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)
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

- Cappelli et al. IET 2018
- Johnson et al. CVPR 2013
- Zhao et al. BTAS 2012
- AGa et al. SMC 2019
- Bontrager et al. BTAS 2018
- Cao and Jain ICB 2018
- Attia et al. SMC 2019
- Proposed: Rolled
- Proposed: Plain
Contribution 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

Step 1: Training CAE

- **Real Fingerprint** $x$
- **Genc**
- **Gdec**
- **Reconstruction Loss** $\mathcal{L}_{\text{CAE}}$
- **Latent Vector** $z'$
- **CAE Output** $x'$

Step 2: Training I-WGAN

- **Real Fingerprints**
- **Synthetic Fingerprints** $\hat{x}$
- **Fixed-length Representations**
- **G**
- **D**
- **Adversarial Loss** $\mathcal{L}_{\text{adv}}$
- **Real or Fake?**
- **Identity Loss** $\mathcal{L}_{\text{identity}}$

Proposed Approach

**Step 1: Training CAE**

- Input $z \in \mathbb{R}^{512}$
- **CAE Output** $x'$
- **Reconstruction Loss** $\mathcal{L}_{\text{CAE}}$
- **Genc**
- **Gdec**
- **Latent Vector** $z'$
- **Real Fingerprint** $x$

**Step 2: Training I-WGAN**

- **Mutiivariate Normal Distribution**
- **Input** $z \in \mathbb{R}^{512}$
- **Synthetic Fingerprints** $\hat{x}$
- **Fixed-length Representations**
- **Adversarial Loss** $\mathcal{L}_{\text{adv}}$
- **Real or Fake?**
- **Identity Loss** $\mathcal{L}_{\text{identity}}$
Convolutional Autoencoder (CAE)

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

Identity Loss

- Use DeepPrint (Engelsma et al. *PAMI 2019*) as \( F(\hat{x}) \) to extract fixed-length representations

- For each latent pair \((z_i, z_j)\):

\[
L_{\text{identity}} = \frac{1}{\sum ||F(G(z_i)) - F(G(z_j))||, (z_i \neq z_j)}
\]
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

Step: 00049
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:** Kolmogorov-Smirnov test (Massey JASA 1951) used to compute difference between the distributions of each metric
## Fingerprint Realism

### Datasets:

<table>
<thead>
<tr>
<th></th>
<th>Plain Fingerprint datasets</th>
<th>Rolled Fingerprint datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real</strong></td>
<td>CASIA-Fingerprint v5 (2000 fingerprints)</td>
<td>NIST SD4 (2000 enrollment fingerprints)</td>
</tr>
<tr>
<td></td>
<td>NIST SD302L (1951 fingerprints)</td>
<td>NIST SD14 (last 2000 enrollment fingerprints)</td>
</tr>
<tr>
<td></td>
<td>NIST SD302M (1979 fingerprints)</td>
<td>NIST SD302U (2000 fingerprints)</td>
</tr>
<tr>
<td><strong>Synthetic</strong></td>
<td>SFinGe (2000 fingerprints)</td>
<td>Cao and Jain ICB 2018 (2000 fingerprints)</td>
</tr>
</tbody>
</table>
Fingerprint Realism

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

- 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

Imposter scores distribution computed using real rolled prints (NIST SD4) and synthetic rolled prints from Cao and Jain, and proposed approach.
DeepPrint Search against 1 Million Fingerprints

- 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]

Confidence interval for rank-N search accuracy on NIST SD4 using DeepPrint
COTS Search against 1 Million Fingerprints

- Fingerprint search using NIST SD4 on synthetic rolled fingerprints

- Rank-1 search accuracies using Innovatrics SDK:

<table>
<thead>
<tr>
<th>Gallery Size</th>
<th>Proposed Approach</th>
<th>Cao and Jain</th>
</tr>
</thead>
<tbody>
<tr>
<td>250K Gallery</td>
<td>91.45%</td>
<td>90.85%</td>
</tr>
<tr>
<td>1M Gallery</td>
<td>90.35%</td>
<td>90.40%</td>
</tr>
</tbody>
</table>

- Uniqueness of our synthetic rolled prints becomes more evident at large gallery sizes
Search against 100 Million Fingerprints

- **Challenges:**
  - Synthesis
  - Search experiment

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

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