

DeltaMask: Minimizing Communication in Federated Fine-Tuning of Vision Foundation Models

Vasileios Tsouvalas · Yuki M. Asano · Aaqib Saeed

Eindhoven University of Technology, The Netherlands | University of Technology Nuremberg, Germany

IEEE BigData 2024

01 Motivation

The problem

Fine-tuning large vision models across clients creates substantial communication costs. Even one-bit updates can be expensive, so adaptation must preserve useful predictions while sending less information each round.

Contributions

- **Frozen-backbone adaptation:** Learn task-specific stochastic masks for pretrained vision models.
- **Selective updates:** Rank mask changes and compress their indices with probabilistic filters.
- **Empirical comparison:** Evaluate accuracy and bitrate across eight datasets and multiple pretrained architectures.

Federated adaptation setting

In federated learning (FL), clients adapt a shared pretrained model without collecting their images centrally. Model size drives communication cost, even when pretrained representations already transfer well to local classification tasks.

Difference from complete-mask coding

FedPM compresses complete sampled masks with arithmetic coding. *DeltaMask* instead exploits changes relative to a reproducible server mask; communication savings must be assessed together with classification accuracy.

Scope

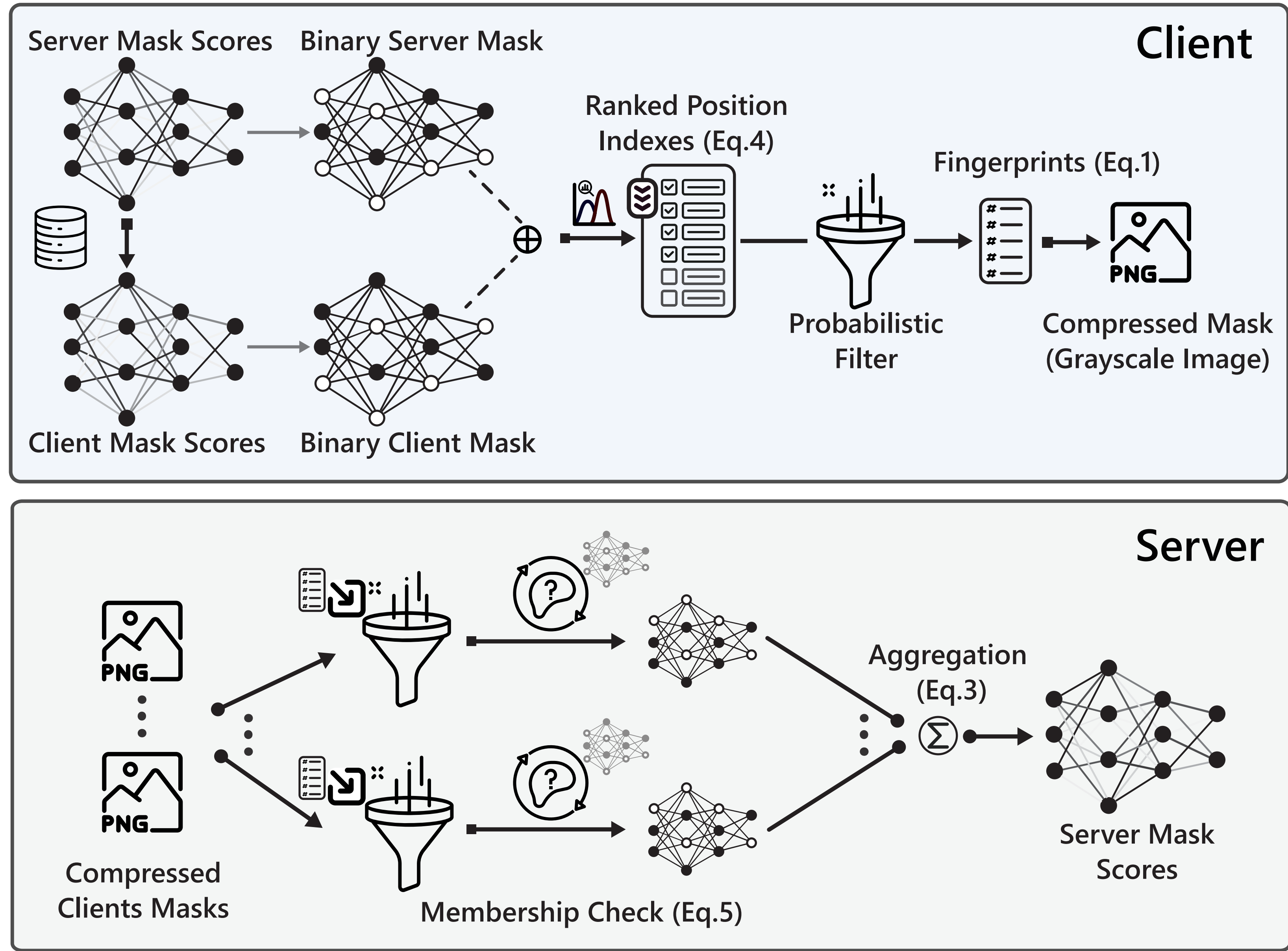
Probabilistic filters can introduce false-positive mask changes. Experiments cover image classification, and accuracy remains below unrestricted fine-tuning.

Notation

Dirichlet distribution with concentration a (Dir(a)) · independent and identically distributed (IID) · Kullback-Leibler divergence (KL) · bits per parameter (bpp) · vision transformer (ViT) · percentage points (pp)

02 DeltaMask

Communicate selected mask changes



Technical core

$$\tilde{w} = m \odot w_{\text{pre}}, \quad m_i \sim \text{Bernoulli}(\theta_i)$$

w_{pre} : frozen pretrained weights; θ : learned keep probabilities; m : sampled binary mask. Adapt which weights are active.

$$\Delta = \{i : m_i^{\text{local}} \neq m_i^{\text{server}}\}, \quad \Delta' = \text{TopK}_{\text{KL}}(\Delta)$$

Δ : changed mask indices; K : retained count. Rank changes by local-to-server Bernoulli KL divergence, then send selected indices.



Paper



Code

03 Evaluation

0.151 bpp

Average transmitted bitrate

8 datasets · CLIP ViT-B/32 · IID

85.44%

Average accuracy

Same setting · full participation

Low bitrate with a measured accuracy trade-off

Method	Avg. accuracy (%)	Avg. bits/parameter
Fine-tuning	88.08	32
FedMask	79.23	1.0
EDEN	84.14	0.691
FedPM	86.07	0.872
DeltaMask (Ours)	85.44	0.151

CLIP ViT-B/32 · 8 datasets · IID Dir(10) · 30 clients · full participation · 100 rounds. Means of 3 runs.

Non-IID data with 20% participation

Method	Avg. accuracy (%)	Avg. bits/parameter
Fine-tuning	83.19	32
EDEN	78.56	0.715
FedPM	81.64	0.948
DeltaMask (Ours)	81.13	0.233

Same 8 datasets and backbone · non-IID Dir(0.1) · 30 clients · 20% participation · 300 rounds. Means of 3 runs.

Takeaway

DeltaMask sends **selected mask changes**, reaching **0.151 bits/parameter** at **85.44% accuracy** across eight IID tasks. Compared with FedPM at 0.872 bits/parameter, it reduces communication while trailing accuracy by 0.63 pp.