Machine
Learning
Laboratory

Gradient Inversion with Generative Image Prior

Jinwoo Jeon[*1], Jaechang Kim[*2], Kangwook Lee[3], Sewoong Oh[4], Jungseul Ok[12]

[*]contributed equally

[1]Department of Computer Science and Engineering, POSTECH

[2]Graduate School of Artificial Intelligence, POSTECH

[3]Department of Electrical and Computer Engineering, University of Wisconsin-Madison

[4]Paul G. Allen School of Computer Science & Engineering, University of Washington

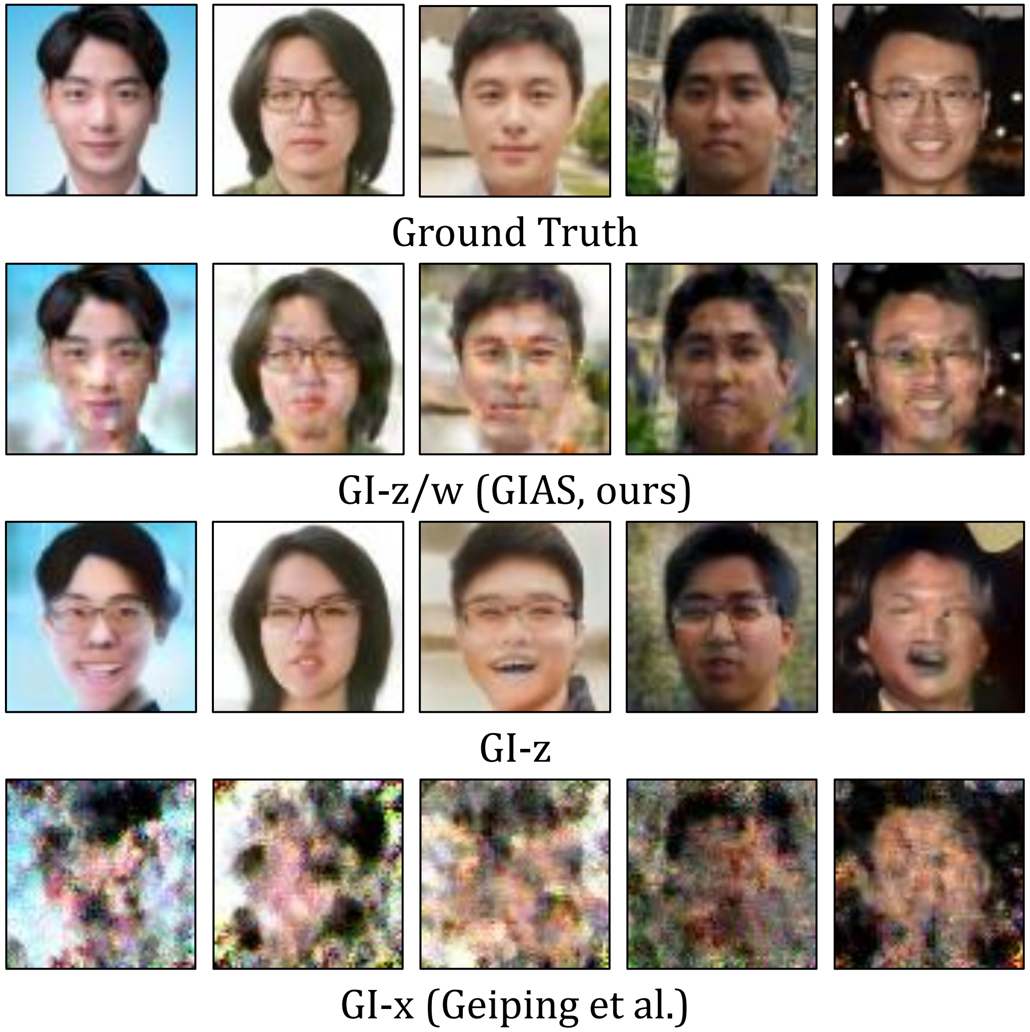
Advantages of Federated Learning

Privacy Concern in Federated Learning

Advantage in privacy is debatable

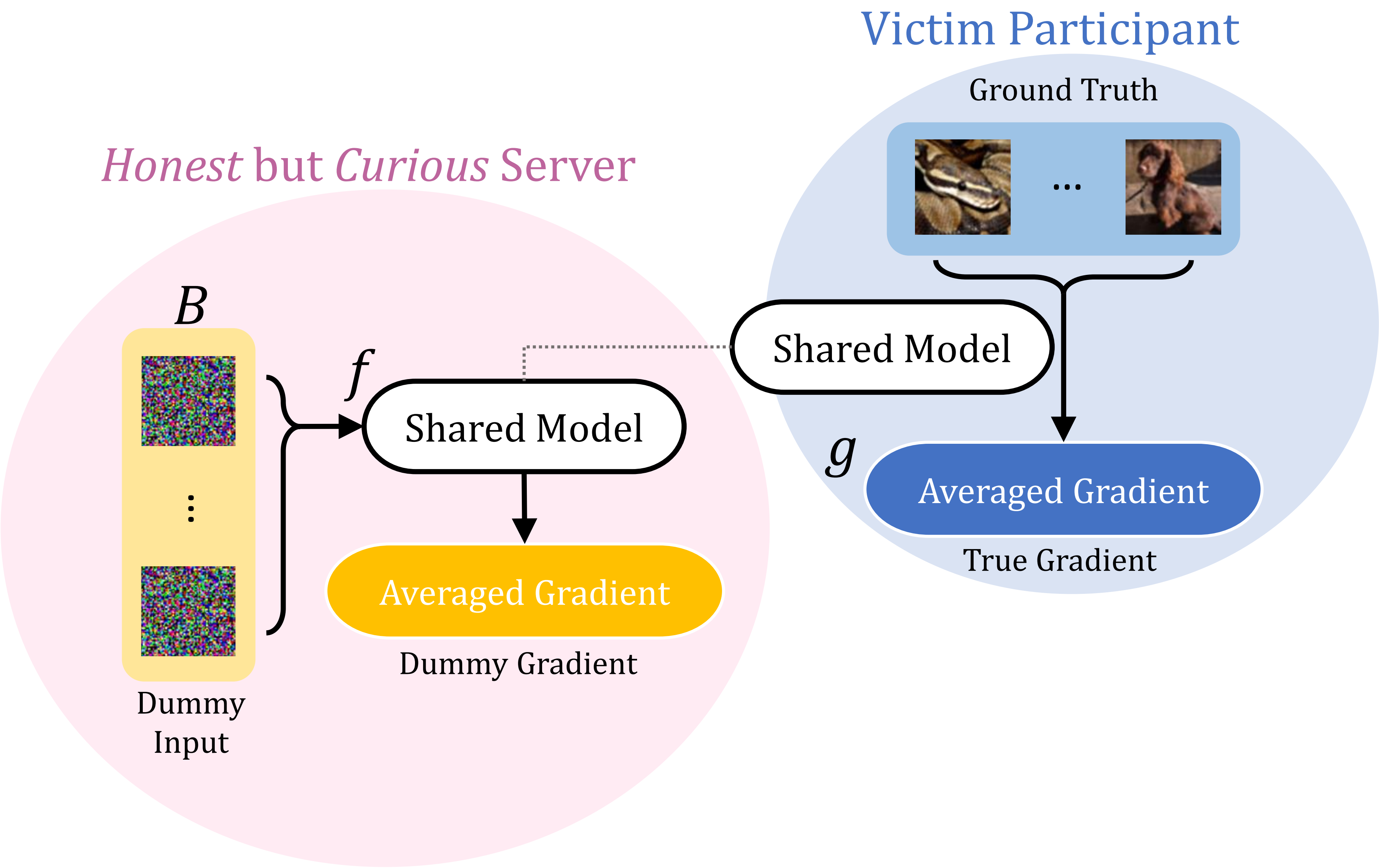
Gradient Inversion with Generative Image Prior

Gradient Inversion Attack using Prior

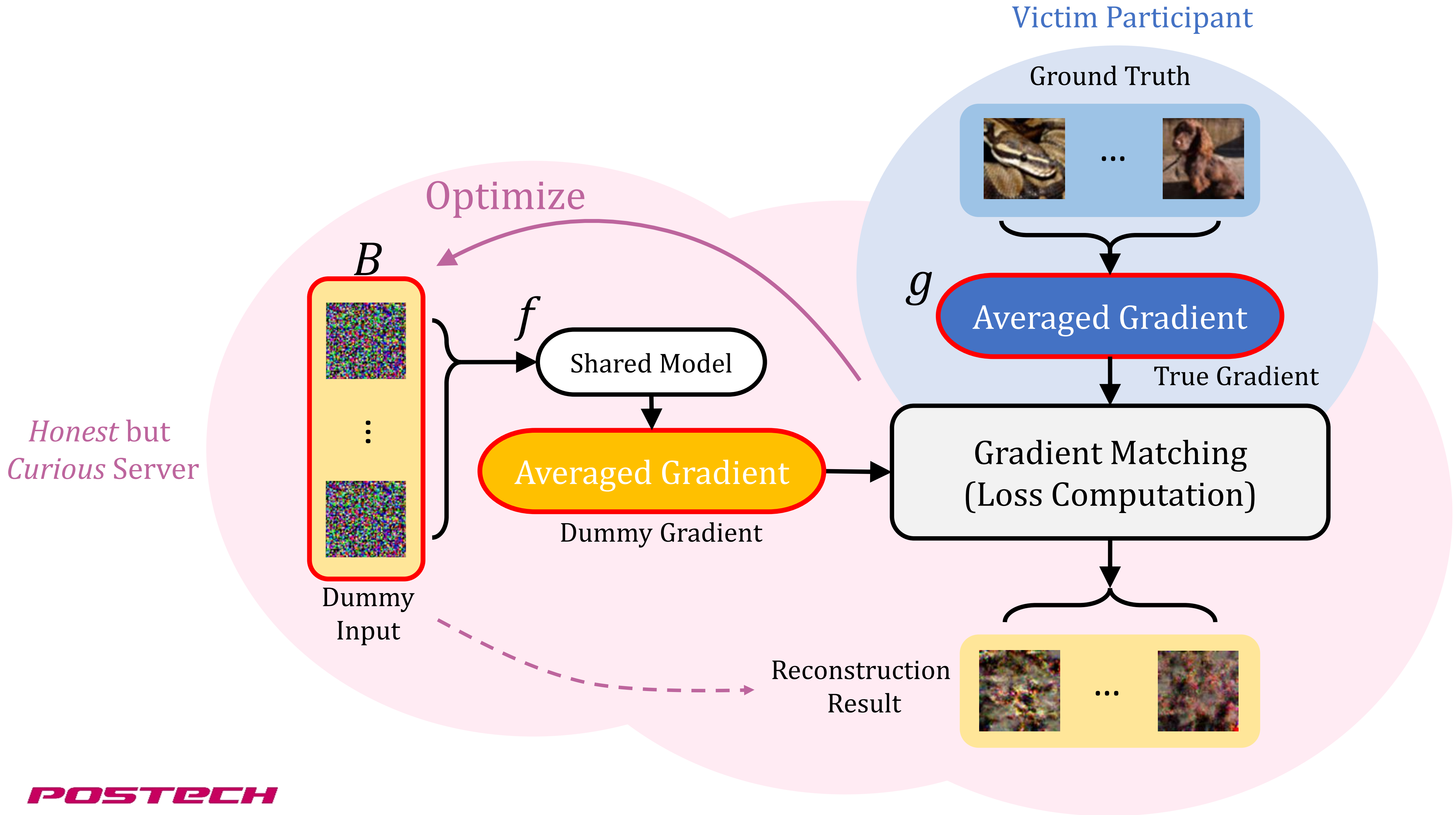


Learning Prior via Gradient Inversion

What is Gradient Inversion?



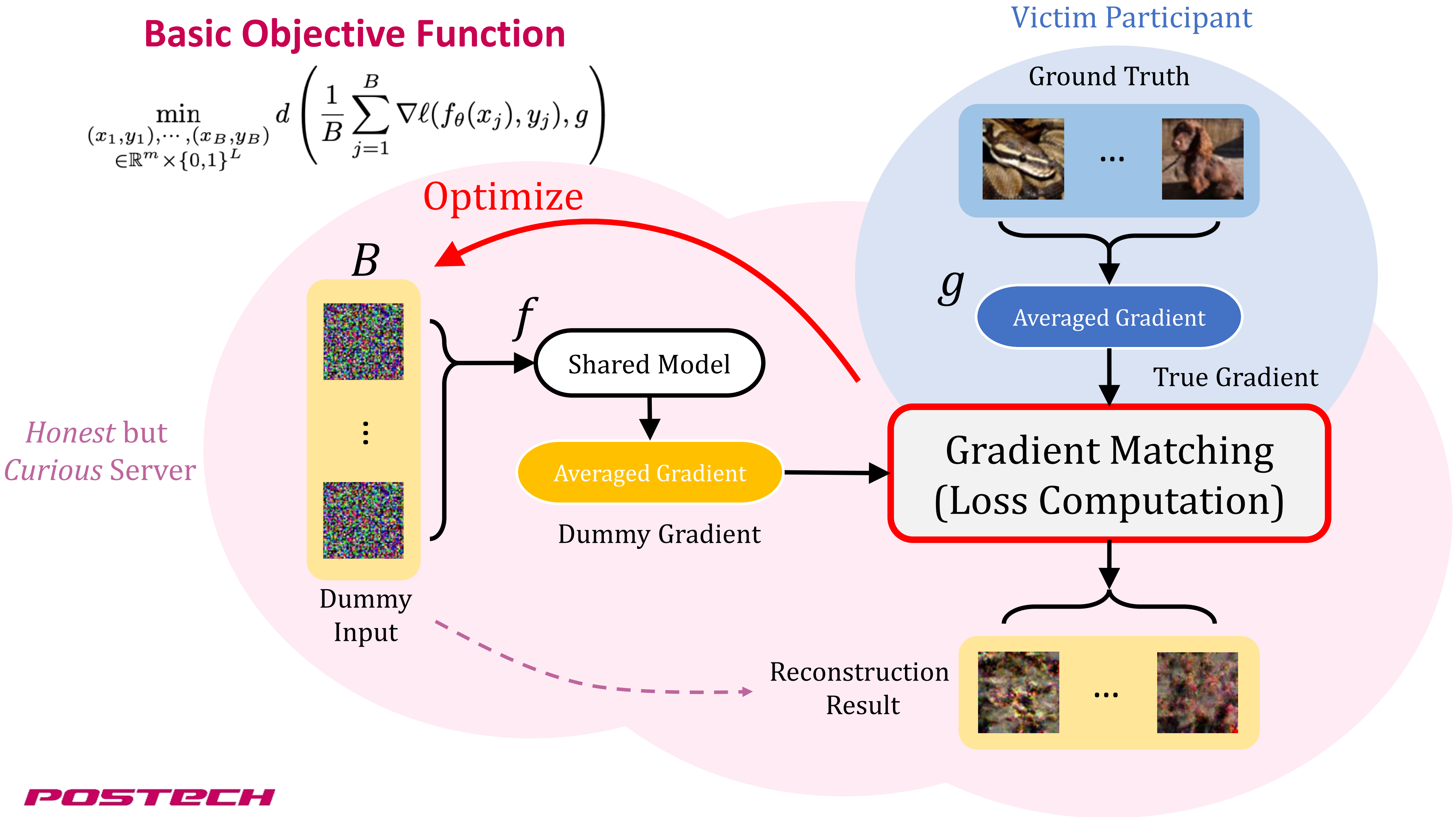
What is Gradient Inversion?



What is Gradient Inversion?

Basic Objective Function

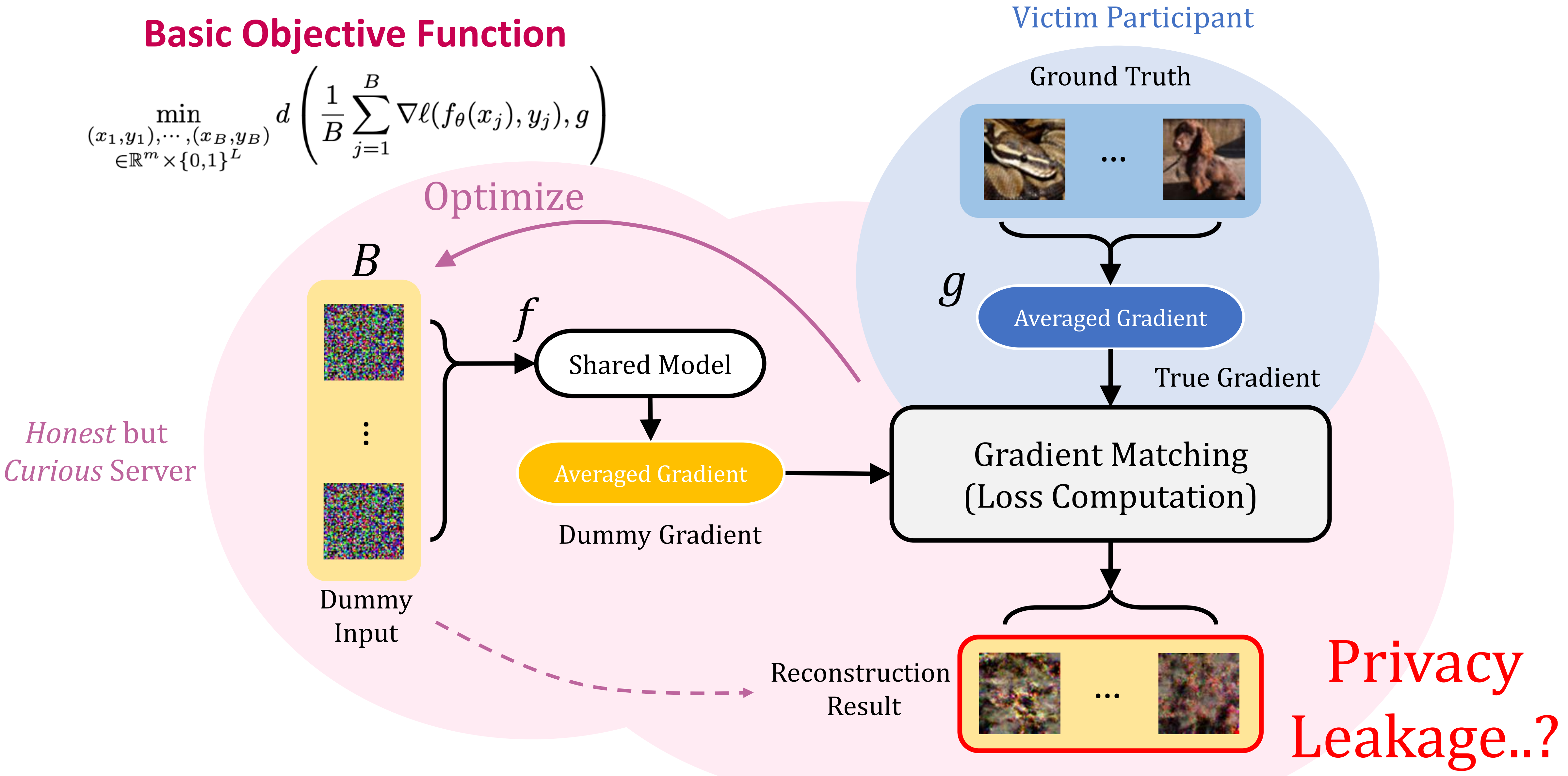
$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B)\\ \in\mathbb{R}^m\times\{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B}\nabla\ell(f_\theta(x_j),y_j),\, g\right)$$



What is Gradient Inversion?

Basic Objective Function

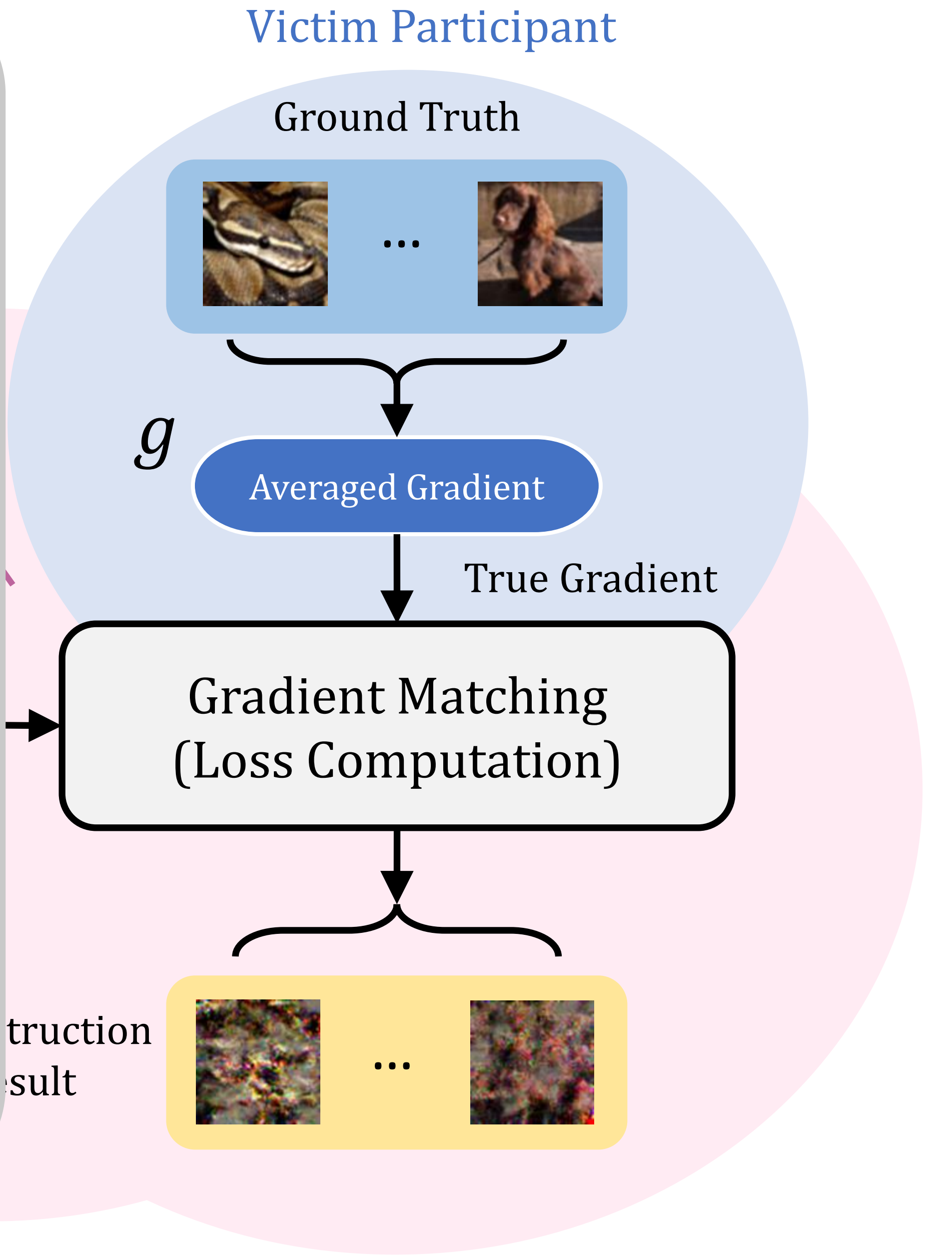
$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B)\\ \in\mathbb{R}^m\times\{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B}\nabla\ell(f_\theta(x_j),y_j),g\right)$$



What is Gradient Inversion?

Why Gradient Inversion is hard?

$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B)\\ \in\mathbb{R}^m\times\{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B}\nabla\ell(f_\theta(x_j),y_j),g\right)$$



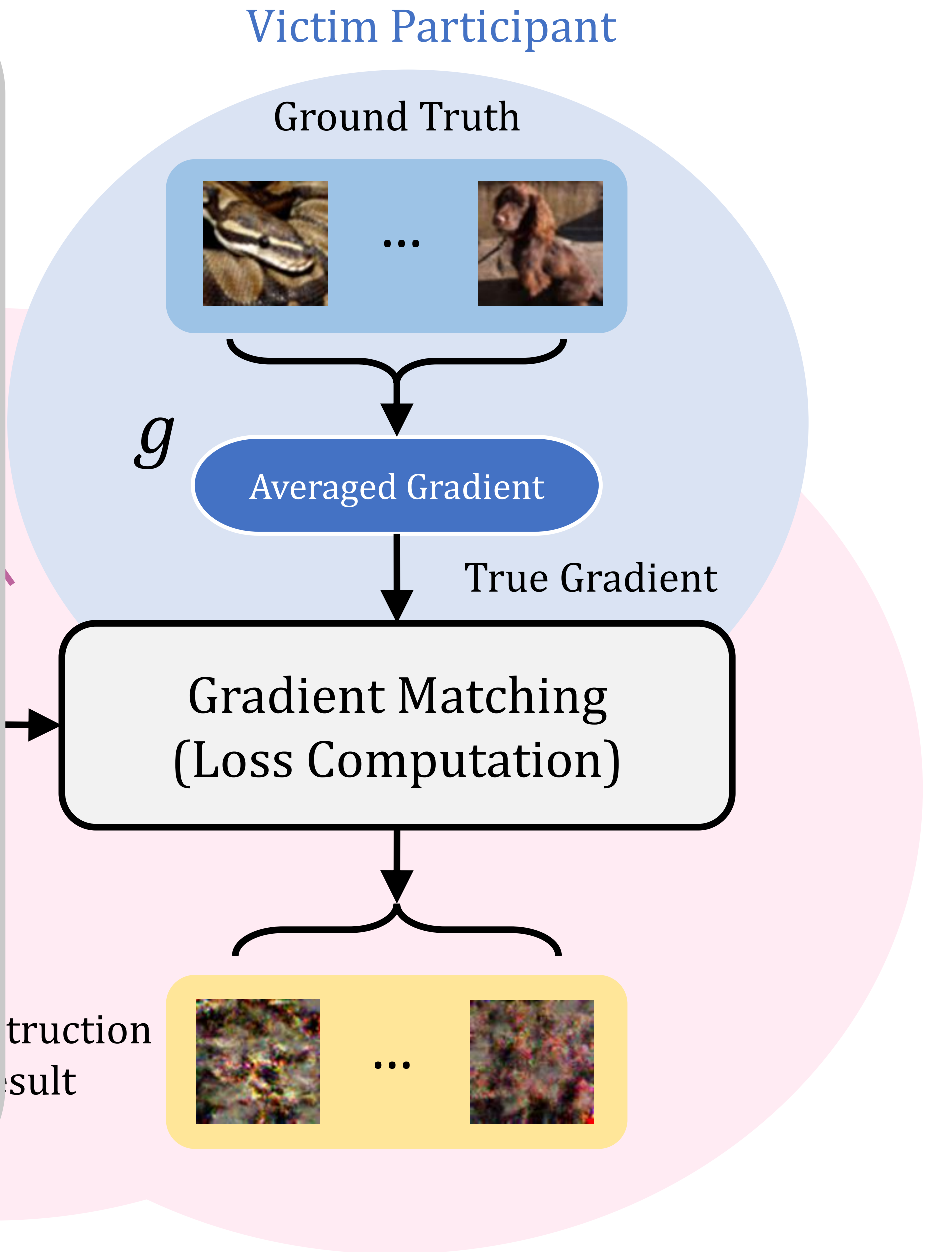
What is Gradient Inversion?

Why Gradient Inversion is hard?

$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B)\\ \in\mathbb{R}^m\times\{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B}\nabla\ell(f_\theta(x_j),y_j), g\right)$$

Source Separation + **Compressed Sensing**

→ Under-determined Problem!



What is Gradient Inversion?

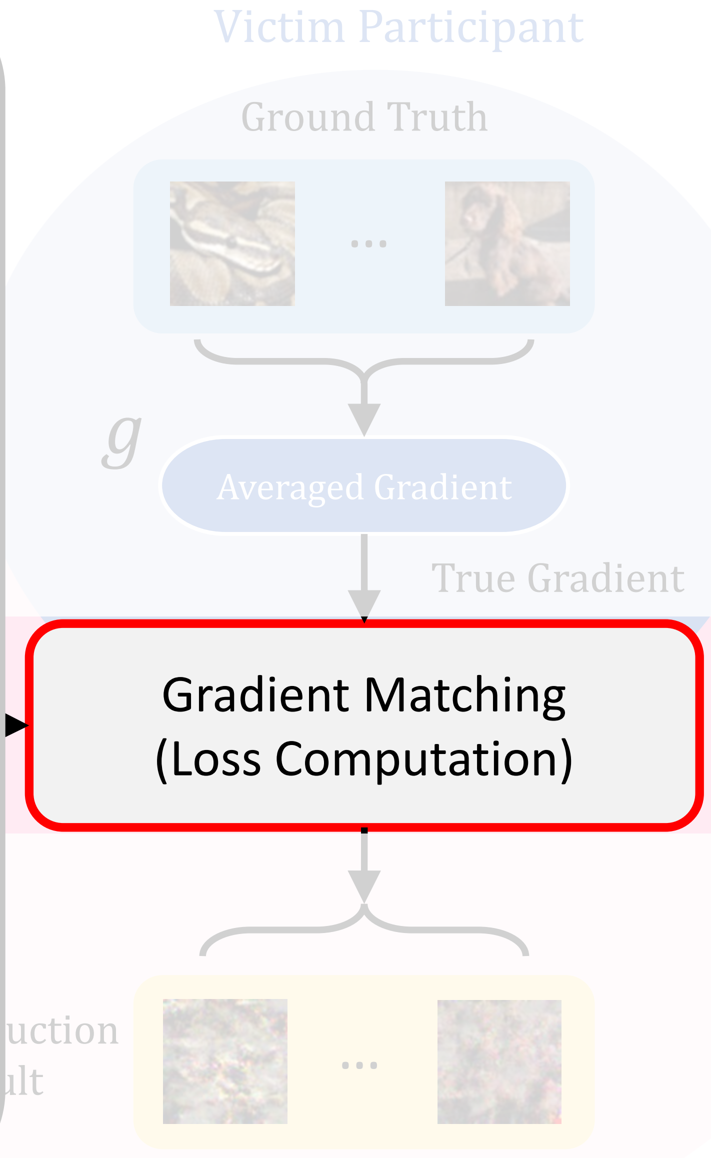
Why Gradient Inversion is hard?

$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B) \\ \in \mathbb{R}^m \times \{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B} \nabla \ell(f_\theta(x_j), y_j), g\right)$$

Source Separation

+ Compressed Sensing

→ Under-determined Problem!



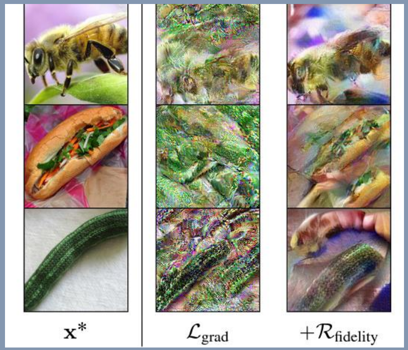
Priors can help attacks

Total Variation (Geiping et al.)

$$R_{\mathrm{TV}}(x) := \sum_{(i,j)} \sum_{(i',j') \in \partial(i,j)} \|x(i,j) - x(i',j')\|^2$$

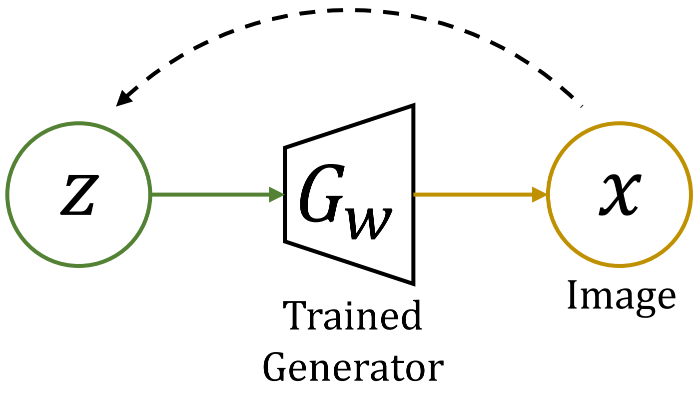
BN statistics (Yin et al.)

$$R_{\mathrm{BN}}(x_1, \ldots, x_B; \theta) := \sum_l \|\mu_l - \mu_{l,\mathrm{exact}}\|_2 + \|\sigma_l^2 - \sigma_{l,\mathrm{exact}}^2\|_2$$

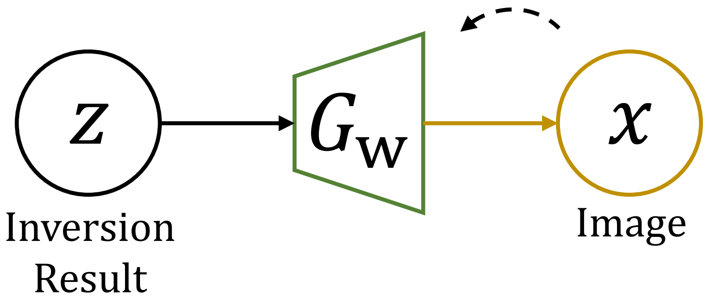


Gradient Inversion on Alternative Space

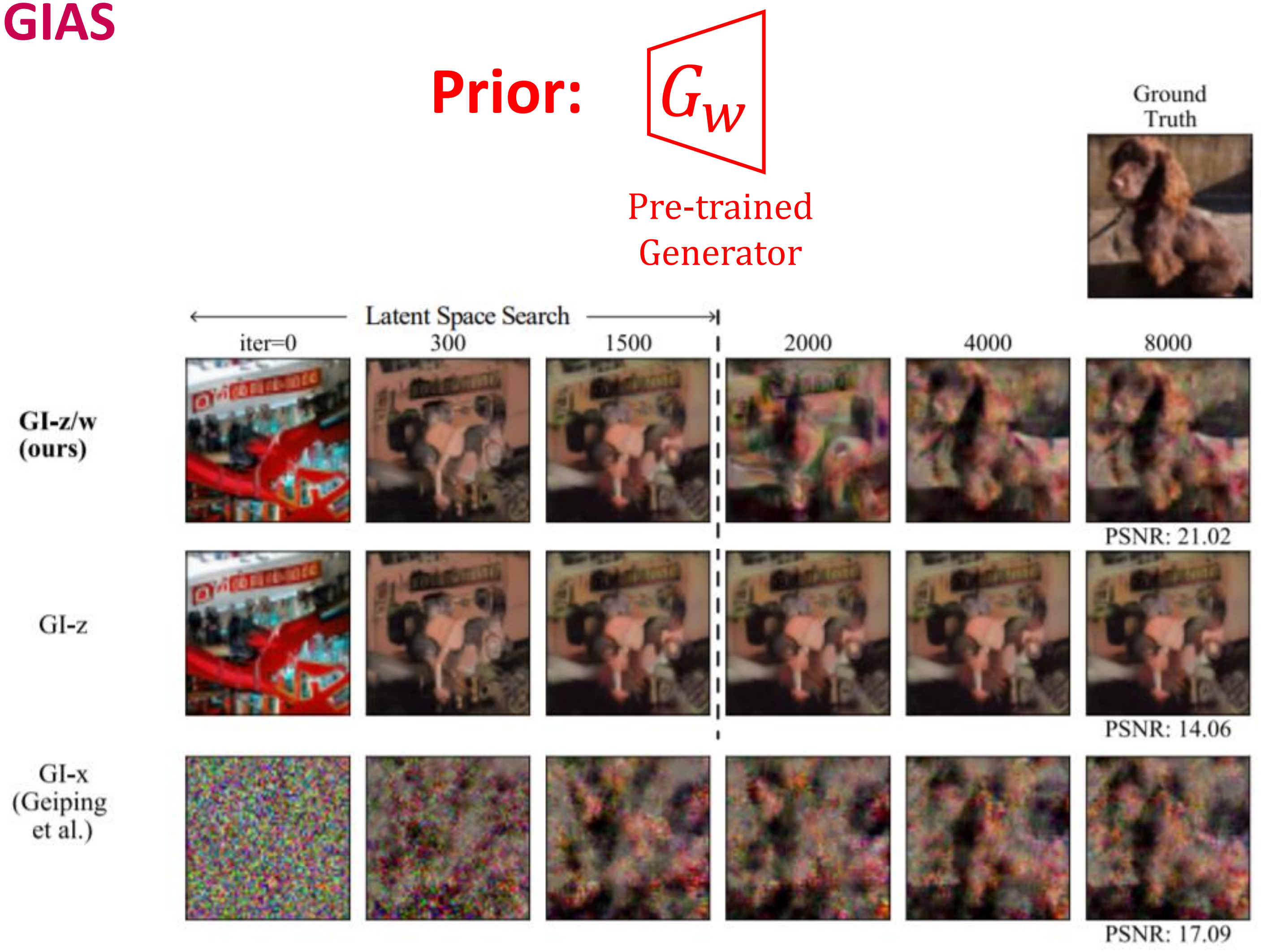
GAN Inversion



Generative Prior



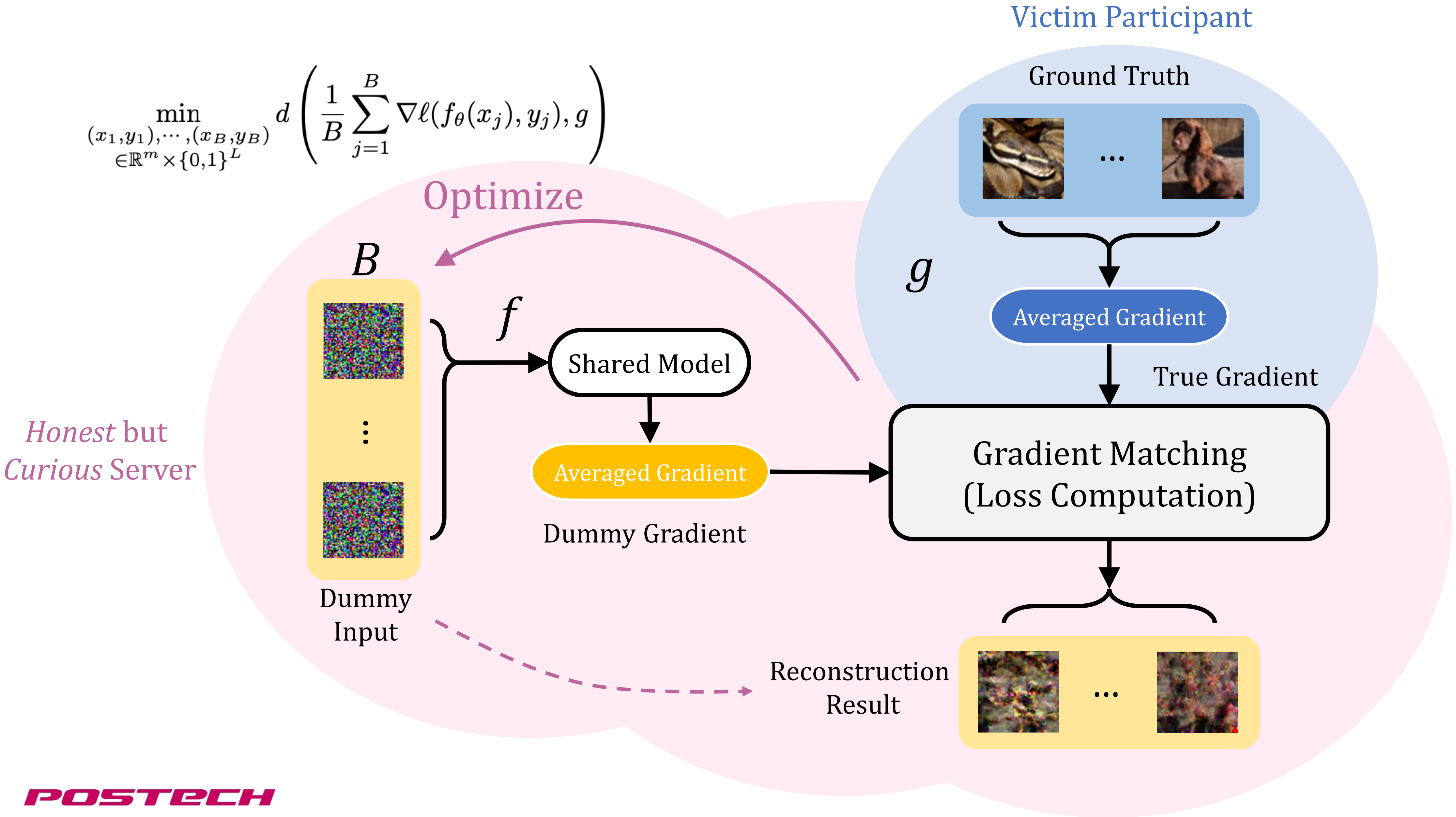
GIAS



POSTECH

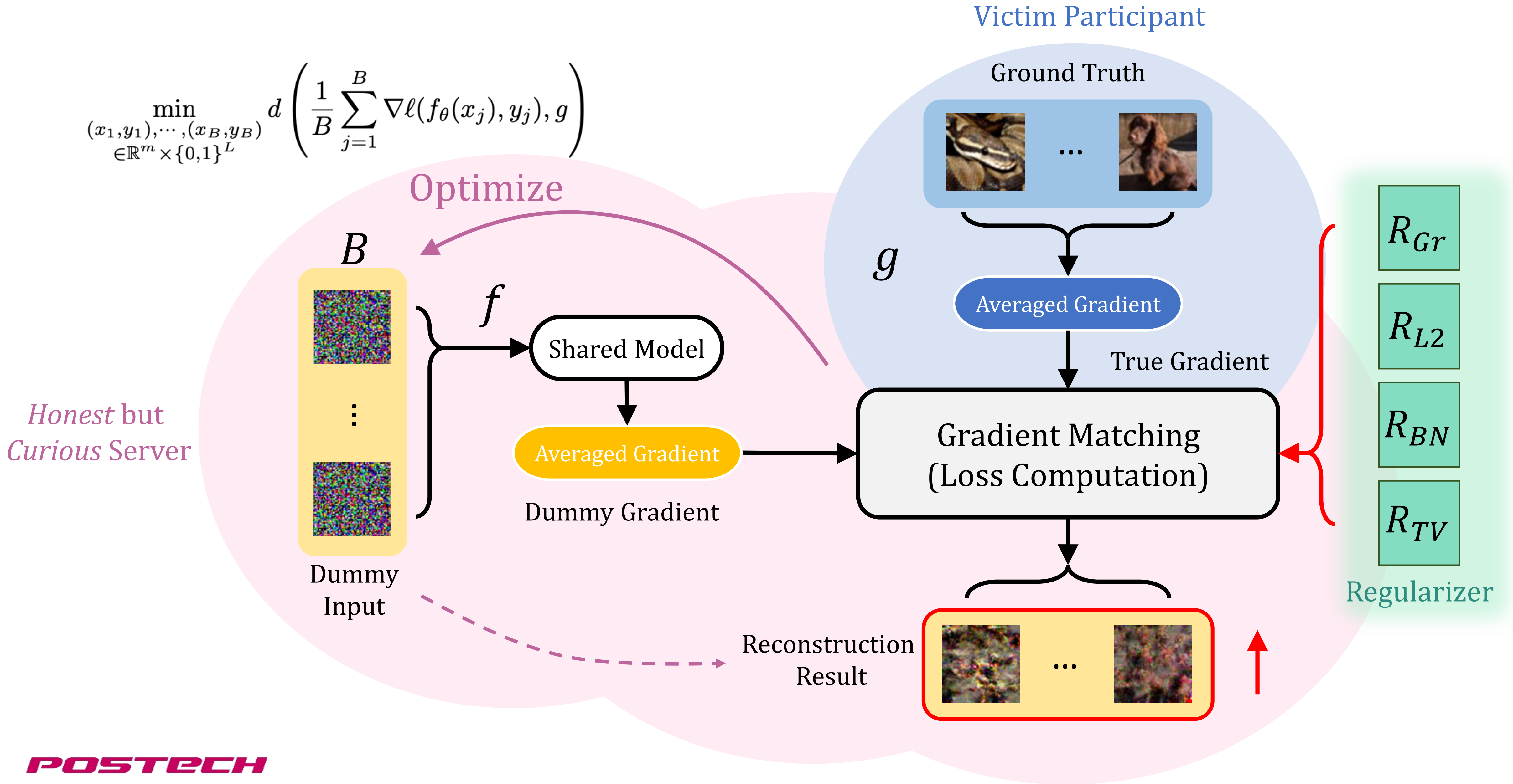
Basic Gradient Inversion

$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B) \\ \in \mathbb{R}^m \times \{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B} \nabla \ell(f_\theta(x_j), y_j), g\right)$$



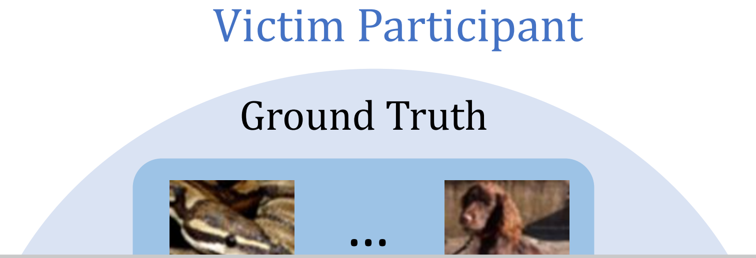
Gradient Inversion with Regularizers

$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B)\\ \in \mathbb{R}^m \times \{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B} \nabla \ell(f_\theta(x_j), y_j), g\right)$$

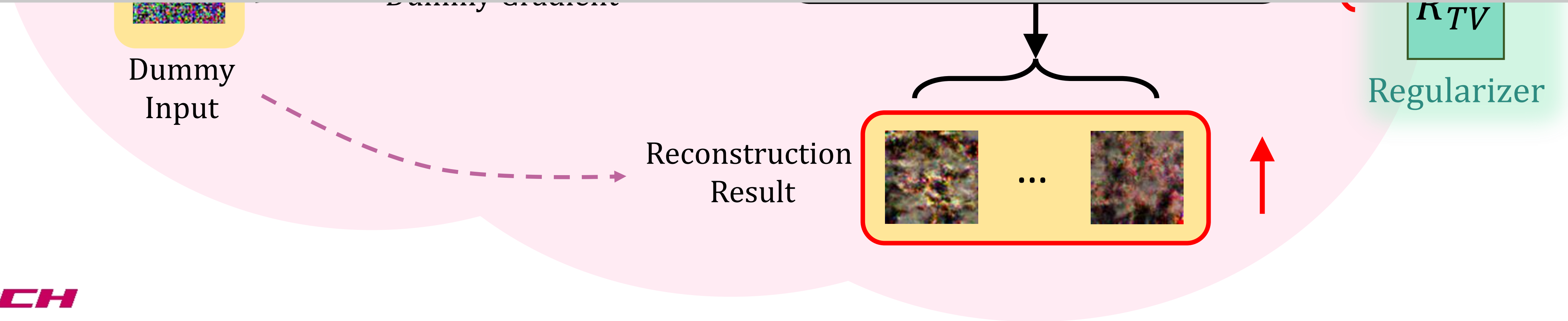


Gradient Inversion with Regularizers

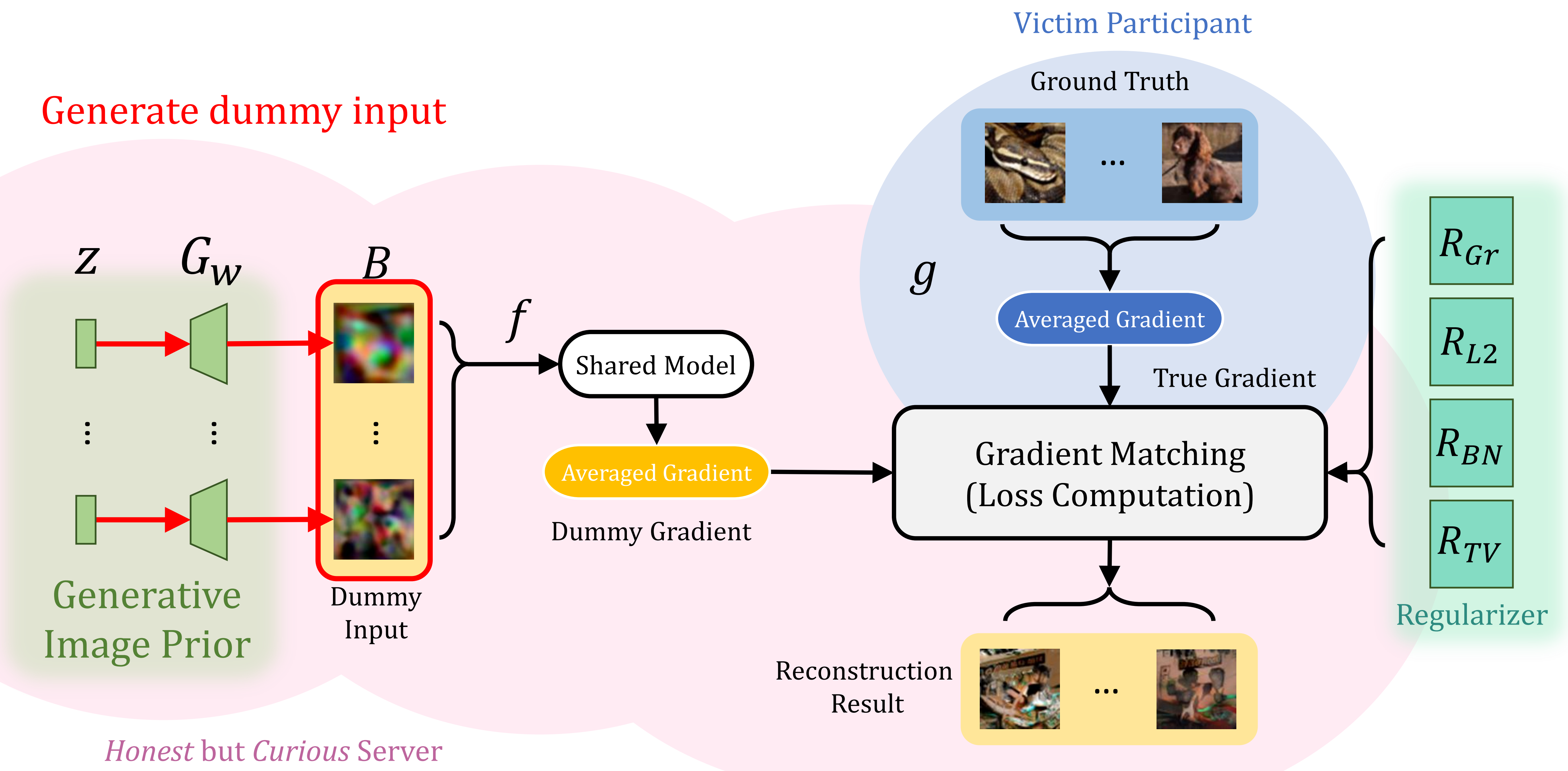
$$\min_{\substack{(x_1,y_1),\cdots,(x_B,y_B) \\ \in\mathbb{R}^m\times\{0,1\}^L}} d\left(\frac{1}{B}\sum_{j=1}^{B}\nabla\ell(f_\theta(x_j),y_j), g\right)$$



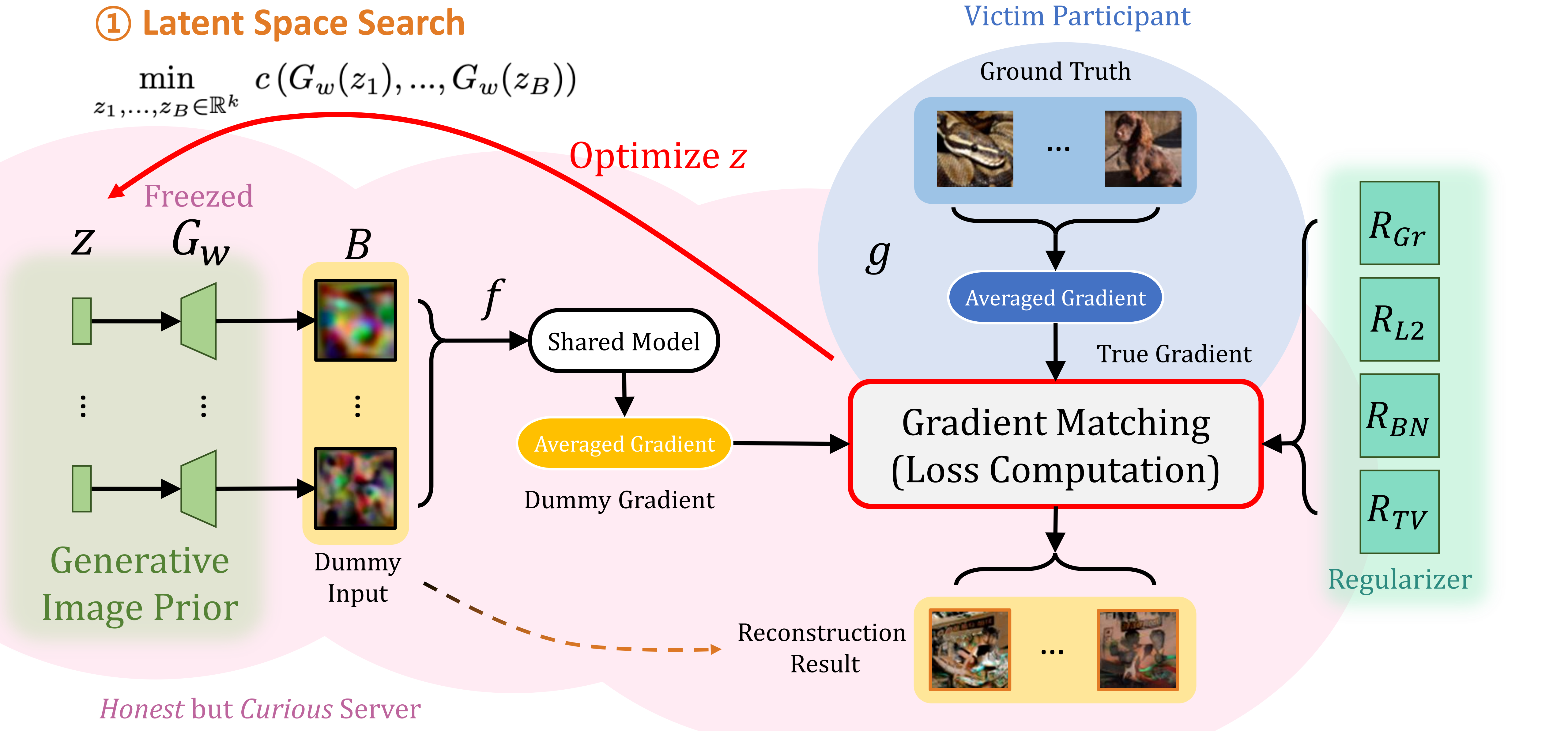
*What if we make the algorithm **only** explore among the **natural images**?*



Gradient Inversion on Alternative Space



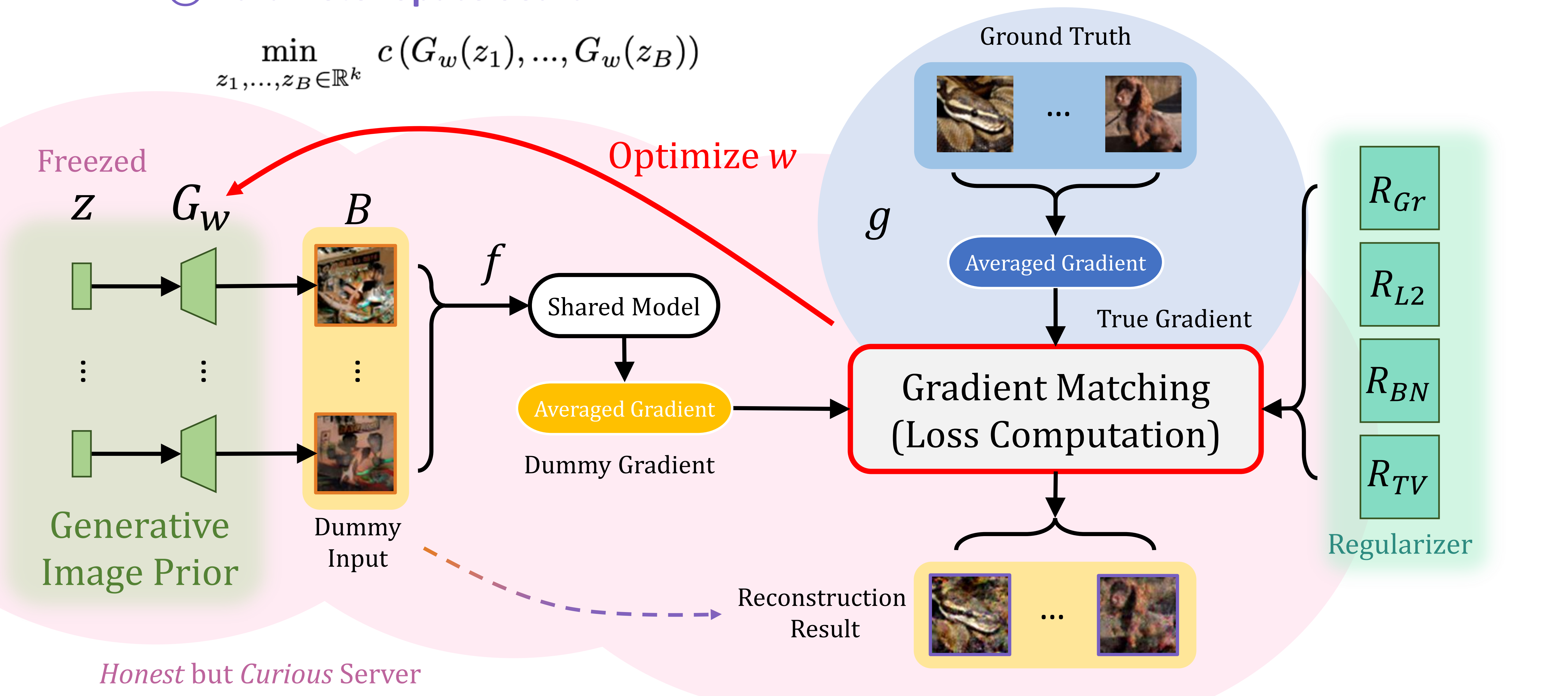
Gradient Inversion on Alternative Space



Gradient Inversion on Alternative Space

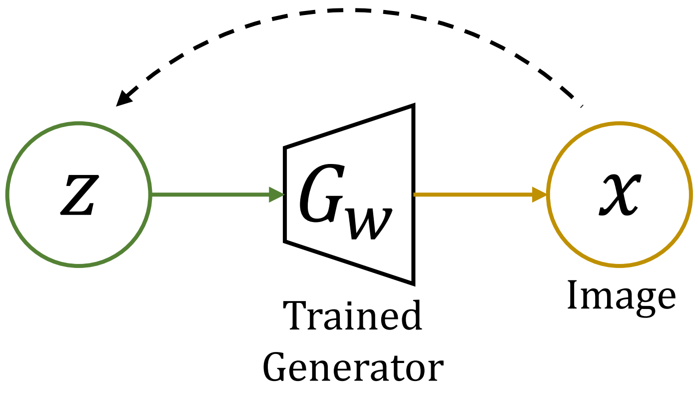
② Parameter Space Search

$$\min_{z_1,\dots,z_B \in \mathbb{R}^k} c\left(G_w(z_1), \dots, G_w(z_B)\right)$$

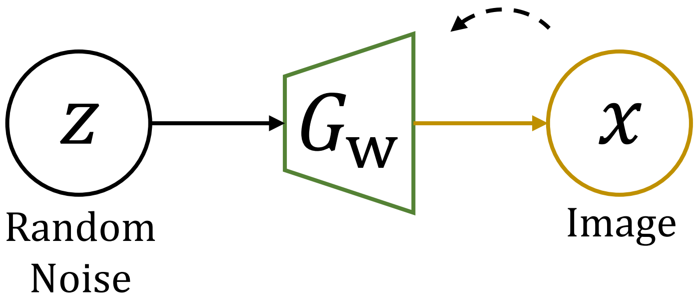


Gradient Inversion with Generative Image Prior

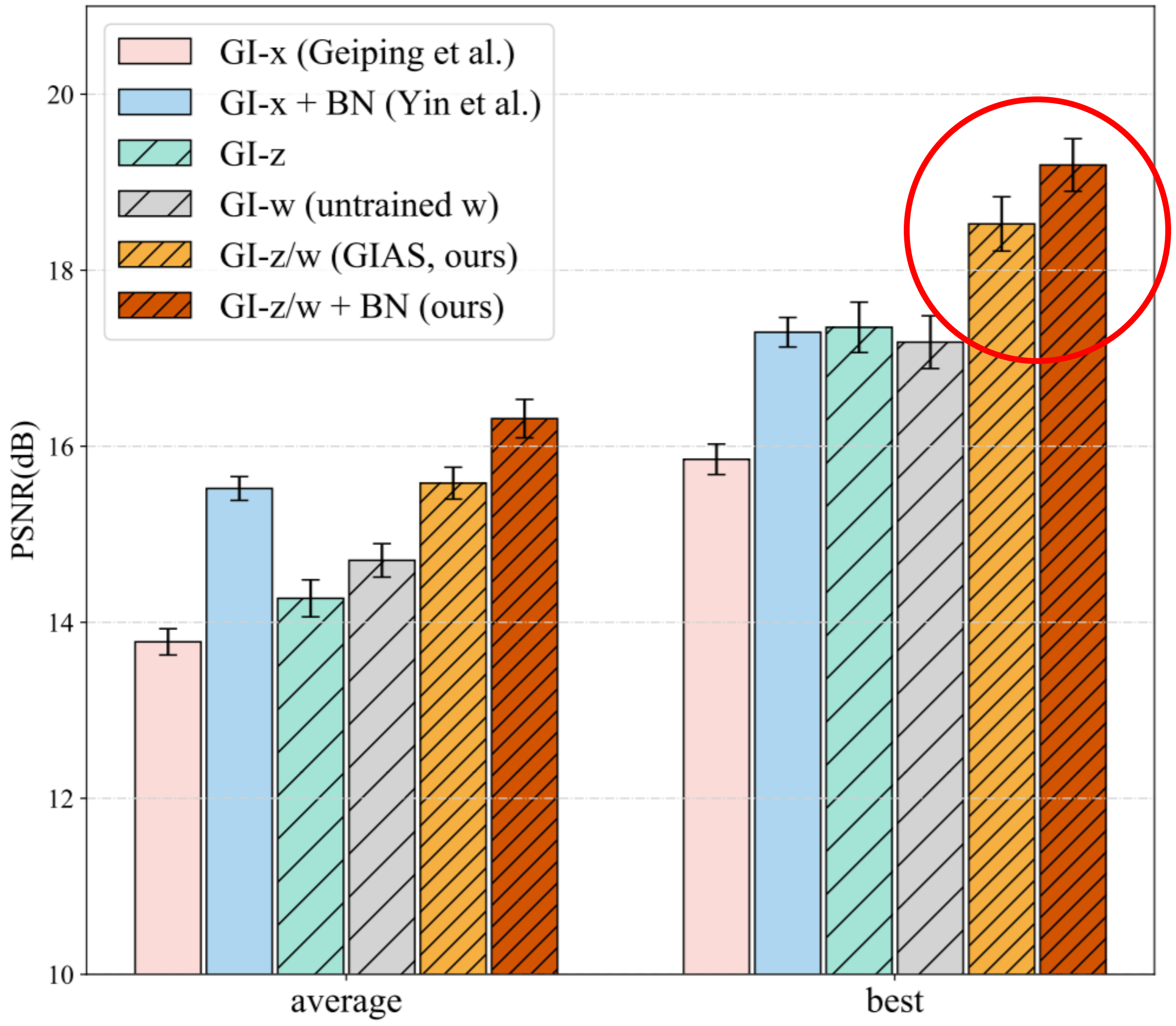
GAN Inversion



Deep Image Prior



GIAS

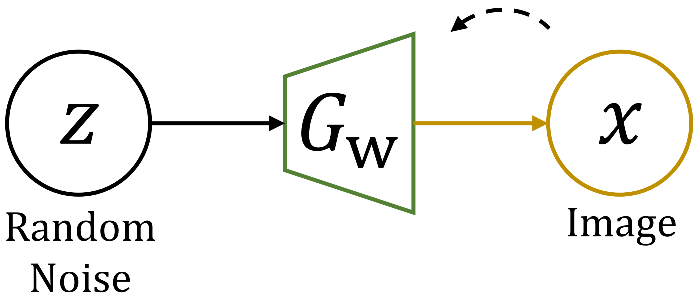


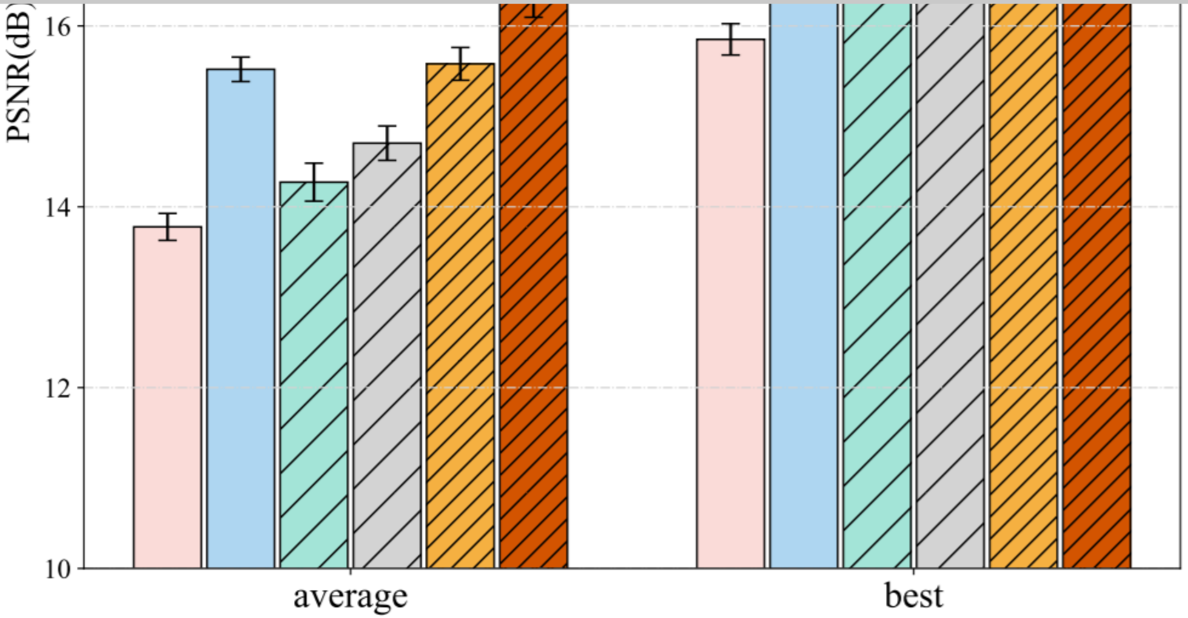
POSTECH

Gradient Inversion with Generative Image Prior

*How to **get a pre-trained model** with **unknown dataset**?*

Deep Image Prior





POSTECH

Gradient Inversion with Generative Image Prior

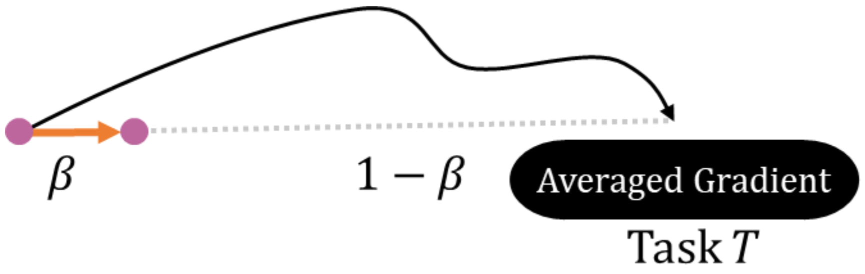
*How to **get a pre-trained model** with **unknown dataset**?*

*Is there any **advantage** of doing **a series of Gradient Inversions**?*

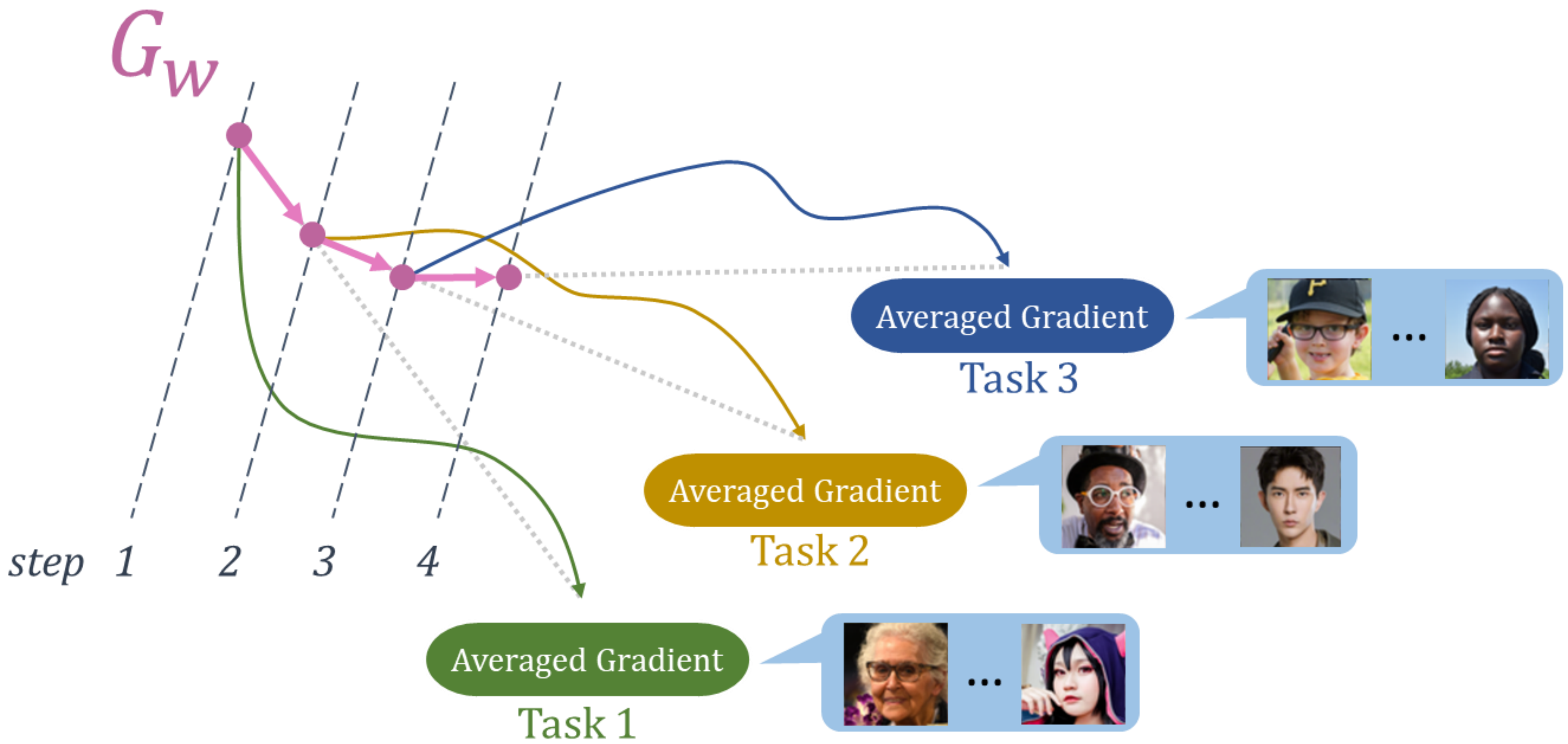
Gradient Inversion to Meta-Learn

First-Order Meta-Learning: Reptile

(A. Nichol et al.)



GIML

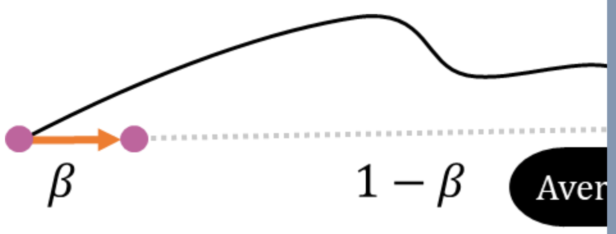


POSTECH

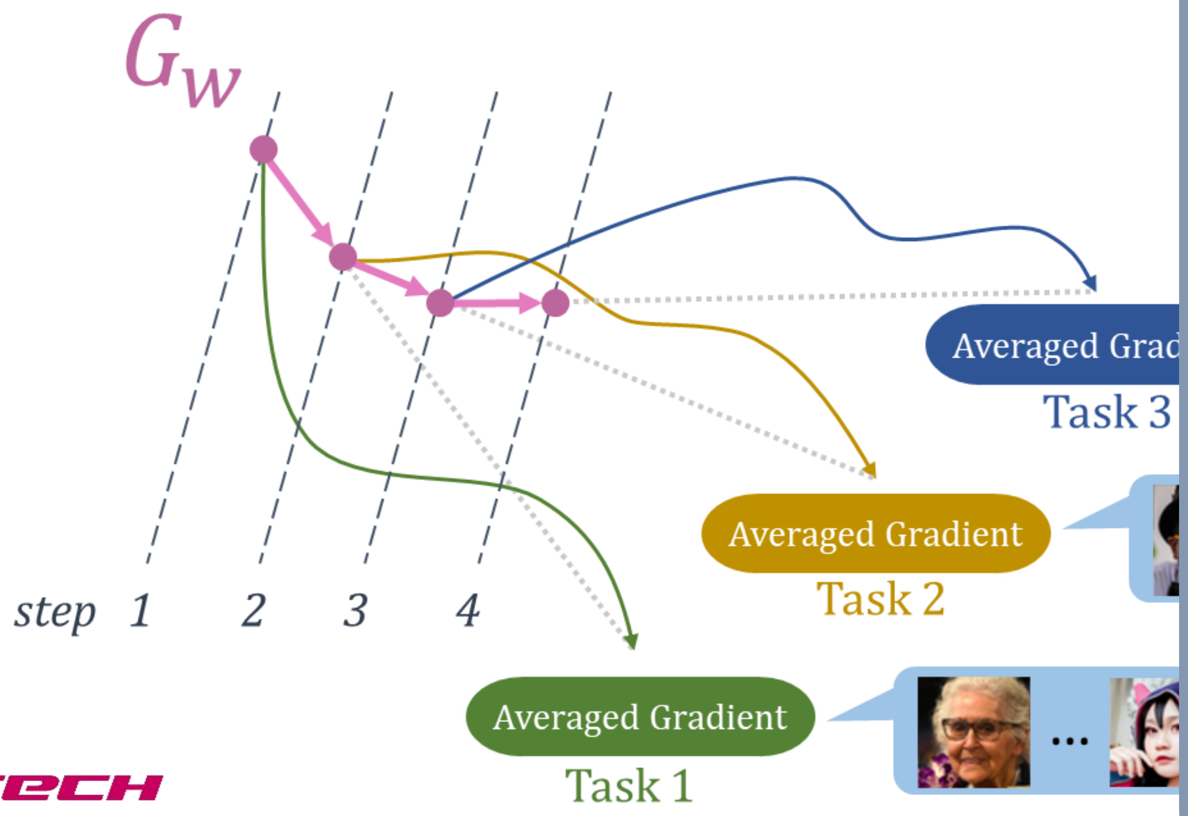
Gradient Inversion to Meta-Learn

First-Order Meta-Learning: Reptile

(A. Nichol et al.)

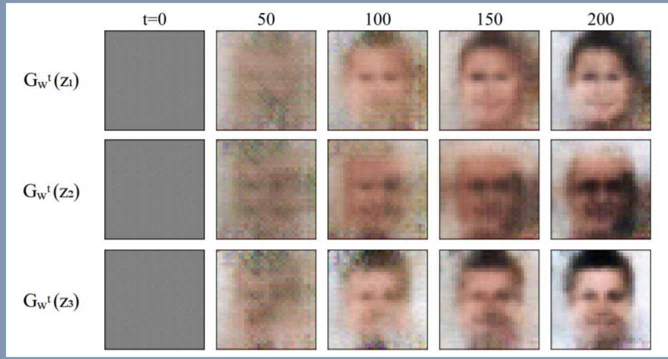


GIML



GIAS using GIML

Prior:

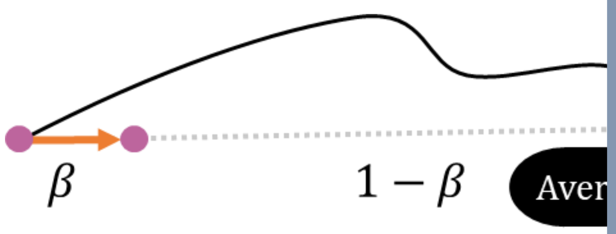


Meta-Learned Generator

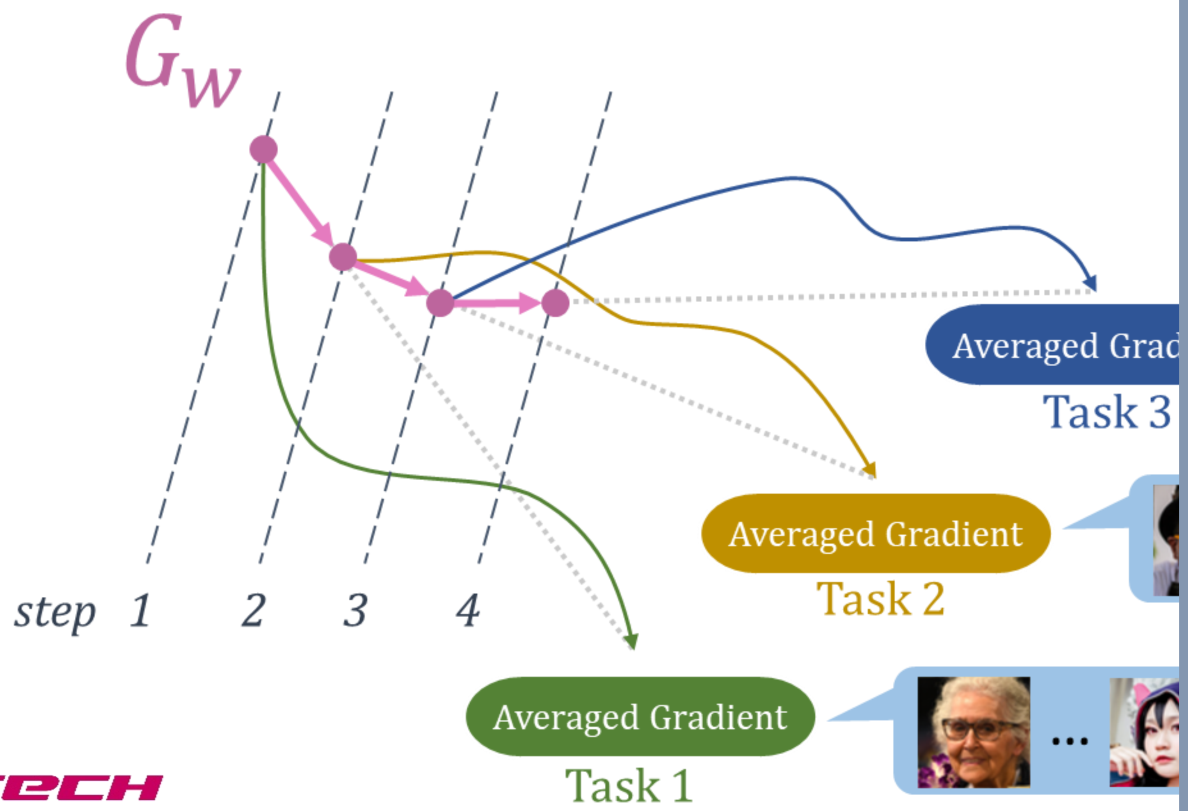
Gradient Inversion to Meta-Learn

First-Order Meta-Learning: Reptile

(A. Nichol et al.)

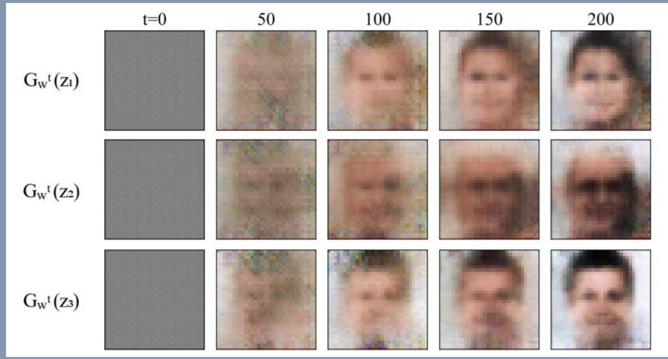


GIML



GIAS using GIML

Prior:



Meta-Learned Generator

Task T

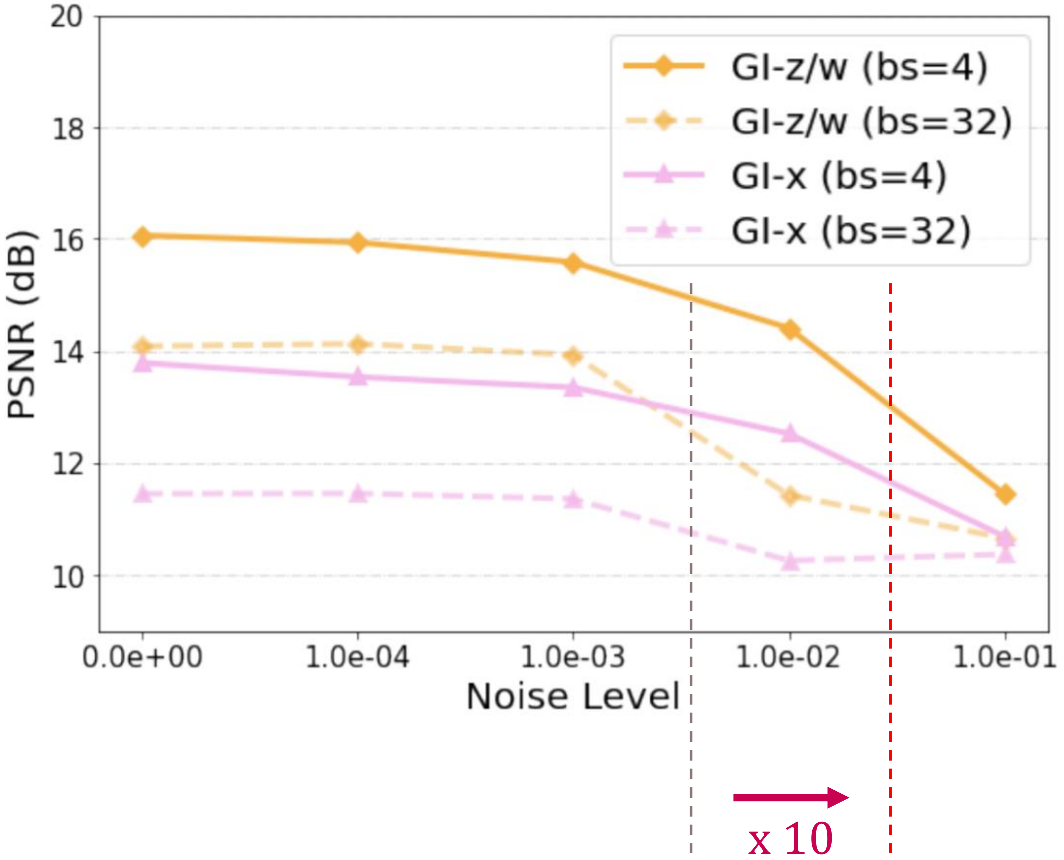
Not Trained

Well Trained

Wrong Prior

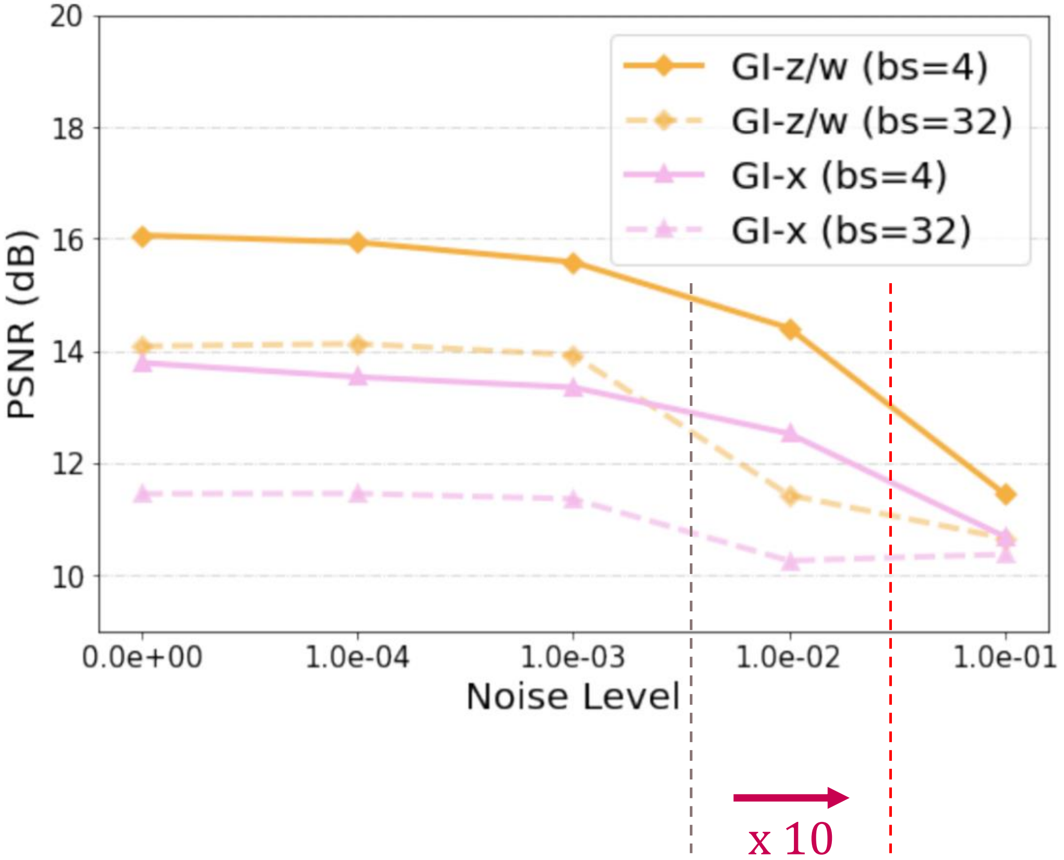
How to Protect Gradient Inversion Attacks?

Differential Privacy – Noisy Gradient

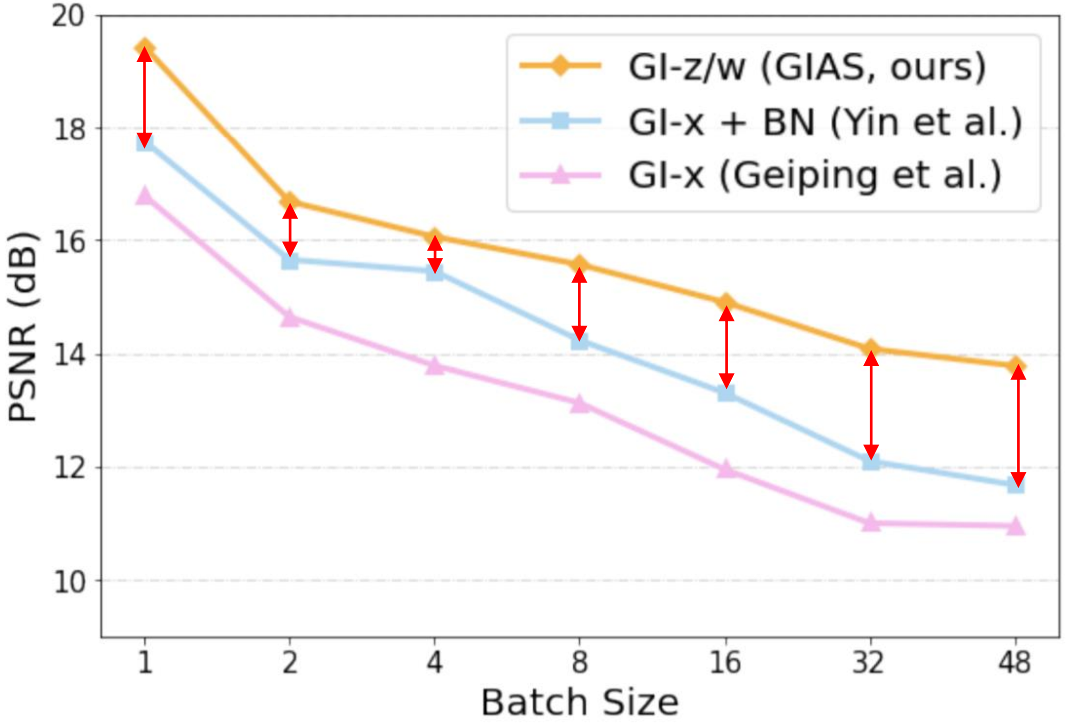


How to Protect Gradient Inversion Attacks?

Differential Privacy – Noisy Gradient



Large batch size



How to Protect Gradient Inversion Attacks?

Takeaways

Conclusion

- Priors help to solve under-determined problems like Gradient Inversion
- Utilizing priors gives us significant benefits to accomplish our goal
- Learning prior of the data via Gradient Inversion is possible

Our Contribution

- Propose the necessity of a higher standard on privacy
- Warn FL practitioners to choose more conservative choice of defense mechanisms

Machine
Learning
Laboratory

Thank You :)

Jinwoo Jeon[*1], Jaechang Kim[*2], Kangwook Lee[3], Sewoong Oh[4], Jungseul Ok[12]

[*]contributed equally
[1]Department of Computer Science and Engineering, POSTECH
[2]Graduate School of Artificial Intelligence, POSTECH
[3]Department of Electrical and Computer Engineering, University of Wisconsin-Madison
[4]Paul G. Allen School of Computer Science & Engineering, University of Washington