.03
4	0.806	0.1	0.805	0.009	0.894	0.02	0.868	0.02
1	0.810	0.03	0.778	0.008	0.889	0.02	0.869	0.009
0.8	0.804	0.03	0.769	0.008	0.882	0.02	0.864	0.008
0.6	0.799	0.02	0.775	0.005	0.880	0.0008	0.835	0.003
0.2	0.788	0.02	0.711	0.003	0.839	0.0004	0.831	0.002
0.05	0.755	0.018	0.624	0.003	0.834	0.0004	0.827	0.0009
0.002	0.607	0.017	0.512	0.003	0.832	0.0004	0.823	0.0005
0.0002	0.485	0.005	0.497	0.003	0.808	0.0004	0.820	0.0004

Qualitative comparison

Image	Ground Truth (GT)	NC scheme prediction	SplitFedZip prediction (utilizing AE in the FG scheme)				
			$\lambda = 0.0002$	$\lambda = 0.05$	$\lambda = 1$	$\lambda = 100$	$\lambda = 10^{10}$
Blastocyst Dataset							
							
182.BMP		0.892	0.474	0.652	0.795	0.839	0.843
HAM10K Dataset							
							
ISIC-0025193.JPG		0.892	0.663	0.804	0.871	0.895	0.903

References

[1] Thapa, C.; Arachchige, P. C. M.; Camtepe, S.; and Sun,L. 2022. "SplitFed: When Federated Learning Meets SplitLearning". In Proc. AAAI, volume 36, 8485–8493.

[2] Ball´e, J.; Minnen, D.; Singh, S.; Hwang, S. J.; and Johnston,N. 2018. "Variational image compression with a scale hyper-prior". In Proc. ICLR.

[3] Cheng, Z.; Sun, H.; Takeuchi, M.; and Katto, J. 2020. "Learned image compression with discretized gaussian mixture likelihoods and attention modules". In Proc. CVPR,7939–7948.

[4] Ayad, A.; Renner, M.; and Schmeink, A. 2021. "Improving the communication and computation efficiency of split learning for iot applications". In Proc. IEEE GLOBECOM,01–06.