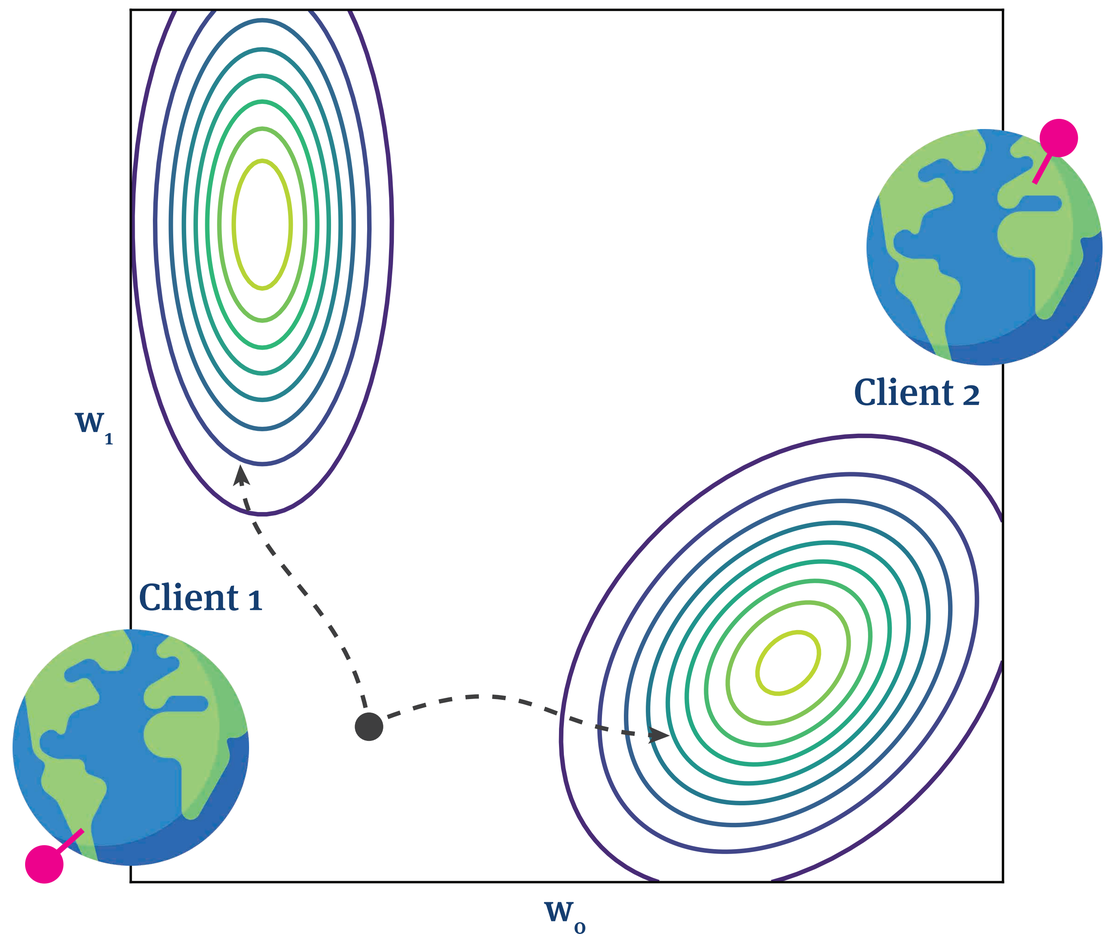
Federated Learning with Mixture-of-Experts
A solution to data heterogeneity among clients

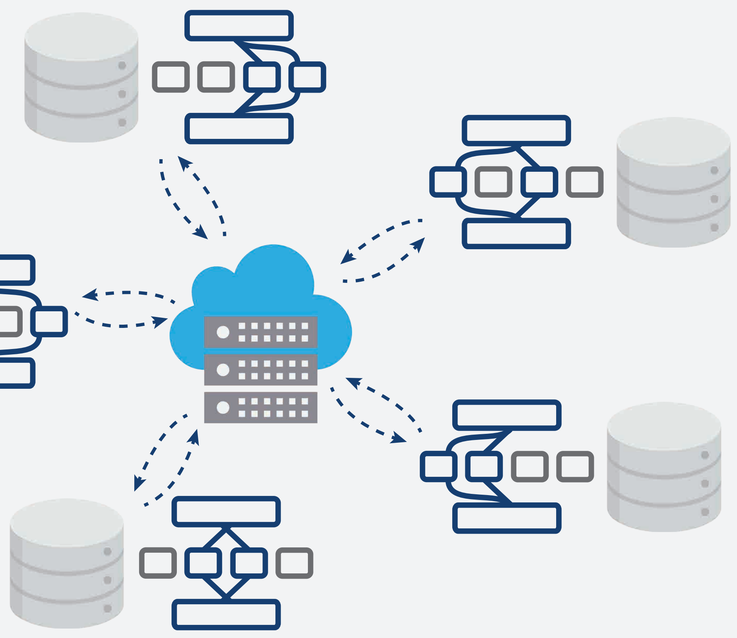
Juan I. Pisula, Katarzyna Bozek

Universität zu Köln

BOZEK LAB
DATA SCIENCE OF BIOIMAGES

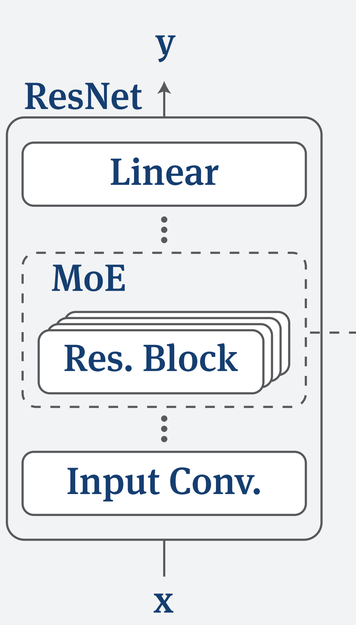
In Federated Learning scenarios with domain shifts between clients, local optimization trajectories may be inconsistent with the global empirical risk minimum direction.



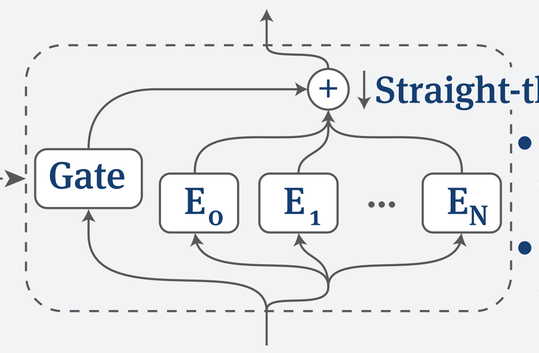


We propose using Mixture-of-Experts models to address this issue. This way, the federated model can automatically process different inputs with different network parameters.

Model



We replace the Residual Blocks of a ResNet with MoE versions of themselves

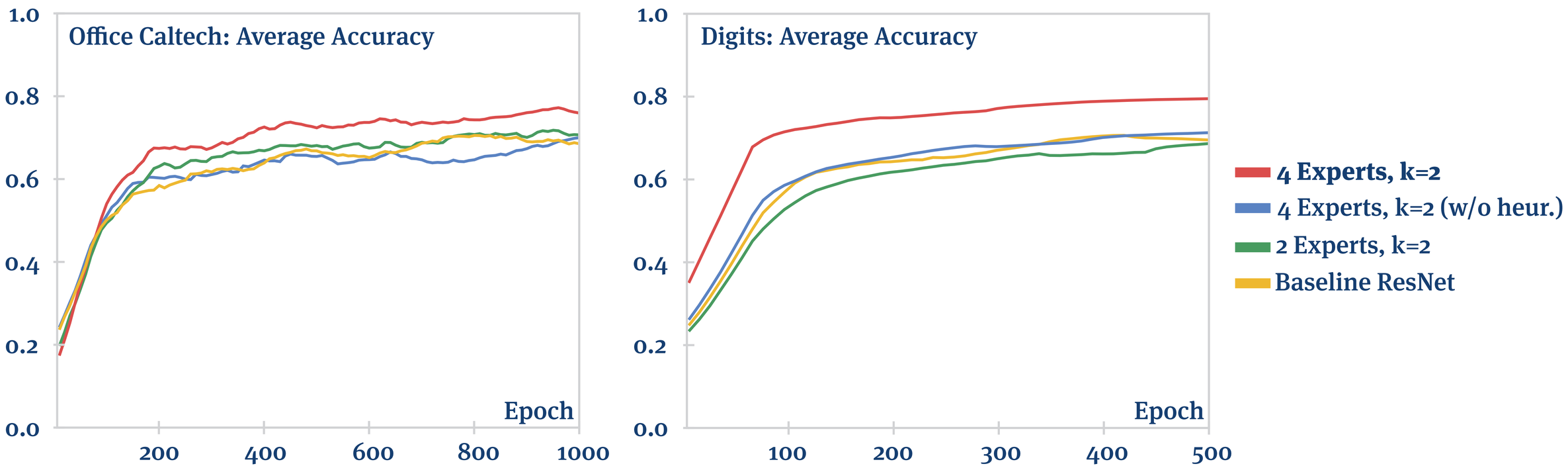


Straight-through gradients:
- **Training: all experts** are updated during backward.
- **Inference:** computation of unused experts is spared.

Gate network:
- Outputs a non-zero weight for the **top-k experts.**
- Modified "**noisy gating**" from [1].

Experiments

Federated Average Performance



[1] Shazeer, N. et al. (2017). Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. In *International Conference on Learning Representations*.