IEEE 5th International Conference on Cryptography, Security and Privacy

Dimensionality-reduced Secure Outlier Detection on Union of Subspaces

Kunzan Liu

Dept. of Electronic Engineering, Tsinghua University

Email: lkz18@mails.tsinghua.edu.cn

Homepage: liukunzan.github.io

Co-Authors: Yuchen Jiao, Ye Jin, Xu Xiang, and Yuantao Gu

Jan. 8-10, 2021

Overview

IEEE 5th International Conference on Cryptography, Security and Privacy

Dimensionality-reduced Secure Outlier Detection on Union of Subspaces

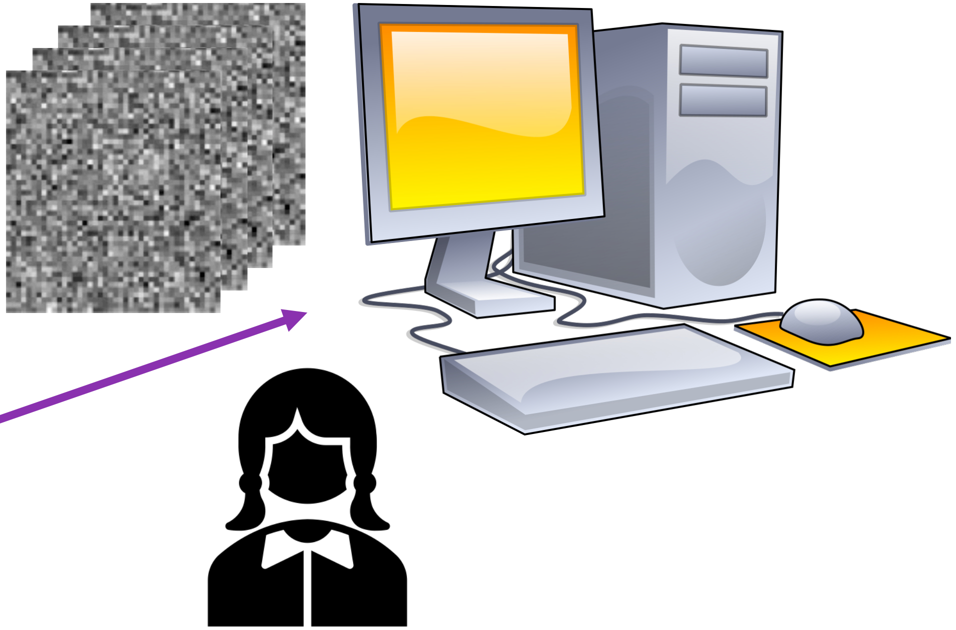
We propose a secure machine learning algorithm to process high-dimensional data.
Theoretical analysis is our main contribution.

Background

Accuracy

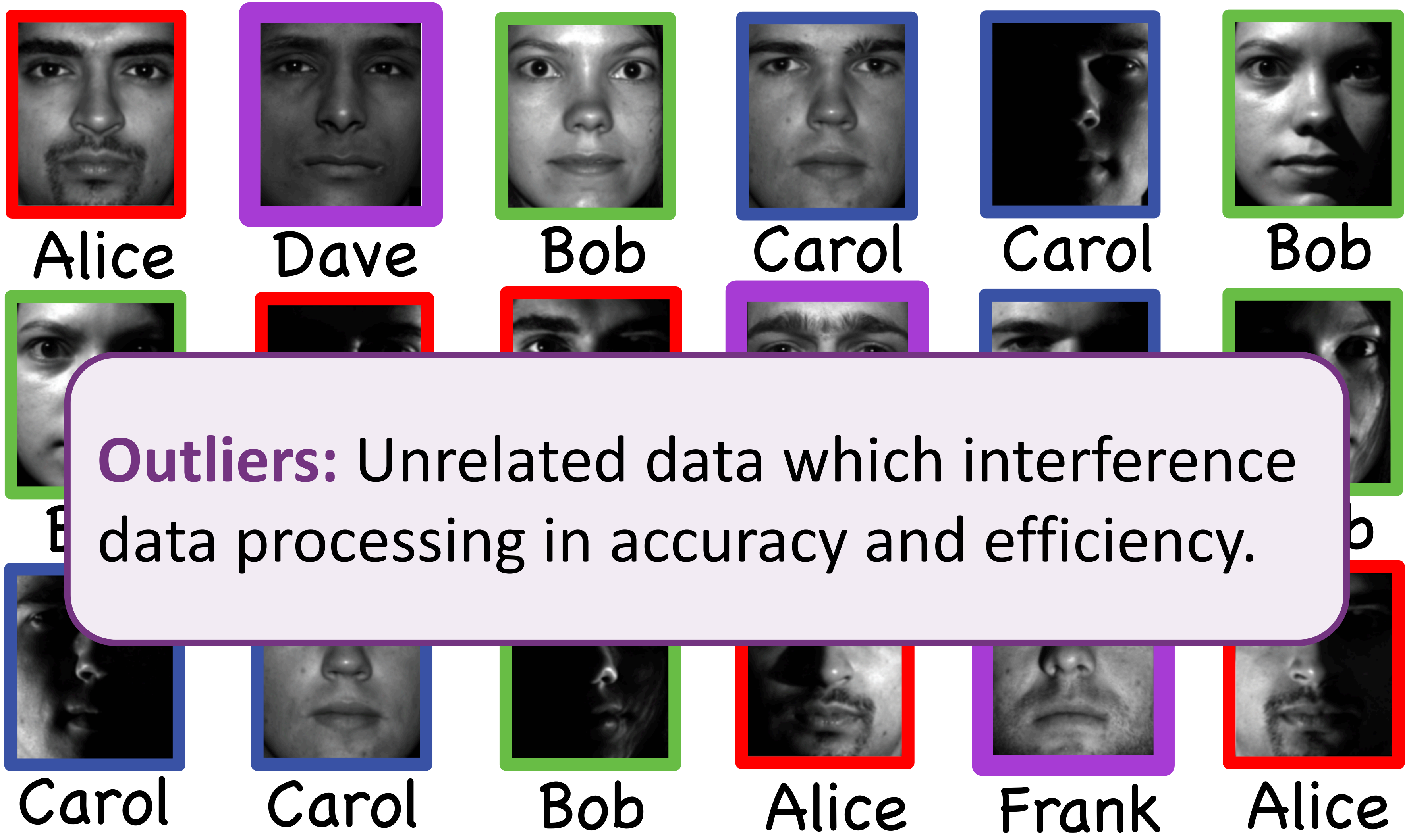
contradiction

Privacy



Can we achieve data processing with high accuracy under an encrypted situation?

What is Outlier detection (OD)?

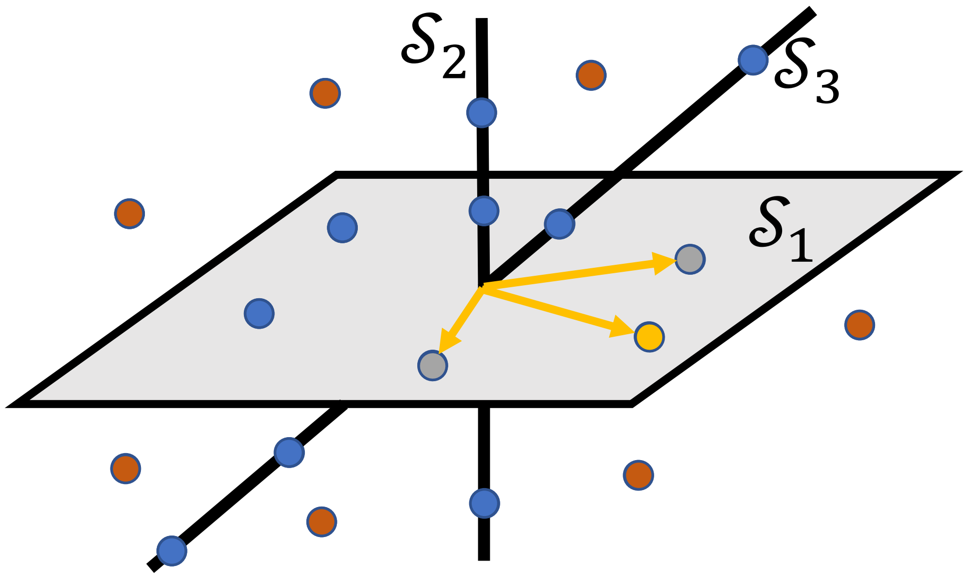


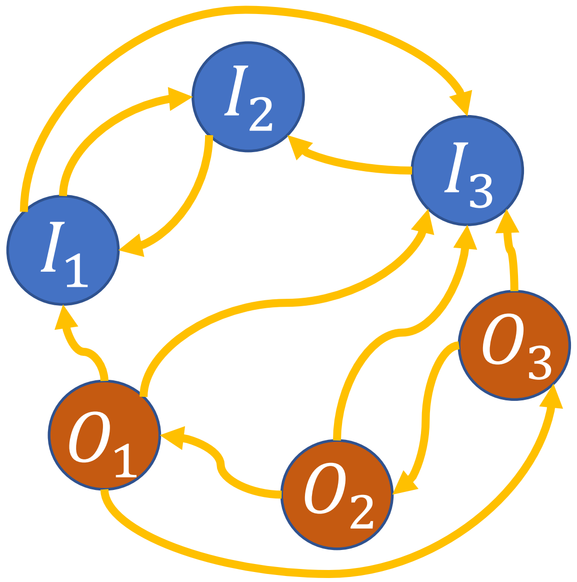
Outliers: Unrelated data which interference data processing in accuracy and efficiency.

- **Preliminaries**
- Algorithm
- Theorem
- Experiments
- Conclusion

Existing Methods?

$$\boldsymbol{r}_j = \arg\min_{c} \lambda \|\boldsymbol{c}\|_1 + \frac{1-\lambda}{2}\|\boldsymbol{c}\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{x}_j - \boldsymbol{X}\boldsymbol{c}\|_2^2$$





Step 1: Self-representation to obtain feature

Step 2: Random walk to identify outliers

Our Contribution

- Preliminaries
- Algorithm
- Theorem
- Experiments
- Conclusion

Dimensionality-reduced Secure Outlier Detection

in real-world dataset

- **Cons:** Efficiency declines significantly when dimension is high

Question: Can we cut down computational cost while preserve high accuracy?

[1] C. You, D. P. Robinson, and R. Vidal, "*Provable self-representation based outlier detection in a union of subspaces*," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4323–4332, 2017.

Random Projection

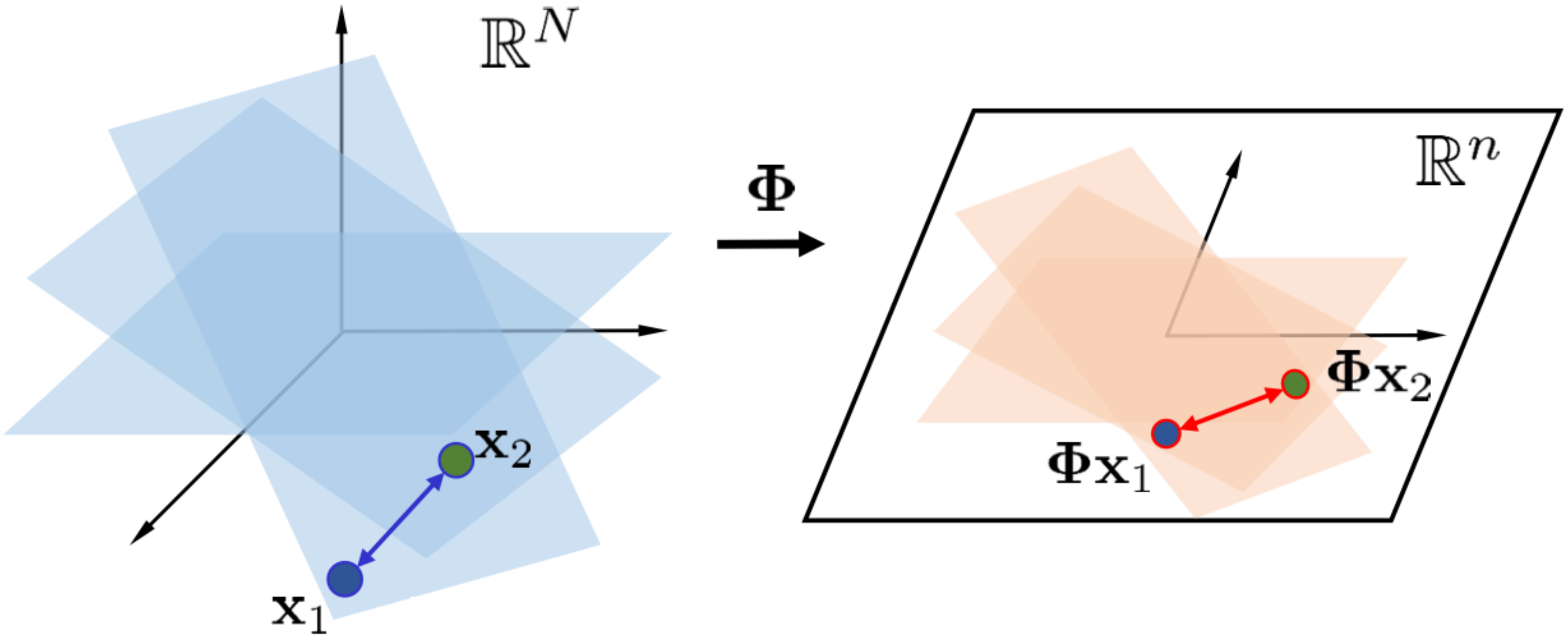


Fig: Random projection. It projects data from high-dim space $\mathbb{R}^N$ to low-dim space $\mathbb{R}^n$, and it is suitable for processing high-dim dataset which can preserve structure of dataset.

[2] G. Li, Q. Liu and Y. Gu, "*Rigorous restricted isometry property of low-dimensional subspaces*," Applied and Computational Harmonic Analysis, 2019.

Proposed Algorithm

Algorithm Dimensionality-reduced Secure Outlier Detection

Input: Dataset and parameters

Step 1: Random projection.

Step 2: Self-representation.

Step 3: Random walk.

Output: Ouliers.

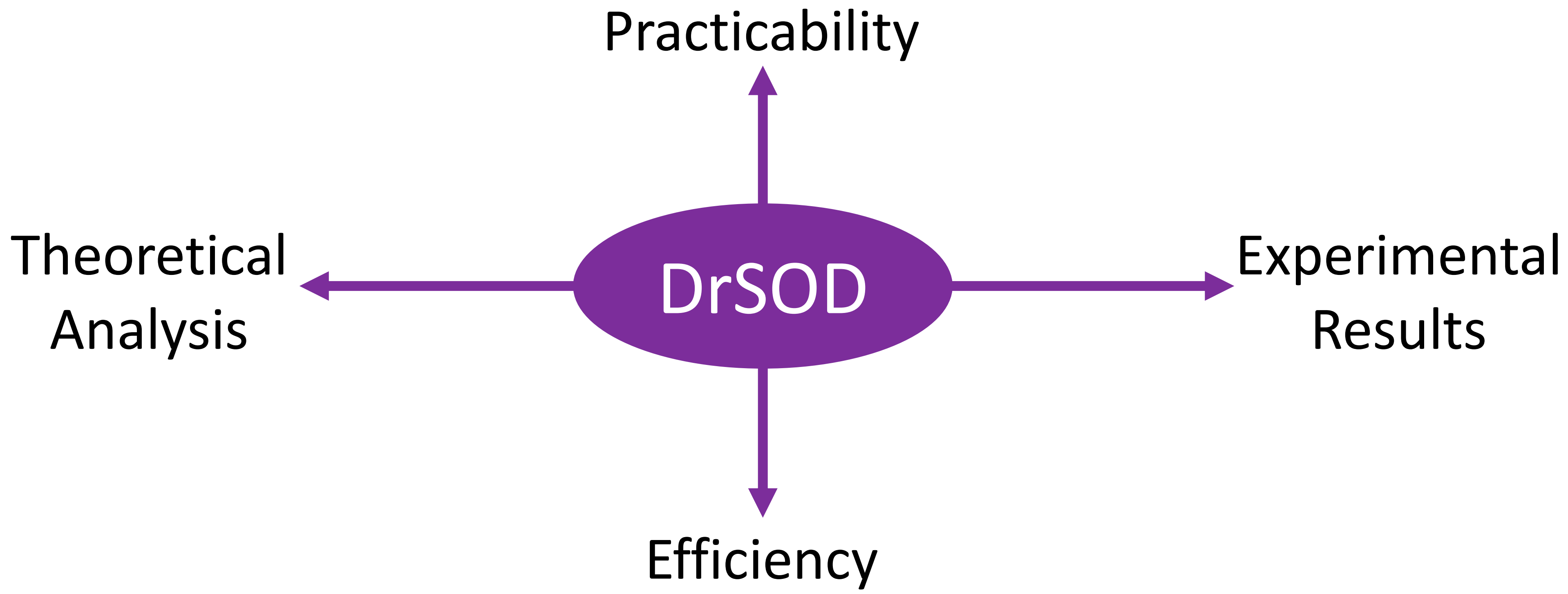
Representation Vector

Dataset

$$\boldsymbol{r}_j = \arg\min_{c} \lambda \|\boldsymbol{c}\|_1 + \frac{1-\lambda}{2}\|\boldsymbol{c}\|_2^2 + \frac{\gamma}{2}\left\|\boldsymbol{x}_j - \boldsymbol{X}\boldsymbol{c}\right\|_2^2$$

Comparison with SOTA Work

Question: Can we cut down computational cost while preserve high accuracy?



Theoretical Analysis

Notations N −Original Dim, n −Compressed Dim, M −Quantity
$\varepsilon \to 0$, c_1 and c_2 are positive constants

	SOTA		Ours
Algorit			ed OD
Assump			
Condit			$\lambda - \tau$ $M)$
Success Rate	1	$\approx$	$1 - \exp\left(-c_2\varepsilon^2 n\right) \to 1$
Complexity	$\mathcal{O}\left(N^2 M^3\right)$	$\gg$	$\mathcal{O}\left(n^2 M^3\right)$

Conclusion: Accuracy is preserved with probability 100% while computational cost is compressed significantly.

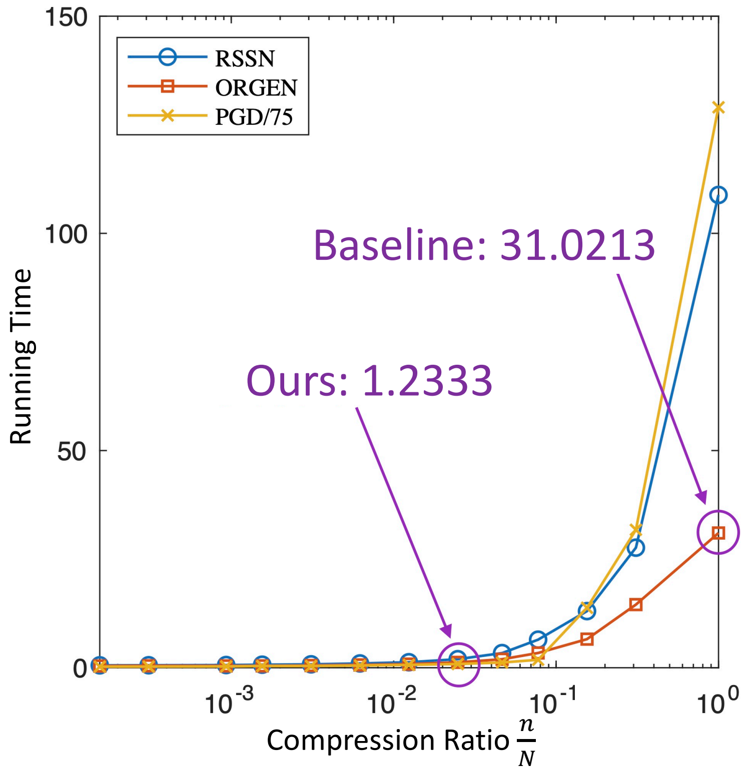
$$\tau := \max_j \left\{\varepsilon\left(1 + \|\boldsymbol{\delta}_j\|_2^2\right) + \sqrt{6\varepsilon}\,(1+\varepsilon)\,\gamma\,\|\boldsymbol{\delta}_j\|_2\right\} \to 0$$

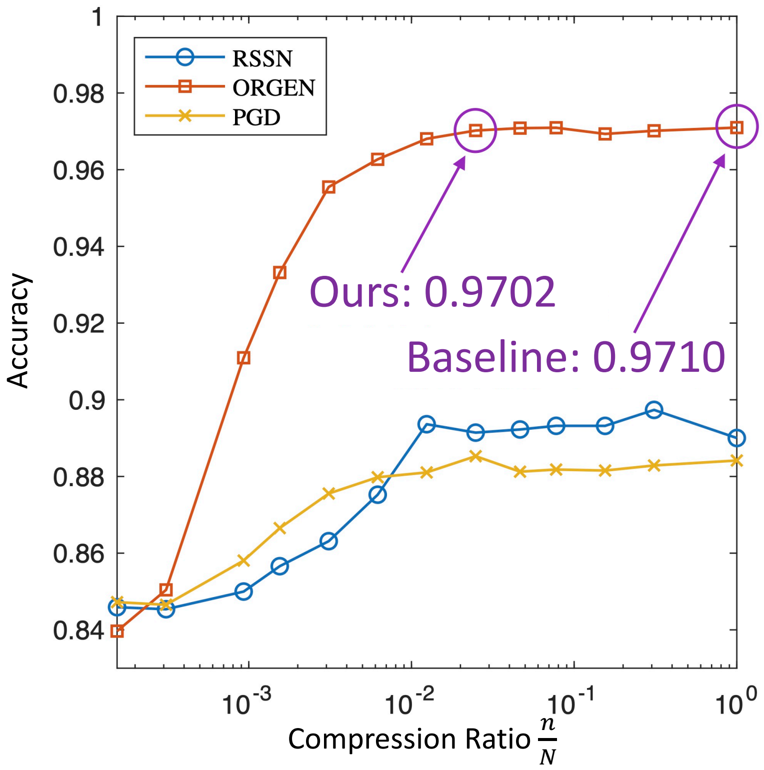
Experimental Results

Dataset Extended Yale Face B

32256×262

$$\boldsymbol{r}_j = \arg\min_{c} \lambda\|\boldsymbol{c}\|_1 + \frac{1-\lambda}{2}\|\boldsymbol{c}\|_2^2 + \frac{\gamma}{2}\left\|\boldsymbol{x}_j - \boldsymbol{X}\boldsymbol{c}\right\|_2^2$$





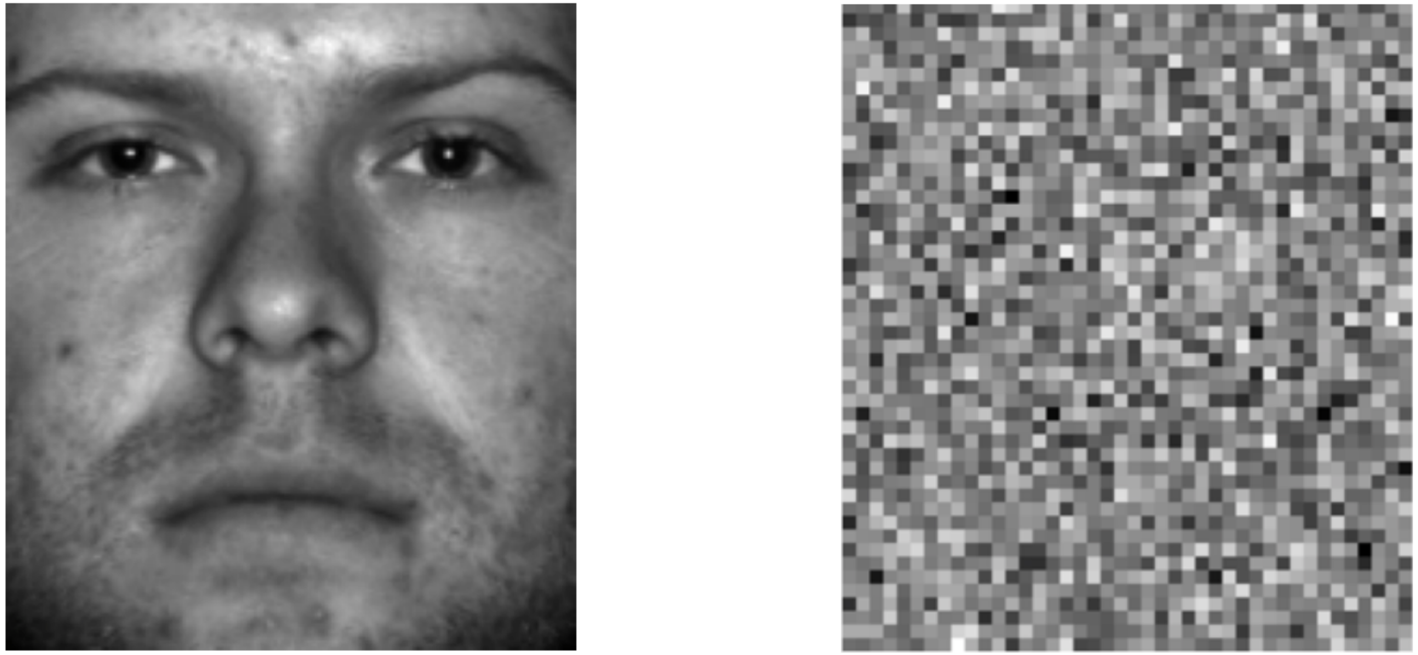
Running Time ↓ **96.03%**

Accuracy ↓ **0.08%**

- Preliminaries
- Algorithm
- Theorem
- **Experiments**
- Conclusion

Experimental Results

Dataset Extended Yale Face B



(a) original image (b) preprocessed image

Preprocessed image become unrecognizable!

Innovation

Theoretical

- First time: Dim-reduction is introduced to OD
- First time: Subspace preserving property is proved

Practical

- Motion segmentation
- Face clustering
- High acc + Fast + Less storage

Potential

Network compression / Damaged data processing

IEEE 5th International Conference on Cryptography, Security and Privacy

Dimensionality-reduced Secure Outlier Detection on Union of Subspaces

Kunzan Liu

Dept. of Electronic Engineering, Tsinghua University

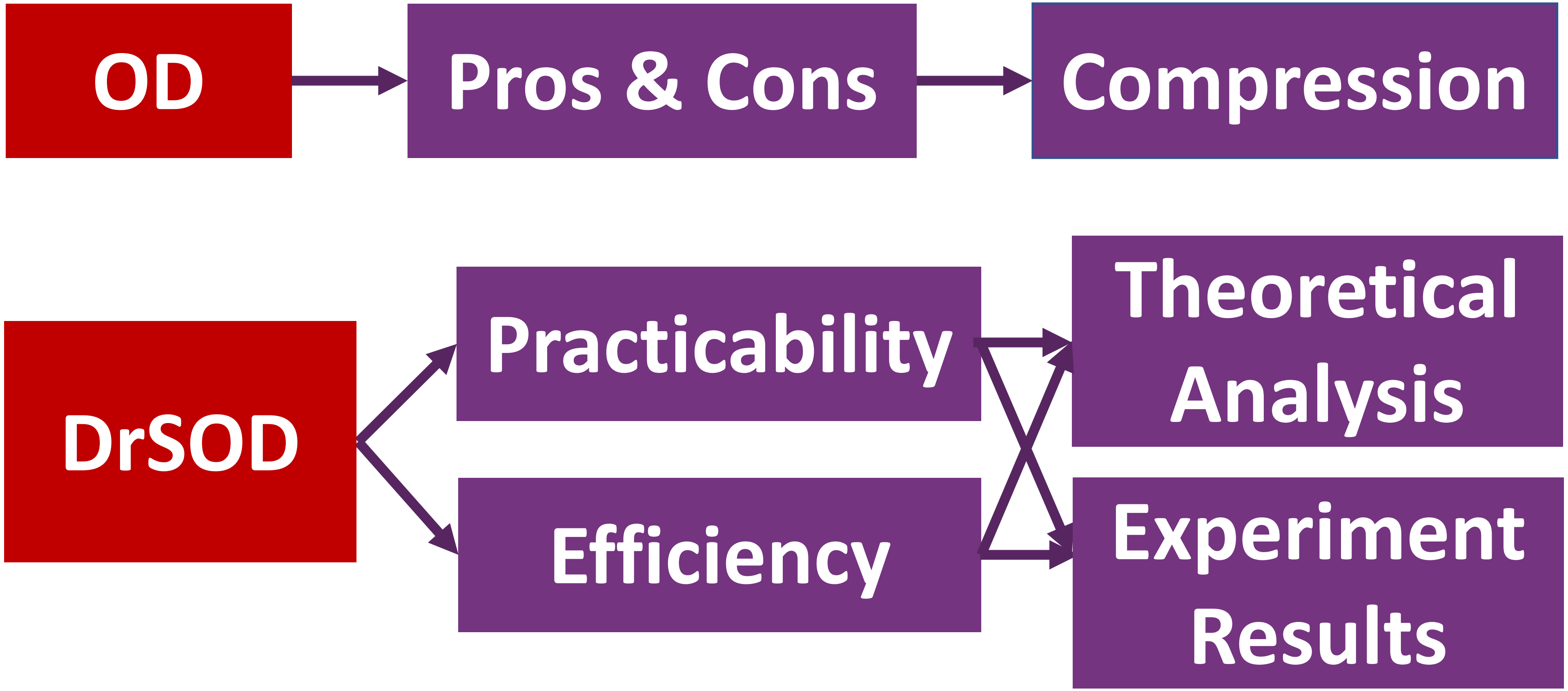
Email: lkz18@mails.tsinghua.edu.cn

Homepage: liukunzan.github.io

Co-Authors: Yuchen Jiao, Ye Jin, Xu Xiang, and Yuantao Gu

Jan. 8-10, 2021

Wrap Up


OD
Pros & Cons
Compression
DrSOD
Practicability
Efficiency
Theoretical Analysis
Experiment Results