Private Image Reconstruction from System Side Channels using Generative Models

Yuanyuan Yuan[1,2], Shuai Wang[1], Junping Zhang[2]

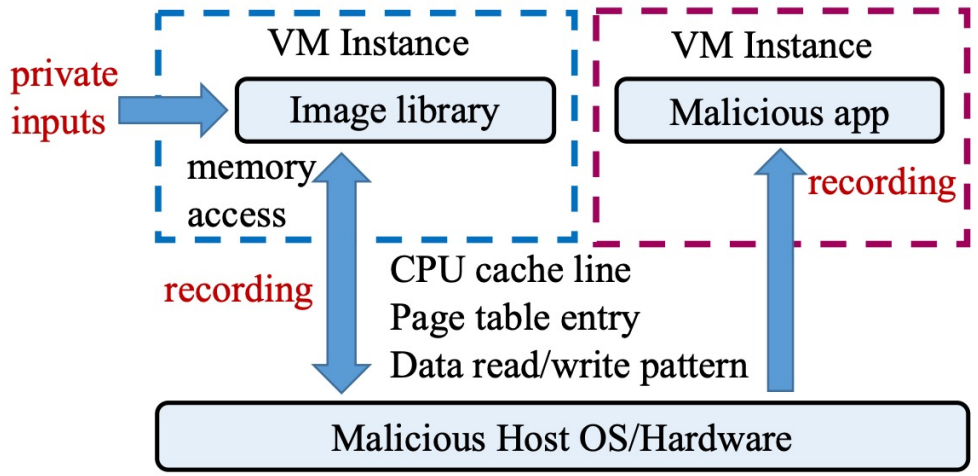
[1]The Hong Kong University of Science and Technology, [2]Fudan University

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FUDAN UNIVERSITY 1905

What is Side Channel Analysis?

Recover privacy from program's non-functional characteristics



Online Phase

Can be launched automatically!

Comprehension of victim software:

1. Source code
2. Analyze manually

- ❖ Program specific
- ❖ Domain expert

Logged side channel trace are:

1. Lengthy
2. Noisy
3. Sparse

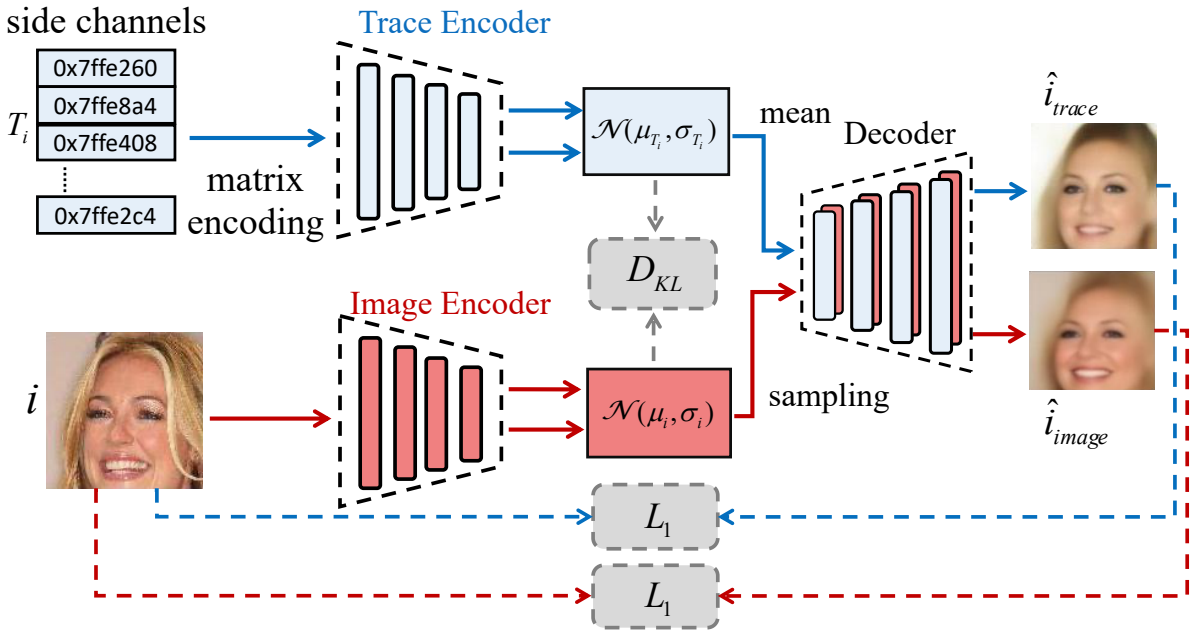
- ❖ Error prone
- ❖ Low quality
- ❖ Small scale

Offline Phase

Can be performed in an effective and fully automated way?

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

The Proposed Framework



VAE-LP Module

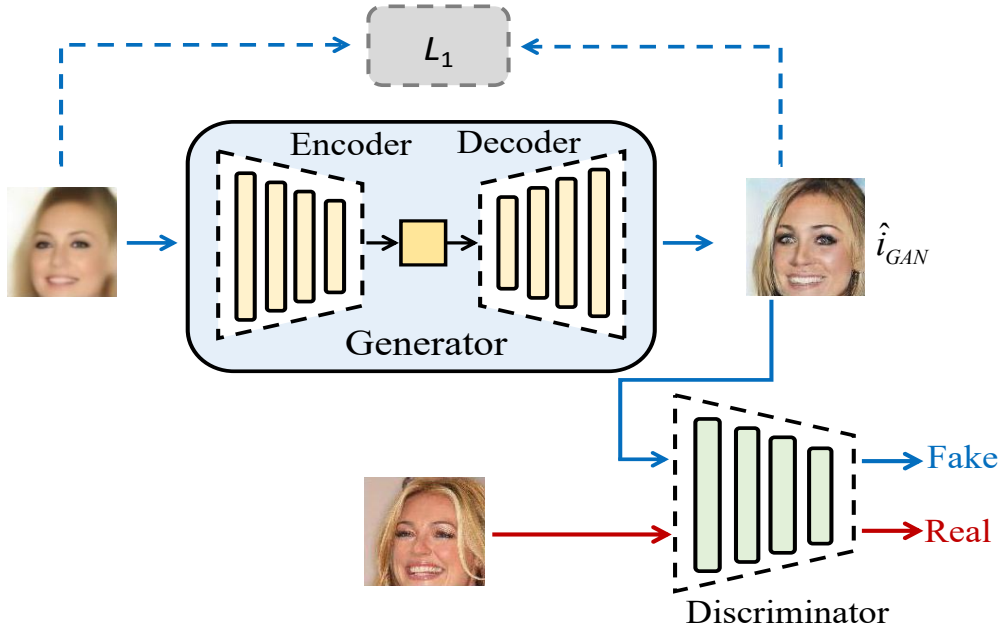
1. The upper dataflow extracts secret-dependent records from side channels.

2. The lower dataflow encodes and decodes reference image. The shared decoder works as a guideline.

Testing phase:

- Only retains the upper dataflow marked in blue.
- Side channels → Private images

The Proposed Framework

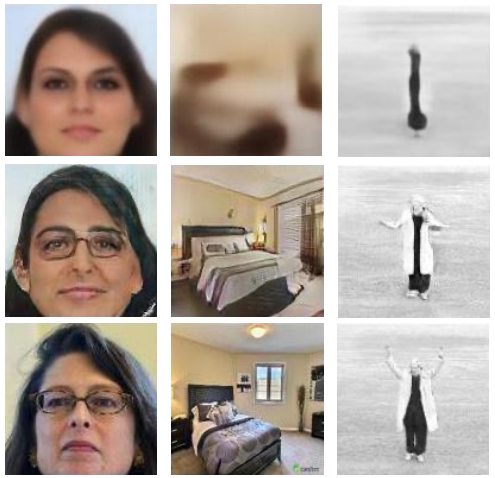


GAN Module

Accepts output from VAE-LP module and refine it.

Why refine?

Human adversaries!

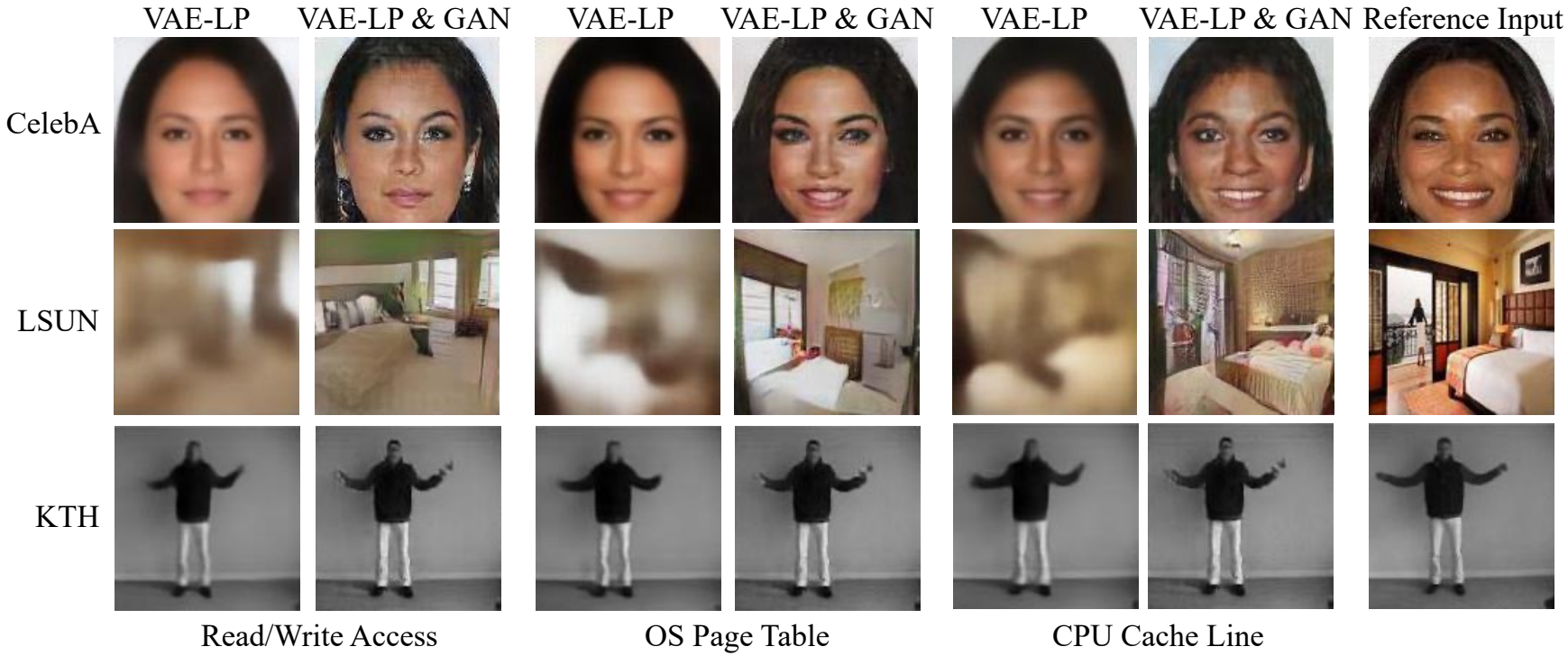


VAE-LP Module

GAN Module

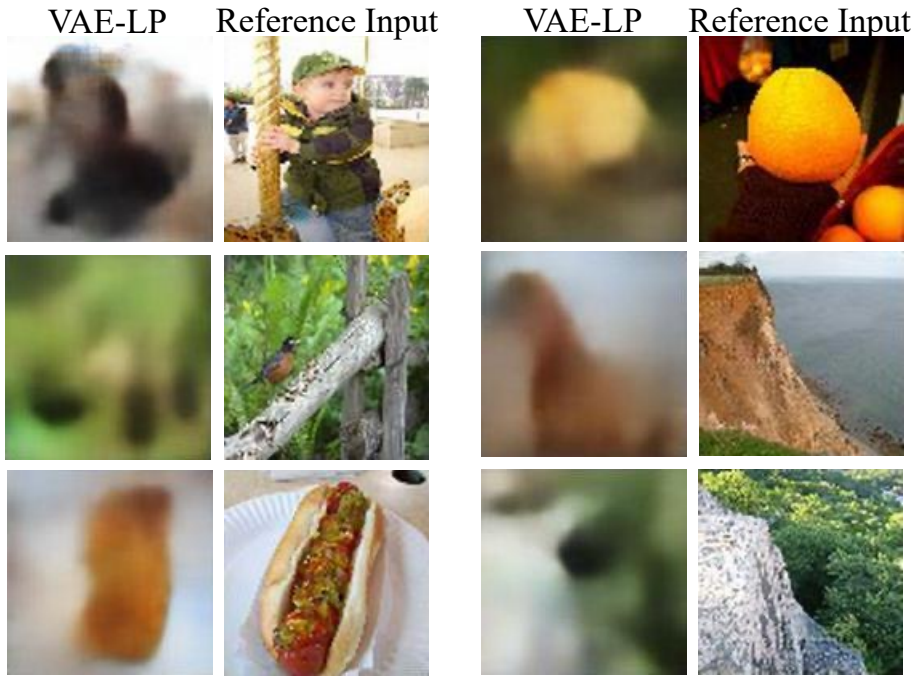
Reference Input

Qualitative Evaluation



Reconstructed images on three datasets using three side channels.

Qualitative Evaluation



1. Images are mixed
2. No class label is provided

Reconstructed images on mini-ImageNet using CPU cacheline.

Quantitative Evaluation

System side channel	$K = 1, N = 100$			$K = 5, N = 100$			$K = 20, N = 100$		
	CelebA	KTH	LSUN	CelebA	KTH	LSUN	CelebA	KTH	LSUN
CPU Cache Line	47.28%	38.22%	10.00%	74.98%	63.56%	28.74%	90.70%	83.72%	58.88%
OS Page Table	16.16%	35.86%	2.26%	37.96%	61.28%	8.62%	65.14%	82.22%	26.42%
Read/Write	19.58%	33.44%	2.24%	42.44%	58.30%	8.28%	67.82%	80.14%	26.64%

mini-ImageNet/(K, N)	(1, 100)	(5, 100)	(20, 100)
CPU Cache Line	18.8%	40.6%	66.0%

Given N images, whether reference input is in the top-k similar images using SSIM score.