A. Further Information on Software Implementation and Datasets

In order to train scaling factors separately, an additional optimizer is instantiated, which only updates the scaling factors $\mathcal{S}$. All $s \in \mathcal{S}$ are initialized with the value 1 and then optimized with Adam or SGD with a momentum of 0.9. From a software perspective, we implemented a wrapper function, which detects all convolutional and dense layers within the respective neural network and replaces them with a scaled version of the respective module class. Note that this procedure changes the computational graph, however scaling factors have no effect on the original graph output if set to 1. All experiments were conducted on a homogeneous GPU cluster employing NVIDIA Ampere A100 GPUs (40 GB RAM). We use PyTorch 1.8.1 and torchvision 0.9.1 as deep learning framework and CUDA 11.1 for NN GPU acceleration.

The *Pascal Visual Object Classes Challenge 2012* [7] provides $11,540$ images categorized into 20 classes. The *Chest X-Ray* dataset [12] consists of $5,856$ images which are categorized into two classes: "pneumonia" and "normal". *CIFAR10* [15] consists of $60,000$ images with a resolution of 32×32 pixels, containing 10 classes. The data sets are split into approximately $60 : 20 : 20\%$, $75 : 15 : 10\%$ and $70 : 15 : 15\%$ for training:validation:testing purposes for Pascal VOC, Chest X-Ray and CIFAR10, respectively. We applied normalization and random horizontal flipping to all samples. Additionally, VOC samples are center cropped to 224×224 pixels and Chest X-Ray samples to 150×150 pixels.
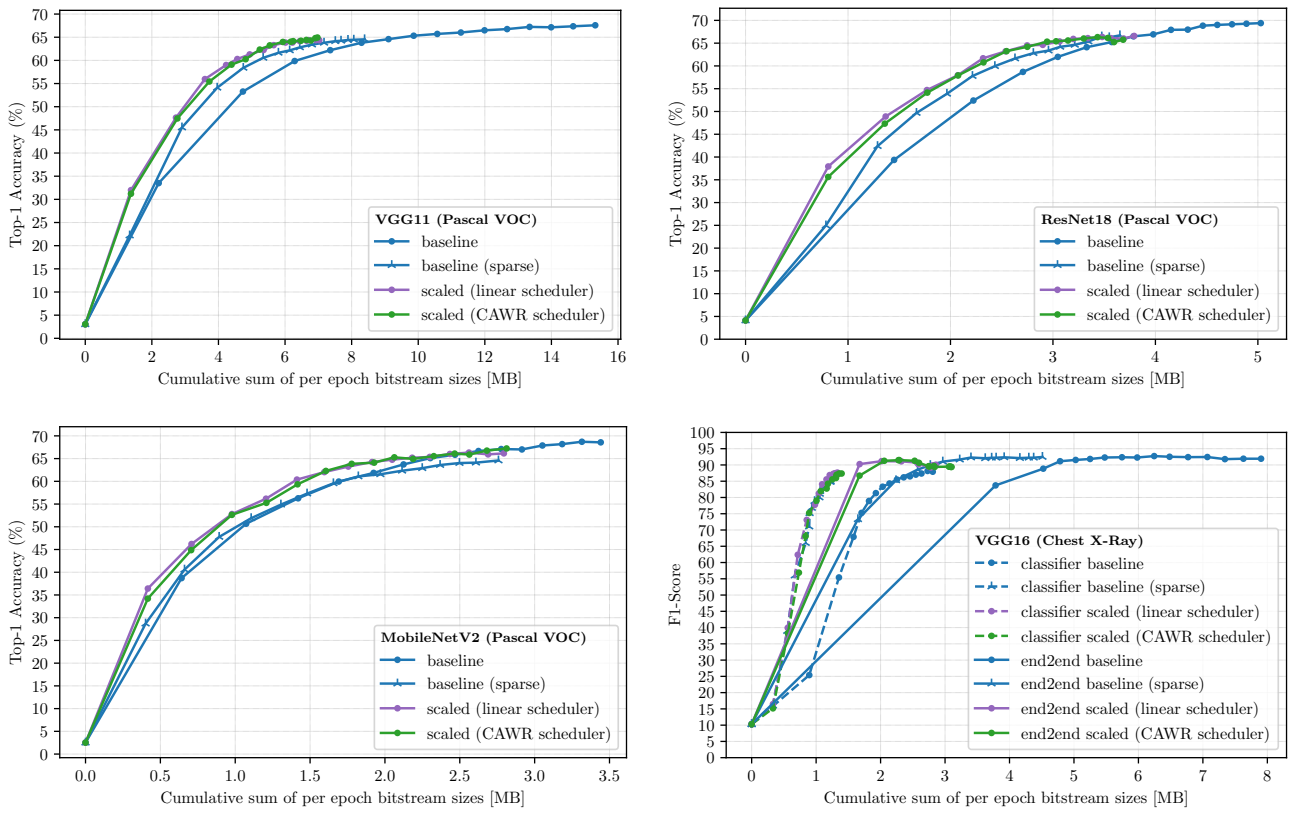
B. SGD Results

Figure B.1. SGD optimization for filter-scaled sparse federated training of VGG11, ResNet18, MobileNetV2 and VGG16 (row-major order) solving the Pascal VOC and chest x-ray classification tasks, respectively.
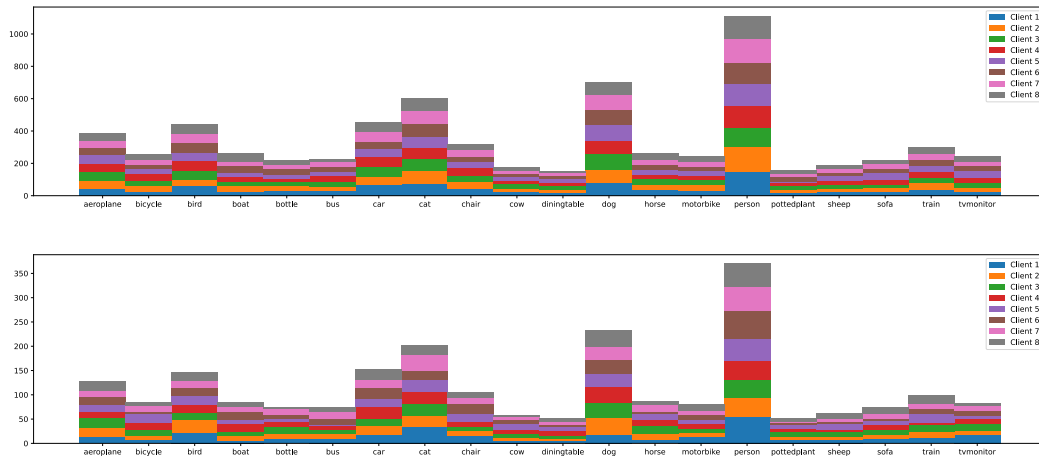
C. Data Distribution

Figure C.1. Training (top) and validation (bottom) data distribution of the Pascal VOC scenario with 8 clients.
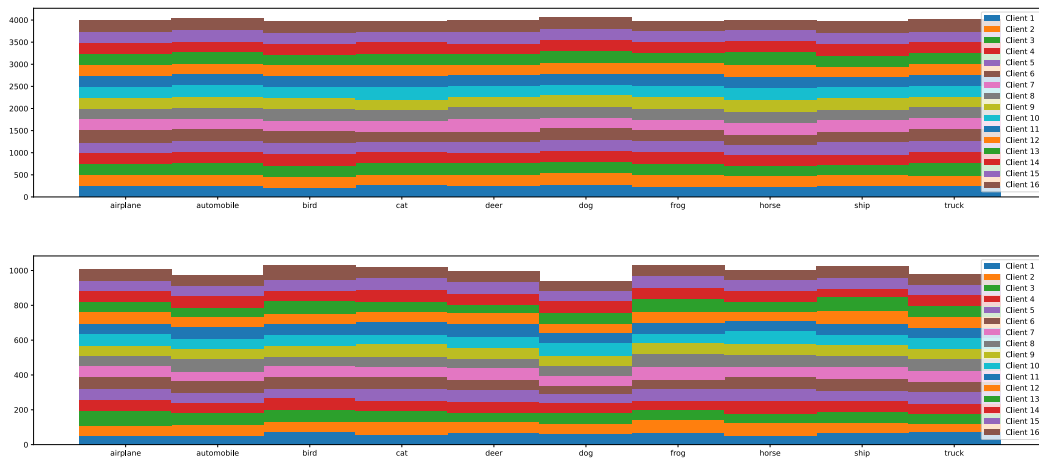

Figure C.2. Training (top) and validation (bottom) data distribution of the CIFAR10 scenario with 16 clients.