An In-Depth Analysis of Distributed Training of Deep Neural Networks

IEEE IPDPS'21
May 20, 2021

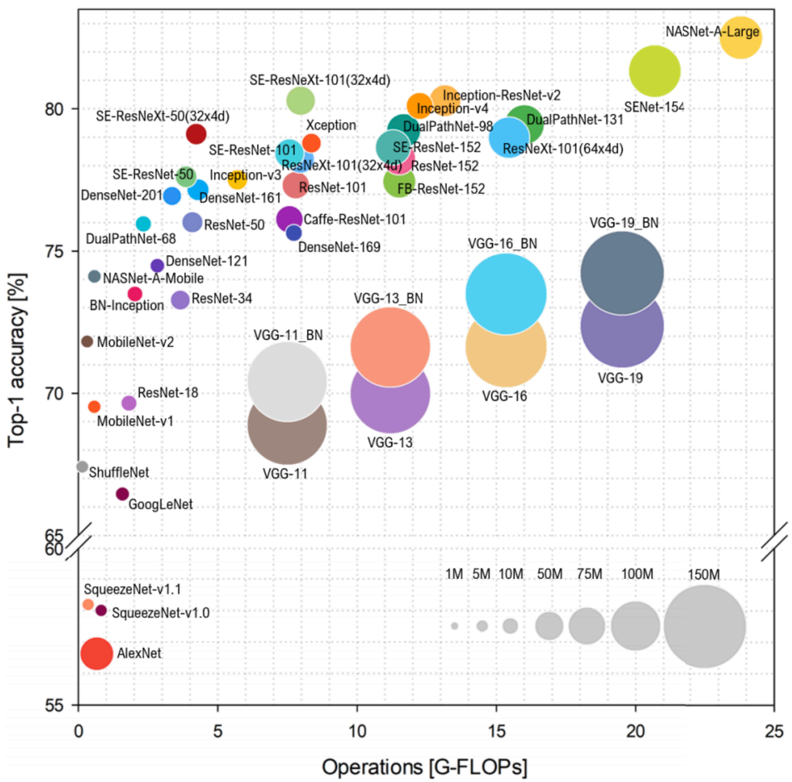
Yunyong Ko, Kibong Choi, Jiwon Seo, and Sang-Wook Kim
Department of Compute Science
Hanyang University, Republic of Korea

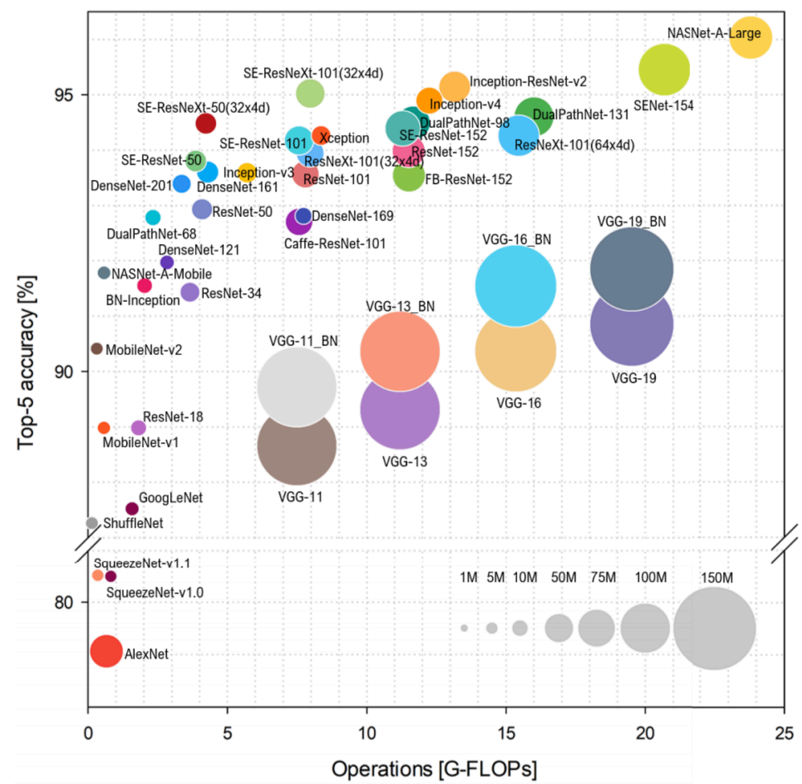
Table of Contents

Background

☐ Training of DNNs requires massive time

- DNN models are becoming more complicated
  - ☐ VGG (150M params.), BERT (345M params.), GPT-3 (175B params.), …
  - ☐ Ex) Training BERT with 8 NVIDIA V100 GPUs takes 2 weeks

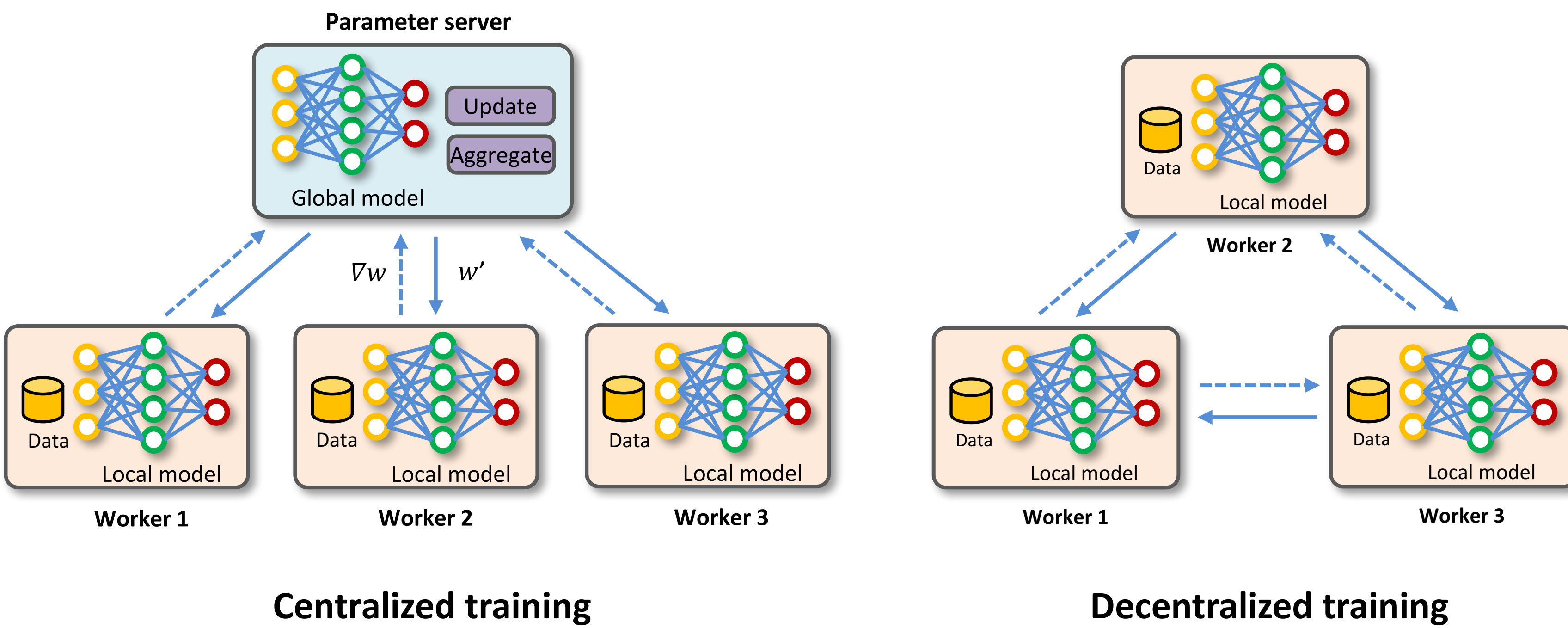




Distributed Deep Learning

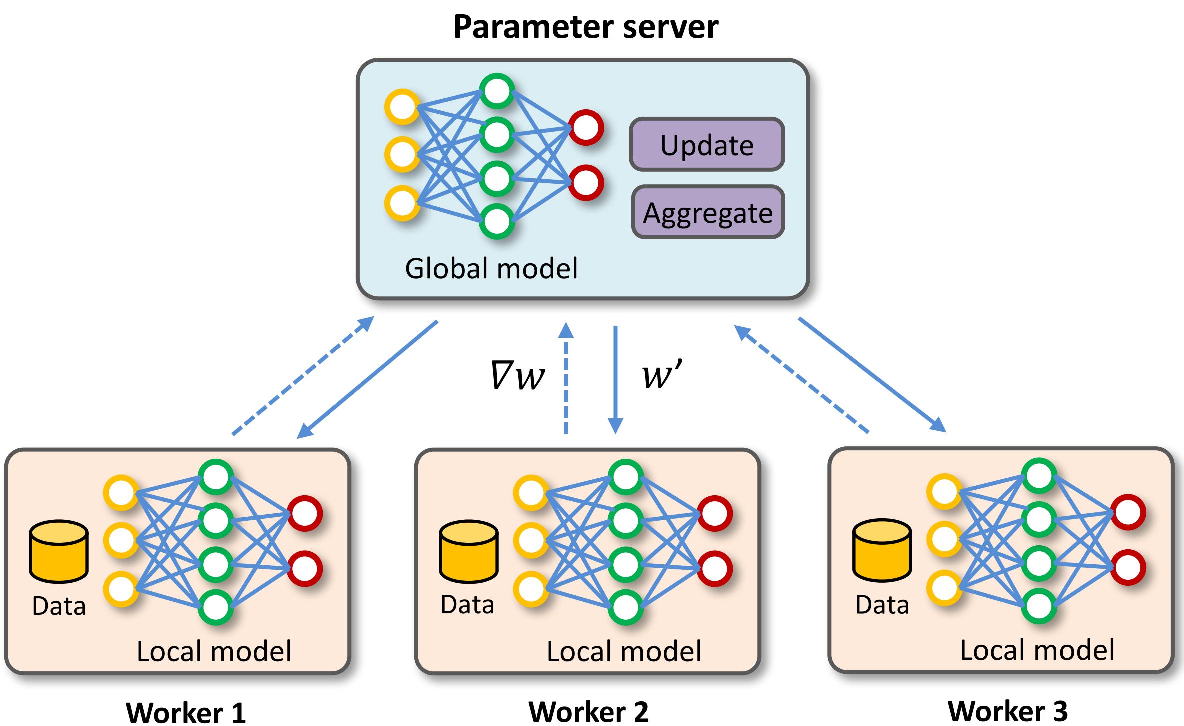
Train a DNN model using multiple workers

- Each worker trains its local model based on its local data
- The training results are aggregated via communication



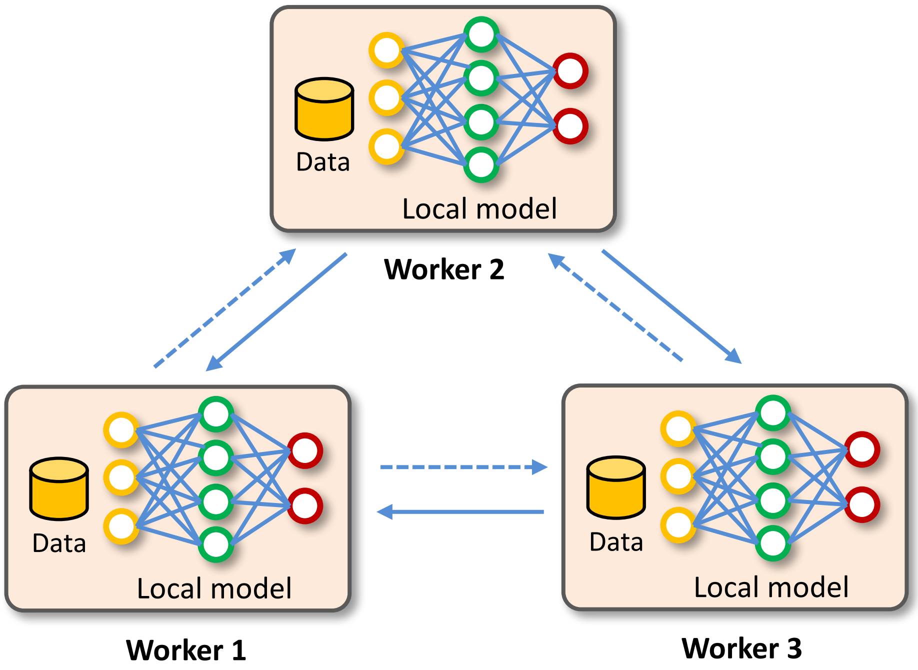
Centralized Training

- ***Parameter server (PS)*** **aggregates the training results from workers and manages the global model**
  - There is explicit global model in PS
  - PS can be a bottleneck of the training



Decentralized Training

- **The training results of workers are aggregated via** ***peer-to-peer communication***
  - To avoid the problem of *PS being bottleneck*
  - There is no explicit global model



Goals of This Work

Motivation

- Trade-off (Performance vs Accuracy) depends on
  - Models, # of workers, computing power of GPUs, network BW, …

Goals

- Evaluate existing training algorithms in a fair way
- Conduct comprehensive analysis

Evaluation & Analysis

Algorithms for Evaluation

Distributed Training Algorithms		
	Centralized	**Decentralized**
Synchronous	Bulk synchronous parallel SGD (BSP) (*Gerbessiotis et al., JPDC'94*)	Allreduce SGD (AR-SGD) (*Goyal et al., arXiv'17*)
Asynchronous	Asynchronous parallel SGD (ASP) (*Recht et al., NeurIPS'12*)	Gossip-based SGD (GoSGD) (*Blot et al., arXiv'16*)
	Stale synchronous parallel SGD (SSP) (*Ho et al., NeurIPS'13*)	Asynchronous decentralized parallel SGD (AD-PSGD) (*Lian et al., ICML'18*)
	Elastic averaging SDG (EASGD) (*Zhang et al., NeurIPS'15*)	

Evaluation Aspects

- Model Accuracy
- Hyperparameter Sensitivity
- Scalability
- Effects of Optimization Techniques

Experiments setup

☐ Models

- ResNet-50: 25M parameters (**computation**-intensive)
- VGG-16: 128M parameters (**communication**-intensive)

☐ Dataset

- ImageNet-1K: 1.28M training images and 5K test images

☐ Software

- DL framework: TensorFlow 1.12
- Communication library: MPICH 3.1.4

☐ System resources

- CPU: Intel Xeon CPU E5-2698 v4 (with 256 GB memory)
- GPU (worker): NVIDIA Titan V (24 GPUs in total)
  - 14.90 TFLOPS, 12GB memory
- Network bandwidth (BW)
  - 10Gbps Ethernet (low BW) and 56Gbps InfiniBand (high BW)

Interesting Findings

1. **Intermittent communication to reduce the aggregation overhead has highly negative impact on the model accuracy**

2. **Centralized training has a strength in model accuracy, compared to decentralized training**

3. **Centralized training algorithms suffer from the problem of PS being a bottleneck**

F1. Intermittent Communication

1. **Intermittent communication to reduce the aggregation overhead has highly negative impact on the model accuracy**
   - It causes large variance among parameters of workers
     - In EASGD and GoSGD, significant loss occurs in model accuracy
   - The accuracies of ASP and AD-PSGD are comparable to that of BSP
     - Aggregating the training results (gradients/parameters) of all workers at every iteration

TABLE III
TEST ACCURACY OF ASYNCHRONOUS ALGORITHMS FOR RESNET-50 ON IMAGENET-1K.

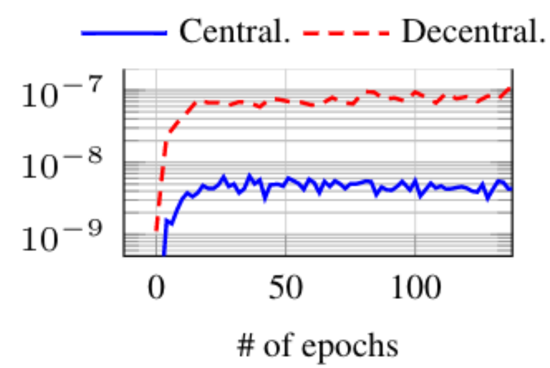
	BSP	**ASP**	**SSP**		**EASGD**		**GoSGD**			**AD-PSGD**
workers	-	-	$s = 3$	$s = 10$	$\tau = 4$	$\tau = 8$	$p = 1$	$p = 0.1$	$p = 0.01$	-
4	0.7514	0.7508	0.7480	0.7462	0.7028	0.7027	0.7160	0.6892	0.6775	0.7483
8	0.7509	0.7482	0.7450	0.7412	0.6357	0.6269	0.6529	0.6173	0.5845	0.7447
16	0.7496	0.7447	0.7393	0.7147	0.5416	0.5237	0.5492	0.5135	0.4922	0.7439
24	0.7511	0.7459	0.7282	0.6448	0.4709	0.4528	0.4641	0.4475	0.3938	0.7411

F2. Strength of Centralized Training

2. Centralized training has a strength in model accuracy, compared to decentralized training

- ASP always outperforms AD-PSGD in model accuracy
  - The difference tends to get larger as the number of workers increases
  - (Aggregation via PS)vs (Aggregation via peer-to-peer communication)

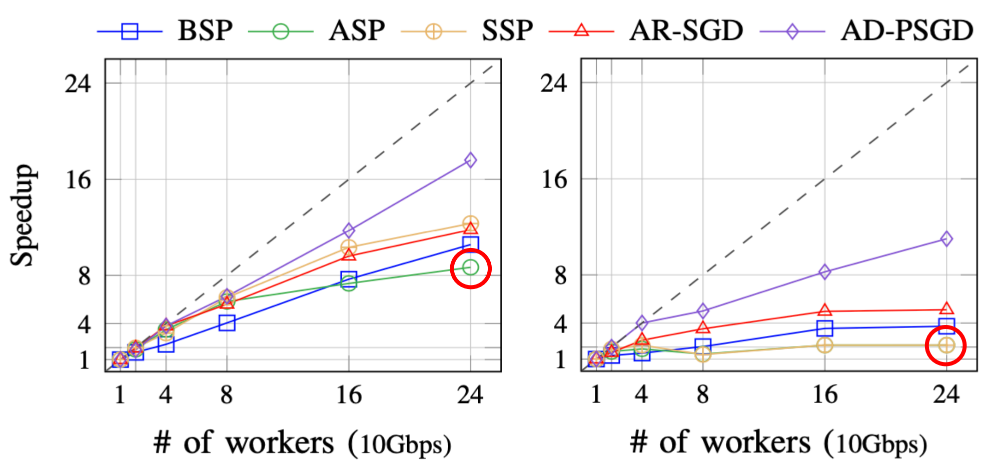
	ASP	AD-PSGD
4 workers	0.7508 (0.25%)	0.7483
8 workers	0.7482 (0.35%)	0.7447
16 workers	0.7447 (0.08%)	0.7439
24 workers	0.7459 (**0.48**%)	0.7411



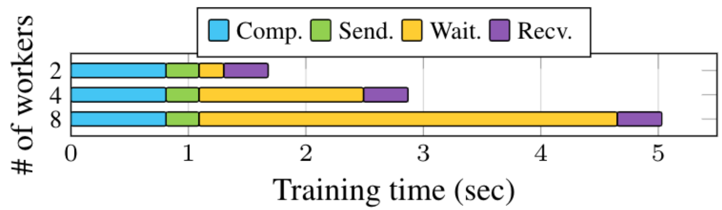
F3. Problem of Centralized Training

3. Centralized training algorithms suffer from the problem of PS being a bottleneck

- The '*waiting time*' increases as the number of workers increases
  - All centralized algorithms (BSP, ASP, SSP) show poor scalability
  - ASP shows always the worst speed-up results
- Decentralized training successfully avoids the communication bottleneck



(a) ResNet-50 (b) VGG-16



Breakdown of training time
for VGG-16 training in centralized training

Conclusions

- **We fairly evaluated seven distributed training algorithms in terms of various aspects**

- **We conducted in-depth analysis of the evaluation results and reported some interesting findings**
  1. Intermittent communication has negative impact on the model accuracy
  2. Centralized training has a strength in model accuracy
  3. Centralized training algorithms suffer from the PS bottleneck problem

- **We believe that our findings can be useful at industry and academia in applying and designing distributed training algorithms**

Thank You !