Fingerprint Synthesis: Search with 100 Million Prints

Vishesh Mistry*
Michigan State University
East Lansing, MI, USA
mistryvi@msu.edu

Joshua J. Engelsma
Michigan State University
East Lansing, MI, USA
engelsm7@cse.msu.edu

Anil K. Jain
Michigan State University
East Lansing, MI, USA
jain@cse.msu.edu

***Presenter**

Supported by grant from the NIST Image Group

Need for a Large-scale Synthetic Fingerprint Dataset

- **Automated Fingerprint Identification Systems (AFIS)**
  - India's Aadhaar Project (1.26 billion ten-prints)
  - FBI's Next Generation Identification System (145.3 million ten-prints)

Person enrolling in Aadhaar

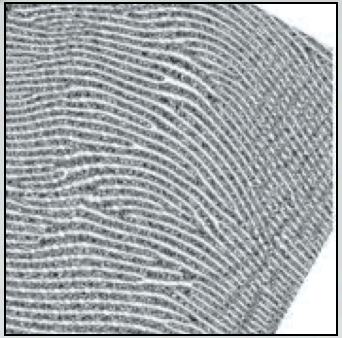
FBI's NGI system

Need for a Large-scale Synthetic Fingerprint Dataset

- Fingerprint search algorithms **evaluated on small-scale datasets**

- Evaluating against a gallery of 100 million fingerprints:
  - **Collection** of large-scale fingerprints
  - Obtain fingerprints from **forensic/government agencies**
  - **Synthesize** fingerprint images

Related Work in Fingerprint Synthesis



Zhao et al.
BTAS 2012

Johnson et al.
CVPR 2013

Cappelli et al.
IET 2018

Proposed: Rolled

Bontrager et al.
BTAS 2018

Cao and Jain
ICB 2018

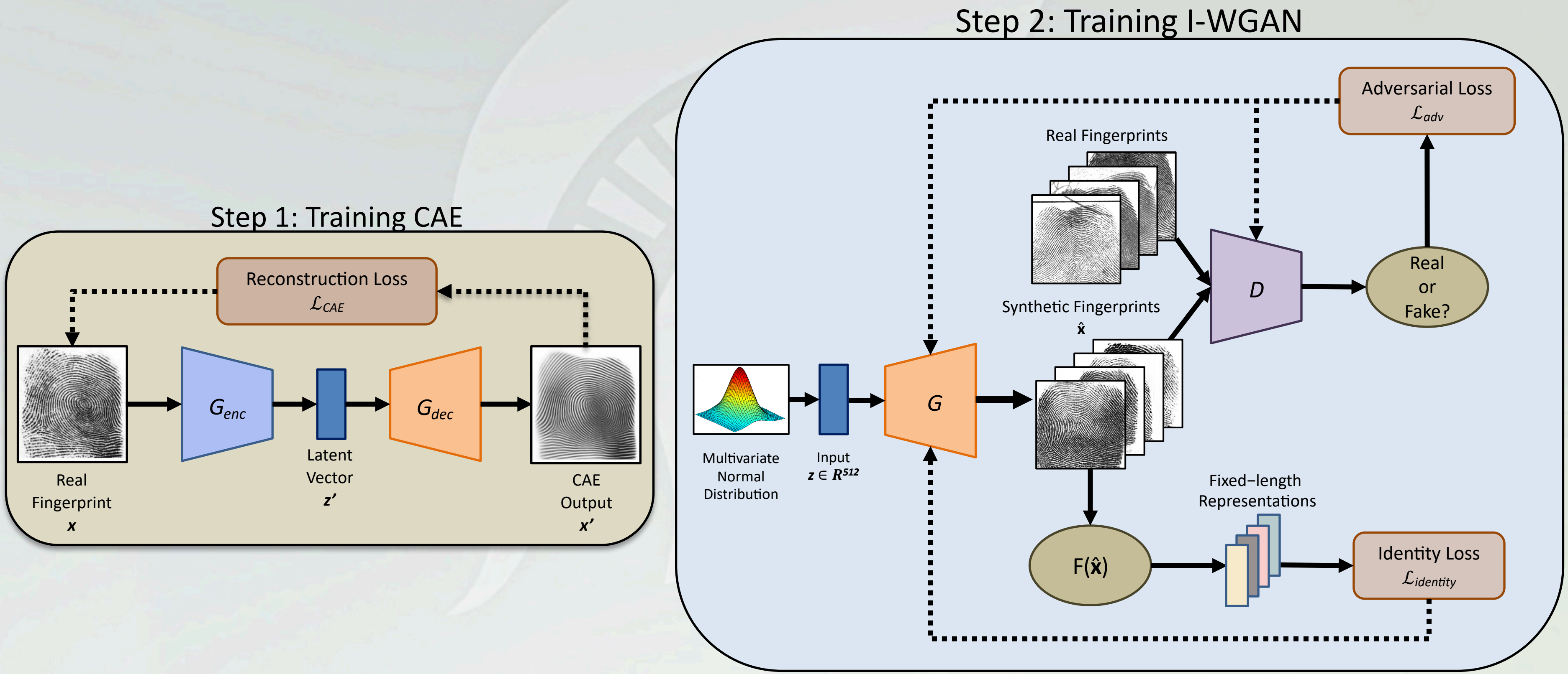
Attia et al.
SMC 2019

Proposed: Plain

Contributions of our Proposed Approach

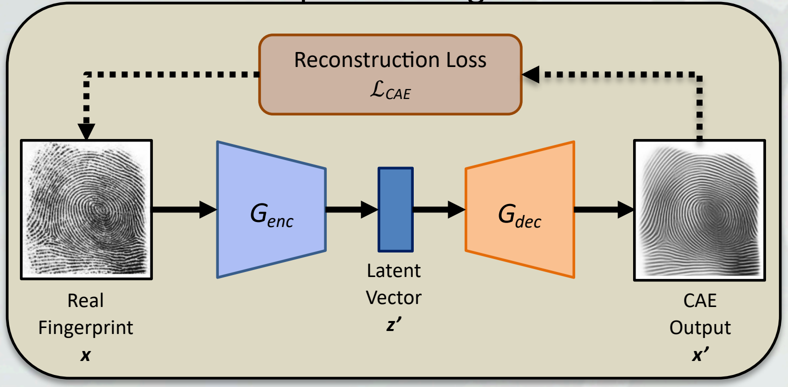
- **Fingerprint synthesis algorithm based on GANs**
- **Identity Loss** to generate fingerprints of more unique identities
- Synthesis of **100 million fingerprint images**
- **Large-scale search evaluation** against 100 million synthetic prints

Proposed Approach



Convolutional Autoencoder (CAE)

Step 1: Training CAE



- Training CAE in an unsupervised fashion
- Reconstruction loss:
$$\mathcal{L}_{CAE} = ||x - x'||_2^2$$
- Weights of G_{dec} used to initialize Generator G of I-WGAN

Improved-WGAN

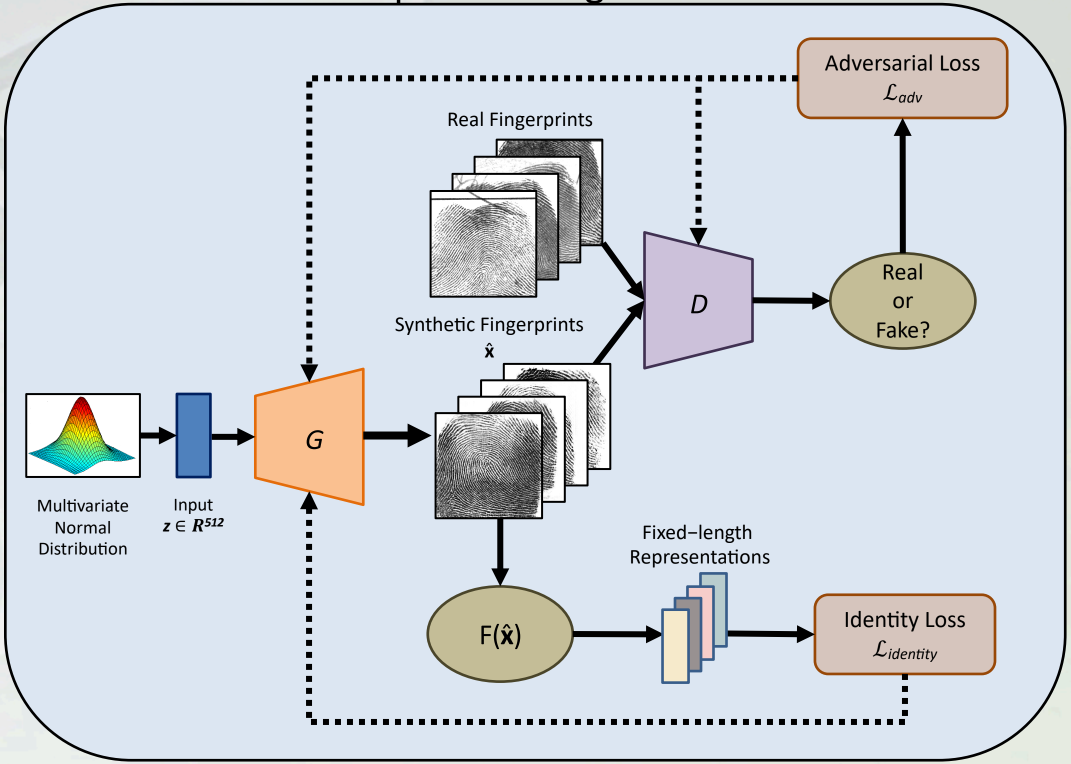
Identity Loss

- Use DeepPrint (Engelsma *et al. PAMI 2019*) as F(x̂) to extract fixed-length representations

- For each latent pair (z_i, z_j):

$$\mathcal{L}_{identity} = \frac{1}{\sum ||F(G(z_i)) - F(G(z_j))||}, (z_i \neq z_j)$$

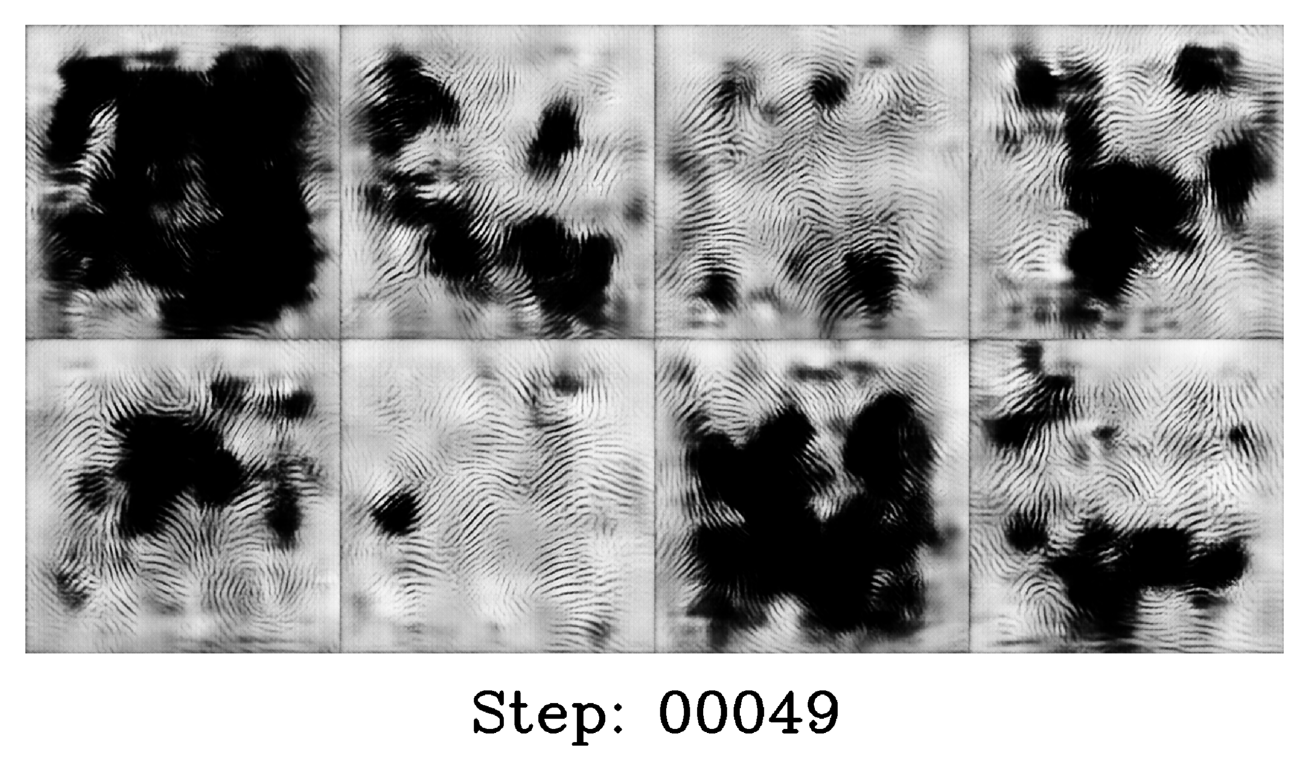
Step 2: Training I-WGAN



Training and Synthesis

- CAE and I-WGAN **trained using 280,000 rolled fingerprint images**
- **Fine-tuned for synthesizing plain fingerprints** using 84K plain prints
- **Synthesis of 100 million rolled fingerprints** using HPCC
  - 100 jobs in parallel, each job generating 1 million prints
  - Total time taken: 51 CPU hours; 1.8ms/image

Generated Fingerprints while Training



Experimental Results

Fingerprint Realism

- **Metrics:**
  - Minutiae count – template and block,
  - Minutiae direction – template and block
  - Minutiae convex hull area
  - Minutiae spatial distribution (2D minutiae histogram from Gottschlich *et al.* IET 2014)
  - Block minutiae quality
  - NFIQ 2.0 quality scores

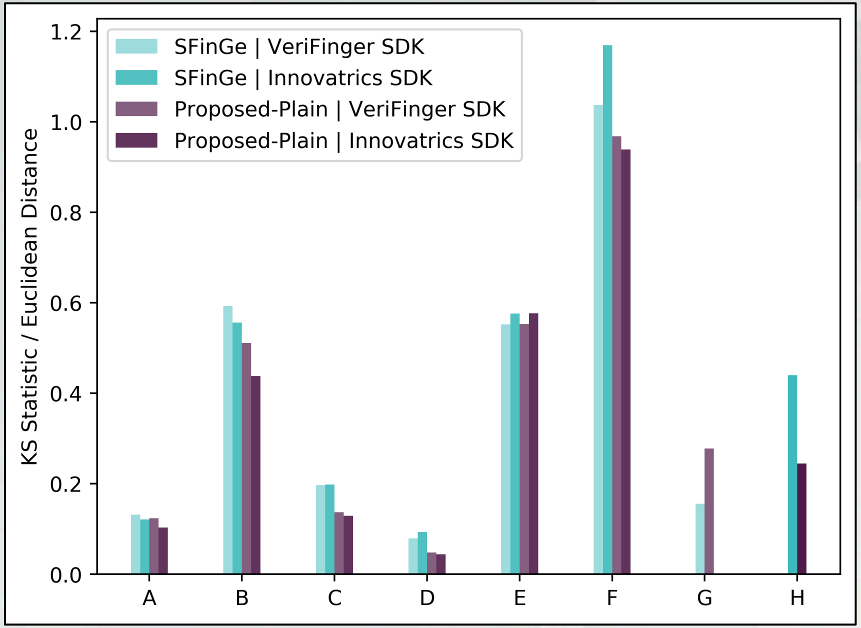
- **Statistical Test:** Kolmogorv-Smirnov test (Massey JASA 1951) used to compute difference between the distributions of each metric

Fingerprint Realism

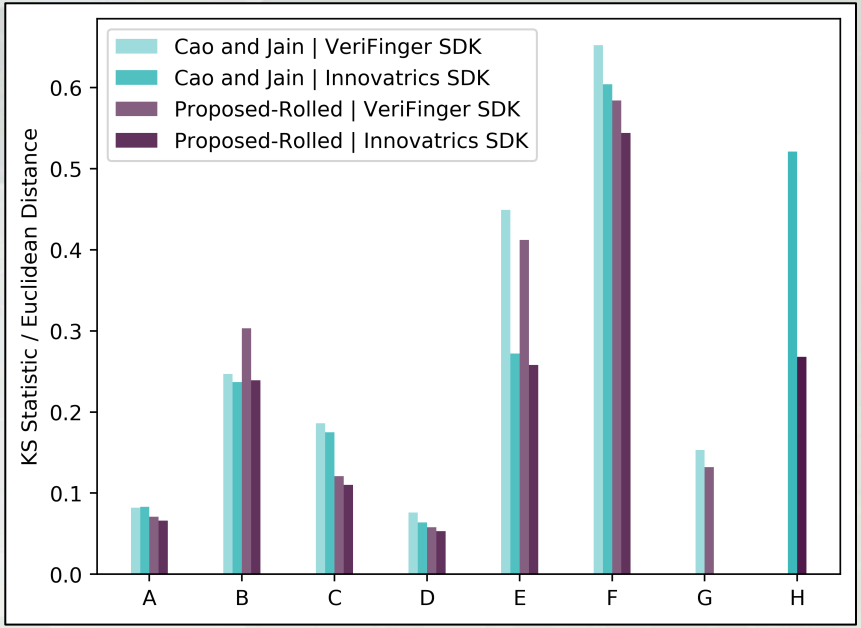
- Datasets:

	Plain Fingerprint datasets	Rolled Fingerprint datasets
Real	CASIA-Fingerprint v5 (2000 fingerprints)	NIST SD4 (2000 enrollment fingerprints)
	NIST SD302L (1951 fingerprints)	NIST SD14 (last 2000 enrollment fingerprints)
	NIST SD302M (1979 fingerprints)	NIST SD302U (2000 fingerprints)
Synthetic	SFinGe (2000 fingerprints)	Cao and Jain ICB 2018 (2000 fingerprints)
	Proposed Approach (2000 fingerprints)	Proposed Approach (2000 fingerprints)

Fingerprint Realism



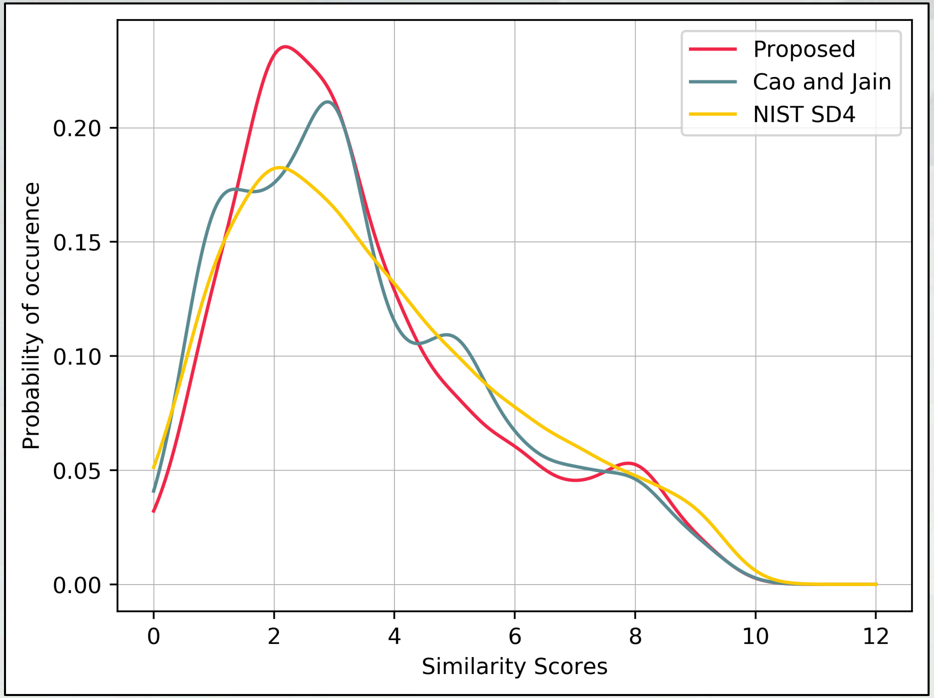
(a)



(b)

Comparison of synthetic plain (a) and rolled (b) fingerprints to real fingerprints using 8 metrics: minutiae count – block [A] and template [B], direction – block [C] and template [D], convex hull area [E], spatial distributions [F], block minutiae quality [G], and NFIQ 2.0 quality [H].

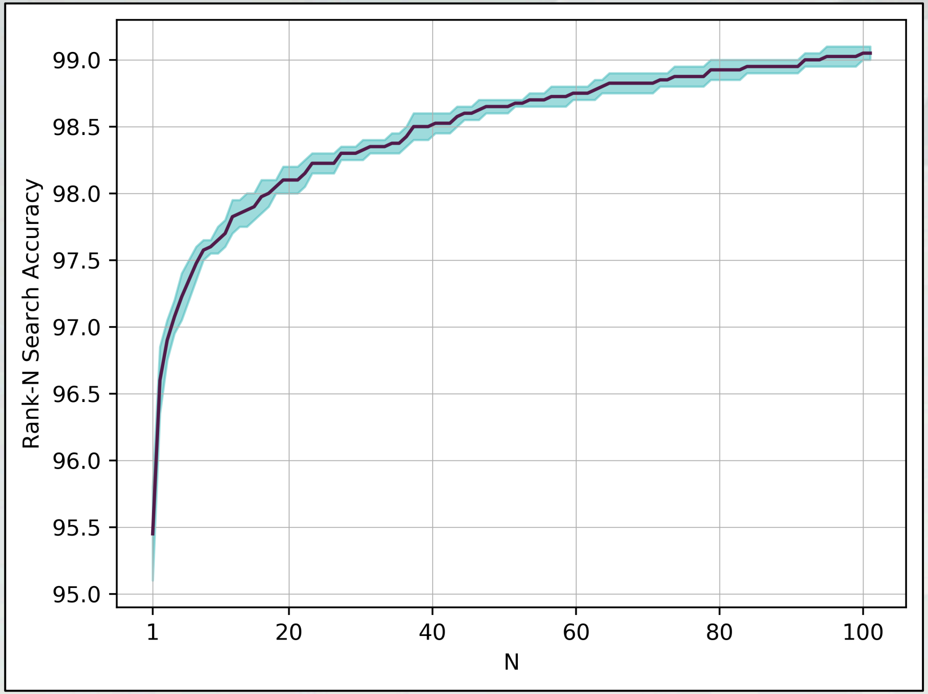
Imposter Scores Distribution



Imposter scores distribution computed using real rolled prints (NIST SD4) and synthetic rolled prints from Cao and Jain, and proposed approach

- 500K imposter scores computed using VeriFinger
- **(Mean, STD):**
  - **(3.47, 2.13) - Proposed**
  - **(3.48, 2.18) - Cao and Jain**
- Identity loss helped generate more diverse fingerprints

DeepPrint Search against 1 Million Fingerprints



Confidence interval for rank-N search accuracy on NIST SD4 using DeepPrint

- Gallery of 1 million synthetic fingerprints
- Confidence intervals for rank-N search accuracies
- Mean rank-1 search acc: **95.53%** with conf. interval of **[95.1, 95.8]**

COTS Search against 1 Million Fingerprints

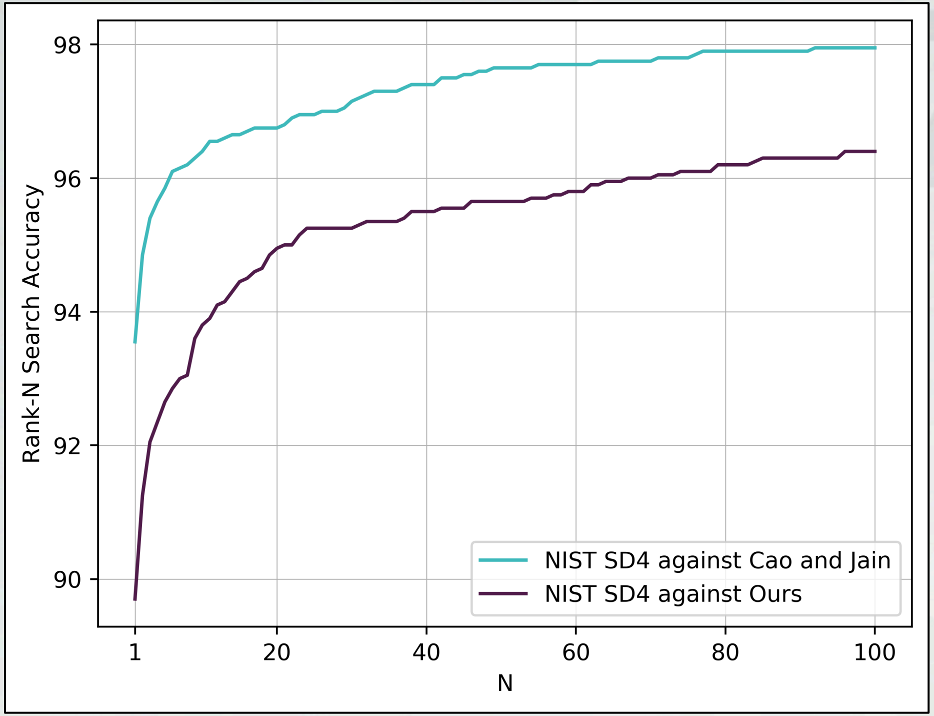
- Fingerprint search using NIST SD4 on synthetic rolled fingerprints

- Rank-1 search accuracies using Innovatrics SDK:

	Proposed Approach	Cao and Jain
250K Gallery	91.45%	90.85%
1M Gallery	90.35%	90.40%

- Uniqueness of our synthetic rolled prints becomes more evident at large gallery sizes

Search against 100 Million Fingerprints



Rank-N fingerprint search accuracies on NIST SD4 for galleries of 100 million synthetic fingerprints using DeepPrint as the matcher

- **Challenges:**
  - Synthesis
  - Search experiment
- Rank-1 search accuracy on gallery:
  - Proposed approach: 89.7%
  - Cao and Jain: 93.55%

Conclusions

- Propose a **fingerprint synthesis algorithm** based on I-WGAN and an identity loss to generate **diverse and realistic fingerprints**
- Show fingerprint **search performance at a scale of 100 million**
- **Ongoing work:**
  - Scale search to a gallery of 1 billion fingerprints
  - Further improve realism and diversity of prints

Thank you!