Learning a Fixed-Length Fingerprint Representation

Joshua J. Engelsma, Kai Cao, and Anil K. Jain, *Life Fellow, IEEE*

Abstract—We present DeepPrint, a deep network, which learns to extract fixed-length fingerprint representations of only 200 bytes. DeepPrint incorporates fingerprint domain knowledge, including alignment and minutiae detection, into the deep network architecture to maximize the discriminative power of its representation. The compact, DeepPrint representation has several advantages over the prevailing variable length minutiae representation which (i) requires computationally expensive graph matching techniques, (ii) is difficult to secure using strong encryption schemes (e.g. homomorphic encryption), and (iii) has low discriminative power in poor quality fingerprints where minutiae extraction is unreliable. We benchmark DeepPrint against two top performing COTS SDKs (Verifinger and Innovatrics) from the NIST and FVC evaluations. Coupled with a re-ranking scheme, the DeepPrint rank-1 search accuracy on the NIST SD4 dataset against a gallery of 1.1 million fingerprints is comparable to the top COTS matcher, but it is significantly faster (**DeepPrint:** 98.80% in 0.3 seconds vs. **COTS A:** 98.85% in 27 seconds). To the best of our knowledge, the DeepPrint representation is the most compact and discriminative fixed-length fingerprint representation reported in the academic literature.

Index Terms—Fingerprint Matching, Minutiae Representation, Fixed-Length Representation, Representation Learning, Deep Networks, Large-scale Search, Domain Knowledge in Deep Networks

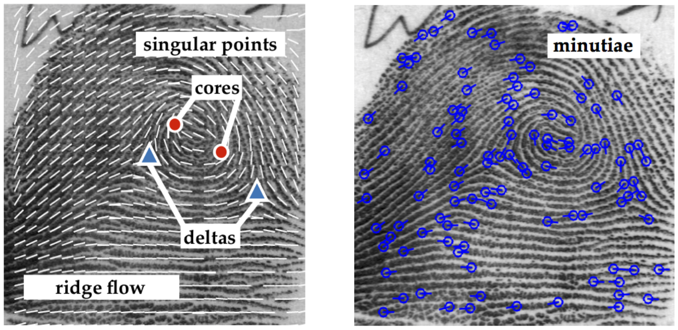
1 Introduction

OVER 100 years ago, the pioneering giant of modern day fingerprint recognition, Sir Francis Galton, astutely commented on fingerprints in his 1892 book titled "Finger Prints":

> *"They have the unique merit of retaining all their peculiarities unchanged throughout life, and afford in consequence an incomparably surer criterion of identity than any other bodily feature."* [1]

Galton went on to describe fingerprint *minutiae*, the small details woven throughout the papillary ridges on each of our fingers, which Galton believed provided uniqueness and permanence properties for accurately identifying individuals. Over the 100 years since Galton's ground breaking scientific observations, fingerprint recognition systems have become ubiquitous and can be found in a plethora of different domains [2] such as forensics [3], healthcare, mobile device security [4], mobile payments [4], border crossing [5], and national ID [6]. To date, virtually all of these systems continue to rely upon the location and orientation of minutiae within fingerprint images for recognition (Fig. 1).

Although automated fingerprint recognition systems based on minutiae representations (*i.e.* handcrafted features) have seen tremendous success over the years, they have several limitations.

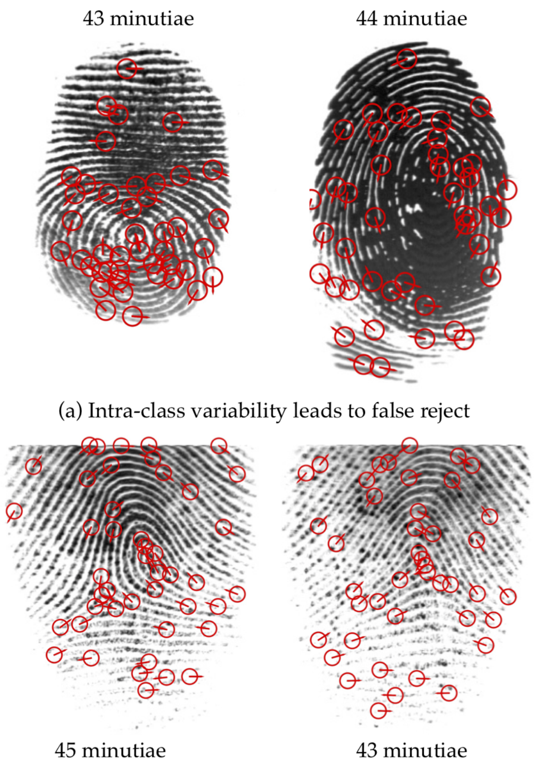


(a) Level-1 features (b) Level-2 features

Fig. 1. The most popular fingerprint representation consists of (a) global level-1 features (ridge flow, core, and delta) and (b) local level-2 features, called minutiae points, together with their descriptors (e.g., texture in local minutiae neighborhoods). The fingerprint image illustrated here is a rolled impression from the NIST SD4 database [7]. The number of minutiae in NIST4 rolled fingerprint images range all the way from 12 to 196.

- Minutiae-based representations are of variable length, since the number of extracted minutiae (Table 1) varies amongst different fingerprint images even of the same finger (Fig. 2 (a)). Variations in the number of minutiae originate from a user's interaction with the fingerprint reader (placement position and applied pressure) and condition of the finger (dry, wet, cuts, bruises, etc.). This variation in the number of minutiae causes two main problems: (i) pairwise fingerprint comparison is computationally demanding and varies with number of minutiae and (ii) matching in the encrypted domain, a necessity for user privacy protection, is computationally expensive, and results in loss of accuracy [9].
- In the context of global population registration, fingerprint recognition can be viewed as a **75 billion class problem** ($\approx$ 7.5 billion living persons, assuming nearly all with 10 fingers) with large intra-class variability and large inter-class similarity (Fig. 2). This necessitates extremely discriminative yet compact representations that are complementary and at least as discriminative as the traditional minutiae-based representation. For example, India's civil registration system, Aadhaar, now has a database of ≈ 1.3 billion residents who are enrolled based on their 10 fingerprints, 2 irises, and face image [6].
- Reliable minutiae extraction in low quality fingerprints (due to noise, distortion, finger condition) is problematic, causing false rejects in the recognition system (Fig. 2 (a)). See also NIST fingerprint evaluation FpVTE 2012 [10].

- *J. J. Engelsma and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824 E-mail: engelsm7@msu.edu, jain@cse.msu.edu*
- *K. Cao is a Senior Biometrics Researcher at Goodix, San Diego, CA*



(a) Intra-class variability leads to false reject

(b) Inter-class similarity leads to false accept

Fig. 2. Failures of the COTS A minutiae-based matcher (minutiae annotated with COTS A). The genuine pair (two impressions from the same finger) in (a) was falsely rejected at 0.1% FAR (score of 9) due to heavy non-linear distortion and moist fingers. The imposter pair (impressions from two different fingers) in (b) was falsely accepted at 0.1% FAR (score of 38) due to the similar minutiae distribution in these two fingerprint images (the score threshold for COTS A @ FAR = 0.1% is 34). In contrast, DeepPrint is able to correctly match the genuine pair in (a) and reject the imposter pair in (b). These slap fingerprint impressions come from public domain FVC 2004 DB1 A database [8]. The number of minutiae in FVC 2004 DB1 A images range from 11 to 87.

TABLE 1
Comparison of variable length minutiae representation with fixed-length DeepPrint representation

Matcher	(Min, Max) # of Minutiae[1]	(Min, Max) Template Size (kB)
COTS A	(12, 196)	(1.5, 23.7)
COTS B	(12, 225)	(0.6, 5.3)
Proposed	N.A.[2]	0.2[†]

[1] Statistics from NIST SD4 and FVC 2004 DB1.
[2] Template is not explicitly comprised of minutiae.
[†] Template size is fixed at 200 bytes, irrespective of the number of minutiae (192 bytes for the features and 8 bytes for 2 decompression scalars).

To overcome the limitations of minutiae-based matchers, we present a reformulation of the fingerprint recognition problem. In particular, rather than extracting varying length minutiae-sets for matching (*i.e.* handcrafted features), we design a deep network embedded with fingerprint domain knowledge, called **DeepPrint**, to *learn* a fixed-length representation of 200 bytes which discriminates between fingerprint images from different fingers (Fig. 4).

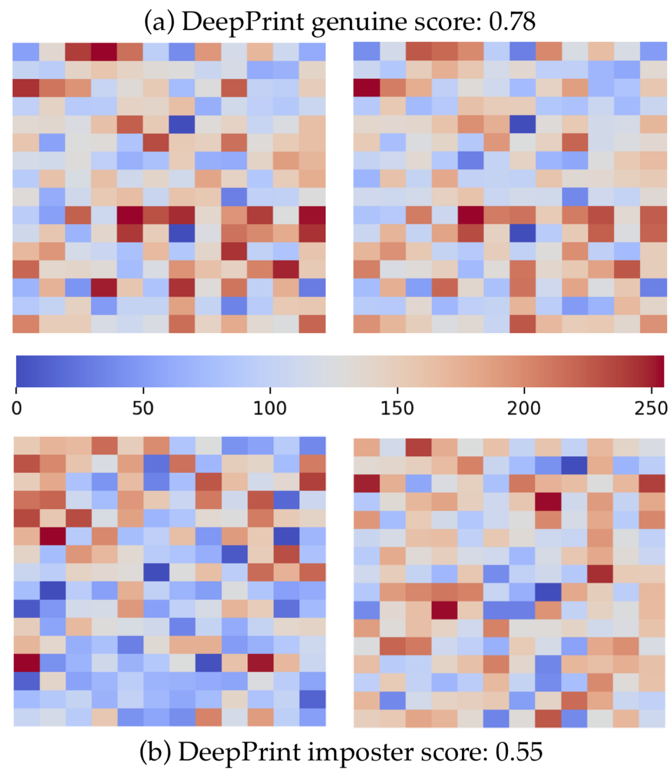


Fig. 3. Fixed-length, 192-dimensional fingerprint representations extracted by DeepPrint (shown as 16×12 feature maps) from the same four fingerprints shown in Figure 2. Unlike COTS A, we correctly classify the pair in (a) as a genuine pair, and the pair in (b) as an imposter pair. The score threshold of DeepPrint @ FAR = 0.1% is 0.76

face recognition systems which have almost entirely abandoned traditional handcrafted features in favor of deep features extracted by deep networks with remarkable success [11], [12], [13]. However, unlike deep network based face recognition systems, we do not completely abandon handcrafted features. Instead, we aim to integrate handcrafted fingerprint features (minutiae [1]) into the deep network architecture to exploit the benefits of both deep networks and traditional, domain knowledge inspired features.

While prevailing minutiae-matchers require expensive graph matching algorithms for fingerprint comparison, the 200 byte representations extracted by DeepPrint can be compared using simple distance metrics such as the cosine similarity, requiring only d multiplications and $d-1$ additions, where d is the dimensionality of the representation (for DeepPrint, $d = 192$)[2]. Another significant advantage of this fixed-length representation is that it can be matched in the encrypted domain using fully homomorphic encryption [14], [15], [16], [17]. Finally, since DeepPrint is able to encode features that go beyond fingerprint minutiae, it is able to match poor quality fingerprints when reliable minutiae extraction is not possible (Figs. 2 and 3).

To arrive at a compact and discriminative representation of only 200 bytes, the DeepPrint architecture is embedded with

1. Note that we do not *require* explicitly storing minutiae in our final template. Rather, we aim to guide DeepPrint to extract features related to minutiae during training of the network.

2. The DeepPrint representation is originally 768 bytes (192 features and 4 bytes per float value). We compress the 768 bytes to 200 by scaling the floats to integer values between [0,255] and saving the two compression parameters with the features. This loss in precision (which saves significant disk storage

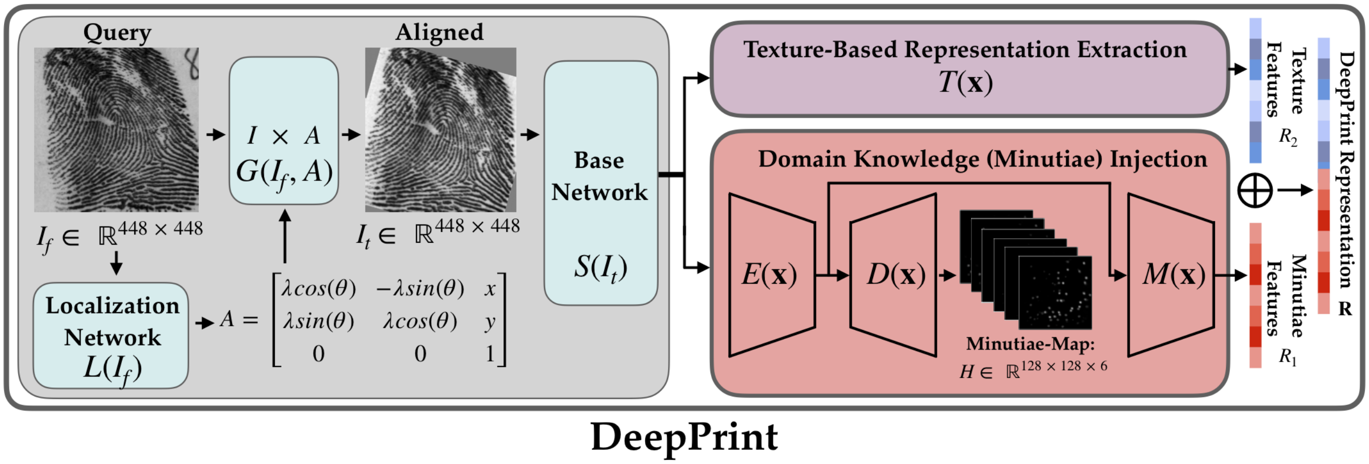


Fig. 4. Flow diagram of DeepPrint: (i) a query fingerprint is aligned via a Localization Network which has been trained end-to-end with the Base-Network and Feature Extraction Networks (no reference points are needed for alignment); (ii) the aligned fingerprint proceeds to the Base-Network which is followed by two branches; (iii) the first branch extracts a 96-dimensional texture-based representation; (iv) the second branch extracts a 96-dimensional minutiae-based representation, guided by a side-task of minutiae detection (via a minutiae map which does not have to be extracted during testing); (v) the texture-based representation and minutiae-based representation are concatenated into a 192-dimensional representation of 768 bytes (192 features and 4 bytes per float). The 768 byte template is compressed into a 200 byte fixed-length representation by truncating floating point value features into integer value features, and saving the scaling and shifting values (8 bytes) used to truncate from floating point values to integers. The 200 byte DeepPrint representations can be used both for authentication and large-scale fingerprint search. The minutiae-map can be used to further improve system accuracy and interpretability by re-ranking candidates retrieved by the fixed-length representation.

fingerprint domain knowledge via an automatic alignment module and a multi-task learning objective which requires minutiae-detection (in the form of a *minutiae-map*) as a side task to representation learning. More specifically, DeepPrint automatically aligns an input fingerprint and subsequently extracts both a *texture representation* and a *minutiae-based representation* (both with 96 features). The 192-dimensional concatenation of these two representations, followed by compression from floating point features to integer value features comprises a 200 byte fixed-length representation (192 bytes for the feature vector and 4 bytes for storing the 2 compression parameters). As a final step, we utilize Product Quantization [18] to further compress the DeepPrint representations stored in the gallery, significantly reducing the computational requirements and time for large-scale fingerprint search.

Detecting minutiae (in the form of a minutiae-map) as a side-task to representation learning has several key benefits:

- We guide our representation to incorporate domain inspired features pertaining to minutiae by sharing parameters between the minutiae-map output task and the representation learning task in the multi-task learning framework.
- Since minutiae representations are the most popular for fingerprint recognition, we posit that our method for guiding the DeepPrint feature extraction via its minutiae-map side-task falls in line with the goal of "Explainable AI" [19].
- Given a probe fingerprint, we first use its DeepPrint representation to find the top k candidates and then re-rank the top k candidates using the minutiae-map provided by DeepPrint [3]. This *optional* re-ranking add-on further improves both accuracy and interpretability.

The primary benefit of the 200 byte representation extracted by DeepPrint comes into play when performing mega-scale search against millions or even billions of identities (e.g., India's Aadhaar [6] and the FBI's Next Generation Identification (NGI) databases [3]). To highlight the significance of this benefit, we benchmark the search performance of DeepPrint against the latest version SDKs (as of July, 2019) of two top performers in the NIST FpVTE 2012 (Innovatrics[4] v7.2.1.40 and Verifinger[5] v10.0[6]) on the NIST SD4 [7] and NIST SD14 [21] databases augmented with a gallery of nearly 1.1 million rolled fingerprints. Our empirical results demonstrate that DeepPrint is competitive with these two state-of-the-art COTS matchers in accuracy while requiring only a fraction of the search time. Furthermore, a given DeepPrint fixed-length representation can also be matched in the encrypted domain via homomorphic encryption with minor loss to recognition accuracy as shown in [14] for face recognition.

More concisely, the primary contributions of this work are:

- A customized deep network (Fig. 4), called DeepPrint, which utilizes fingerprint domain knowledge (alignment and minutiae detection) to learn and extract a discriminative fixed-length fingerprint representation.
- Demonstrating in a manner similar to [29] that Product Quantization can be used to compress DeepPrint *fingerprint* representations, enabling even faster mega-scale search (51 ms search time against a gallery of 1.1 million fingerprints vs. 27,000 ms for a COTS with comparable accuracy).
- Demonstrating with a two-stage search scheme similar to [29] that candidates retrieved by DeepPrint representations can be re-ranked using a minutiae-matcher in conjunction with the DeepPrint minutiae-map. This further

3. The $128 \times 128 \times 6$ DeepPrint minutiae-map can be easily converted into a minutiae-set with n minutia: $\{(x_1, y_1, \theta_1), \ldots, (x_n, y_n, \theta_n)\}$ and passed to

4. https://www.innovatrics.com/
5. https://www.neurotechnology.com/
6. We note that Verifinger v10.0 performs significantly better than earlier

TABLE 2
Published Studies on Fixed-Length Fingerprint Representations

Algorithm	Description	HR @ PR = 1.0%[1] (NIST SD4)[2]	HR @ PR = 1.0% (NIST SD14)[3]	Template Size (bytes)	Gallery Size[4]
Jain *et al.* [22], [23]	**Fingercode**: Global representation extracted using Gabor Filters	N.A.	N.A.	640	N.A.
Cappelli *et al.* [24]	**MCC**: Local descriptors via 3D cylindrical structures comprised of the minutiae-set representation	93.2%	91.0%	1,913	2,700
Cao and Jain [25]	**Inception v3:** Global deep representation extracted via Alignment and Inception v3	98.65%	98.93%	8,192	250,000
Song and Feng [26]	**PDC**: Deep representations extracted at different resolutions and aggregated into global representation	93.3%	N.A.	N.A.	2,000
Song *et al.* [27]	**MDC**: Deep representations extracted from minutiae and aggregated into global representation	99.2%	99.6%	1,200	2,700
Li *et al.* [28]	**Finger Patches:** Local deep representations aggregated into global representation via global average pooling	**99.83%**	99.89%	1,024	2,700
Proposed	**DeepPrint**: Global deep representation extracted via multi-task CNN with built-in fingerprint alignment	99.75%	**99.93%**	**200**[†]	**1,100,000**

[1] In some baselines we estimated the data points from a Figure (specific data points were not reported in the paper).
[2] Only 2,000 fingerprints are included in the gallery to enable comparison with previous works. (HR = Hit Rate, PR = Penetration Rate)
[3] Only last 2,700 pairs (2,700 probes; 2,700 gallery) are used to enable comparison with previous works.
[4] Largest gallery size used in the paper.
[†] The DeepPrint representation can be further compressed to only 64 bytes using product quantization with minor loss in accuracy.

improves system interpretability and accuracy and demonstrates that the DeepPrint features are complementary to the traditional minutiae representation.

- Benchmarking DeepPrint against two state-of-the-art COTS matchers (Innovatrics and Verifinger) on NIST SD4 and NIST SD14 against a gallery of 1.1 million fingerprints. Empirical results demonstrate that DeepPrint is comparable to COTS matchers in accuracy at a significantly faster search speed.
- Benchmarking the authentication performance of DeepPrint on the NIST SD4 and NIST SD14 rolled-fingerprints databases and the FVC 2004 DB1 A slap fingerprint database [8]. Again, DeepPrint shows comparable performance against the two COTS matchers, demonstrating the generalization ability of DeepPrint to both rolled and slap fingerprint databases.
- Demonstrating that homomorphic encryption can be used to match DeepPrint templates in the encrypted domain, in real time (1.26 ms), with minimal loss to matching accuracy as shown for fixed-length face representations [14].
- An interpretability visualization which demonstrates our ability to guide DeepPrint to look at minutiae-related features.

2 Prior Work

Several early works [22], [23], [24] presented fixed-length fingerprint representations using traditional image processing techniques. In [22], [23], Jain *et al.* extracted a global fixed-length representation of 640 bytes, called Fingercode, using a set of Gabor Filters. Cappelli *et al.* introduced a fixed-length minutiae cylindrical structures computed with minutiae points [24]. While both of these representations demonstrated success at the time they were proposed, their accuracy is now significantly inferior to state-of-the-art COTS matchers

Following the seminal contributions of [22], [23] and [24], the past 10 years of research on fixed-length fingerprint representations [31], [32], [33], [34], [35], [36], [37], [38], [39] has not produced a representation competitive in terms of fingerprint recognition accuracy with the traditional minutiae-based representation. However, recent studies [25], [26], [27], [28] have utilized deep networks to extract highly discriminative fixed-length fingerprint representations. More specifically, (i) Cao and Jain [25] used global alignment and Inception v3 to learn fixed-length fingerprint representations. (ii) Song and Feng [26] used deep networks to extract representations at various resolutions which were then aggregated into a global fixed-length representation. (iii) Song *et al.* [27] further learned fixed-length minutiae descriptors which were aggregated into a global fixed-length representation via an aggregation network. Finally, (v) Li *et al.* [28] extracted local descriptors from predefined "fingerprint classes" which were then aggregated into a global fixed-length representation through global average pooling.

While these efforts show tremendous promise, each method has some limitations. In particular, (i) the algorithms proposed in [25] and [26] both required computationally demanding global alignment as a preprocessing step, and the accuracy is inferior to state-of-the-art COTS matchers. (ii) The representations extracted in [27] require the arduous process of minutiae-detection, patch extraction, patch-level inference, and an aggregation network to build a single global feature representation. (iii) While the

June 2001 January 2003 January 2005 August 2007 April 2010 July 2013

Fig. 5. Fingerprint impressions from one subject in the DeepPrint training dataset [30]. Impressions were captured longitudinally, resulting in the variability across impressions (contrast and intensity from environmental conditions; distortion and alignment from user placement). Importantly, training with longitudinal data enables learning compact representations which are invariant to the typical noise observed across fingerprint impressions over time, a necessity in any fingerprint recognition system.

(with small gallery size), the accuracy was not reported for slap fingerprints. Since [28] aggregates local descriptors by averaging them together, it is unlikely that the approach would work well when areas of the fingerprint are occluded or missing (often times the case in slap fingerprint databases like FVC 2004 DB1 A), and (v) all of the algorithms, suffer from lack of interpretability compared to traditional minutiae representations.

In addition, existing studies targeting deep, fixed-length fingerprint representations all lack an extensive, large-scale evaluation of the deep features. Indeed, one of the primary motivations for fixed-length fingerprint representations is to perform orders of magnitude faster large scale search. However, with the exception of Cao and Jain [25], who evaluate against a database of 250K fingerprints, the next largest gallery size used in any of the aforementioned studies is only 2,700.

As an addendum, deep networks have also been used to improve *specific sub-modules* of fingerprint recognition systems such as segmentation [40], [41], [42], [43], orientation field estimation [44], [45], [46], minutiae extraction [47], [48], [49], and minutiae descriptor extraction [50]. However, these works all still operate within the conventional paradigm of extracting an unordered, variable length set of minutiae for fingerprint matching.

Algorithm 1 Extract DeepPrint Representation

$L(I_f)$: Shallow localization network, outputs x, y, θ
A: Affine matrix composed with parameters x, y, θ
$G(I_f, A)$: Bilinear grid sampler, outputs aligned fingerprint
$S(I_t)$: Inception v4 stem
$E(\mathbf{x})$: Shared minutiae parameters
$M(\mathbf{x})$: Minutia representation branch
$D(\mathbf{x})$: Minutiae map estimation
$T(\mathbf{x})$: Texture representation branch

Input: Unaligned 448×448 fingerprint image I_f
$A \leftarrow (x, y, \theta) \leftarrow L(I_f)$
$I_t \leftarrow G(I_f, A)$
$F_{map} \leftarrow S(I_t)$
$M_{map} \leftarrow E(F_{map})$
$R_1 \leftarrow M(M_{map})$
$H \leftarrow D(M_{map})$
$R_2 \leftarrow T(F_{map})$
$\mathbf{R} \leftarrow R_1 \oplus R_2$
Output: Fingerprint representation $\mathbf{R} \in \mathbb{R}^{192}$ and minutiae-map H. (H can be optionally utilized for (i) visualization and (ii) fusion of DeepPrint scores obtained via $\mathbf{R}$ with minutiae-matching scores.)

3 DEEPPRINT

In the following section, we (i) provide a high-level overview and intuition of DeepPrint, (ii) present how we incorporate automatic alignment into DeepPrint, and (iii) demonstrate how the accuracy and interpretability of DeepPrint is improved through the injection of fingerprint domain knowledge.

3.1 Overview

A high level overview of DeepPrint is provided in Figure 4 with pseudocode in Algorithm 1. DeepPrint is trained with a longitudinal database (Fig. 5) comprised of 455K rolled fingerprint images stemming from 38,291 unique fingers [30]. Longitudinal fingerprint databases consist of fingerprints from distinct subjects captured over time (Fig. 5) [30]. It is necessary to train DeepPrint with a large, longitudinal database so that it can learn compact, fixed-length representations which are invariant to the differences introduced during fingerprint image acquisition at different times and in different environments (humidity, temperature, user interaction with the reader, and finger injuries). The primary task during training is to predict the finger identity label $c \in [0, 38291]$ fingerprint images ($\approx$ 12 fingerprint impressions / finger). The last fully connected layer is taken as the representation for fingerprint comparison during authentication and search.

The input to DeepPrint is a 448×448 [7] grayscale fingerprint image, I_f, which is first passed through the alignment module (Fig. 4). The alignment module consists of a localization network, L, and a grid sampler, G [51]. After applying the localization network and grid sampler to I_f, an aligned fingerprint I_t is passed to the base-network, S.

The base-network is the stem of the Inception v4 architecture (Inception v4 minus Inception modules). Following the base-network are two different branches (Fig. 4) comprised primarily of the three Inception modules (A, B, and C) described in [52]. The first branch, $T(x)$, completes the Inception v4 architecture [8] as

7. Fingerprint images in our training dataset vary in size from $\approx 512 \times 512$ to $\approx 800 \times 800$. As a pre-processing step, we do a center cropping (using Gaussian filtering, dilation and erosion, and thresholding) to all images to $\approx 448 \times 448$. This size is sufficient to cover most of the rolled fingerprint area without extraneous background pixels.

8. We selected Inception v4 after evaluating numerous other architectures

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.