Active Property Inference Attack

Gradients on Main Loss

$L(x, y; \theta)$

Main Task (Gender)

Layer N

Layer 2

Layer 1

Upload local updates

Server

Download joint model

Gradients on Joint Loss

$\alpha * L(x, y; \theta) + (1 - a) * L(x, p; \theta)$

Main Task (Gender)

Inference Task (Facial ID)

Layer N

Layer 2

Layer 1
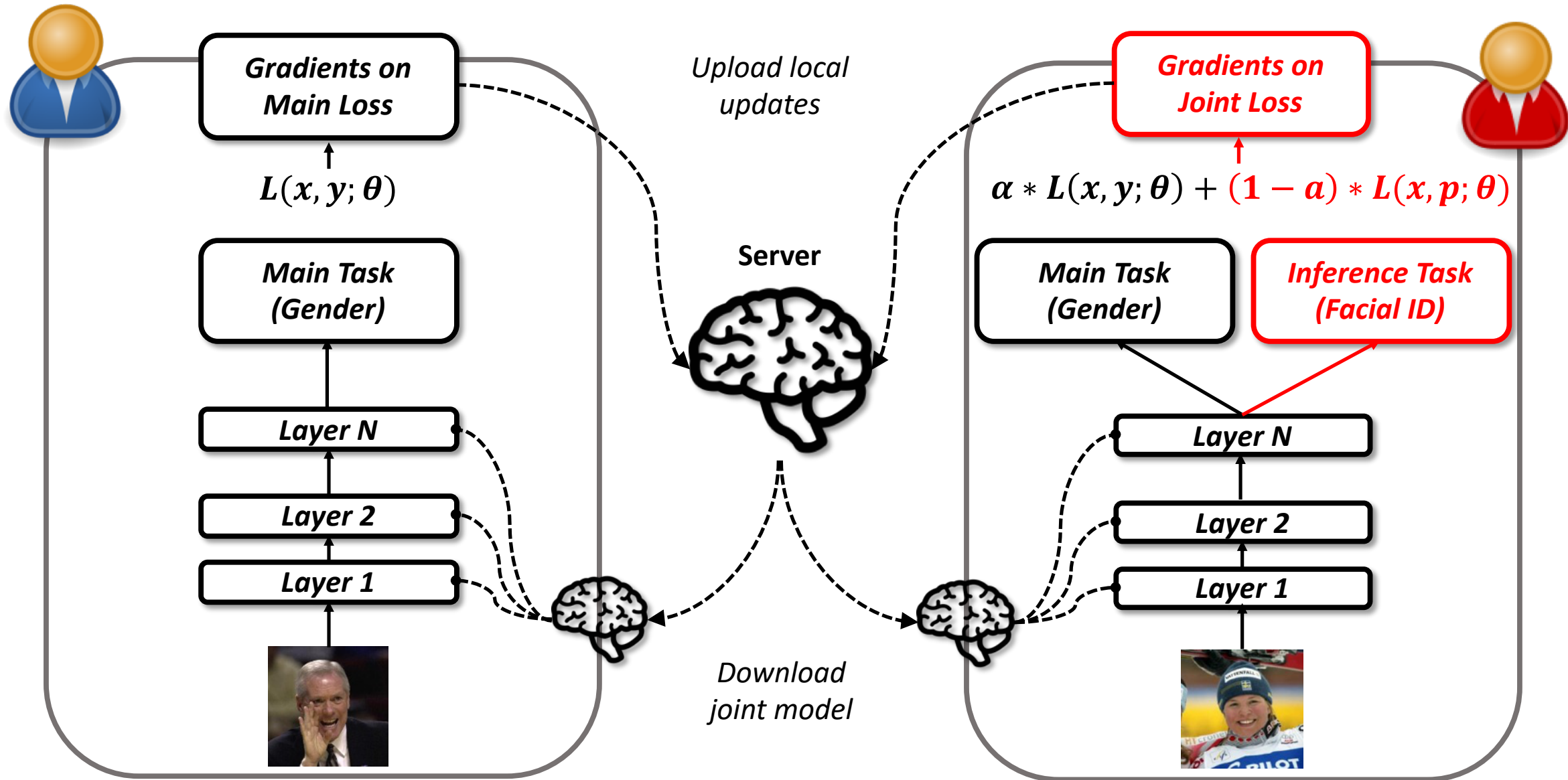
Dataset	Type	Main Task	Inference Task
LFW	Images	Gender/Smile/Age Eyewear/Race/Hair	Race/Eyewear
FaceScrub	Images	Gender	Identity
PIPA	Images	Age	Gender
FourSquare	Locations	Gender	Membership
Yelp-health	Text	Review Score	Membership Doctor specialty
Yelp-author	Text	Review Score	Author
CSI	Text	Sentiment	Membership Region/Gender/Veracity