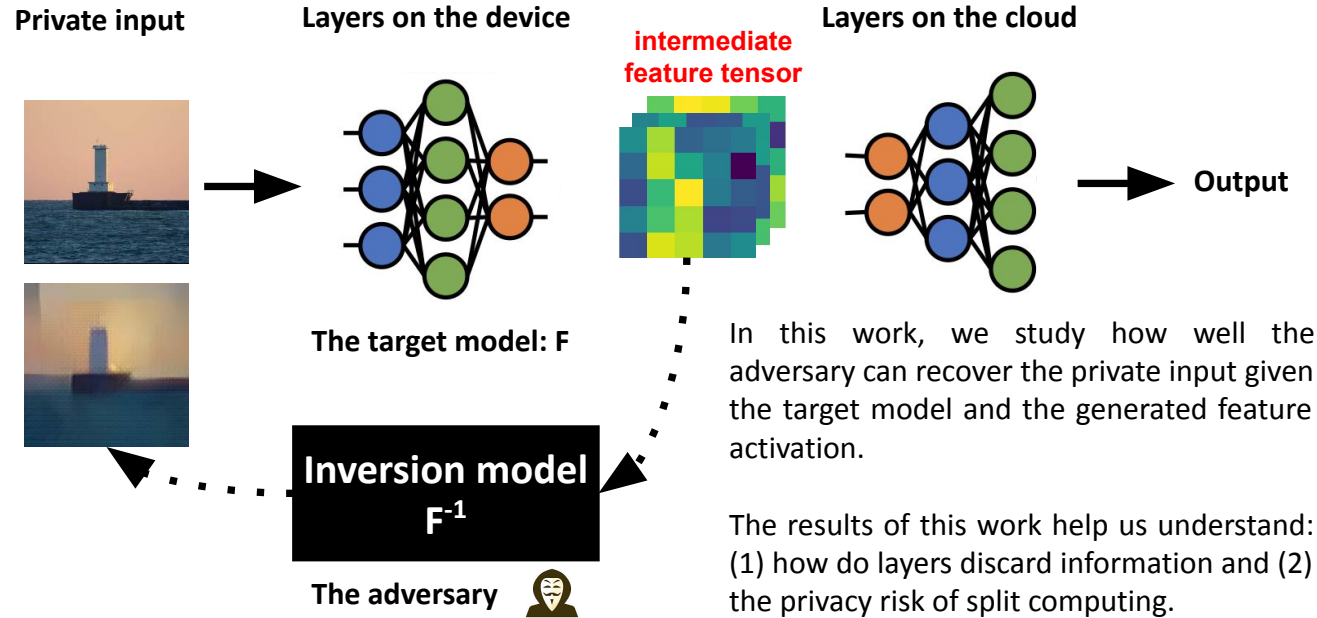
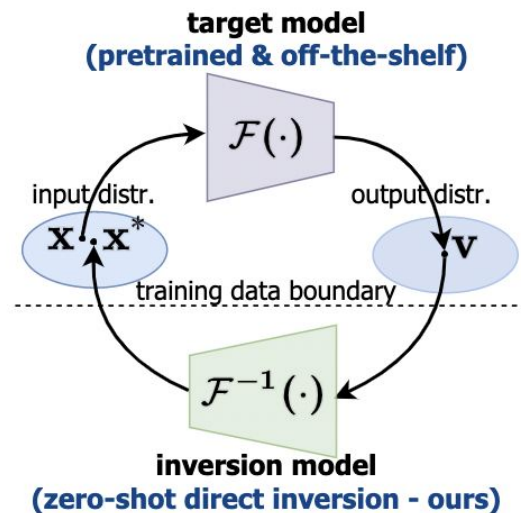


Privacy Risk in Split Computing



Problem Setting

- Generalizability:** off-the-shelf pre-trained networks (e.g., ResNet, RepVGG, ResNet-SelfSup and SNGAN)
- Efficiency:** we consider direct model inversion which takes feature activation as input and directly generates the recovered image.
- No Real Data:** the adversary does NOT have access to any real data to train the inversion model.
- Scale:** deep target model (with 20+ layers) and large-scale datasets (ImageNet and CelebA).

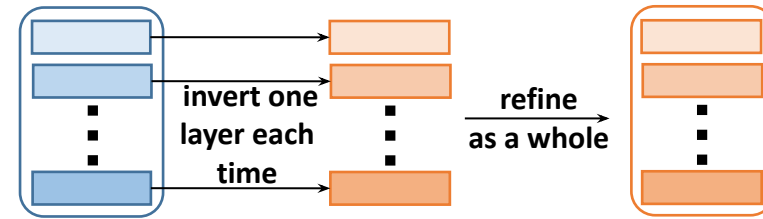


Method: Divide-and-Conquer Inversion (DCI)

- DCI **partitions** the overall inversion problem into several block-wise inversion sub-problems before integrating them together for refinement,

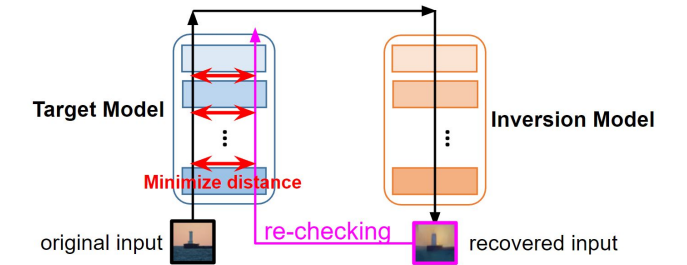
$$\mathcal{L}_{\text{layer}} = \|\mathbf{u}_k - \mathcal{F}_k^{-1}(\mathcal{F}_k(\mathbf{u}_k))\|_1, \quad \mathbf{u}_k = \mathcal{F}_{1:(k-1)}(\mathbf{x})$$

$$\mathcal{L}_{\text{img}}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|_1, \quad \mathbf{x}' = \mathcal{F}_{1:k}^{-1}(\mathcal{F}_{1:k}(\mathbf{x}))$$



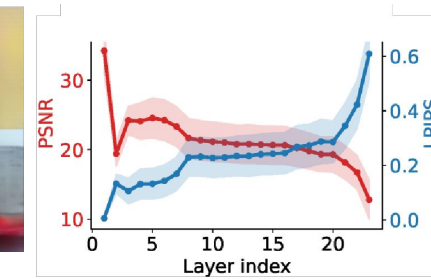
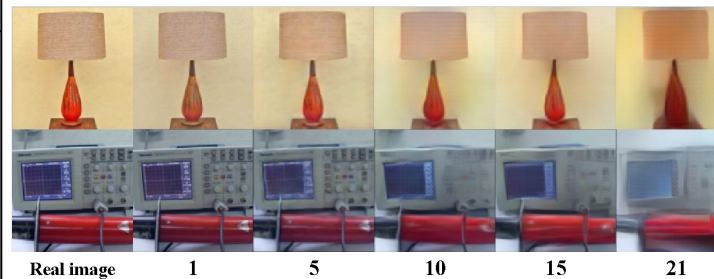
- we explore **cycle consistency** to measure the quality of recovered inputs by re-checking them with the target model,

$$\mathcal{L}_{\text{cyc}}(\mathbf{x}, \mathbf{x}') = \sum_{l=1}^L \|\mathcal{F}_{1:l}(\mathbf{x}) - \mathcal{F}_{1:l}(\mathbf{x}')\|_1$$

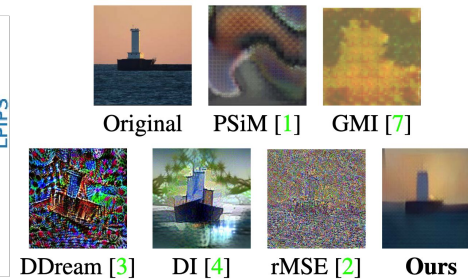


Results

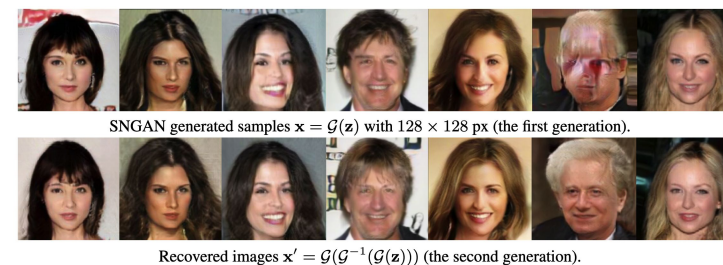
RepVGG-A0 (Up to 21 Conv Blocks) Inversion - ImageNet



Comparison on RepVGG-A0



GAN Inversion - CelebA



ResNet50 Inversion - ImageNet

