QIFK2023 **123**

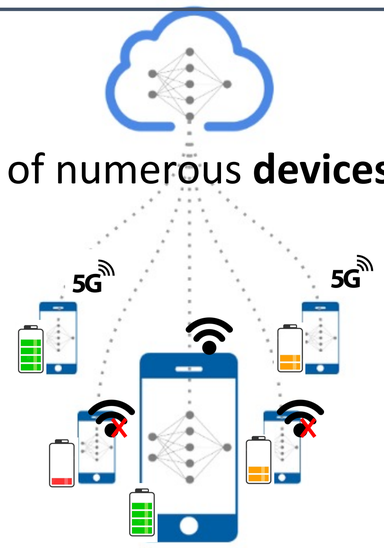
NeFL: Nested Federated Learning for Heterogeneous Clients

Honggu Kang, Seohyeon Cha, Jinwoo Shin, Jongmyeong Lee and Joonhyuk Kang

School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

Federated Learning & On-device AI

- Federated Learning
  - Employing **computing resources & data** of numerous **devices**
  - AI is being employed on devices
- System Heterogeneity
  - Computation (CPU/GPU/ASIC)
  - Communication (4G/5G/Wi-Fi)
  - Battery (low/high)
  - Memory (trust-zone)

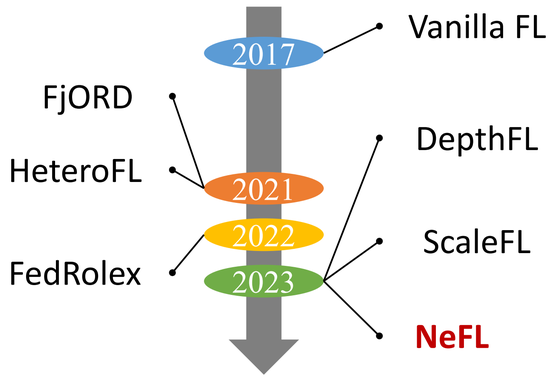


Model Scaling Approaches

Widthwise scaling
- FjORD/HeteroFL/FedRolex

Depthwise scaling
- DepthFL

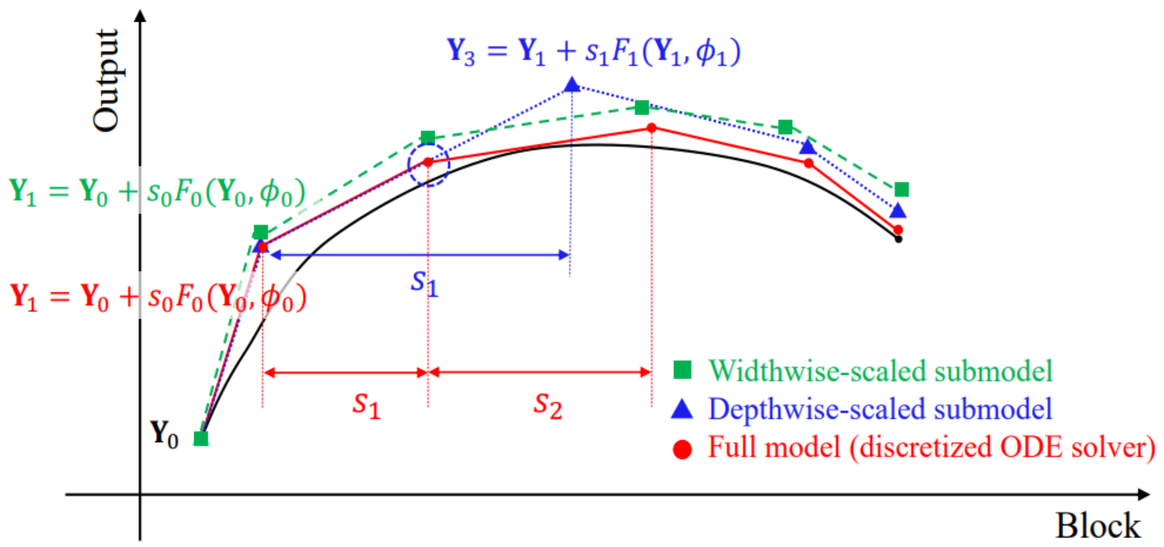
Widthwise & depthwise scaling
- ScaleFL/**NeFL (ours)**

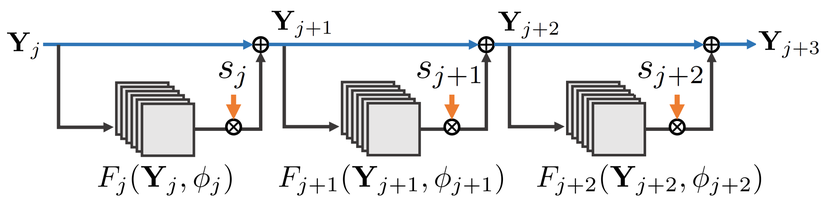


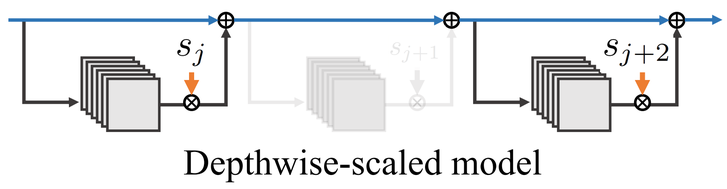
Contributions

- We propose a **general model scaling method** employing the concept of ODE solver to deal with the **system heterogeneity**
- We propose a method for **parameter averaging** across generally scaled submodels
- We evaluate the performance of NeFL through a series of experiments and verify the applicability of NeFL over recent studies (**pre-training & statistical heterogeneity**)

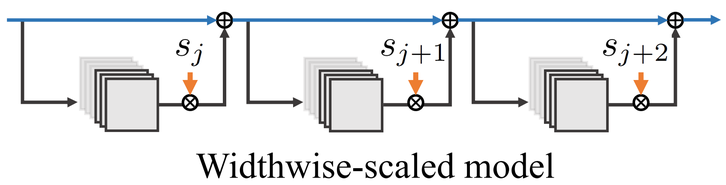
ODE Interpretation for Model Scaling





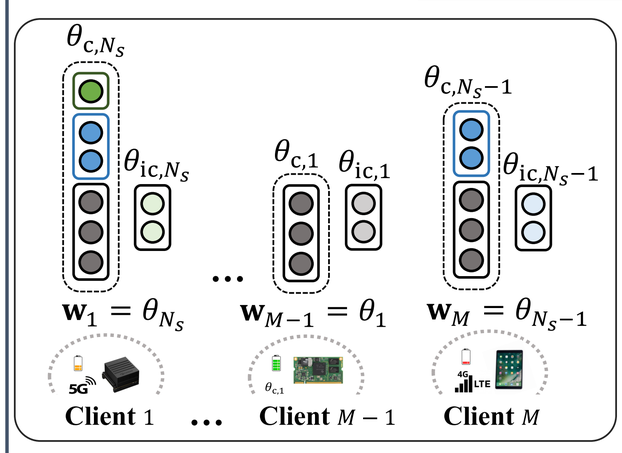


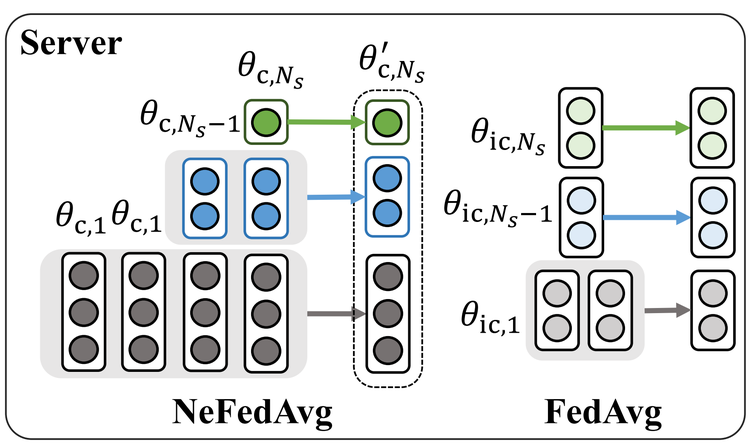
Depthwise-scaled model



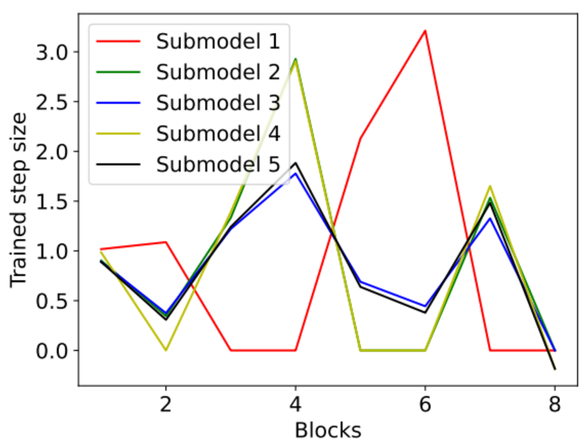
Widthwise-scaled model

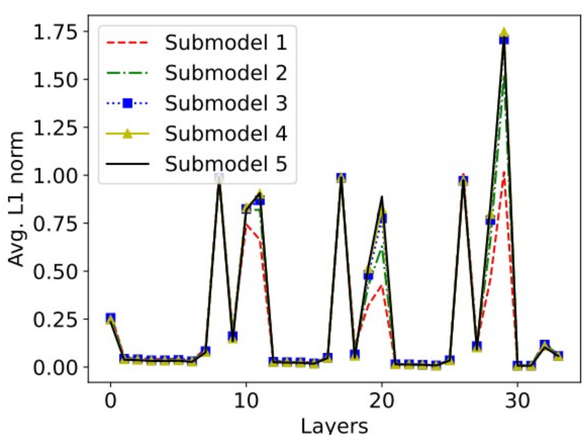
Parameter Averaging





Inconsistent/consistent Parameters





Results

Model	Method	IID		non-IID	
		Worst	Avg	Worst	Avg
ResNet18	HeteroFL	80.62 (± 0.24)	84.26 (± 1.95)	76.25 (± 1.05)	80.11 (± 2.03)
	FjORD	85.12 (± 0.22)	87.32 (± 1.21)	75.81 (± 5.65)	77.99 (± 6.50)
	DepthFL	64.80 (± 10.49)	82.44 (± 10.17)	59.61 (± 5.16)	76.89 (± 9.60)
	NeFL (ours)	**86.86 (± 0.22)**	**87.88 (± 0.68)**	**81.26 (± 2.44)**	**81.71 (± 3.14)**

Model	Method	IID		non-IID	
		Worst	Avg	Worst	Avg
Pre-trained ResNet18	HeteroFL	78.26 (± 0.15)	84.48 (± 3.04)	71.95 (± 1.32)	76.17 (± 3.39)
	FjORD	86.37 (± 0.18)	88.91 (± 1.37)	81.81 (± 1.10)	81.96 (± 5.76)
	DepthFL	47.76 (± 8.54)	82.86 (± 17.98)	39.78 (± 3.74)	67.71 (± 16.88)
	NeFL (ours)	**88.61 (± 0.08)**	**89.60 (± 0.70)**	**82.91 (± 0.47)**	**85.85 (± 2.43)**

Model	Param. #	IID		non-IID	
		Worst	Avg	Worst	Avg
Pre-trained ViT	86.4M	93.02 (± 0.06)	95.96 (± 2.10)	87.56 (± 0.16)	92.74 (± 3.95)
Pre-trained Wide ResNet101	124.8M	90.9 (± 0.16)	91.35 (± 0.39)	87.17 (± 0.04)	87.74 (± 1.06)

	Depthwise	Widthwise	Step size
DepthFL	✓		
FjORD / HeteroFL		✓	
NeFL-D	✓		✓
NeFL-W		✓	✓
NeFL-WD	✓	✓	✓

Method	Worst	Avg.
HeteroFL	80.62	84.26
FjORD	85.12	87.32
NeFL-W (ours)	**85.13**	**87.36**

Method	Worst	Avg.
DepthFL	64.80	82.44
NeFL-D (ours)	**86.06**	**87.94**

Method	Worst	Avg.
NeFL-WD (ours)	**86.86**	**87.88**

Model	Metric	Method		
		Width/Depthwise scaling	Widthwise scaling	Depthwise scaling
ResNet18	Param #	6.71M	6.71M	6.68M
	FLOPs	87.8M	85M	102M