COMPUTER SCIENCE

Unveiling intra-person fingerprint similarity via deep contrastive learning

Gabe Guo[1]*, Aniv Ray[1], Miles Izydorczak[2], Judah Goldfeder[1], Hod Lipson[3], Wenyao Xu[4]

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Fingerprint biometrics are integral to digital authentication and forensic science. However, they are based on the unproven assumption that no two fingerprints, even from different fingers of the same person, are alike. This renders them useless in scenarios where the presented fingerprints are from different fingers than those on record. Contrary to this prevailing assumption, we show above 99.99% confidence that fingerprints from different fingers of the same person share very strong similarities. Using deep twin neural networks to extract fingerprint representation vectors, we find that these similarities hold across all pairs of fingers within the same person, even when controlling for spurious factors like sensor modality. We also find evidence that ridge orientation, especially near the fingerprint center, explains a substantial part of this similarity, whereas minutiae used in traditional methods are almost nonpredictive. Our experiments suggest that, in some situations, this relationship can increase forensic investigation efficiency by almost two orders of magnitude.

INTRODUCTION

Fingerprints have stood the test of time as the gold standard biometric modality, having achieved ubiquity in digital authentication and forensic science. For instance, there are billions of mobile devices worldwide that rely on fingerprint identification technology (*1*–*4*). As it relates to forensics, the FBI has more than 150 million fingerprints on record (*5*, *6*), while crime laboratories analyze hundreds of thousands of fingerprints per year (*7*). Recently, there has been great research activity on fingerprints, spanning the development of artificial intelligence (AI)–based recognition techniques (*8*, *9*), the creation of hardware and chemical sensing modalities (*10*–*14*), reliability analysis in criminal justice scenarios (*15*–*17*), and genetic origins (*18*).

However, fingerprint biometrics are based on the traditional assumption that no two fingerprints, even from the same person, are alike (*9*, *19*, *20*). (Although technically, this assumption has never been definitively proven, multiple models estimate the probability of a given fingerprint configuration randomly occurring to be orders of magnitude less than the probability of randomly selecting a given person—say, the president of the United States—from the population of all human beings alive (*20*).) Thus, they only work for matching two samples of the same exact finger (*8*, *21*, *22*). This renders them useless in crime scenes or authentication scenarios where the presented fingerprints are from different fingers than the fingerprints on record.

For instance, imagine that detectives have obtained two fingerprints (e.g., right index and right middle) from crime scene A. From crime scene B, they obtained two other fingerprints (e.g., left pinky and left thumb). They have a list of 1000 potential suspects from scene A. A different list of 1000 potential suspects is available for scene B. Given the current information, traditional fingerprint biometrics are unable to discern whether the fingerprints from scenes A and B are related and involve the same person unless all fingerprints of all suspects are readily available on file. However, if intra-person fingerprint similarity can be established, as proposed in this paper, then the suspect lists from both scenes could be substantially reduced. Using the findings reported in this paper, the list of 1000 suspects could potentially be prioritized to about 40 higher-likelihood candidates.

Our central claim is that we can sidestep the same-finger limitation by exploiting nontraditional fingerprint features. Past studies provided evidence that fingerprint patterns may be partially genetically determined (*18*, *23*–*26*), which implies that there could be similarities among fingerprints from the same person. Furthermore, recent research shows that partial fingerprints from different users have common features that can be exploited to fool authentication systems (*27*). Last, liveness detection (i.e., whether a physical fingerprint is real or a spoof copy made from synthetic materials like rubber or silicone) systems perform better when trained on samples from the user whose fingerprints they test on, even when those training fingers are different than the testing fingers (*28*, *29*).

In this work, our main discovery is that fingerprints from different fingers of the same person share strong similarities; these results hold across all combinations of fingers, even from different hands of the same person. These similarities can mostly be explained by fingerprint ridge orientation.

We found this relationship by training twin deep neural networks (Fig. 1A) to predict whether two fingerprint samples (not necessarily from the same finger) were from the same person (*30*–*32*).

The neural network's outputted representation vectors for same-person pairs showed statistically significant differences from its representations for different-person pairs ($P < 10^{-4}$, with paired one-sided t test), providing very strong evidence of the intra-person fingerprint similarity. We systematically ruled out spurious sources of similarity by controlling for sensor modality, sample source, image background, and image brightness, leaving us confident that the similarities are due to intrinsic fingerprint patterns. To further promote confidence in our results, we extensively interrogated the extracted features from our deep neural networks (*33*, *34*) and found that they were fingerprint-like features—in particular, the ridge orientation near the center of the fingerprint heavily contributes to the similarity.

We also validated the usefulness of our results by using them to improve the efficiency of a simulated criminal justice lead-generation process by more than an order of magnitude. Last, we investigated the biases and generalization ability of our model, as it relates to demographics.

[1]Department of Computer Science, Columbia University, New York, NY 10027, USA. [2]Department of Computer Science, Tufts University, Medford, MA 02155, USA. [3]Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA. [4]Department of Computer Science and Engineering, SUNY Buffalo, Buffalo, NY 14260, USA.
*Corresponding author. Email: gzg2104@columbia.edu

RESULTS

Unveiling the similarity

General similarity analysis

We assess the degree to which pairs of fingerprints from different fingers of the same person are similar, with the expectation that they should be much more similar than pairs of fingerprints from different people, no matter which fingers they come from.

To do this, we conduct a one-sided paired *t* test (*35*, *36*) with $\alpha = 10^{-4}$ on (i) the average representation vector (obtained from twin neural network) distance between two fingerprints from the same person and (ii) the average representation vector distance between two fingerprints from different people. Furthermore, using ROC AUC (which stands for receiver operating characteristic—area under the curve) (*37*), we quantify the ability of these deeply learned fingerprint representations to discriminate between same-person fingerprint pairs and different-person fingerprint pairs, based on representation vector distance (ROC AUC ranges from 0 → 1, where values above 0.5 indicate better discriminative ability).

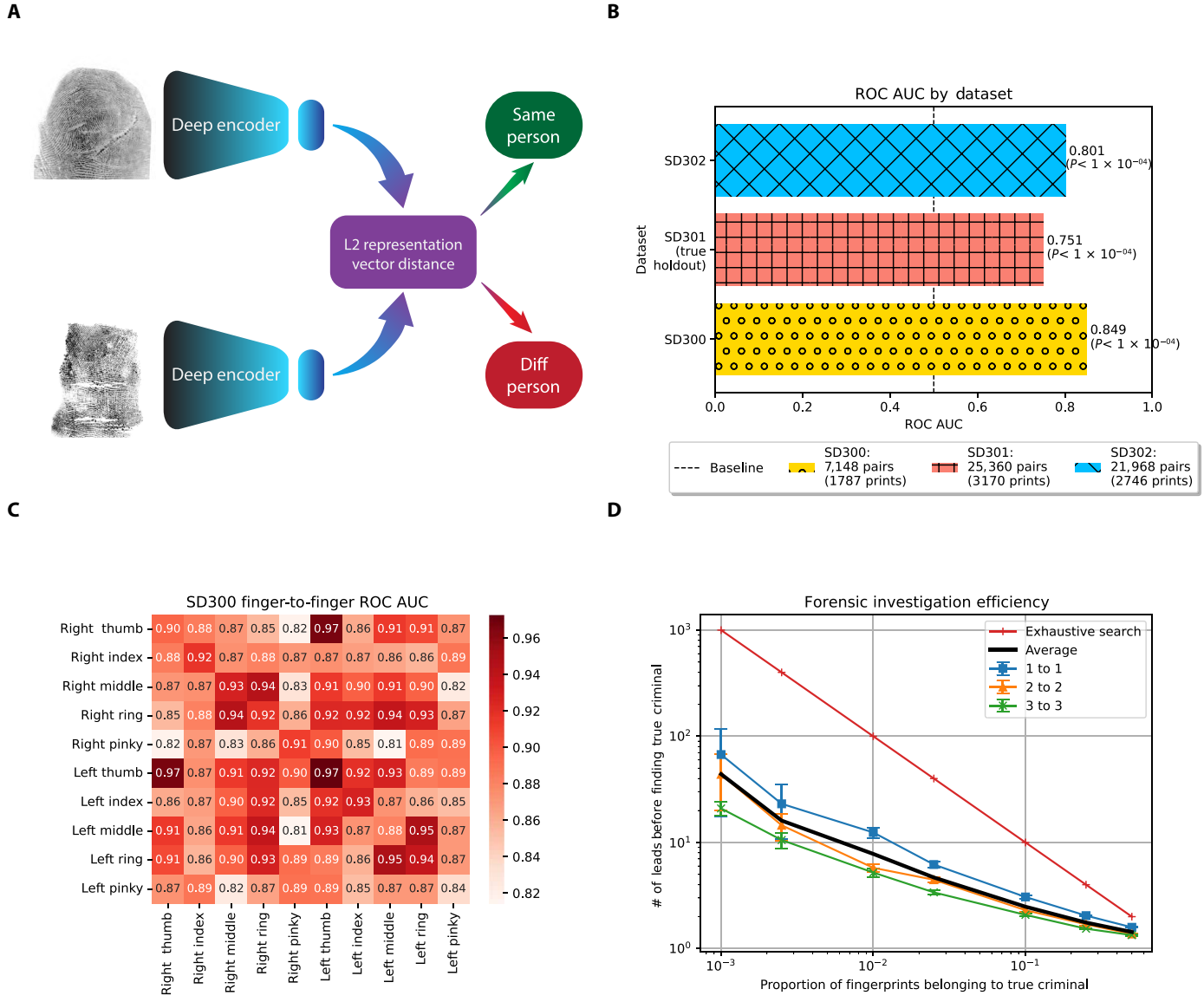


Fig. 1. Overview of cross-finger similarity analysis pipeline and results. Panel (**A**) shows how we use twin neural networks to analyze fingerprints and discover the cross-finger similarity. Panel (**B**) shows that our fingerprint representation vectors reveal a statistically significant difference (as determined by one-sided paired *t* tests) between pairs of distinct fingerprints (e.g., left index versus right pinky) that originate from the same person and from different people. (Please refer to figs. S1 to S3 for prediction examples.) Panel (**C**) shows that the cross-finger similarities (measured by ROC AUC, which stands for receiver operating characteristic—area under the curve, where higher is better) hold true across every conceivable pair of fingers, even from different hands. (Results displayed are from SD300, fig. S4 shows that this result holds across datasets, and fig. S5 shows that these finger-by-finger similarities are statistically significant.) Panel (**D**) (showing means and SEM of a number of leads) validates the usefulness of our discovery, as the cross-finger similarity can reduce the work in criminal justice investigations by more than an order of magnitude. (See fig. S8B for an alternate representation of this data.) Note that we use exhaustive search as our baseline, as there are no existing methods for cross-finger matching.

We trained and validated our model on NIST SD302 (*38*), NIST SD300 (*39*), UB RidgeBase (*40*, *41*), and MSU PrintsGAN (*42*) (more details in Materials and Methods). We tested on NIST SD301 (23 people) (*43*) and different subsets (i.e., containing previously unseen people) of NIST SD302 (20 people) (*38*) and NIST SD300 (90 people) (*39*) than were used in training/validation. We are aware that SD301 was created as a precursor trial run to SD302, with both datasets being collected in Maryland (*38*, *43*). With that in mind, we contacted the National Institute of Standards and Technology (NIST) to find out if there was any overlap between the datasets. The project leader at NIST said that overlap was highly unlikely. For additional verification, we ran NIST's NBIS software (*44*) to conduct a 10-fingerprint comparison of the person from SD301 to each person from SD302. On the basis of this analysis, we found out that there was only one common person between the two datasets: Person 00002239 from SD301, a.k.a. Person 00002455 from SD302. To prevent data leakage, we removed this person from SD301. We randomly selected pairs from each dataset, with an exactly equal number of pairs from the same person and pairs from different people. We ensured that the two fingerprints in each pair were from different fingers (e.g., right index and left pinky) since same-finger matching (e.g., two samples of right index) is already a solved problem (*8*, *9*, *19*, *20*). Furthermore, we forced the two fingerprints in each pair to come from different sampling events [as defined by device code (*38*, *43*) and impression type (*39*)] to minimize the effect of spurious similarities (e.g., user behavior during a data collection session and characteristics of the specific sensor).

Results are shown in Fig. 1B. Our fingerprint representation vectors exhibit statistically significant differences between fingerprints from the same person and fingerprints from different people (as measured by the one-sided paired *t* test), with discriminative power well above the baseline level (as measured by ROC AUC). We note that while the participants in the testing subsets of NIST SD302 and SD300 were never seen during model training, their fingerprint samples come from a similar distribution as the training items from SD302 and SD300, due to them being collected by the same experimental protocol (*38*, *39*). On the contrary, since NIST SD301 was collected in a different experiment than the other two datasets (albeit as a practice run for SD302) (*43*), its distribution is different from that of the training set. Thus, it is expected that the similarities appear to be stronger in SD302 and SD300 than in SD301. It also means that our success in finding similarities in SD301 is very strong proof that our fingerprint similarities are robust and universal.

Finger-by-finger similarities

We further investigate the specific finger-by-finger similarities. For this experiment, we make sure to train our model with equal numbers of all fingers, so that it is not biased toward discerning the similarity for some fingers that happen to appear more frequently in the dataset. As in the previous experiment, we used ROC AUC, but instead of the paired *t* test, we used Welch's one-sided two-sample *t* test (reasons explained in Materials and Methods) (*45*).

Figure 1C displays the results (with additional information in figs. S4 to S6). We see that regardless of which pair of fingers we consider, the similarity is still statistically significant, and the discriminative ability is well above the baseline. This result holds even across pairs of fingers from different hands.

Understanding the similarity

Important features

Now that we have established the existence of a strong similarity among a person's 10 fingerprints, regardless of which fingers we consider, we examine the specific features that contribute to this similarity. Specifically, we analyze the binary patterns, ridge orientation, ridge density, and minutiae, which are commonly used in traditional fingerprint analysis (*8*, *9*, *20*). We analyze the similarity using the same deep learning pipeline as in the general similarity analysis—we train, validate, and test on the feature extraction maps for SD302 (Fig. 2B); and we use the original fingerprint images as a baseline.

Figure 2A shows that all the feature maps exhibit a statistically significant ability to distinguish between pairs of distinct fingerprints from the same person and different people. However, some are clearly better than others. In general, the more fingerprint-like a feature map looks, the more strongly it shows the similarity. We highlight that the binarized images performed almost as well as the original images, meaning that the similarity is due mostly to inherent ridge patterns, rather than spurious characteristics (e.g., image brightness, image background noise, and pressure applied by the user when providing the sample). Furthermore, it is very interesting that ridge orientation maps perform almost as well as the binarized and original images—this suggests that most of the cross-finger similarity can actually be explained by ridge orientation. The most unexpected result was that the minutiae maps barely outperformed random guessing, contrary to the common use of minutiae in traditional same-finger-to-same-finger matching (*8*, *9*, *20*). We believe that this disconnect stems from the fact that minutiae represent the peculiarities of a single fingerprint. This rareness is what makes them so powerful for same-finger matching. However, due to this rareness, we believe that they are unlikely to occur across fingers and therefore may not be as useful for cross-finger matching.

Interrogating the neural network

To further illustrate that the similarity originates from genuine fingerprint patterns rather than spurious similarities (e.g., image background and image brightness), we interrogate the feature maps extracted by our neural networks, with the expectation that these features should resemble fingerprint patterns. We visualize the convolutional filters of our embedder by optimizing the input image via gradient ascent to maximize their activation (*33*, *34*).

The results for the first 16 filters of each selected layer are shown in Fig. 3. Filters were chosen from layers 5, 11, and 17. We observe a trend in the filter visualizations going from the beginning to the end of the network: filters in earlier layers exhibit simpler ridge/minutia patterns, the middle layers show more complex multidirectional patterns, and filters in the last layer display high-level patterns that look much like fingerprints—this increasing complexity is expected of deep neural networks that process images. Furthermore, the ridge patterns in the filter visualizations are all generally the same shade of gray, meaning that we can rule out image brightness as a source of similarity. Overall, each of these visualizations resembles recognizable parts of fingerprint patterns (rather than random noise or background patterns), bolstering our confidence that the similarity learned by our deep models is due to genuine fingerprint patterns, and not spurious similarities.

Pinpointing the crucial areas

To gain insight into what areas of fingerprints contribute to cross-finger similarity, we generate saliency maps that explain our neural network decisions. We use GradCAM (*46*, *47*) with a contrastive scoring function (*30*) to generate the saliency maps.

Figure 4 shows the saliency maps generated from our networks. We observe that our twin neural networks focus primarily on singular regions: areas where the ridge orientation rapidly changes, e.g., deltas (*20*). This explains why orientation maps (Fig. 4D)—which preserve singular

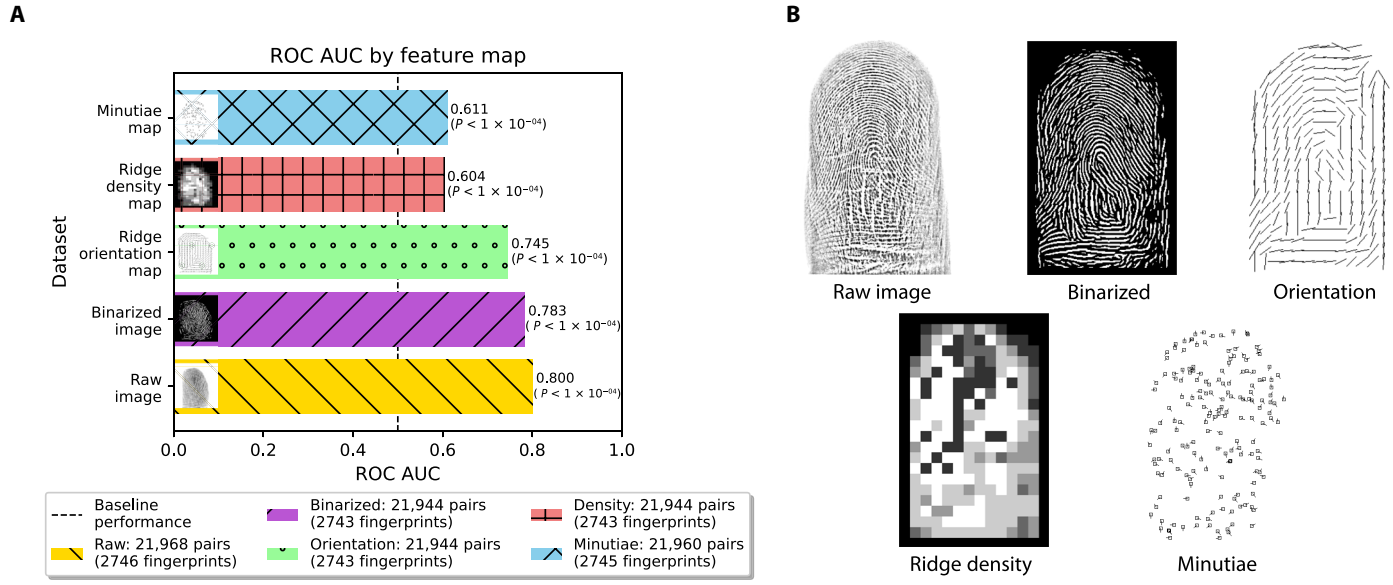


Fig. 2. Feature similarity analysis. As shown in (**A**), the cross-finger similarity can be explained almost exclusively by the binary ridge patterns (rather than spurious effects, like background or image brightness). We also find that ridge orientation (without considering thickness) explains a majority of the cross-finger similarity. However, minutiae are virtually noninformative, belying their widespread prominence in traditional fingerprint matching. (Statistical significance is determined by one-sided paired *t* tests.) Panel (**B**) shows examples of each of the compared feature maps.

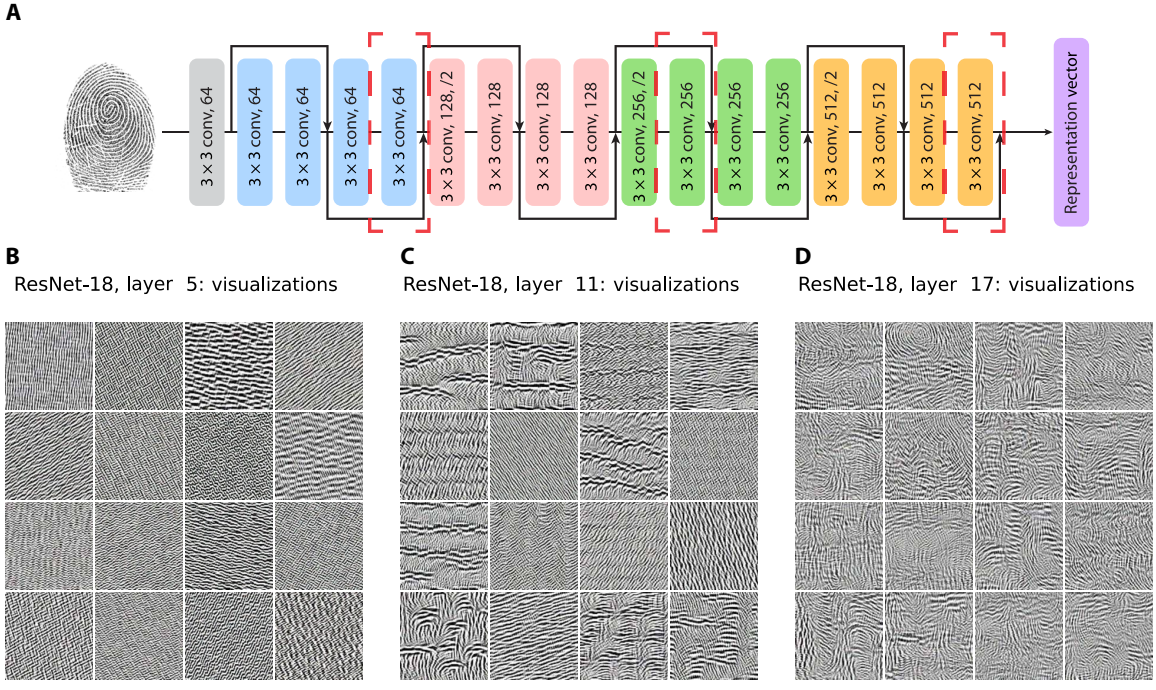


Fig. 3. Visualization of relevant deep-learned features. Our deep twin networks consider features that highly resemble fingerprint patterns, thus allowing us to rule out spurious factors for the similarity. Panel (**A**) shows our network architecture, with dashed boxes surrounding visualized layers. Panel (**B**) shows 16 out of 64 feature maps considered in the low-level layer, panel (**C**) shows 16 out of 256 feature maps considered in the mid-level layer, and panel (**D**) shows 16 out of 512 feature maps considered in the high-level layer.

regions—were shown in Fig. 2 to substantially outperform ridge density (Fig. 4E) and minutiae maps (Fig. 4F). Looking at the high-level features in Fig. 3D, we see that the filters in the final convolutional layer do seem to be searching for singular regions: This shows that the different interpretability methods we use lead to complementary conclusions. We also observe that fingerprint pattern type (e.g., arch, tented arch, left loop, right loop, central pocket loop, double loop, whorl, and accidental) (*20*, *48*, *49*) is considered, but alone cannot explain the intra-person fingerprint similarity: We found numerous examples of correctly matched fingerprints of different pattern types (Fig. 4, B and C, and fig. S1A) and the

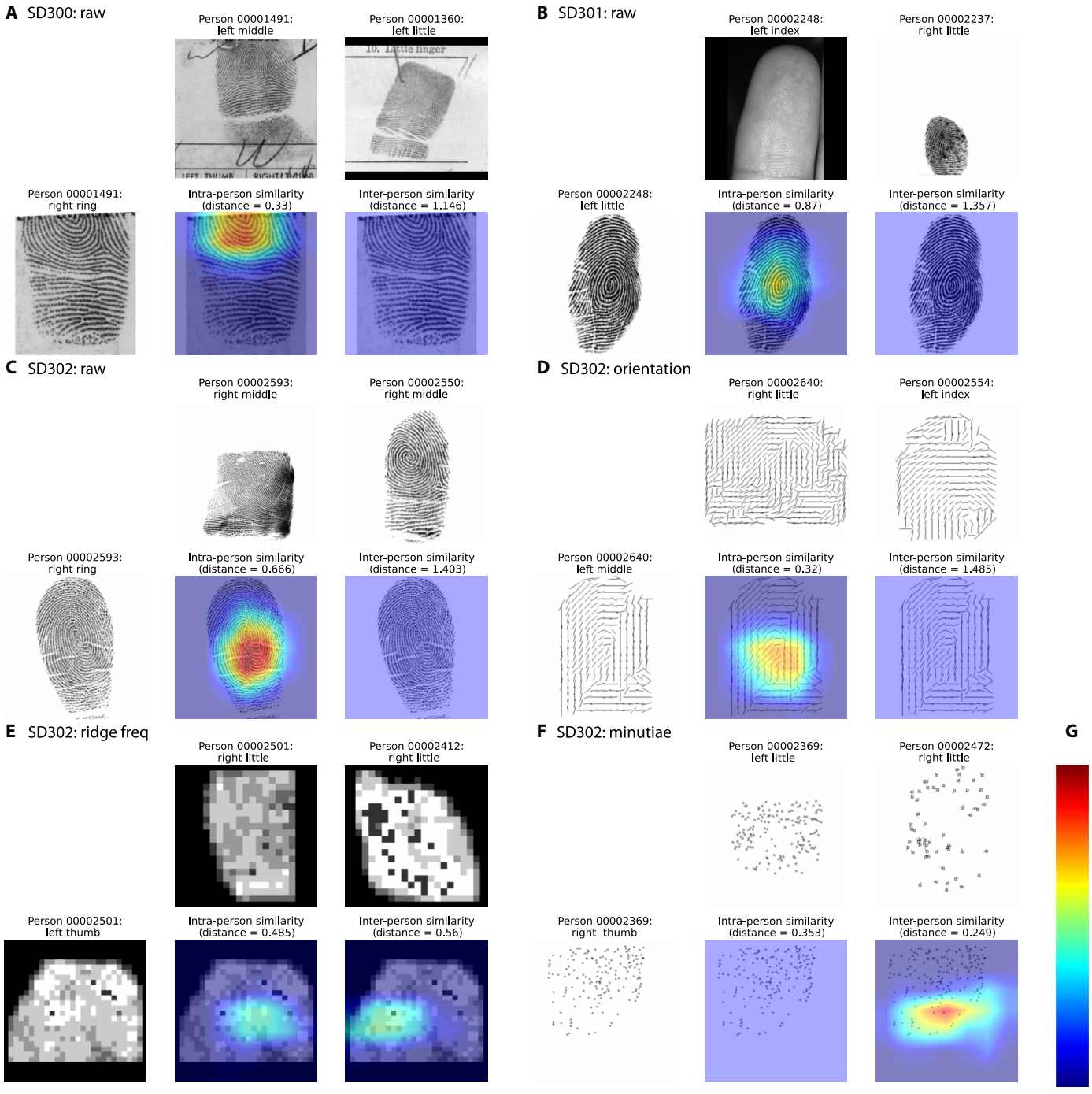


Fig. 4. Saliency maps for cross-finger fingerprint matching. Panels (**A** to **F**) are generated via GradCAM (*46*, *47*) with modified triplet loss (*30*). The bottom leftmost fingerprint in each panel is the reference sample, the top left fingerprint is from the same person, and the top right fingerprint is from a different person. The left saliency map highlights areas that contribute to the similarity between the two fingerprints from the same person, while the right saliency map highlights areas that contribute to the similarity between the two fingerprints from different people. Violet represents unimportant regions, and red represents highly important regions, as shown in (**G**).

same pattern type (Fig. 4A and fig. S3D). More examples are available in the appendix (figs. S1 to S3).

Lead efficiency analysis

One of the most obvious uses for our fingerprint recognition system is to generate leads for law enforcement investigations—thus, it is crucial to understand how many false positives our system generates before producing a true positive.

We model this with a geometric distribution where the probability parameter is the probability of a sample being a true positive, given that the test result is positive. We select the match threshold with the highest f1 score (*50*) and find the expected number of trials

to find the true positive. As this probability depends on the prior (i.e., the proportion of positive examples in the test dataset), we modify our test dataset and repeat the experiment with different proportions. We ran this experiment three times and reported the mean and SEM.

For this experiment, we test N-to-N fingerprint matching (for $N = 1,2,3$), where we have N fingerprints from person A and N fingerprints from person B: This multi-finger scenario is important because criminals may leave behind multiple fingerprints at crime scenes (*51*). We impose the following restrictions: Person A's fingerprints (e.g., right index and left pinky) are from different fingers than person B's fingerprints (e.g., left middle and left ring), and Person A's fingerprints were all collected at the same sampling event, which is a different sampling event than person B's fingerprints [as defined by SD302 sensor code letter (*38*)]. To get the similarity score, we average the representation vector distances for all $N \times N$ pairs of fingers.

We see from Fig. 1D that our method reduces the number of leads required by more than an order of magnitude, as compared to the exhaustive search baseline. This highlights the potential of this test as an important tool in forensic investigations.

Demographic experiments

We also conduct exploratory experiments on the impact of demographic characteristics, such as race and gender, on the intra-person fingerprint similarity. In particular, we want to understand how the intra-person fingerprint similarity generalizes across demographic groups, i.e., if the similarities in one demographic group also apply to another demographic group.

To do this, we create subsets of SD302 based on the demographic group: gender (male versus female) and race (white versus person of color). For a fair comparison, we make sure that each demographic subset has an equal number of people and a roughly equal number of fingerprint images (which means that many of the samples from the majority group are not used in this experiment). Specifically, each gender group has 54 training IDs (i.e., people), 6 validation IDs, and 8 testing IDs; and each racial group has 49 training IDs, 6 validation IDs, and 7 testing IDs. We understand that the number of testing IDs may be lower than we would like to conclude. To combat this, we reshuffle the train-val-test splits five times (similar to cross-validation), such that we have five testing sets with no overlapping participants, making for a total of 40 testing IDs for gender and 35 testing IDs for race.

We train and validate the model from scratch on only one demographic group. Then, we test it on the same demographic group, and we also test it separately on the other demographic group (in both cases, we continue to compare fingerprints from different fingers, like thumb versus pinky). Our rationale for this protocol is that if the performance is better when testing on the same demographic group the model was trained on than testing on the other demographic group, we can infer that at least some of the intra-person fingerprint similarity differs between these demographic groups. Conversely, if the performance is the same (or even better) when testing on the demographic group the model was not trained on, then we have evidence to suggest that the intra-person fingerprint similarity does not differ between these demographic groups. In addition, we train the model on the combined demographic groups and test on individual demographic groups, with the expectation that performance should increase over training on any single demographic group, given that the intra-person fingerprint similarity is generalizable.

Figure 5 shows the ROC AUC (sample means and sample SD) results. For both the gender (Fig. 5A) and racial (Fig. 5B) splits, we clearly see that the model performs the best when it is trained on all demographic groups—this indicates that the similarity is generalizable to a great degree. We also see that in some cases (e.g., male train, male test; white train, white test), the model performs slightly better when tested on the same demographic group it was trained and validated on (e.g., train, validate, and test on whites)—as opposed to either training or testing on a different demographic group—indicating that there may be a slight group-specific portion to the similarity (e.g., there may be a specific kind of cross-finger similarity that males have, but females do not have).

DISCUSSION

We suggest that the intra-person fingerprint similarities are of interest not only because they challenge long-held beliefs but also because this similarity could help improve the ability to find leads for investigations when the fingerprints obtained from crime scenes are from different fingers than the fingerprints already on file. We hope this additional information could help prioritize leads when many possibilities exist, help exonerate innocent suspects, or even help create leads for cold cases.

In a similar vein, our discovery can also help narrow down the candidate list generated by automated fingerprint identification systems (AFIS). Particularly when AFIS draws from large databases, many fingerprints that are not from the culprit may often be returned (*20*). However, with the intra-person fingerprint similarity detected by our work, for every suspect in the list, we can verify if their other fingerprints also satisfy the similarity to the queried fingerprint—if not, then we can eliminate those suspects, thereby reducing the number of close non-matches.

In addition, our work can be useful in digital authentication scenarios. Using our fingerprint processing pipeline, a person can enroll into their device's fingerprint scanner with one finger (e.g., left index) and unlock it with any other finger (e.g., right pinky). This increases convenience, and it is also useful in scenarios where the original finger a person enrolled with becomes temporarily or permanently unreadable (e.g., occluded by bandages or dirt, ridge patterns have been rubbed off due to traumatic event), as they can still access their device with their other fingers.

Limitations and future work

The performance of our system is still markedly below that of state-of-the-art systems designed for same-finger matching (*8*, *9*, *42*), due to cross-finger matching being a substantially harder problem. Our current system, as is, would therefore not be appropriate for use as deciding evidence in court or in authentication situations. Furthermore, even when our system is used only for lead generation, the demographic fairness experiments reveal the risk of certain demographic groups being falsely investigated more often than others—we urge future users to be aware of these biases.

Moreover, since the goal of our work was to validate whether intra-person fingerprint similarities exist at all, we generally used high-quality images of full fingerprints. However, in many real-world scenarios, only low-quality samples of partial fingerprints may be available, e.g., latent fingerprints collected at a crime scene

(*22*). While our model showed some degree of robustness against low-quality (fig. S2D) and partial (figs. S1D, S3B, and S4A) images, many of the failure cases were on low-quality images (figs. S2F and S3E). Furthermore, our analysis was done with a dataset of sufficient size (~60,000 fingerprints across training, validation, and testing; ~7000 fingerprints from 133 people in testing alone) to statistically validate our central claim, but smaller than we would have liked to build a strong real-world system.

We suggest that if the proposed system is trained using very large governmental databases, including partial fingerprints, then substantially more useful performance could likely be obtained. Furthermore, we explored here only a single neural network architecture, one which was sufficient to support our claim but not necessarily optimal. Now that we have established that intra-person fingerprint similarity exists, alternative architectures should be explored, commensurate with the amount of data available.

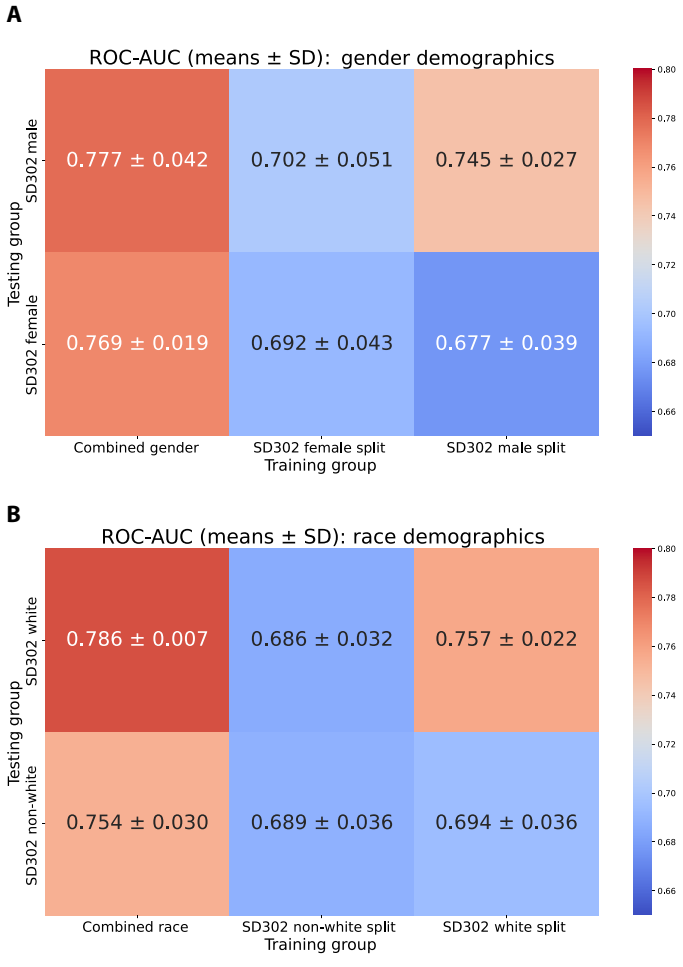


Fig. 5. Generalizability of similarities across demographic groups. Expressed in ROC AUC (sample means and sample SD). We reshuffle the demographic datasets five times, to get five mutually exclusive testing sets per demographic group (similar to cross-validation). Sample means and sample SDs (i.e., *dof* = 5 − 1 = 4) shown in the diagram are calculated from these five reshuffles. Panel (**A**) shows how similarity features learned from one gender generalize to fingerprints from another gender, with 8 × 5 = 40 distinct male test participants and 8 × 5 = 40 distinct female test participants. Panel (**B**) shows how similarity features learned from one racial group generalize to fingerprints from another racial group, with 7 × 5 = 35 distinct white test participants and 7 × 5 = 35 distinct non-white test participants.

MATERIALS AND METHODS

Model architecture

What the architecture should do

As previously stated, we want a system that takes as input two images of fingerprints and outputs whether they are from the same person. To do this, we require a model that takes an image of a fingerprint as input and outputs a latent embedding (i.e., fingerprint representation) vector. The idea is that we can run this model on the two fingerprint images of interest, and then compare the distance between their embeddings. If the distance between the embeddings is lower than some threshold, then we conclude that the two fingerprints are from the same person; otherwise, we predict that they are from different people. In designing this system, we drew inspiration from previous work on face recognition, which used a similar structure to identify with high accuracy whether two images of a face are from the same person (*30*).

Specific technical details

To satisfy these criteria, we use a ResNet-18 model as a feature extractor (*52*). We remove the final output layer, using the penultimate 512-dimensional layer to generate our embedding (i.e., feature representation) vectors. We then normalize the embedding layer to have a magnitude of 1 (this increases the ease and stability of training). To compare the outputted embedding vectors, we do a threshold calculation on squared L2 distance.

Training process

Transfer learning

We train the model in two passes. First, we pre-train on synthetic fingerprint data (MSU PrintsGAN) (*42*). We pre-train because the synthetic fingerprint dataset has substantially more samples than the real fingerprint datasets, so this synthetic dataset is a good source of data for teaching our model how to extract fingerprint features. However, the synthetic dataset only has one finger per person, but our target task is to match different fingers from the same person. Thus, we fine-tune real fingerprint datasets that contain multiple fingers per person (NIST SD302, NIST SD300, and UB RidgeBase) (*38*–*41*).

Training loop

In both pre-training and fine-tuning, we use triplet loss with a margin of 0.2 (*30*), Adam optimizer (*53*) with an initial learning rate of 0.001, and batch size 64. Pre-training has a maximum of 25 epochs, and fine-tuning has a maximum of 250 epochs, subject to early stopping (if the validation loss exceeds the running average over the last 85 epochs). Over the course of the training cycle, we use a cosine decay schedule to decay the learning rate to 10^{-7} by the 250th epoch (*54*).

To train with triplet loss, each sample passed into the model is a tuple: anchor image, positive example, and negative example; where the anchor image is a fingerprint from a given person, the positive example is another fingerprint from the same person, and the negative example is a fingerprint from a different person. We note that this means that the model sees an equal number of examples of fingerprint pairs from the same person and fingerprint pairs from different people.

Then, we iteratively (via gradient descent) minimize anchor-positive distance and maximize anchor-negative distance, using the objective function L for triplet loss [i.e., we want min(L)]: $L(\mathbf{a}, \mathbf{p}, \mathbf{n}) = \max\{d(\mathbf{a}, \mathbf{p}) - d(\mathbf{a}, \mathbf{n}) + \alpha, 0\}$ where $\mathbf{a}$ is the embedding vector for the anchor image, $\mathbf{p}$ is the embedding vector for the positive example, $\mathbf{n}$ is the embedding vector for the negative example, α is a hyperparameter that represents the desired margin between positive and negative examples, and $d()$ is Euclidean distance.

We randomly chose the triplets for training, with a few restrictions. First, we ensure that within a triplet, all the fingerprints are from the same dataset. The reason for this is that images from different datasets generally have different backgrounds and textures, so if we were to train our model with triplets from different datasets, the model would learn to compare the image background and texture, although we want to find the similarities in the actual fingerprint ridge patterns. Second, we ensured that the anchor and positive images were taken from different sampling events. We do this so that the model learns to capture the similarities between fingerprints, regardless of the random noise that may be in a particular sample, e.g., the pressure exerted by the user, characteristics of the sensor, and image background. We also select the validation dataset randomly, but we try to keep the finger pairings (e.g., right index to left pinky) in the dataset relatively balanced in regards to the type of finger.

Dataset

Pre-training

We pre-train our model to improve its ability to extract fingerprint features. For this purpose, we use MSU's PrintsGAN dataset. It is a synthetically generated dataset, consisting of 525,000 fingerprint images from 35,000 made-up "identities"; each "identity" has 15 samples (*42*). Specifically, the PrintsGAN images were pre-created (*42*) using generative adversarial networks (*55*) trained on the Michigan State Police longitudinal database (*15*). We use a 95-3-2 train-val-test split, making sure that there is no overlap among the people in each set.

We pre-train on PrintsGAN rather than ImageNet (*56*) because the samples in PrintsGAN much more closely resemble the samples from our target task. However, PrintsGAN has two limitations: (i) the fingerprints are not from real people, so they do not reflect the variety present in real life (in particular, the images have very little background noise) and thereby may introduce bias into the model and (ii) it does not have all 10 different fingerprints from the same person, it only has multiple samples of one fingerprint from the same "person." These limitations are why we must fine-tune other datasets (*38*). We emphasize that we do not test on PrintsGAN.

Fine-tuning

We fine-tune using 53,315 fingerprint samples from 927 people (and use 5819 fingerprint samples from 114 different people for validation), from the SD302 (*38*), SD300 (*39*), and UB RidgeBase (*40*, *41*) datasets. All of these datasets contain samples from multiple fingers (e.g., left index, right pinky) per person. We do not use the low-quality "latent" fingerprint (e.g., collected from arbitrary surfaces) samples from these databases, since our goal is to discern the cross-finger similarities, and it is harder to extract features from such samples.

Testing

Just as in training, we generate triplets for testing: anchor, positive, and negative. We do this so that we can get an exactly equal number of same-person and different-person pairs, where the same-person pair is obtained by taking the anchor and positive fingerprints, and the different-person pair is obtained by taking the anchor and negative fingerprints.

We test on a holdout portion of SD302 that contains 20 different people than those seen in training and validation; this has 2746 distinct fingerprints. We also test on a holdout portion of SD300 that contains 90 different people than those seen in training and validation; this has 1787 distinct fingerprints. While it is valid to test on these holdout portions of SD302 and SD300, these holdouts are still drawn from similar distributions as the portions of SD302 and SD300 that were used in training since they were collected using the same respective experimental protocols.

Therefore, to test the robustness of our fingerprint representations, we also use SD301 (*43*), a dataset that was collected in a different experiment than SD302 and SD300, and was not used at all during training. Since SD301 was created as a trial run for SD302 (*38*, *43*), we took care to verify that there were no overlapping participants in the data we used—after contacting NIST and running NBIS (*44*) on the datasets, we found a singular overlapping participant, which we removed from SD301. Thus, our clean version of SD301 has 23 distinct people with 3170 fingerprint samples.

Data augmentations

For all images, we add padding to make them square-shaped, resize to 224 × 224, and convert to grayscale. For the training set only, we randomly apply the following transformations: horizontal flip (over the y axis), rotation, translation, shearing, cropping, aspect ratio resizing, Gaussian blurring, noise addition, and brightness scaling.

Metrics

t Test statistical analysis

We conduct one-sided t tests (*35*, *36*) with $\alpha = 10^{-4}$ on (i) the average representation vector (obtained from the twin neural network) distance between two fingerprints from the same person and (ii) the average representation vector distance between two fingerprints from different people. Our alternative hypothesis H_a is that the average representation distance between two fingerprints from the same person is less than that between two fingerprints from different people. For the general similarity and feature similarity experiments, $n_1 = n_2 = \frac{\#\,\text{pairs}}{2}$ can be obtained from the bar charts shown in the main text. For the finger-by-finger similarity, n_1 and n_2 can be obtained from the matrices shown in the appendix. Calculations are done in SciPy (*57*).

For most of our analyses, we conducted paired t tests (Figs. 1B and 2) (*35*, *36*). This is because the testing pairs are generated as triplets ($\mathbf{r_A}$, $\mathbf{f_A}$, $\mathbf{f_B}$), where $\mathbf{r_A}$ is the reference fingerprint from person A, $\mathbf{f_A}$ is another fingerprint from person A, and $\mathbf{f_B}$ is a fingerprint from person B. Thus, we have natural correspondences between samples in the same-person set and samples in the different-person set, i.e., ($\mathbf{r_A}$, $\mathbf{f_A}$) is the analog of ($\mathbf{r_A}$, $\mathbf{f_B}$).

However, for the finger-by-finger similarity analysis (Fig. 1C and figs. S4 to S6), we have to conduct Welch's two-sample t test (*45*). This is because we partition the pairs of fingerprints by which specific fingers they came from, e.g., compare only the pairs that contain a right index and left pinky. Since $\mathbf{f_A}$ and $\mathbf{f_B}$ in the aforementioned triplets may not come from the same fingers (e.g., $\mathbf{f_A}$ is from a left

pinky and $\mathbf{f_B}$ is from a right thumb), each same-person sample no longer has a natural partner in the different-person set.

ROC AUC

ROC AUC (receiver operating characteristic—area under the curve) measures the discriminative ability of our representation vectors between (i) pairs of fingerprints from the same person and (ii) pairs of fingerprints from different people. As for interpretation, ROC AUC ranges between 0 and 1: Higher values mean that we can correctly identify more same-person fingerprint pairs without mislabeling different-person fingerprint pairs to be from the same person, i.e., better discriminative ability. Specifically, a value of 1 indicates that our representations can perfectly discriminate between same-person and different-person fingerprint pairs, a value of 0.5 indicates that our representations show no difference between same-person and different-person fingerprint pairs, and a value of 0 indicates that our representations are completely incorrect (i.e., different-person pairs are always labeled to be from the same person, and vice versa). It is calculated as follows: First, we calculate the finger-to-finger representation vector distance for each selected pair of fingerprints in the testing set (keeping in mind that each pair either contains two fingerprints from the same person or two fingerprints from different people). Then, for every possible classification threshold on the pairwise distance (where a value below the threshold indicates that the fingerprints are both from the same person, and a value above the threshold indicates that the fingerprints are from different people), we calculate $(x, y) = (\text{FPR}, \text{TPR}) = \left(\frac{\text{FP}}{\text{FP}+\text{TN}}, \frac{\text{TP}}{\text{TP}+\text{FN}} \right)$. (TPR stands for true positive rate, which is the percentage of same-person fingerprint pairs correctly identified as such; FPR stands for false positive rate, which is the percentage of different-person fingerprint pairs that are incorrectly identified as coming from the same person.) To get the final ROC AUC value, we take the area under the (FPR, TPR) curve.

Thus, it makes sense why higher ROC AUC values are better: They mean that we are able to achieve higher rates of true matches than false matches. Moreover, the advantage of ROC AUC is that it is threshold-independent, as opposed to metrics like accuracy, which depend on a predetermined threshold. This is especially important in fingerprint recognition, where different tasks require different thresholds, e.g., when generating leads for investigations, we would favor a threshold that maximizes TPR, but when providing evidence in a court of law, we would favor a threshold that minimizes FPR.

Pinpointing the crucial areas with saliency maps

To generate the saliency maps in Fig. 4, we need to give GradCAM a differentiable scoring function, in which higher values indicate more confidence in the prediction. [GradCAM uses the gradients of the scoring function with respect to the convolutional feature maps to pinpoint the most important regions of the image (*46*).] A natural candidate for the scoring function is the triplet loss (*30*) since it is what we use to optimize our model.

We set $\alpha = 0$, resulting in $S(\mathbf{r}, \mathbf{f_A}, \mathbf{f_B}) = \max\{d(\mathbf{r}, \mathbf{f_A}) - d(\mathbf{r}, \mathbf{f_B}), 0\}$, where $\mathbf{r}$ is the reference fingerprint embedding, $\mathbf{f_A}$ and $\mathbf{f_B}$ are embeddings for two query fingerprints, and d is squared Euclidean distance. When S is positive, B is more similar to the reference than A (since B's embedding has a lower distance), i.e., B is predicted to be from the same person as the reference fingerprint. Likewise, when S is 0, A is predicted to be from the same person as the reference fingerprint (since A's embedding now has a lower distance).

For the saliency maps between two fingerprints from the same person (e.g., "intra-person similarity" between person 00001491's right ring and left middle from Fig. 4A), we use $S(\mathbf{r}, \mathbf{f_{diff}}, \mathbf{f_{same}})$. For the saliency maps between two fingerprints from different people (e.g., "inter-person similarity" between person 00002369's right thumb and person 00002472's right little in Fig. 4F), we use $S(\mathbf{r}, \mathbf{f_{same}}, \mathbf{f_{diff}})$. We swap the order of $\mathbf{f_{same}}$ and $\mathbf{f_{diff}}$ to ensure that positive S values correspond to higher saliency.

We use the features generated by the final ResNet-18 block to calculate the saliency map, as is standard practice (*46*, *47*). To reduce noise in the saliency maps, we use the data augmentations and eigen-smoothing recommended by Gildenblat *et al.* (*47*).

Lead efficiency analysis

We perform a statistical analysis to evaluate our model for lead generation in a hypothetical forensic investigation. We modify the test dataset by changing the proportion of positive samples (matching pairs), i.e., the prior. From the model predictions, we calculate the probability (as defined by a geometric distribution) of a true positive, given that the test result is positive. This value is calculated assuming a theoretically infinite dataset (suspect pool) so that the results are invariant to the size of our specific dataset (2260 samples). At low priors (e.g., 0.1% positive samples), the total number of positive samples becomes negligible, so we pass through the dataset again, generating more pairs each time. However, there is a trade-off between the number of samples and the computation time of the experiment. Thus, we select the number of passes such that the total number of positive samples remains at 1130 (half the dataset) or the number of passes reaches 20. We cap the number of passes at 20 but would like to increase it further, given more time and computing in future work.

A graph of the number of leads generated by our method, as compared to an exhaustive search baseline, is provided in Fig. 1D. An alternative representation of these data (as percent reduction in leads compared to an exhaustive search baseline) is provided in fig. S8B. We see that our method is beneficial at low priors where an exhaustive search is difficult.

Experiment-specific modifications

Finger-by-finger similarities

In this experiment, we wanted to give our twin neural networks fair chances to learn the similarities for every finger type (e.g., right thumb and left index). Thus, we take a subset of the SD302 dataset with equal numbers of all finger types: 22,660 samples from 160 people in training, 2290 samples from 20 people in validation, and 2260 samples from 20 people in testing. We also use samples from SD300 with roughly equal (within ±20) numbers of all finger types: 14,118 samples from 710 people in training, 1754 samples from 88 people in validation, and 1787 samples from 90 people in testing. We exclude RidgeBase because it does not contain any thumbs. We also use a balanced subset (i.e., has equal numbers of all finger types) of SD301 in testing (again, excluding the overlapping participant from SD302), with 2490 samples—this acts as an unbiased holdout. In addition, we do not pre-train on PrintsGAN—while it does not adversely affect the performance of our model, it does bias our scientific analysis of which fingers are most similar by giving the model a large number of samples for same-finger matching (which is already a solved problem).

In addition, as noted earlier, we use Welch's two-sample *t* test (*45*) rather than the paired *t* test (*36*) since partitioning the pairs by finger type (e.g., considering only pairs of right thumbs and left middle fingers) means that each same-person fingerprint pair may no longer be in the same group as its corresponding different-person fingerprint pair.

Feature similarities

In this experiment, we train, validate, and test the various feature extraction maps of SD302. We use only SD302 because it is widely regarded as the gold standard publicly available 10-finger dataset (*9*, *42*), and extracting and running deep learning on the feature maps for all the other datasets would be too computationally expensive given our limited resources. Again, we do not pre-train on PrintsGAN because we did not have the computational resources to extract feature maps for all 525,000 images, and to pre-train on the original PrintsGAN images would bias our model to prefer feature maps that more closely resemble full fingerprints, which is contrary to this experiment's purpose of finding the most similar features.

We base our fingerprint binarization, orientation, and ridge density extraction methods on the algorithms described by Hong *et al.* (*58*), Rathore *et al.* (*59*), and open-source code (*60*, *61*). For minutiae extraction, we use NIST's NBIS software, specifically MINDTCT (*44*).

The binarized, orientation, and ridge density maps each have 28,270 training samples from 160 people, 2803 validation samples from 20 people, and 2743 testing samples from 20 people. The minutiae density maps have 28,285 training samples from 160 people, 2805 validation samples from 20 people, and 2745 testing samples from 20 people. The regular SD302 dataset has 28,285 training samples from 160 people, 2805 validation samples from 20 people, and 2746 testing samples from 20 people. This discrepancy in the number of images is because some of the images were too low-quality to extract reliable feature maps from, but those images make up a negligible portion of the dataset.

Lead efficiency

We test on the balanced SD302 dataset for two reasons: (i) SD302 uses a wide variety of sampling modalities, as would be present in a criminal investigation, and (ii) we need to use sets of multiple fingers from the same person in lead generation, and the balanced dataset makes it computationally easier to select sets of fingerprints from distinct fingers of the same person. SD302 was the only dataset tested in this experiment.

Demographics

For the demographic generalizability experiment, we increased the early stopping interval from 85 to 105. This is because since there are fewer training samples in these experiments, the model takes more epochs to start converging, so we do not want to stop it too early.

Supplementary Materials

This PDF file includes:

Sections S1 to S7
Figs. S1 to S10
References

REFERENCES AND NOTES

1. "Set up TouchID on iPhone." https://support.apple.com/guide/iphone/set-up-touch-id-iph672384a0b/ios [accessed 25 August 2021].
2. "Use the fingerprint sensor on your Galaxy phone or tablet."www.samsung.com/us/support/answer/ANS00082563/.
3. C. Gartenberg, "Samsung has sold 2 billion galaxy phones in less than a decade." www.theverge.com/2019/2/20/18233599/samsung-galaxy-phones-sold-2-billion-users-unpacked-2019.
4. J. Kastrenakes, "Apple says there are now over 1 billion active iPhones," 2021.
5. "CJIS digitizes millions of files in modernization push." www.fbi.gov/news/stories/fbi-digitizes-millions-of-files-in-modernization-push.
6. "NGI fact sheet." https://le.fbi.gov/file-repository/ngi-fact-sheet.pdf/view.
7. M. Durose, A. M. Burch, K. Walsh, E. Tiry, Publicly funded forensic crime laboratories: Resources and services, 2014. *Impressions* **50**, 40 (2014).
8. J. J. Engelsma, K. Cao, A. K. Jain, Learning a fixed-length fingerprint representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 1981, 1997 (2019).
9. S. A. Grosz, J. J. Engelsma, R. Ranjan, N. Ramakrishnan, M. Aggarwal, G. G. Medioni, A. K. Jain, Minutiae-guided fingerprint embeddings via vision transformers. arXiv:2210.13994 [quant-ph] (2022).
10. J. Lee, M. Pyo, S. Lee, J. Kim, M. Ra, W.-Y. Kim, B. J. Park, C. W. Lee, J.-M. Kim, Hydrochromic conjugated polymers for human sweat pore mapping. *Nat. Commun.* **5**, 3736 2014.
11. Q. Hao, X.-R. Ren, Y. Chen, C. Zhao, J. Xu, D. Wang, H. Liu, A sweat-responsive covalent organic framework film for material-based liveness detection and sweat pore analysis. *Nat. Commun.* **14**, 578 (2023).
12. B. W. An, S. Heo, S. Ji, F. Bien, J.-U. Park, Transparent and flexible fingerprint sensor array with multiplexed detection of tactile pressure and skin temperature. *Nat. Commun.* **9**, 2458 (2018).
13. H. Jung, S. Sim, H. Lee, Biometric authentication security enhancement under quantum dot light-emitting diode display via fingerprint imaging and temperature sensing. *Sci. Rep.* **13**, 794 (2023).
14. Z. Zhang, X. Zhao, X. Zhang, X. Hou, X. Ma, S. Tang, Y. Zhang, G. Xu, Q. Liu, S. Long, In-sensor reservoir computing system for latent fingerprint recognition with deep ultraviolet photo-synapses and memristor array. *Nat. Commun.* **13**, 6590 (2022).
15. S. Yoon, A. K. Jain, Longitudinal study of fingerprint recognition. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8555–8560 (2015).
16. L. M. Wein, M. Baveja, Using fingerprint image quality to improve the identification performance of the U.S. visitor and immigrant status indicator technology program. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7772–7775 (2005).
17. B. T. Ulery, R. A. Hicklin, J. Buscaglia, M. A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7733–7738 (2011).
18. J. Li, J. D. Glover, H. Zhang, M. Peng, J. Tan, C. B. Mallick, D. Hou, Y. Yang, S. Wu, Y. Liu, Q. Peng, S. C. Zheng, E. I. Crosse, A. Medvinsky, R. A. Anderson, H. Brown, Z. Yuan, S. Zhou, Y. Xu, J. P. Kemp, Y. Y. W. Ho, D. Z. Loesch, L. Wang, Y. Li, S. Tang, X. Wu, R. G. Walters, K. Lin, R. Meng, J. Lv, J. M. Chernus, K. Neiswanger, E. Feingold, D. M. Evans, S. E. Medland, N. G. Martin, S. M. Weinberg, M. L. Marazita, G. Chen, Z. Chen, Y. Zhou, M. Cheeseman, L. Wang, L. Jin, D. J. Headon, S. Wang, Limb development genes underlie variation in human fingerprint patterns. *Cell* **185**, 95–112 (2022).
19. S. Pankanti, S. Prabhakar, A. K. Jain, On the individuality of fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1010–1025 (2002).
20. D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, (Springer, 2009) vol. 2.
21. A. S. Rathore, W. Zhu, A. Daiyan, C. Xu, K. Wang, F. Lin, K. Ren, and W. Xu, Sonicprint: A generally adoptable and secure fingerprint biometrics in smart devices, in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, Toronto, Canada, 16 to 19 June 2020, pp. 121–134.
22. K. Cao, A. K. Jain, Automated latent fingerprint recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 788–800 (2019).
23. H. Cummins, C. Midlo, *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*, vol. 319 (Dover Publications, 1961).
24. A. K. Jain, S. Prabhakarb, S. Pankanti, On the similarity of identical twin fingerprints. *Pattern Recognit.* **35**, 2653–2663 (2002).
25. T. Reed, R. J. Viken, S. A. Rinehart, High heritability of fingertip arch patterns in twin-pairs. *Am. J. Med. Genet. A* **140**, 263–271 (2006).
26. S. B. Holt, Quantitative genetics of finger-print patterns. *Br. Med. Bull.* **17**, 247–250 (1961).
27. A. Roy, N. Memon, A. Ross, Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. *IEEE Trans. Inf. Forensics Secur.* **12**, 2013–2025 (2017).
28. L. Ghiani, G. L. Marcialis, F. Roli, P. Tuveri, User-specific effects in fingerprint presentation attacks detection: Insights for future research, in *2016 International Conference on Biometrics (ICB)*, Halmstad, Sweden, 13 to 16 June 2016 (IEEE, 2016), pp. 1–6.
29. L. Ghiani, G. L. Marcialis, F. Roli, Fingerprint presentation attacks detection based on the user-specific effect, in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, CO, 1 to 4 October 2017 (IEEE, 2017), pp. 352–358.
30. F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 07 to 12 June 2015, (IEEE, 2015), pp. 815–823.
31. J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a "Siamese" time delay neural network. *Int. J. Pattern Recognit. Artif. Intell.* **7**, 669–688 (1993).
32. S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, San Diego, CA, 20 to 25 June 2005 (IEEE, 2005), pp. 539–546.

33. D. Erhan, Y. Bengio, A. Courville, P. Vincent, "Visualizing higher-layer features of a deep network." (Tech. Rep. 1341, Univ. of Montreal, vol. 1341, no. 3, p. 1, 2009.
34. U. Ozbulak, "Pytorch cnn visualizations." https://github.com/utkuozbulak/pytorch-cnn-visualizations, 2019.
35. H. Hsu, P. A. Lachenbruch, Paired t test. *Wiley StatsRef: Statistics Reference Online*, (2014).
36. X. Manfei, D. Fralick, J. Z. Zheng, B. Wang, X. M. Tu, C. Feng, The differences and similarities between two-sample t-test and paired t-test. *Shanghai Arch. Psychiatry* **29**, 184–188 (2017).
37. T. Fawcett, An introduction to roc analysis. *Pattern Recogn. Lett.* **27**, 861–874 (2006).
38. G. Fiumara, P. Flanagan, J. Grantham, K. Ko, K. Marshall, M. Schwarz, E. Tabassi, B. Woodgate, C. Boehnen, "Nist special database 302: Nail to nail fingerprint challenge" (Tech. Rep. 2007, National Institute of Standards and Technology, 2019).
39. G. Fiumara, P. Flanagan, J. Grantham, B. Bandini, K. Ko, J. Libert, "National institute of standards and technology special database 300: Uncompressed plain and rolled images from fingerprint cards" (Technical Note 1993, National Institute of Standards and Technology, 2018).
40. B. Jawade, A. Agarwal, S. Setlur, and N. Ratha, Multi loss fusion for matching smartphone captured contactless finger images, in *2021 IEEE International Workshop on Information Forensics and Security (WIFS)*, Montpellier, France, 7 to 10 December 2021 (IEEE, 2021), pp. 1–6.
41. B. Jawade, D. Mohan, S. Setlur, N. Ratha, V. Govindaraju, RidgeBase: A cross-sensor multi-finger contactless fingerprint dataset, in *2022 IEEE International Joint Conference on Biometrics (IJCB)*, Abu Dhabi, 10 to 13 October 2022 (IEEE, 2022).
42. J. J. Engelsma, S. A. Grosz, A. K. Jain, Printsgan: Synthetic fingerprint generator. arXiv:2201.03674 [quant-ph] (2022).
43. G. Fiumara, P. Flanagan, M. Schwarz, E. Tabassi, C. Boehnen, "National institute of standards and technology special database 301: Nail to nail fingerprint challenge dry run" (Technical Note 2002, National Institute of Standards and Technology, 2018).
44. C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, K. Ko, "User's guide to NIST biometric image software (NBIS)." https://www.nist.gov/publications/users-guide-nist-biometric-image-software-nbis, 2007.
45. B. L. Welch, The generalisation of student's problems when several different population variances are involved. *Biometrika* **34**, 28–35 (1947).
46. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in *Proceedings of the IEEE International Conference on COMPUTER VISION,* Venice, Italy, 22 to 29 October 2017, (IEEE, 2017), pp. 618–626.
47. J. Gildenblat and contributors, "Pytorch library for cam methods." https://github.com/jacobgil/pytorch-grad-cam, 2021.
48. M. Galar, J. Derrac, D. Peralta, I. Triguero, D. Paternain, C. Lopez-Molina, S. García, J. M. Benítez, M. Pagola, E. Barrenechea, H. Bustince, F. Herrera, A survey of fingerprint classification part i: Taxonomies on feature extraction methods and learning models. *Knowl. Based Syst.* **81**, 76–97 (2015).
49. M. M. Chong, H. N. Tan, L. Jun, R. K. Gay, Geometric framework for fingerprint image classification. *Pattern Recognit.* **30**, 1475–1488 (1997).
50. "sklearn.metrics.f1_score." https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html, [accessed 25 September 2023].
51. J. G. Barnes, Fingerprint sourcebook-chapter 1: History, (U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, pp. 1–16, 2010).
52. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, NV, USA, 27 to 30 June 2016, (IEEE, 2016), pp. 770–778.
53. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [quant-ph] (2014).
54. I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts. arXiv:1608.03983 [quant-ph] (2016).
55. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets. arXiv:1406.2661 [stat.ML] (2014).
56. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition,* Miami, FL, USA, 20 to 25 June 2009, (IEEE, 2009), pp. 248–255.
57. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt; SciPy 1.0 Contributors, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261–272 (2020).
58. L. Hong, Y. Wan, A. Jain, Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 777–789 (1998).
59. A. S. Rathore, Y. Shen, C. Xu, J. Snyderman, J. Han, F. Zhang, Z. Li, F. Lin, W. Xu, K. Ren, Fakeguard: Exploring haptic response to mitigate the vulnerability in commercial fingerprint anti-spoofing, in *29th Annual Network and Distributed System Security Symposium*, *NDSS 2022*, San Diego, CA, 24 to 28 April 2022 (The Internet Society, 2022).
60. "Fingerprint-enhancement-python." https://github.com/Utkarsh-Deshmukh/Fingerprint-Enhancement-Python .
61. "FingerprintFeatureExtraction." https://github.com/Utkarsh-Deshmukh/Fingerprint-Feature-Extraction.
62. "Ridgebase benchmark dataset." www.buffalo.edu/cubs/research/datasets/ridgebase-benchmark-dataset.html.
63. "gabeguo/fingerprintmatching: Publicly available source code, to support reproducibility efforts of our cross-finger matching discovery." https://doi.org/10.5281/zenodo.8415788, 2023.
64. D. M. Diez, C. D. Barr, M. Cetinkaya-Rundel, *OpenIntro Statistics* (OpenIntro Boston, MA, 2012).
65. G. C. Foster, D. Lane, D. Scott, M. Hebl, R. Guerra, D. Osherson, H. Zimmer, "An introduction to psychological statistics." https://irl.umsl.edu/cgi/viewcontent cgi?article=1000&context=oer. 2018.

Acknowledgments: We thank R. Zheng, A. S. Rathore, J. Snyderman, C. Xu, K. Wang, and T. Dear for helpful discussions and insights. **Funding:** This work was in part supported by the U.S. National Science Foundation under grant no. CNS-2050910 to W.X. and by the U.S. National Science Foundation under AI Institute for Dynamical Systems under grant no. 2112085 to H.L. **Author contributions:** Conceptualization: G.G., J.G., M.I., H.L., and W.X. Data curation: G.G., A.R., and W.X. Formal analysis: G.G., A.R., J.G., and H.L. Funding acquisition: W.X. and H.L. Investigation: G.G. and A.R. Methodology: G.G., A.R., M.I., J.G., W.X., and H.L. Project administration: W.X., G.G., and H.L. Software: G.G., M.I., and A.R. Supervision: W.X. and H.L. Validation: G.G., A.R., J.G., and W.X. Visualization: G.G., A.R., J.G., and W.X. Writing—original draft: G.G., A.R., J.G., and W.X. Writing—review and editing: G.G., A.R., J.G., M.I., W.X., and H.L. **Competing interests:** G.G., A.R., W.X., and H.L. are inventors on a U.S. patent application 63/516,302 submitted by Columbia University and SUNY Buffalo that covers "Methods and Systems for Cross-Finger Biometric Verification." The other authors declare that they have no competing interests. **Data and materials availability:** Data can be obtained by contacting NIST (*38*, *39*, *43*) (https://nigos.nist.gov/datasets/sd300/request, https://nigos.nist.gov/datasets/sd301/request, and https://nigos.nist.gov/datasets/sd302/request) and SUNY Buffalo (*40*, *41*, *62*) (www.buffalo.edu/cubs/research/datasets/ridgebase-benchmark-dataset.html). Code is available at https://github.com/gabeguo/FingerprintMatching/tree/refactor_training and https://doi.org/10.5281/zenodo.8415788 (*63*).

Submitted 29 March 2023
Accepted 15 December 2023
Published 12 January 2024
10.1126/sciadv.adi0329