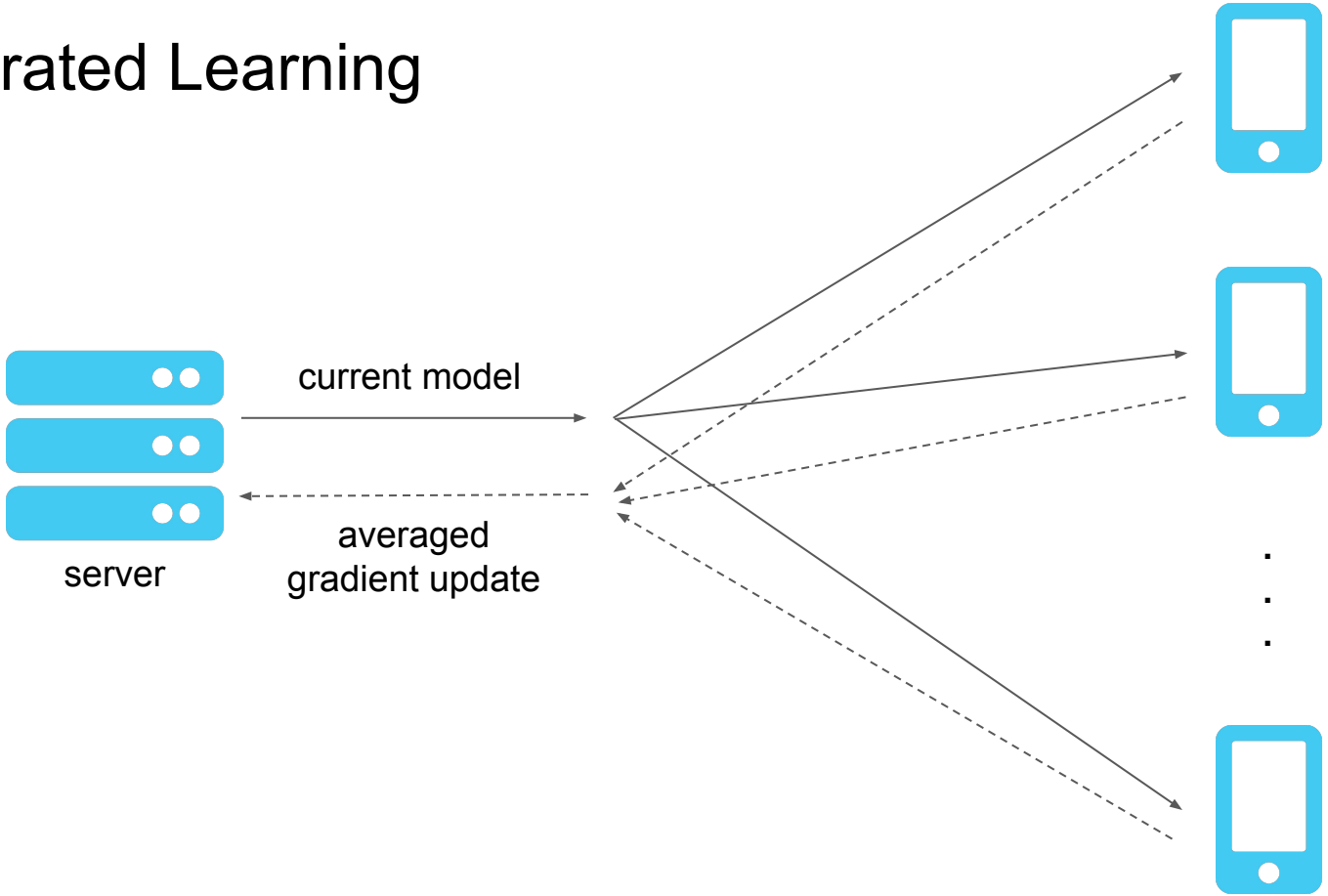
Fishing for User Data in Large-Batch Federated Learning via Gradient Magnification

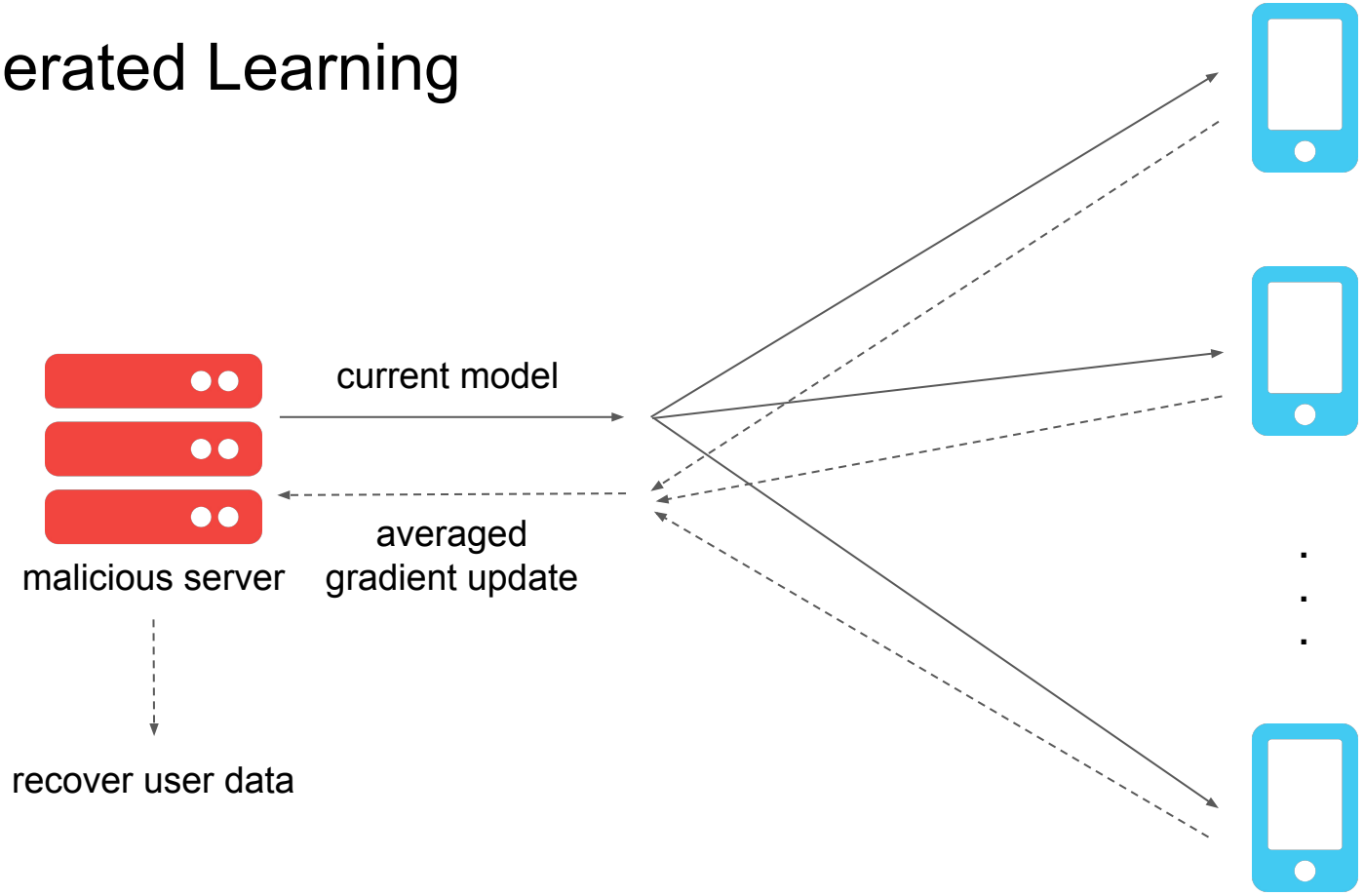
Yuxin Wen*, Jonas Geiping*, Liam Fowl*
Micah Goldblum, Tom Goldstein

University of Maryland & New York University

Federated Learning



Federated Learning



Previous Attacks and Our Contributions

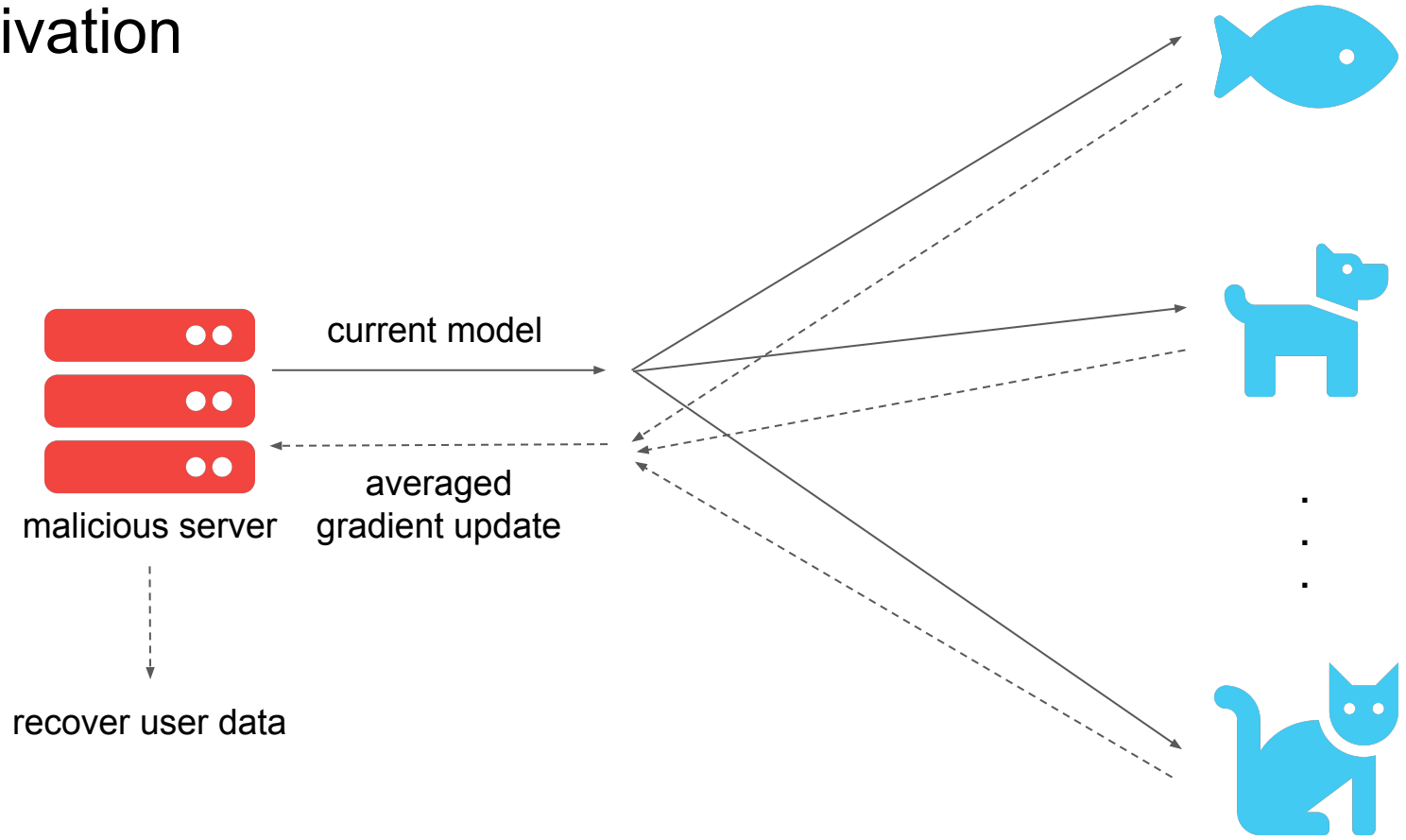
Previous Attacks:

- require settings with very small batch sizes
- require unrealistic and conspicuous architecture modifications

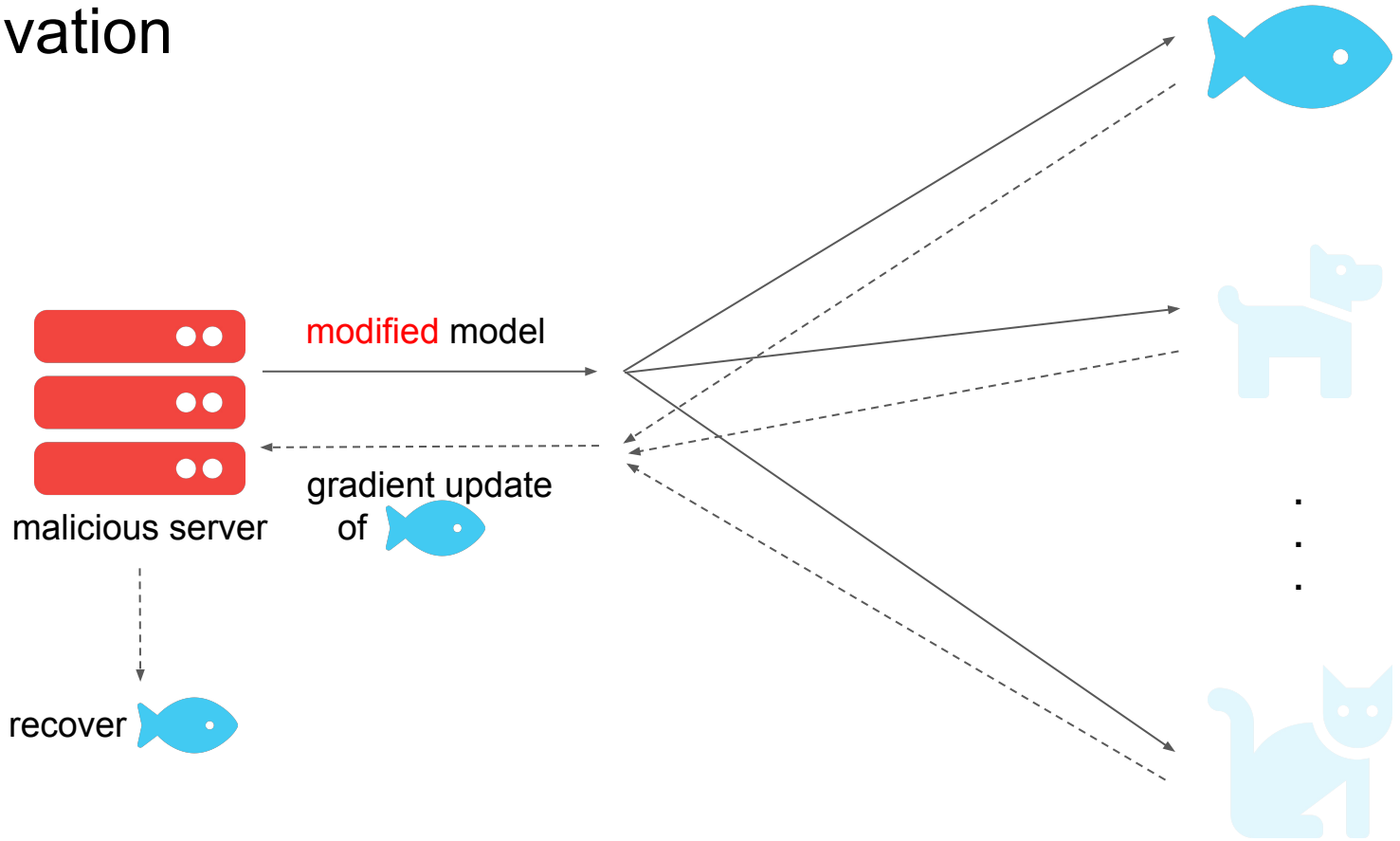
Our Attack:

- works in large batch settings
- only needs modifications of model parameters
- is scalable to both cross-device and cross-silo settings

Motivation


current model
averaged
gradient update
malicious server
recover user data

Motivation


modified model
gradient update
of
malicious server
recover

Class Fishing Strategy

$$W_{i,j} = \begin{cases} W_{i,j}, & \text{if } i = c \\ 0, & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} b_i, & \text{if } i = c \\ \alpha, & \text{otherwise} \end{cases},$$

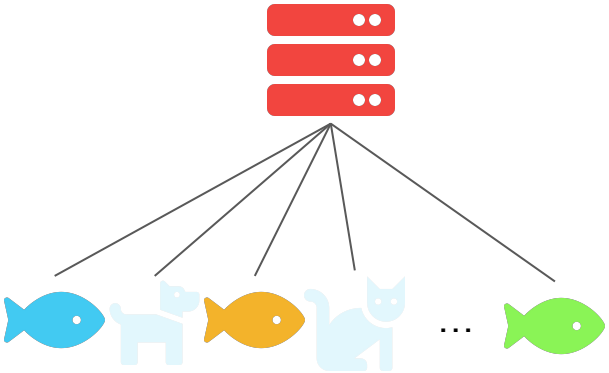
Feature Fishing Strategy

$$W_{i,j} = \begin{cases} \beta, & \text{if } i = c \text{ and } j = k \\ 0, & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} -\beta * \theta, & \text{if } i = c \\ 0, & \text{otherwise} \end{cases},$$

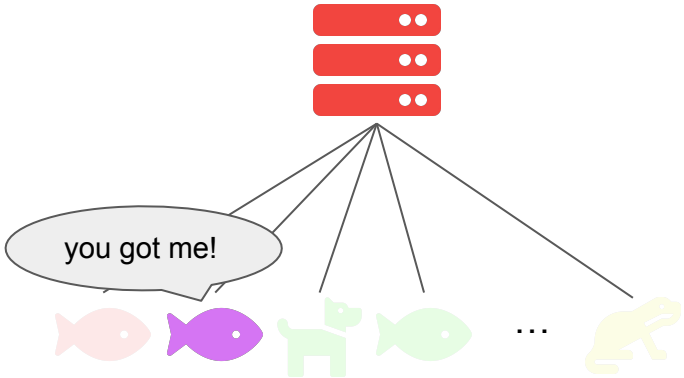
Cross-Device

Estimating



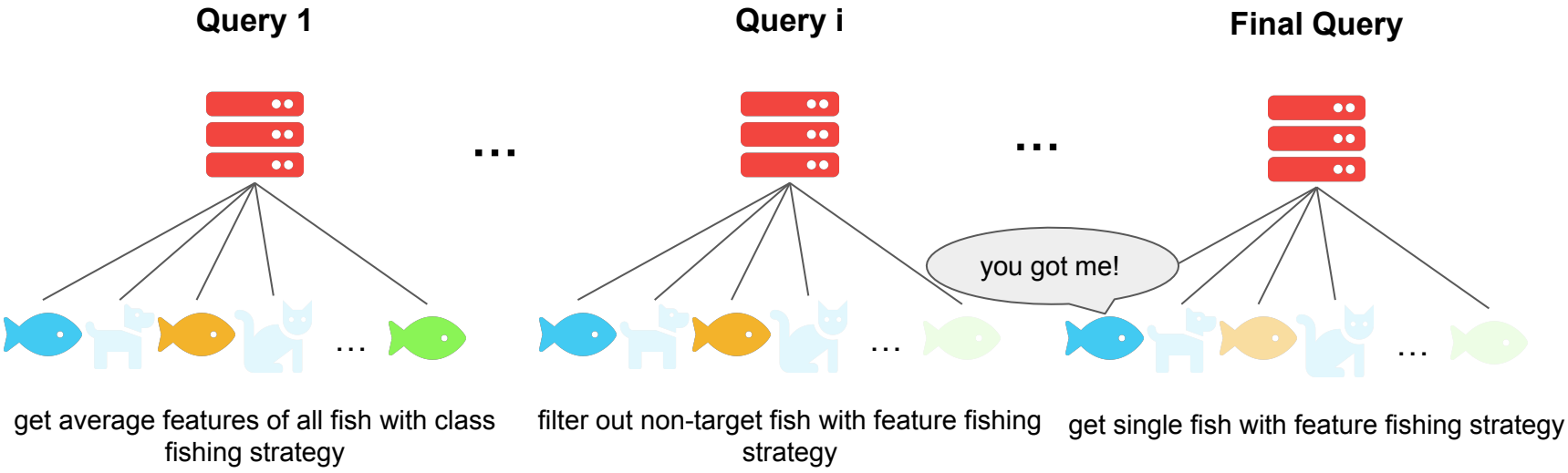
collect features and estimate the feature distributions of the target class with class fishing strategy

Fishing



set a proper feature cutoff and fish out a single data point with feature fishing strategy

Cross-Silo

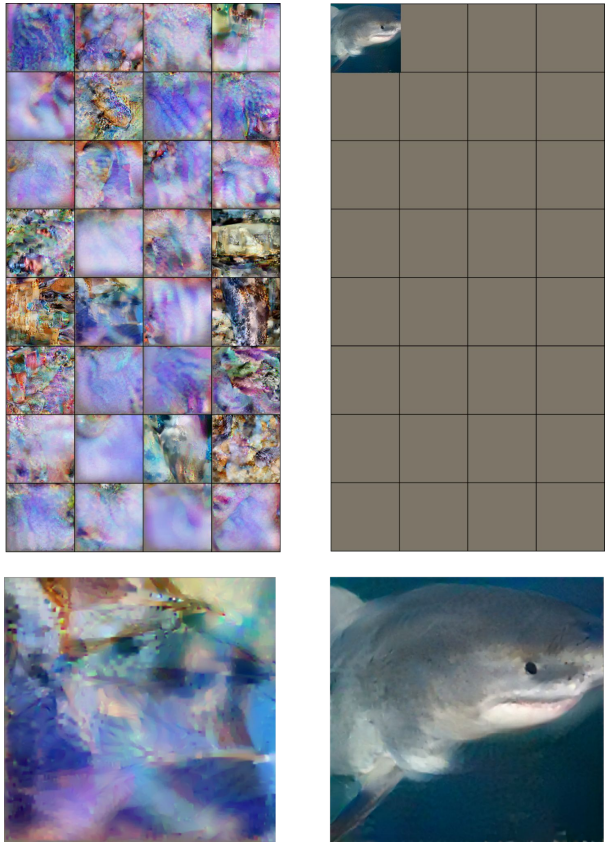


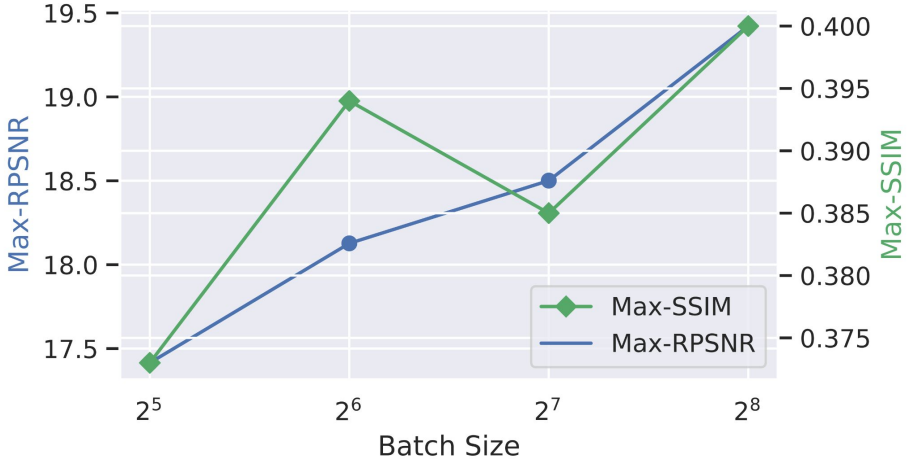
get average features of all fish with class fishing strategy

filter out non-target fish with feature fishing strategy

get single fish with feature fishing strategy

Result




19.5
19.0
18.5
18.0
17.5
Max-RPSNR
0.400
0.395
0.390
0.385
0.380
0.375
Max-SSIM
Max-SSIM
Max-RPSNR
2^5
2^6
2^7
2^8
Batch Size

Thanks!

https://github.com/JonasGeiping/breaching