

# Security in Federated Learning

Privacy-Preserving Benchmarking Tool for Gradient Inversion Attacks

RESULT: 23 dB PSNR drop = Attack Blocked

## BASELINE



Reconstructed Private Data (12 FL Clients)

**Attack Quality Metrics**  
PSNR: 27.48 +/- 1.37  
SSIM: 0.92 +/- 0.03  
LabelMatch: 100%

## DIFFERENTIAL PRIVACY

Epsilon = 8.0

LPIPS: 0.807  
SSIM: -0.001  
PSNR: 6.7 dB  
dPSNR: -22.7 dB  
dLPIPS: +0.690

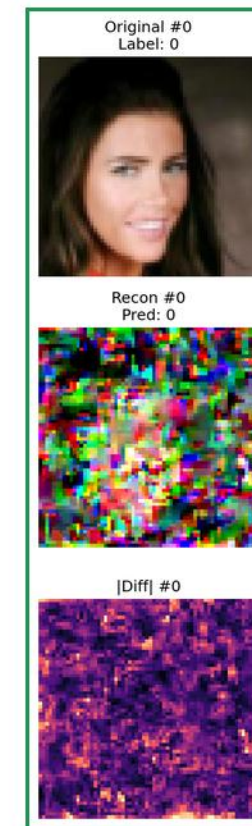
Epsilon = 1.0

LPIPS: 0.747  
SSIM: -0.001  
PSNR: 6.3 dB  
dPSNR: -23.1 dB  
dLPIPS: +0.629

Epsilon = 0.1

LPIPS: 0.806  
SSIM: -0.001  
PSNR: 6.4 dB  
dPSNR: -23.0 dB  
dLPIPS: +0.689

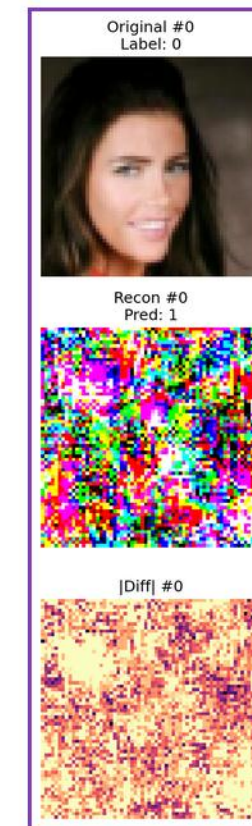
## HOMOMORPHIC ENCRYPTION



LPIPS: 0.635 SSIM: 0.343  
PSNR: 14.0 dB  
LabelMatch: Yes

**Defense Impact:**  
dPSNR = -15.3 dB  
dLPIPS = +0.517

## DP + HE

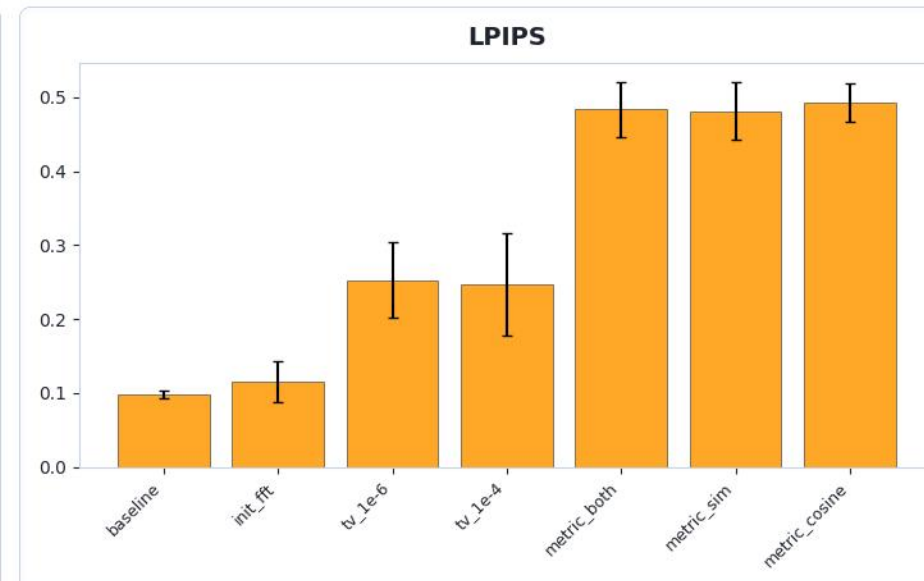
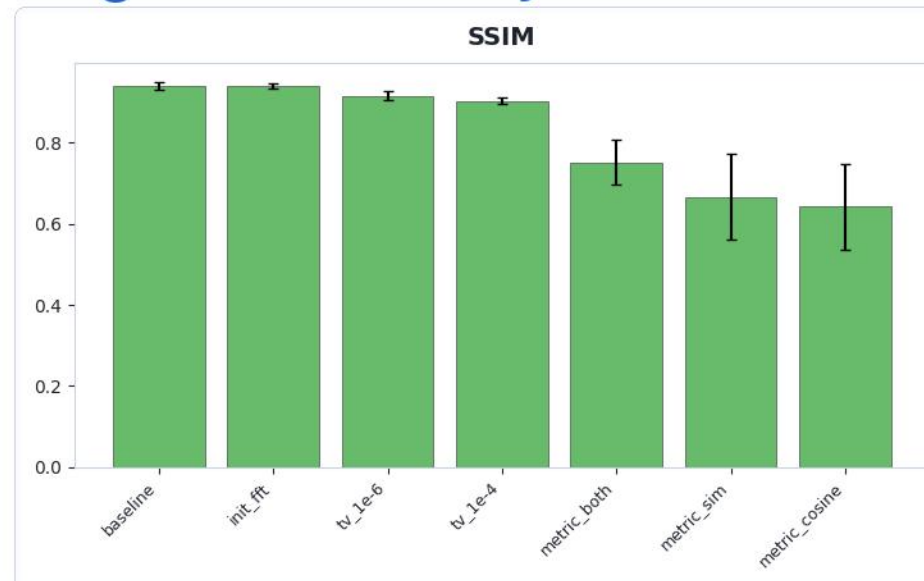
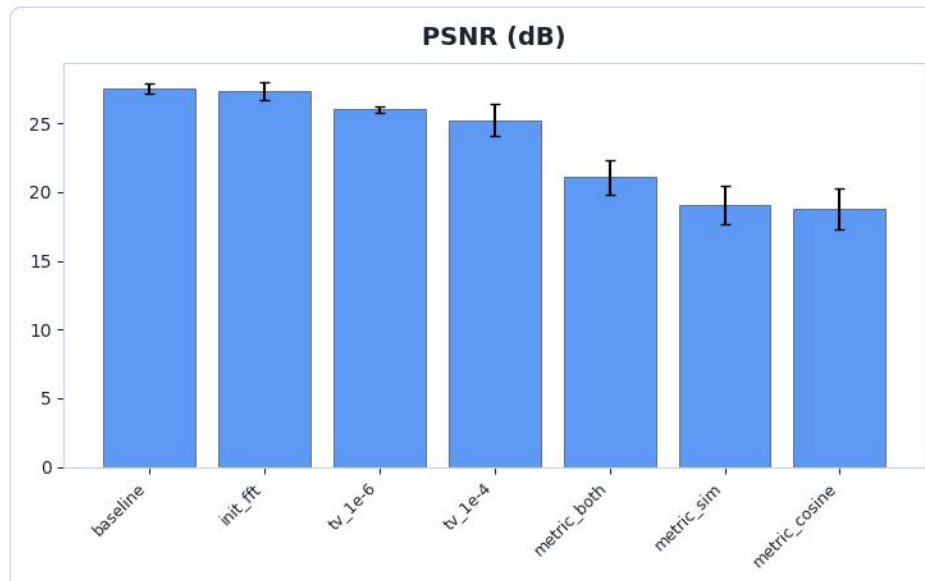


LPIPS: 0.824 SSIM: -0.003  
PSNR: 6.4 dB  
LabelMatch: No (Attack Failed)

**IMPACT:**  
dPSNR = -23.0 dB  
dLPIPS = +0.707

Privacy Preserved: Attack Fully Mitigated

## ABLATION STUDY: Attack Configuration Analysis



### BEST ATTACK SETTINGS

baseline: 27.5 dB PSNR  
init\_fft: 27.4 dB PSNR  
tv\_1e-6: 26.0 dB PSNR

### RESEARCH INSIGHTS

Best PSNR: baseline  
Best LPIPS: baseline  
MSE metric outperforms cosine

Optimal Config: init\_fft/c3