SuperFedNAS: Cost-Efficient Federated Neural Architecture Search for On-Device Inference

Georgia Institute of Technology · SAIL · Cisco

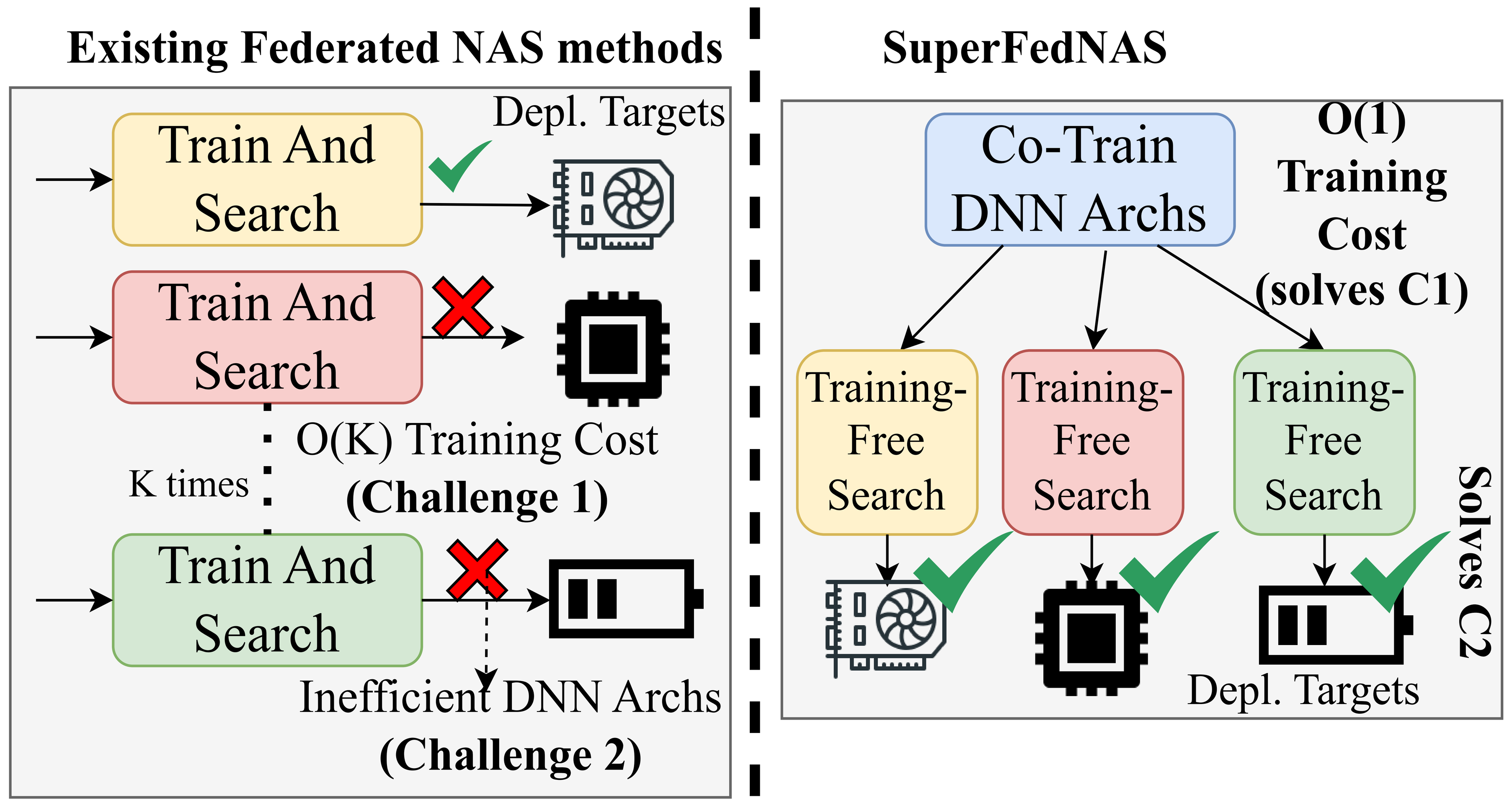
Alind Khare[1], Animesh Agrawal[1], Aditya Annavajjala[1], Payman Behnam[1], Myungjin Lee[2], Hugo Latapie[2], Alexey Tumanov[1]

[1]Georgia Tech & [2]Cisco Research

Challenges in Federated NAS

C1: Prohibitively Expensive to Satisfy Multiple Inference Deployments

C2: Don't Produce Optimal DNNs that satisfy deployment targets



The contributions are:

1. SuperFedNAS: A federated NAS method that searches/trains rich diversity of DNN archs for efficient inference.
2. Satisfies k deployment targets in O(1) cost. Decouples train and search in federated NAS.
3. Maxnet: A novel training algorithm optimizes a novel objective to train supernets in FL and reduce interference.

Related Works

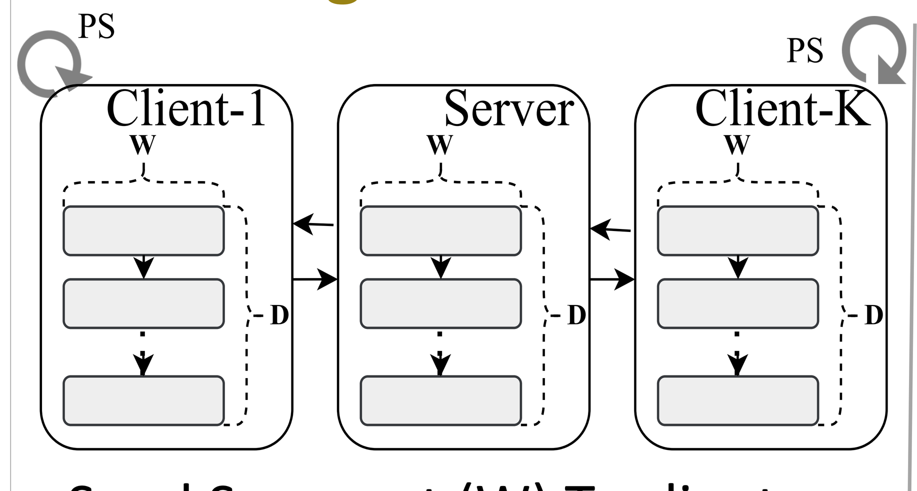
Feature	FedNAS	FedPNAS	ScaleFL	Inco.	SuperFedNAS
Weight-Sharing					✓
Uses NAS	✓	✓			✓
Training Cost for N depl.	O(N)	O(N)	O(N)	O(N)	O(1)
Satisfies Depl. @ Inference					✓

1. FedNAS: Federated Deep Learning via Neural Architecture Search. CVPR'20 Workshop. He et al.
2. FedPNAS: Personalized Neural Architecture Search for Federated Learning. Hoang et al.
3. ScaleFL: Resource-Adaptive Federated Learning with Heterogeneous Clients. CVPR'23. Ilhan et al.
4. Inco.: Internal Cross-Layer Gradients For Extending Homogeneity to Heterogeneity in FL. ICLR'23. Chan et al.
5. Once-for-All: Train One Network & Specialize it For Efficient Deployment. ICLR'20. Cai et al.

Federated Training of Supernet

Naïve Approaches

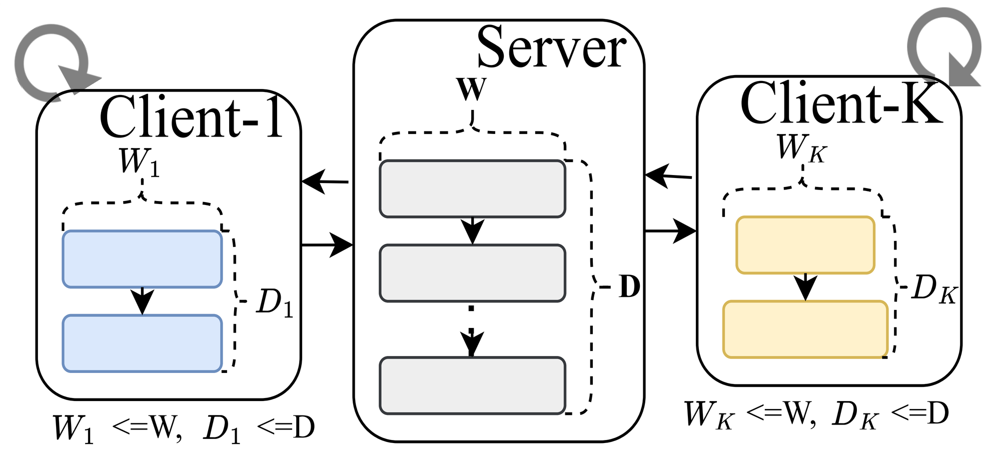
Multi-Stage



Send Supernet (W) To clients

1. Clients use Progressive Shrinking (OFA ICLR'20) to train Supernets locally.
2. Aggregation is FedAvg of Supernet Weights (W)

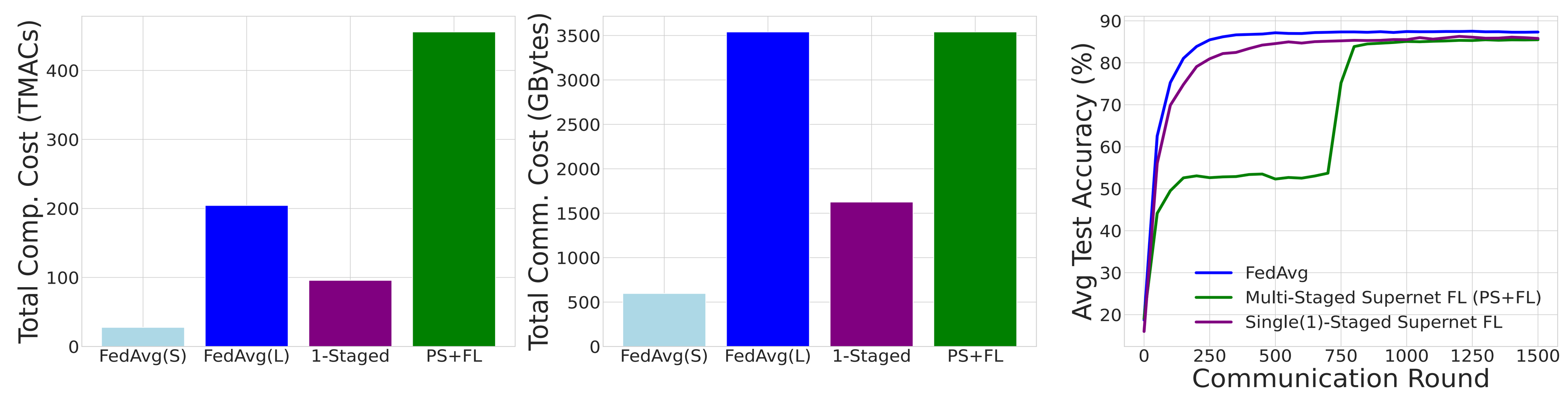
Single-Stage



Send Subnets $\mathcal{G}(W, \alpha_i) \sqsubset W$ To clients

1. Sampling and Assignment of subnets to clients is random.
2. Aggregation with parameter-wise cardinal averaging based on overlap of $\mathcal{G}(W, \alpha_i)$.

Naïve Approaches vs FedAvg



Takeaways

1. Multi-Stage Supernet FL Training has high training costs.
2. Both naïve methods suffer from slow convergence and sub-optimal accuracy due to interference. Don't Solve **C1 and C2.**

Maxnet: Novel Supernet FL-Training Technique

Single-Stage Paradigm (Solves C1). Optimizes a novel objective that:

1. Improves worst-performing DNN arch on each data partition.
2. Samples DNN archs based on overlap $\mathcal{G}(W, \alpha_i) \sqsubset W$ & DNN arch. Loss

DNN Arch. $\in \mathcal{A}$	Method	Test Accuracy (%)		
		non-iid=100	non-iid=1	non-iid=0.1
Smallest	FedAvg	85.25 ± 0.46	83.42 ± 0.19	77.15 ± 2.5
	Single-Staged Supernet FL	84.6 ± 0.19	83.17 ± 0.12	76.28 ± 1.31
	Multi-Staged Supernet FL	84.53 ± 0.58	82.82 ± 0.34	76.26 ± 2.35
	MaxNet	$\mathbf{89.42 \pm 0.11}$	$\mathbf{88.69 \pm 0.2}$	$\mathbf{81.81 \pm 1.59}$
Largest	FedAvg	89.44 ± 0.67	87.88 ± 0.7	81.24 ± 1.99
	Single-Staged Supernet FL	87.14 ± 0.2	86.03 ± 0.26	80.02 ± 2.07
	Multi-Staged Supernet FL	86.45 ± 0.53	85.02 ± 0.32	78.57 ± 2.48
	MaxNet	$\mathbf{91.34 \pm 0.3}$	$\mathbf{90.91 \pm 0.15}$	$\mathbf{84.72 \pm 1.78}$

Evaluation

Image Datasets

Billion MACs	Method	Test Accuracy (%)		
		CIFAR10	CIFAR100	CINIC10
0.45-0.95	FedAvg	85.25 ± 0.46	43.19 ± 0.54	61.76 ± 0.78
	FedNAS	77.33 ± 0.31	40.92 ± 2.21	58.15 ± 0.18
	FedPNAS	88.83 ± 0.5	45.77 ± 0.68	64.3 ± 0.98
	SuperFedNAS	$\mathbf{89.42 \pm 0.11}$	$\mathbf{56.35 \pm 0.3}$	$\mathbf{73.12 \pm 0.77}$
0.95-1.45	FedAvg	86.36 ± 0.22	43.92 ± 0.57	63 ± 0.17
	FedPNAS	89.27 ± 0.51	47.8 ± 26	66.74 ± 0.32
	SuperFedNAS	$\mathbf{90.22 \pm 0.31}$	$\mathbf{57.16 \pm 0.23}$	$\mathbf{74.5 \pm 0.74}$
1.45-2.45	FedAvg	87.59 ± 0.27	44.4 ± 0.56	64 ± 0.07
	FedNAS	86.41 ± 0.1	55.82 ± 0.29	69.97 ± 0.27
	SuperFedNAS	$\mathbf{90.93 \pm 0.23}$	$\mathbf{57.85 \pm 0.31}$	$\mathbf{75.08 \pm 0.7}$
2.45-3.75	FedAvg	89.44 ± 0.67	45 ± 0.27	66.02 ± 0.13
	FedNAS	89.43 ± 0.36	58.39 ± 0.23	71.93 ± 0.13
	SuperFedNAS	$\mathbf{91.34 \pm 0.3}$	$\mathbf{58.25 \pm 0.39}$	$\mathbf{75.38 \pm 0.73}$

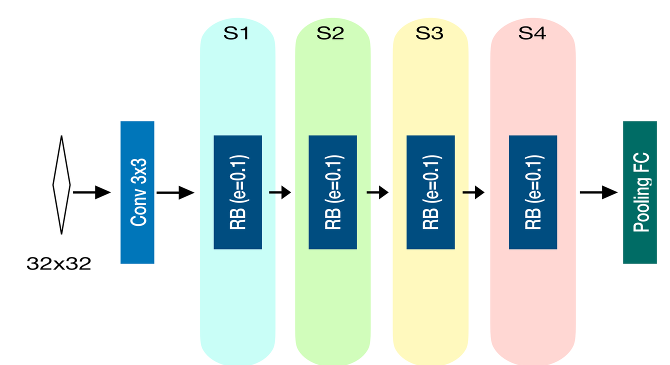
Text Dataset

Million MACs	Method	Test Accuracy (%)
0-0.5	FedAvg	48.52 ± 0.11
	SuperFedNAS	48.22 ± 0.27
0.5-1	FedAvg	49.17 ± 0.02
	SuperFedNAS	$\mathbf{49.81 \pm 0.16}$
1-1.5	FedAvg	51.94 ± 0.03
	SuperFedNAS	$\mathbf{53.26 \pm 0.06}$
1.5-2.75	FedAvg	53.48 ± 0.09
	SuperFedNAS	$\mathbf{54.59 \pm 0.15}$
2.75-4.0	FedAvg	53.62 ± 0.1
	SuperFedNAS	$\mathbf{54.61 \pm 0.13}$

Tough FL Setting.

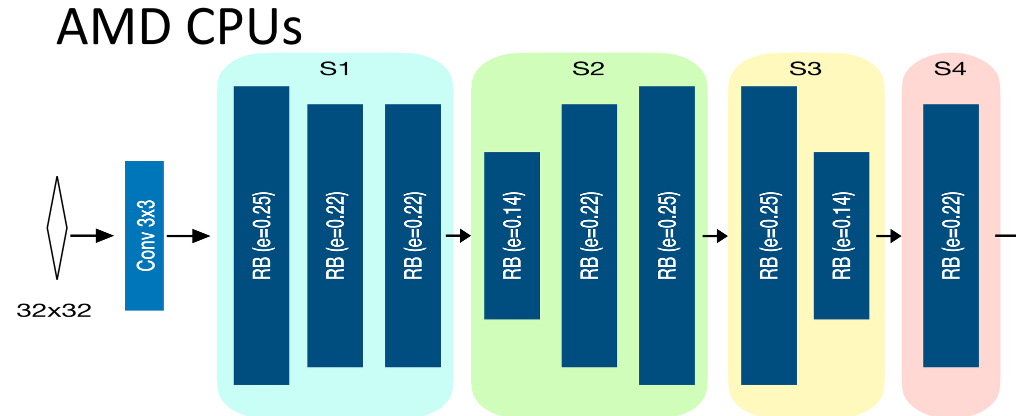
1. Shakespeare dataset (LEAF benchmark)
2. Non-iid, & 660 clients

Specialized DNNs for Target Depl.



112 ms latency target on AMD CPUs

- SuperFedNAS specializes DNN archs for different deployment targets
- **Solves C2**
- Wide/Deep DNN for GPU
- Thin/Shallow DNN for CPU



18 ms latency target on NvidiaRTX 2080 Ti