



Article Privacy-Friendly Datasets of Synthetic Fingerprints for Evaluation of Biometric Algorithms

Andrey Makrushin *,[†], Venkata Srinath Mannam [†] and Jana Dittmann

* Correspondence: andrey.makrushin@ovgu.de

⁺ These authors contributed equally to this work.

Featured Application: Getting around privacy concerns that arise when evaluating biometric systems with real biometric samples by introducing privacy-friendly datasets of synthetic fingerprints.

Abstract: The datasets of synthetic biometric samples are created having in mind two major objectives: bypassing privacy concerns and compensating for missing sample variability in datasets of real biometric samples. If the purpose of generating samples is the evaluation of biometric systems, the foremost challenge is to generate so-called mated impressions-different fingerprints of the same finger. Note that for fingerprints, the finger's identity is given by the co-location of minutiae points. The other challenge is to ensure the realism of generated samples. We solve both challenges by reconstructing fingerprints from pseudo-random minutiae making use of the pix2pix network. For controlling the identity of mated impressions, we derive the locations and orientations of minutiae from randomly created non-realistic synthetic fingerprints and slightly modify them in an identitypreserving way. Our previously trained pix2pix models reconstruct fingerprint images from minutiae maps, ensuring that the realistic appearance is transferred from training to synthetic samples. The main contribution of this work lies in creating and making public two synthetic fingerprint datasets of 500 virtual subjects with 8 fingers each and 10 impressions per finger, totaling 40,000 samples in each dataset. Our synthetic datasets are designed to possess characteristics of real biometric datasets. Thus, we believe they can be applied for the privacy-friendly testing of fingerprint recognition systems. In our evaluation, we use NFIQ2 for approving the visual quality and Verifinger SDK for measuring the reconstruction success.

Keywords: synthetic biometrics; fingerprint synthesis; synthetic datasets; generative adversarial networks; pix2pix

1. Introduction

The range of applications that make use of biometric user authentication based on fingerprints is very broad. On the one end, there are uncritical applications, e.g., mobile phones can be unlocked by scanning a fingertip. On the other hand, there are very sensitive access-control systems that scan fingers to grant access to bank accounts or governmental services.

The recent trend of developing fingerprint processing and recognition algorithms is an application of deep learning, or to be more precise, training of deep convolutional neural networks (DCNN). In fact, deep learning-based approaches recently outperformed almost all traditional approaches, even in the domain of fingerprint processing. However, the training of DCNN models is data-greedy, meaning that a huge number of samples is needed. This makes collecting a large-scale dataset of biometric samples a prerequisite for the development of up-to-date fingerprint recognition systems.

Considering interpretability and explainability issues related to DCNN, training networks with real biomertic data raises serious privacy concerns because it is hard or almost



Citation: Makrushin, A.; Mannam, V.S.; Dittmann, J. Privacy-Friendly Datasets of Synthetic Fingerprints for Evaluation of Biometric Algorithms. *Appl. Sci.* 2023, *13*, 10000. https:// doi.org/10.3390/app131810000

Academic Editor: Andrew Teoh Beng Jin

Received: 19 June 2023 Revised: 1 September 2023 Accepted: 2 September 2023 Published: 5 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Department of Computer Science, Otto von Guericke University, 39106 Magdeburg, Germany

impossible to ensure that no training samples can be induced from the outcomes of a trained DCNN model [1]. Hence, training datasets need to be privacy-friendly, meaning that they are harmonized with local regulations on the protection of private data. It is important to note that in the European Union, biometric data is a subject of protection by the General Data Protection Regulation (GDPR) and in covered by Article 9. Even if this article might not apply to using real biometric datasets in academic research, biometric data are considered to be a special case of private data, which means that there are strong restrictions on collecting, processing, and sharing this type of data. All in all, the requirements on private data protection make the usage of real biometric data inconvenient or in some cases even impossible. The eventual conflicts with local data protection regulations forced many institutions to withdraw their biometric datasets from public access. For instance, the National Institute of Standards and Technology (NIST) has withdrawn the fingerprint datasets SD4, SD14, and SD27. Instead, the fingerprint dataset SD300 has been made publicly available, for which the Federal Bureau of Investigation (FBI) explicitly confirmed that all subjects presented in the dataset are deceased, see https://www.nist.gov/itl/iad/imagegroup/nist-special-database-300 (accessed on 4 September 2023).

A simple and inexpensive way to get around privacy concerns related to real fingerprints is to introduce virtual individuals along with their synthetic fingerprints. It is important that synthetic fingerprints look like real fingerprints, which implies that they appear realistic to the human eye and that the generated images possess the same statistical characteristics as real fingerprint images. Moreover, synthetic fingerprints should be anonymous, which means that it is impossible to link a synthetic fingerprint to a finger of any natural person that is presented in the dataset used for training of the fingerprint generator.

With the recent development of generative adversarial networks (GAN), the task of realistic image synthesis can be considered solved. As demonstrated in [2,3], the synthesis of *random* fingerprint images that appear realistic to the human eye and possess visual characteristics of fingerprints used for generative model training can be conducted by any advanced NVIDIA GAN architecture such as progressive growing GAN (PGGAN) [4], StyleGAN [5], or StylaGAN2 [6]. However, these architectures neither ensure anonymity nor are capable of generating mated impressions (different fingerprints of the same finger).

If the focus is the evaluation of biometric systems, the synthesis of mated impressions is the most important ability of a generative model. For that, the standard unconditional GAN architecture should be extended by two mechanisms: conditional generation (taking *fingerprint identity* as a condition) and the simulation of intra-class variations. The majority of fingerprint verification algorithms use minutiae as the means of fingerprint matching. Hence, the finger's identity is de facto given by co-allocation of minutiae points. Hence, the straightforward idea to generate a fingerprint for a particular finger is to reconstruct it from minutiae. In our previous paper [7], we have demonstrated that the pix2pix architecture is a powerful and scalable conditional GAN (cGAN) that fits well for reconstructing fingerprint images from minutiae. Reconstruction from minutiae is also a key for fulfilling the requirements on the anonymity and diversity (as these are introduced in [8]) of synthetic fingerprints because the minutiae sets can be created in a pseudo-random way or simply derived from synthetic fingerprints that are known to be anonymous.

The *main contribution* of this paper is in compiling and providing the research community with two datasets of realistic synthetic fingerprints. In order to compile datasets, we combined two concepts: (i) the generation of minutiae sets for mated and non-mated impressions and (ii) the reconstruction of realistic fingerprints from minutiae. The latter concept has been introduced in our previous report [7], while the former concept is the *key novelty* of this paper. In particular, our *novel* approach is in combining model-based and data-driven fingerprint generation techniques to create *realistic* and *anonymous* fingerprints with an option for *mated* impressions. The high-level view of our approach for compiling datasets is depicted in Figure 1. The dashed lines refer to the parts introduced in our



previous report, while the solid lines and text in bold refer to the new parts firstly reported in this paper.

Figure 1. A concept of identity-aware synthesis of fingerprints by reconstructing patterns from minutiae.

We first define 500 virtual subjects with 8 fingers each. For each finger, we define a basic pattern based on global statistics and generate a non-realistic fingerprint with this basic pattern as a source of minutiae (denoted as non-realistic pseudo-random synthetic fingerprint). After extracting minutiae, we slightly manipulate minutiae locations and orientations as well as remove a small subset of minutiae (denoted as pseudo-random minutiae modification) to create nine further sets of minutiae for generating mated impressions. Next, we synthesize ten impressions per finger, making use of the two selected pix2pix models from our previous study [7]. In total, each dataset comprises 40,000 samples.

The synthetic fingerprints are assessed using the NIST fingerprint image quality (NFIQ) 2 tool for approving their visual quality and Verifinger SDK for measuring the reconstruction success. Subjectively, the fingerprints are almost indistinguishable from real ones. The synthetic datasets are designed to possess characteristics of real biometric datasets, so we believe they can be applied for the privacy-friendly testing of fingerprint recognition systems. Both datasets and generative models are publicly available at https://gitti.cs.uni-magdeburg.de/Andrey/gensynth-pix2pix (accessed on 4 September 2023).

Hereafter, the paper is organized as follows: Section 2 provides an overview of related works. In Section 3, we first elaborate on the generation of pix2pix models and then show how these models are applied to a compilation of synthetic datasets with mated and non-mated impressions. In Section 4, we first assess the performance of the generative models and then approve the utility of our synthetic datasets. In Section 5, we critically discuss the evaluation results. Section 6 concludes the paper with the summary of results and future work.

2. Background and Related Work

Having a reliable and robust generator of fingerprint images is a prerequisite for the compilation of synthetic datasets. This section presents an overview of studies related to model-based and data-driven fingerprint generation techniques.

2.1. Model-Based Fingerprint Generation

Model-based approaches can be roughly classified in physical and statistical modeling.

2.1.1. Physical Fingerprint Modeling

Physical modeling approaches strive to explain the process of forming ridge-line patterns on fingertips and simulate the pattern generation process by differential equations. For instance, in [9], the authors consider the basal layer of the epidermis as an elastic sheet influenced by neighboring tissues. By modeling the buckling process using Karman equations, they confirm that ridges are formed in the direction perpendicular to the direction of greatest stress. They define two factors contributing to stress: resistance at the nail furrow and flexion creases, and the regression of volar pads during finger-print development. The fingerprint patterns generated in computer simulations confirm these hypotheses. In [10], Legendre polynomials are used for modeling fingerprint ridge orientations, resulting in a discontinuous orientation field. The method is effective for

modeling singular points by zero-poles of the polynomials but comes with a trade-off in terms of computational complexity. The study in [11] outlines the limitations of modeling fingerprints by the phase portraits of differential equations.

2.1.2. Statistical Fingerprint Modeling

In contrast to physical modeling, statistical modeling approaches build on statistical models trained using publicly available databases of real fingerprints to retain predefined features, including singular points, orientation fields, and minutiae [12]. In [13], the authors propose a process that involves generating multiple master fingerprints and subsequently producing different fingerprint impressions for each master fingerprint. The generation of master fingerprints comprises four steps: the generation of fingerprint shape, directional map, density map, and ridge pattern. Likewise, the process of fingerprint impression generation consists of four steps: ridge average thickness variation, distortion, the addition of noise, and global rotation. The study in [14] introduces a technique for modeling texture characteristics of real fingerprints (incl. ridge intensity, ridge width, ridge cross-sectional slope, ridge noise, and valley noise) to be reproduced in synthetic fingerprints.

The approach from [13] is implemented and advanced in the commercial tool SFinGe [15], which is de facto the state-of-the-art tool for statistical modeling of fingerprints. It starts with defining a fingerprint area and estimating orientation and frequency maps from a given basic pattern and singularity points. Then, Gabor filters are applied to iteratively draw ridge lines along the orientation lines, resulting in a master-fingerprint. Minutiae points are randomly sampled based on constructed ridge lines. Next, the physical characteristics of fingers and of sensor characteristics, as well as the contact between fingers and the sensor surface, are simulated by applying further filters. As a result, the fingerprint patterns appear more realistic. Anguli [16] is an open-source re-implementation of the core functionality of SFinGe. The tool is available at https://dsl.cds.iisc.ac.in/projects/Anguli (accessed on 4 September 2023) and can be perfectly utilized for the synthesis of idealistic non-realistic fingerprint patterns.

Although the aforementioned model-based approaches can be applied for the compilation of large-scale fingerprint databases, they generally suffer from lacking realism, meaning that synthetic fingerprints are often visually distinguishable from real fingerprints [17].

2.2. Data-Driven Fingerprint Generation

In contrast to fingerprint modeling, data-driven approaches rely on learning the appearance of fingerprints from data. The most commonly used technique for realistic data synthesis is the generative adversarial network, or simply GAN.

2.2.1. Basics of GAN

GANs were originally introduced in [18]. GANs belong to the family of deep neural networks as they consist of two deep neural networks called generator (G) and discriminator (D). In traditional GANs, G obtains a random latent vector and produces a synthetic sample that should resemble the training samples, while D decides whether the given instance is original or synthetic. The training of G and D is done interchangeably based on an objective function. Here, the objective of G is to produce synthetic instances that are able to fool D. The objective of D is to make no errors in discriminating between real and synthetic instances. The weights of G and D are updated via back propagation.

An input of *G* is a random latent vector *Z* sampled from the uniform or normal distribution. Given *Z*, *G* generates a synthetic instance. During GAN training, *Z* is not controlled. After training, the latent space can be explored by modifying inputs and observing the changes in outputs. For producing more diverse outputs, the authors of [5] propose to add noise.

The role of D is in guiding G to produce more realistic data. D acts as a traditional binary classifier. It obtains either a training sample (labeled as real) or a synthetic sample generated by G (labeled as fake). The outcome of D is the adversarial loss used to adapt

both G and D. After the GAN training is finished, D is dropped and G is used as a standalone tool.

In the initial work [18], *G* contains only fully connected layers. As suggested in [19], the original GAN struggles to process high dimensional data such as images because *D* always dominates the scene. In order to work with images, the authors of [20] proposed to use deep convolutional neural networks (DCNNs) in the GAN architecture. Recently, NVIDIA research group proposed several advanced architectures such as PGGAN [4], StyleGAN [5], StylaGAN2 [6], and StylaGAN3 [21] that still dominate the field.

In order to control the GAN generated output, the conditional GAN (cGAN) architecture is introduced in [22]. Instead of feeding a random latent vector to the generator, the GAN is fed with some meaningful information called "condition". This is how GAN is guided to generate specific data. CycleGAN [23] and pix2pix [24] are two popular and effective cGAN architectures for image-to-image translation.

2.2.2. Fingerprint Generation via GANs

According to [25], GANs outperform statistical modeling at generating fingerprint images by capturing their underlying probability distribution and replicating the overall appearance from training samples, rather than modeling fingerprint attributes. Although GANs enhance the visual quality of generated samples, they fall short of accurately capturing the crucial fingerprint characteristics due to their reliance on random inputs during synthesis.

Finger-GAN proposed in [26] is the very first study that utilizes a GAN-based framework to create synthetic fingerprint images. The proposed approach uses the deep convolutional generative adversarial network (DCGAN) [20] with a modified loss function. The primary goal is that the lines in created fingerprint images are connected, just like they are in real fingerprints. The network is trained based on two fingerprint databases fingerprint verification competition (FVC) 2006 [27] and The Hong Kong Polytechnic University (PolyU) [28], making use of data augmentation. The performance is validated using the Frechet inception distance (FID) [29].

Research in [30] proposes a GAN model with loss-doping, which helps prevent a mode collapse and improves convergence. The generator incorporates residual connections, enhancing stability throughout the training process. Data augmentation techniques are employed. Due to hardware limitations, the synthesized fingerprints are limited to the size of 256×256 pixels. However, it is stated that the concept can be extended to synthesize bigger images.

In [2,3], it is demonstrated how the common NVIDIA GAN architectures such as PGGAN, StyleGAN, or StyleGAN2 can be applied to the random generation of partial and full realistic fingerprints.

The study in [31] proposes an approach of synthesizing high-resolution fingerprints of up to 512 \times 512 pixels using the improved Wasserstein GAN (IWGAN) [32] and a convolutional autoencoder. The methodology consists of two phases: training the convolutional autoencoder and then training the IWGAN. The autoencoder includes an encoder to embed the input fingerprint and a decoder to reproduce the original fingerprint from the embedded vector. The IWGAN is trained with the generator initialized using the decoder part of the autoencoder. The fingerprints generated in this way lack uniqueness. This issue is addressed in [33] by incorporating the identity loss, guiding the generator to synthesize fingerprints corresponding to more distinct identities.

In [34], it is proposed to combine traditional statistical modeling with CycleGAN to create master-fingerprints with level three features (sweat pores) and make them appear realistic by performing style transfer with the image-to-image translation. This approach allows for producing several instances of one finger.

The most recent approach utilizing GANs for fingerprint synthesis is PrintsGAN [17]. The synthesized samples closely replicate the minutiae quantity, type, and quality distributions observed in target fingerprint datasets. The proposed method gives increased control

over the synthesis process and enhances the realism of rolled fingerprints in high-resolution images with a size of 512 \times 512 pixels.

2.3. Fingerprint Reconstruction from Minutiae

Reconstructing fingerprints from the minutiae templates or the inversion of fingerprint templates was initially addressed in 2001 [35]. However, intensive research in this area started in 2007 by [36,37]. Regarding the model-based approaches, the process is technically similar to fingerprint synthesis with the only difference being that singularity points and the orientation map are estimated from minutiae and not vice versa. The most common approach in [36] estimates the orientation map using a modified model from [38] and applies Gabor filtering iteratively, starting from minutiae locations. The final rendering step enhances realism. In [37], a fingerprint image is generated from a skeleton image, which is reconstructed from minutiae. The process involves estimating an orientation map using minutiae triplets and drawing streamlines starting from minutiae and border points. In [39], a unique method is introduced where a phase image is reconstructed from the minutiae template and subsequently converted into a fingerprint image. In [40], an alternative approach is presented, wherein ridge patterns are reconstructed using patch dictionaries. This method enables the synthesis of idealistic ridge patterns that clearly lack realism.

Relying on the assertion that real fingerprints exhibit non-random distributions of minutiae, with different fingerprint types showcasing distinct patterns, the authors in [25] propose an integrated framework that combines fingerprint reconstruction and synthesis. They begin with StyleGAN2 training from the NIST SD14 database [41] to facilitate fingerprint synthesis with the generator part of the model. Then, DCNN (referred to as the minutiae-to-vector encoder) is trained to produce embeddings that act as an input for the GAN generator. By steering latent vectors, visual characteristics of fingerprints such as dry skin artifacts can be controlled.

The study in [42] proposes approaches for fingerprint reconstruction from minutiae and from deep neural network embeddings and compares them to each other. The effectiveness of inversion attacks is evaluated, and it is suggested that reconstruction from embeddings is less successful than from minutiae.

The application of pix2pix to fingerprint reconstruction from minutiae was first introduced in [43]. The original challenge is transformed into an image-to-image translation task. First, minutiae points are depicted on a minutiae map, an image visually representing all essential minutiae information. Then, a minutiae map is translated to a fingerprint image. Research in [7,44] advances the concept of [43] and demonstrates state-of-the-art results in fingerprint reconstruction. A new minutiae encoding approach is introduced, and it is shown in cross-sensor and cross-dataset experiments that the pix2pix models have potential to generalize [44] effectively. However, the original pix2pix [24] can only handle images of 256×256 pixels and lower. In [45], pix2pixHD is proposed to handle larger images. Our previous study in [7] introduces a simple yet effective method to handle 512×512 pixel images. In contrast to [45], we extend both the discriminator and generator of the original pix2pix architecture by adding one convolutional layer. We also proposed an alternative minutiae encoding approach and trained several generative models based on real and synthetic fingerprint data.

The work in [7] is an important component of the work presented in this paper enabling generation of large-scale datasets of mated and non-mated fingerprints.

Several aforementioned fingerprint synthesis techniques were applied to the compilation of synthetic datasets. However, only a few of them comprise a large amount of realistic full fingerprints at their native resolution of 500 ppi and also contain mated impressions. The most widespread fingerprint datasets used in different years for the fingerprint verification competition (FVC2000 [46], FVC2002 [47], and FVC2004 [48]) contain only 880 synthetic samples each. The L3-SF dataset created in [34] contains 7400 partial fingerprints. The fingerprint dataset from Clarkson University [3] contains 50,000 full non-mated fingerprints. The only existing synthetic dataset surpassing our datasets in terms of a fingerprint number is the MSU PrintsGAN dataset [17] with 525,000 images of 35,000 identities (fingers) with 15 impressions each. However, the fingers in PrintsGAN are not assigned to virtual subjects and the frequency of the occurrence of particular basic patterns is not addressed.

3. Research Methodology

The research presented in this paper comprises two parts: the training of generative models that will further assist us in generating fingerprints of virtual subjects (Section 3.1) and the compilation of two exemplary biometric datasets of synthetic fingerprints (Section 3.2). Note that the former part is mostly derived from our previous work [7] and replicated here to simplify the understanding of the complete approach.

While defining the pipeline for the generation of datasets of synthetic images appropriate for the evaluation of fingerprint matching algorithms, we have in mind four objectives: (1) appropriate image resolution, (2) the realistic appearance of generated fingerprints, (3) the anonymity of generated fingerprints, and (4) the ability to produce mated fingerprints.

The former two objectives are covered by the design of the pix2pix models trained in our previous work [7], which allow for reliable fingerprint reconstruction from minutiae. To cover the latter two objectives, we propose to apply the Anguli tool to generate random non-realistic fingerprints with predefined basic patterns. These non-realistic fingerprints are used as a source of minutiae from which realistic fingerprints are generated using the pix2pix models. In doing so, we ensure the anonymity of the resulting synthetic fingerprints. In order to produce several mated impressions for one finger, we propose to modify the minutiae extracted from Anguli fingerprints to preserve the identity of the finger and use these modified minutiae sets as an input for pix2pix models. In doing so, we ensure that the outcomes of a pix2pix model are the mated fingerprints, if the transformations applied to minutiae locations and orientations are in a certain range.

3.1. Training of Generative Models

In the following, we formally describe the process of constructing our generative models.

3.1.1. Problem Statement

Let *I* be a fingerprint image and $L : L_i = (x_i, y_i, t_i, \theta_i)$ be the set of minutiae where (x_i, y_i) is a location, t_i is a type (either bifurcation or ending), and θ_i is an orientation of the i_{th} minutiae. Our objective is to train a cGAN generator *G* so that it is able to generate a fingerprint image I^* from a minutiae set *L*, or formally $I^* = G(L)$. The resulting synthetic fingerprint image I^* should: (i) depict a full plain fingerprint at a resolution of 500 ppi, (ii) appear realistic to the human eye, and (iii) yield a high matching score when it is biometrically compared to the original fingerprint image *I*.

The first requirement (i) will be automatically fulfilled if we generate images of 512×512 pixels. In our estimation, the realistic appearance (ii) is proven indirectly by checking the NFIQ2 score [49] of a fingerprint image. For the purpose of biometric comparison of fingerprints (iii), we decided for Neurotechnology Verifinger SDK v12.0 [50].

Note that minutiae extraction can be conducted by any arbitrary tool. For the sake of simplicity, we use the same tool as for biometric fingerprint matching (Verifinger SDK). Let MinExt denote the minutiae extraction function; then, the process can be formalized as L = MinExt(I).

3.1.2. Minutiae Map Generation

The most crucial step in pix2pix-based fingerprint reconstruction is the transformation of a minutiae template to an image depicting minutiae. We call such an image a minutiae map. This process is further referred to as minutiae encoding and visualized in Figure 2. Let



M denote a minutiae map and *Enc* the minutiae encoding function; then, M = Enc(L) = Enc(MinExt(I)).

Figure 2. Construction of a minutiae map. Minutiae extraction is followed by minutiae encoding either by "directed lines" (DL) or by "pointing minutiae" (PM). Black color is used for endings and white for bifurcations. Minutiae orientation is given by the angle θ_i . The fingerprint is from the Neurotechnology CrossMatch dataset [51].

In the initial work on application of pix2pix to fingerprints [43], minutiae are encoded by gray squares. Formally, we encode every minutiae L_i in the minutiae list L by drawing a gray square of a fixed size and centered at (x_i, y_i) in an image of the same size as the original fingerprint image. The intensity of a gray value encodes the minutiae orientation θ_i . The distribution of colors and the size of a square may vary depending on an implementation. However, shades of gray used as an orientation encoding are not optimal because the slight difference in colors may dilute during convolutions.

In [7], we focused on two alternative minutiae encoding approaches: "directed lines", also referred to as DL, and "pointing minutiae", also referred to as PM.

Directed Lines (DL). Each minutiae L_i from the minutiae list L is encoded by a directed line, which starts at (x_i, y_i) and is pulled in the direction defined by the angle θ_i . Bifurcations are represented by white lines (intensity value of 255) and endings by black lines (intensity value of 0). The intensity value for the background of a minutiae map is 128. This selection of intensity values is not random. The idea is to underscore the dualism of bifurcations and endings. It is claimed that directed line encoding outperforms gray square encoding for our purpose. For 500 ppi fingerprint images, we use a line with the length of 15 pixels and the width of 4 pixels.

Pointing Minutiae (PM). Pointing minutiae encoding can be seen as a combination of squares and directed lines. For a minutiae L_i from the minutiae list L, there is a square centered at (x_i, y_i) and a line pulled in the direction defined by the angle θ_i . As in directed line encoding, bifurcations are represented by white lines (intensity value of 255) and endings by black lines (intensity value of 0). The intensity value for the background of a minutiae map is 128. As for directed line encoding, pointing minutiae encoding underscore the dualism of bifurcations and endings resulting in invariance to color inversion. For 500 ppi fingerprint images, we use a line with a length of 15 pixels and a width of 4 pixels. The square size is 7×7 pixels.

3.1.3. Pix2pix Architecture

Pix2pix is a conditional GAN architecture proposed for image-to-image translation. Examples are converting satellite views to street views, summer landscape to winter landscape, or daylight pictures to night pictures. In our considerations, the generator G of pix2pix converts the minutia map M derived from a minutiae list L as described in Section 3.1.2 to a synthetic fingerprint image I^* , see Equation (1).

$$I^* = G(M) = G(Enc(L)) = G(Enc(MinExt(I)))$$
(1)

The basic characteristics of the pix2pix are that it uses convolutional neural network (CNN) layers, which consist of strided convolutions. Especially for compressing the image data, the stride, which is a parameter of CNN, helps in determining how many pixels should be passed moving over the given image. Even though pooling can be used for down-sampling the image, striding has the advantage that it can learn the parameters by itself through training. The network also uses the normalization technique, which helps to reduce model training time and an internal covariance shift. The standard approach is using batch normalization in *G* to make a gradient flow smoother for handling overfitting. The network also uses the dropout layer, which randomly drops out some neuron outputs in the layers. The activation functions are either ReLU or Leaky-ReLU for achieving non-linearity.

Generator with Skips. The role of *G* is to take a minutiae map *M* as an input and generate the fingerprint I^* as close to the original fingerprint *I* as possible.

The backbone of G is the U-Net network introduced in [52]. The proposed U-Net mainly contributes to extending the architecture [53], which also converges with good metrics for a small number of training samples. There are other proposed approaches where the network architecture is referred to as Encoder-Decoder [54,55]. In such networks, the encoder takes the input, progressively down-samples it, compresses it into a bottleneck representation, and progressively up-samples it. Here, the key information is the bottleneck representation. In contrast to such networks, the U-Net-based networks have skip connections, meaning that the contracting layers (down-sampling) are concatenated with corresponding expansion layers (up-sampling). The information that is being passed through the contracting path helps for localizing high resolution features. The generator *G* is visualized in Figure 3. The contracting layers apply "Conv 2d", "Instance-Norm", and "Leaky-ReLU" operations, while the expansion layers apply "Conv-Transpose 2d" and "Instance-Norm" operations. The very first and very last layers apply no normalization. The very last layer finishes the processing with "Tahn" activation. Even though the proposed generator G utilizes U-Net, no image modifications such as cropping, corner-cut, etc., are made at the contracting path, meaning that the input and output dimensions are the same.

The generator loss comprises adversarial loss and L1 loss, see Equation (2).

$$Loss_G = Loss_{Adv} + Loss_{L1} \tag{2}$$

The adversarial loss is derived from the discriminator *D* and is computed with binary cross entropy (*BCE*), see Equation (3).

$$Loss_{Adv} = BCE(D(G(M)))$$
(3)

The *L*1 loss is computed by taking the mean absolute error (*MAE*) between the original image *I* and generator synthesized image I^* and multiplied with a penalty factor λ ; see Equation (4). λ helps to reduce the visual artifacts in some applications.

$$Loss_{L1} = \lambda \cdot MAE(I^*, I) \tag{4}$$



Figure 3. Pix2pix generator architecture.

PatchGAN-based Discriminator. The discriminator *D* that is used in pix2pix work is also commonly known as the PatchGAN discriminator (see Figure 4). The architecture of *D* is a sequence of convolutional layers. Each layer applies "Conv 2d", "Batch-Norm", and "Leaky-ReLU" operations. In the very first layer, "Batch-Norm" is not present. While in traditional GAN, the discriminator classifies the given input as fake or real by a single output neuron unit, the PatchGAN discriminator makes a decision based on the majority voting of several decisions made at an image patch level. A single patch of the input is propagated through the network and then mapped to a single output neuron unit. The same happens to all image patches. The advantage of propagating forward in such a way is that we end up with a network that has fewer parameters. Moreover, the inference can be performed faster as patches can be convolved independently in parallel with less memory consumption. In the formal description of *D* (see Equation (5)), *N* denotes the total number of patches in the input, and $D(P_{ij})$ denotes the discriminator's prediction for the patch located at (*i*, *j*) within the input tensor, which comprises a fingerprint and minutiae map.

$$D(I|I^*, M) = \begin{cases} 1 & \text{if } \frac{1}{N} \sum_{i,j} D(P_{ij}) \ge 0.5\\ 0 & \text{otherwise} \end{cases}$$
(5)

The patch size varies with reference to the network parameters. However, the patch size is defined based on the network parameters. The individual patch is also referred to as a receptive field. The overall discriminator D and a single patch convolution can be seen in Figure 4. Please note that in contrast to the original network in [24], which accepts at most 256×256 pixel images, our network accepts 512×512 pixel images as input. Consequently, the patch size is 142×142 , which is much larger than the original paper.



Figure 4. Pix2pix PatchGAN discriminator: Every layer includes "Conv 2d" followed by "Batch-Norm" and "Leaky-ReLU". In the very first layer, "Batch-Norm" is not present.

For every single training instance, the discriminator loss is the sum of binary cross entropy (BCE) values of original and synthesized images. Given that I and M are sets of original images and their minutiae maps, respectively, the overall loss of D is given by Equation (6).

$$Loss_D = BCE(D(I)) + BCE(D(G(M)))$$
(6)

3.1.4. Concept Overview

We finalize the description of the concept of applying a pix2pix network (imageto-image translation) to the task of fingerprint reconstruction from minutiae maps, with Figure 5 depicting a diagram that gives a schematic description of all components and interactions between them.



Figure 5. An overview of the fingerprint reconstruction concept based on a pix2pix network. The fingerprint is from the Neurotechnology CrossMatch dataset [51].

3.1.5. Technical Aspects of Model Training

As a backbone for our generative models, we cloned the pix2pix network from the following repository: https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix (accessed on 4 September 2023) and extended both the generator and discriminator architectures by one convolutional layer. This step enables training and further synthesis with full fingerprints at a native resolution in 512 × 512 pixel images. The modified pix2pix network is available at https://gitti.cs.uni-magdeburg.de/Andrey/gensynth-pix2pix (accessed on 4 September 2023).

For training, we use the desktop PC with the AMD Ryzen 9 3950X 16-Core CPU (3.5 GHz) and 128 GB RAM. It has two NVIDIA Titan RTX GPUs with 24 GB VRAM each.

Training Dataset. Considering the requirements mentioned in the beginning of Section 3.1, we decided to use high-fidelity plain fingerprints for training our generative models. Our primary focus is on fingerprints captured by optical biometric sensors similar to CrossMatch Verifier 300 [56].

There are three subsets in our training dataset:

- 1. The 408 samples from the Neurotechnology CrossMatch dataset. The images are license free and can be downloaded from:
 - https://www.neurotechnology.com/download.html (accessed on 4 September 2023).
- 2. The 880 samples from the FVC2002 DB1 A+B dataset. This dataset has been created for the Second International Fingerprint Verification Competition (FVC2002) back in 2002:

http://bias.csr.unibo.it/fvc2002/databases.asp (accessed on 4 September 2023). Note that in contrast to other two datasets, the fingerprints in FVC2002 DB1 A+B are from Identix TouchView II and not from CrossMatch Verifier 300.

3. The 880 samples from the FVC2004 DB1 A+B dataset. This dataset was created for the Third International Fingerprint Verification Competition (FVC2004) back in 2004: http://bias.csr.unibo.it/fvc2004/databases.asp (accessed on 4 September 2023).

The total number of samples is 2168.

In order to increase the size and diversity of our training dataset, we perform data augmentation. Fingerprint images are horizontally flipped and rotated with eight angles of rotation: $\pm 5^{\circ}$, $\pm 10^{\circ}$, $\pm 15^{\circ}$, and $\pm 20^{\circ}$. Calculating together the original images, flipped images, rotated original, and rotated flipped images, we increase the number of training samples by a factor of 18. After data augmentation, the total number of samples is **39,024**. In the remainder of the paper, we refer to this dataset as "**aug39k**".

Training Hyperparameters. After abundant experimentation with different training hyperparameters, we ended up with a learning rate of 0.002 and 120 epochs, the first 60 epochs with a constant learning rate, and 60 epochs with linear decay of the learning rate. In our very first training runs, we trained models with batch normalization. The best results have been achieved with the batch size of 64, but the resulting fingerprint patterns were quite noisy. Especially the background, which is expected to be blank containing only white pixels, exhibited a lot of noise. For this reason, batch normalization was replaced by instance normalization, as suggested in [57]. The noise does not appear in the background anymore, but, based on a subjective analysis, the realistic appearance of ridge lines has worsened. The reason for the instance normalization outperforming batch normalization might be the insufficient batch size as well as the fact that the inputs of the generator are minutiae maps where pixel statistics do not vary highly in terms of intensity values.

Based on the subjectively best visual performance, we picked one model trained with a batch normalization and a batch size of 64, for which "directed line" minutiae encoding was applied. This model is further referred to as **"aug39k_DL_BN_60+60ep"**.

Based on the experiment from [7], we also picked the model with the best fingerprint reconstruction performance with acceptable visual performance. This model had been

trained with an instance normalization in only 15 epochs, making use of "pointing minutiae" encoding. We further refer to this model as **"aug39k_PM_IN_15ep"**.

3.2. Compilation of Synthetic Datasets

The high-level overview of the dataset compilation process in shown in Figure 1. We generate our synthetic fingerprints using a combination of model-based and data-driven approaches, which allows for more realistic fingerprints with all necessary characteristics. We first utilize a model-based fingerprint generation tool Anguli as a source of minutiae templates. The fact that Anguli fingerprints do not appear realistic to the human eye plays no role because the only information used is the set of minutiae derived from a pattern. Then, we extract minutiae with the Verifinger SDK [50], convert a minutiae template to a minutiae map using either DL or PM encoding, and finally apply one of the two aforementioned generative models to synthesize realistic fingerprints.

The next important step is taking into account the natural distribution of fingerprint patterns among virtual subjects. The intuition behind this is that not all patterns are equally frequent and that some combinations of patterns come together more frequently than other combinations, see, e.g., http://fingerprints.handresearch.com/dermatoglyphics/fingerprints-5-fingers-distributions.htm (accessed on 4 September 2023). It is well known that the most frequent pattern is the ulnar loop (UL), followed by whorl (WH). Arches (AR) and tented arches (TA) are rare patterns that usually appear on all fingers of a subject. Radial loops (RL) are also very rare; they usually appear on index fingers in combination with ulnar loopson all other fingers.

We first define 500 virtual subjects with 8 fingers each. Thumbs are disregarded due to a significantly different shape. Table 1 shows the distribution of the most frequent fingerprint patterns over the fingers. Summing up the patterns in this table, we come up with 44 UL, 4 RL, 16 WH, 8 AR, and 8 TA fingerprints for 10 subjects. Exactly the same distribution of basic patterns is scaled up to 50 subjects, resulting in 4000 fingerprints. The distribution of basic patterns in our synthetic datasets is visualized in Table 2).

Num. of		Left H	and			Right I	Hand	
Subjects	Index	Middle	Ring	Pinky	Index	Middle	Ring	Pinky
50	UL	UL	UL	UL	UL	UL	UL	UL
50	RL	UL	UL	UL	RL	UL	UL	UL
50	WH	UL	UL	UL	WH	UL	UL	UL
50	WH	WH	WH	WH	WH	WH	WH	WH
50	UL	UL	WH	UL	UL	UL	WH	UL
50	AR	UL	UL	UL	AR	UL	UL	UL
50	WH	WH	UL	UL	WH	WH	UL	UL
50	AR	AR	UL	UL	AR	AR	UL	UL
50	AR	TA	TA	TA	AR	TA	TA	TA
50	RL	TA	UL	UL	RL	TA	UL	UL

Table 1. Common distributions of basic fingerprint patterns over fingers. UL—ulnar loop, RL—radialloop, WH—whorl, AR—arch, and TA—tented arch.

Basic Pattern	Ulnar Loop (UL)	Radial Loop (RL)	Whorl (WH)	Arch (AR)	Tented Arch (TA)
Absolute #	$44 \times 50 = 2200$	$4 \times 50 = 200$	$16 \times 50 = 800$	$8 \times 50 = 400$	$8 \times 50 = 400$
Relative #	55%	5%	20%	10%	10%

Table 2. Distribution of basic patterns in our synthetic datasets.

An ulnar loop on a right hand is represented by a left loop fingerprint image just as an ulnar loop on a left hand is represented by a right loop image. This means that considering the equal proportion of ulnar/radial loops on the right and left hands, we need to generate the same number of right and left loop images. In particular, we need 2400 loops in total, split into 1200 right and 1200 left loops.

For the purpose of basic pattern generation, we apply Anguli to generate 1200 right loops, 1200 left loops, 800 whorls, 400 arches, and 400 tented arches. According to Table 1, the fingerprints are distributed between 500 subjects and their eight fingers: right hand index, middle, ring, and pinky fingers; and left hand index, middle, ring, and pinky fingers. This operation is covered by the first two blocks in Figure 1.

Next, we extract minutiae information for all fingerprints using Verifinger SDK [50]. Since captured fingerprints may not be perfectly aligned or positioned, meaning a fingerprint could be shifted from the image center, slightly rotated, or only partially presented during the sensor capture, we simulate the intra-class variations by creating an additional nine impressions for each virtual finger. This operation is called "pseudo-random modification" in Figure 1.

The first fingerprint impression is always the one generated from minutiae extracted from the original Anguli pattern. The other nine impressions are generated from minutiae that undergo random affine transformations or minutiae side cuts. To be more precise, minutiae modifications include:

- Random rotation varying from -20 to +20 degrees with a step of 1 degree
- Random shift varying from -20 to 20 pixels regarding x-axis and y-axis
- Random cut of 0% to 15% of minutiae points at one of eight sides: top-left, top, top-right, left, right, bottom-left, bottom, and bottom-right

Note that the transformations are applied not to minutiae maps but directly to minutiae coordinates and orientations in a minutiae template.

The addition of mated fingerprints expands the dataset of synthetic fingerprints from 4000 samples to **40,000** samples. Note that both minutiae maps for original Anguli patterns as well as minutiae maps for mated impressions are generated very quickly. The same applies for the reconstruction of realistic fingerprints from minutiae maps. This means that the number of fingerprints in a synthetic dataset can be significantly up-scaled in a reasonable amount of time. Our synthetic datasets of 40,000 samples can be seen as an example that demonstrates the validity of the proposed technique for the compilation of large-scale synthetic datasets.

Our first synthetic dataset is created by applying the **"aug39k_DL_BN_60+60ep"** model to the aforementioned set of 40,000 minutiae templates. Minutiae templates are converted to minutiae maps using "directed line" encoding. The dataset is further referred to as **AMSL SynFP P2P v1**.

Our second synthetic dataset is created by applying the "aug39k_PM_IN_15ep" model to the aforementioned set of 40,000 minutiae templates. Minutiae templates are converted to minutiae maps using "pointing minutiae" encoding. The dataset is further referred to as AMSL SynFP P2P v2.

The last processing step is cutting the central region of 300×420 pixels from the generated 512×512 pixel images. In doing so, the irrelevant border regions are dropped, but the resulting images still contain full fingerprints.

The fingerprints in both datasets are anonymous by design. The design of our generative models ensures the realistic appearance of fingerprint patterns, or to be more precise their resemblance with real plain fingerprints captured by a CrossMatch Verifier 300 sensor. The diversity of inter-class samples is ensured by the design of the Anguli tool, and the sufficient variability and non-excessive diversity of intra-class samples will be confirmed in Section 4.2.

4. Experimental Studies

Our experimental studies have two objectives: (i) the assessment of the pix2pix models, including the effectiveness of the minutiae encoding approaches (directed lines vs. pointing minutiae), which is mostly adapted from our previous report [7]; and (ii) the analysis and discussion of the utility of our synthetic datasets, which is one of the main contributions of this paper.

4.1. Evaluation of Generative Models

In our study, the evaluation of generative models is carried out not by analyzing the design and the backbone network architecture establishing theoretical abilities of the models but rather empirically by analyzing the outcomes of the generative models.

4.1.1. Evaluation Metrics

There are two aspects that need to be confirmed in the evaluation of the synthesized samples: their realism and their biometric resemblance to the minutiae origins.

In our considerations, **the realism of synthetic fingerprints** is equal to their visual quality and is estimated by calculating the NFIQ2 scores [49]. Note that the primary objective of NFIQ2 scores is to estimate the utility of fingerprints or, in other words, their effectiveness for the purpose of user authentication. However, it is known that NFIQ2 scores correlate well with the visual quality of fingerprints. The NFIQ2 scores range from 0 to 100, with higher scores indicating higher utility. Fingerprints with scores above 45 are considered perfect, while scores above 35 indicate good fingerprints. Fingerprints with scores lower than 6 are deemed useless due to poor quality.

The biometric resemblance with the minutiae origin is also referred to as reconstruction success. In order to measure reconstruction success, we apply VeriFinger SDK [50] to calculate matching scores between reconstructed synthetic samples and the origins of minutiae. Given a pair of fingerprints, the Verifinger's outcome is a similarity score ranging from 0 to ∞ , with 0 for maximal dissimilarity. The decision threshold for confirming that fingerprints belong to the same identity is set based on the desired security level of a biometric system. The threshold is set in a way that the false acceptance rate (FAR) does not exceed 0.1%, 0.01%, or 0.001%, with lower values indicating higher security. The Verifinger thresholds at the aforementioned FAR levels are 36, 48, and 60, respectively.

For each generative model under evaluation, we calculate the fingerprint reconstruction success rate as the ratio of fingerprint pairs (synthesized vs. origin) for which matching scores are higher than a pre-defined threshold in all tested fingerprint pairs. For the purpose of comparability with other studies, this ratio is also referred to as the true acceptance rate (TAR). In state-of-the-art studies, two types of TAR are estimated: Type 1 and Type 2. Type 1 TAR involves calculating matching scores between reconstructed and original fingerprints, while Type 2 TAR involves calculating matching scores between the reconstructed fingerprint and other impressions of the original fingerprint.

4.1.2. Evaluation Protocol

We use two datasets containing 880 samples each. These datasets are completely detached from datasets used for the training of our generative models.

The first dataset has been created using Anguli [16]. Due to the idealistic ridge line patterns in this dataset, we expect no minutiae localization errors with any minutiae extractor. Hence, the idealistic fingerprint reconstruction performance can be estimated.

The second dataset is the FVC2004 DB2 A+B dataset collected for the Third International Fingerprint Verification Competition, which comprises real fingerprints collected with an optical scanner **URU 4500**. URU fingerprints are very diverse, including many noisy or even partially corrupted patterns. Such fingerprints are very challenging for any minutiae extractor, implying many minutiae localization errors. Moreover, the fingerprints captured with URU and CrossMatch scanners are very different from each other. This fact might lead to additional fingerprint reconstruction errors. All in all, with this dataset we estimate a pessimistic fingerprint reconstruction performance. Samples from Anguli and URU datesets are shown in Figure 6.



Figure 6. (**Top row**): fingerprints generated by Anguli; (**bottom row**): URU fingerprints from the FVC2004 DB2 A+B dataset.

Both datasets are used as sources of minutiae templates from which the generative model has to reconstruct fingerprint images. The NFIQ2 scores of these datasets are seen as a reference for the visual quality of fingerprints. The reconstruction success is estimated by matching fingerprints from these datasets with their reconstructed counterparts.

4.1.3. Evaluation Results

The evaluation result of the **aug39k_DL_BN_60+60ep** and **aug39k_PM_IN_15ep** models are mostly derived from our previous paper [7], whereby the **aug39k_DL_BN_60+60ep** model is not directly addressed in the paper but has been excessively evaluated in experiments conducted to obtain the optimal training hyperparameters. In order to compare "directed line" (DL) and "pointing minutiae" (PM) encoding approaches, we report the evaluation results for three model snapshots (15, 30, and 55 training epochs) with **aug39k_PM_IN** and **aug39k_PM_DL** generative models.

Realism of Synthetic Fingerprints. First, we discuss the NFIQ2 scores of the test datasets, which are shown in Figure 7. The NFIQ2 scores of original Anguli (left column) and URU fingerprints (right column) are taken as a reference. Our general observation is that the visual quality of the reconstructed fingerprints mostly depends on training samples. Since we use high-fidelity CrossMatch fingerprints for training, the mean NFIQ2 values in all cases are higher than 40. The quality of samples from which minutiae were extracted has an indirect influence on the visual quality of the reconstructed fingerprints, meaning that the missed or falsely localized minutiae may worsen the realism of a fingerprint pattern. We can clearly see that the distribution of NFIQ2 scores of Anguli fingerprints is narrower than those of reconstructed fingerprints, but the mean values of all distributions are very close, indicating the similar visual quality of fingerprints. In contrast, the distribution of NFIQ2 scores of URU fingerprints is broader and shifted to the left, indicating that the visual quality of URU fingerprints is significantly lower than that of reconstructed

fingerprints. These observations apply to all generative models independently of the snapshot selection. Based on diagrams, no conclusion can be made about the superiority of the "directed line" or "pointing minutiae" encoding approach. However, depending on the encoding approach, different snapshots should be preferred if the goal is to obtain the lowest number of patterns with low NFIQ2 values. This fact is evident if reconstructing from URU fingerprints. In the case of "directed line" encoding, the model snapshot with 55 training epochs, and in the case of "pointing minutiae" encoding, the model snapshot with 15 training epochs, has no secondary peak in the low range of NFIQ2 values. We thus intuit that models with "directed line" encoding may benefit from further training.



Figure 7. NFIQ2 score distributions for Anguli (**left column**) and URU (**right column**) fingerprints and their reconstructed counterparts with the **aug39k_DL_IN** (**top row**) and **aug39k_PM_IN** (**bottom row**) models, i.e., their snapshots at 15, 30, and 55 training epochs. DL—"directed line" encoding, PM—"pointing minutiae" encoding.

Reconstruction Success. Second, we report the success of fingerprint reconstruction. For that, the original fingerprints from the test sets are compared with their reconstructed counterparts by means of the Verifinger matching algorithm. Referring to the metrics introduced in Section 4.1.1, TAR Type 1 is used. The reconstruction rates are reported in Table 3 for both the idealistic Anguli fingerprints (an expected upper bound of reconstruction rates) and very challenging URU fingerprints taken from the FVC2004 DB2 A+B dataset (the estimation of pessimistic reconstruction rates). An important point to note is that URU samples can be quite challenging for minutiae extractors. Therefore, we have dropped the samples where Verifinger fails to extract even a single minutiae point.

We can also observe that NFIQ2 score distributions of fingerprints reconstructed from URU samples have tails at the lower end. We suspect that this is due to the low quality of some URU samples, where the minutiae extractors may not be able to extract minutiae properly, resulting in incomplete or inaccurate patterns.

The reason for the low reconstruction rates with the **aug39k_DL_BN_60+60** model and the URU fingerprints is the evaluation run in which zero-minutiae samples have not been dropped, causing a high portion of failed matching trials.

	Model Snapshot				Anguli Fingerprints			URU Fingerprints		
				Type1 TAR @ FAR of						
TrainDB	Enc	Norm	Epochs	0.1%	0.01%	0.001%	0.1%	0.01%	0.001%	
aug39k	DL	IN	15 30 55	100.00 100.00 99.43	100.00 99.77 98.52	99.77 99.20 97.50	87.84 83.29 79.09	82.61 75.11 71.81	76.47 65.90 59.31	
aug39k	PM	IN	15 30 55	100.00 100.00 99.88	100.00 100.00 99.43	100.00 100.00 98.86	95.45 95.11 93.52	95.00 94.31 91.36	93.52 93.29 88.86	
aug39k	DL	BN	60+60	97.95	94.09	85.34	60.34	44.66	26.48	

Table 3. Fingerprint reconstruction success (in %); DL—directed line encoding, PM—pointing minutiae encoding; IN—instance normalization, and BN—batch normalization.

Based on the reconstruction rates, we arrive at the following statements, which apply to both test datasets: Anguli and FVC2004 DB2 A+B:

- The "pointing minutiae" encoding outperforms the "directed line" encoding;
- The best reconstruction performance is achieved with the model snapshots at 15 training epochs;
- For the "pointing minutiae" encoding, there is almost no difference between snapshots at 15 epochs and 30 epochs, while there is a notable performance loss with the snapshot at 55 epochs;
- For "directed line" encoding, the additional training epochs worsen the reconstruction performance, i.e., the snapshot at 15 epochs is better than at 30 epochs, which is better than at 55 epochs.

4.2. Utility Evaluation of the Proposed Synthetic Datasets

We have chosen the **aug39k_DL_BN_60+60ep** model for the compilation of the first synthetic dataset **AMSL SynFP P2P v1** as one that subjectively generates the most realistic patterns that are very close to original fingerprints (sources of minutiae). The downside of this model is that the border regions of the generated fingerprint images that are expected to be uniform or even flat (filled with white pixels only) include random noise. However, since we cut the central part of the fingerprint image, this aspect does not influence the fingerprint's appearance in general. The second characteristic property of this model is the lower reconstruction performance in comparison to the best performing models. Although, at first glance, it sounds like a crucial disadvantage, we leverage this fact to increase the diversity of mated fingerprints. Note that, from the subjective perspective, the majority of reconstructed fingerprints appear to belong to the same identity as the source of minutiae. This dataset should be quite challenging for state-of-the-art fingerprint recognition algorithms.

As an alternative to the aforementioned model, we have chosen the best performing model **aug39k_PM_IN_15ep** for the compilation of the second synthetic fingerprint dataset **AMSL SynFP P2P v2**. The choice is explained by its superiority in terms of fingerprints. As we can see in the bottom row of Figure 7, the model snapshot after 15 training epochs leads to NFIQ2 scores that are, on average, lower than those of the model snapshots after 30 and 55 epochs but generates almost no fingerprints with NFIQ2 scores lower than 30. Hence, our second dataset can be seen as a source of high-quality fingerprints with which state-of-the-art fingerprint recognition algorithms should have close to zero verification error rates.

4.2.1. Realistic Appearance

The general impression of our synthetic fingerprints can be taken from Section 4.1.3 and Figure 8. Here, we compare the distributions of the NFIQ2 scores for the compiled datasets. Figure 9 depicts the probability density functions of 40,000 samples included



in each of the **AMSL SynFP P2P v1** and **AMSL SynFP P2P v2** datasets. We also plot the distribution of the original 4000 Anguli fingerprints as a reference.

Figure 8. Results of fingerprint reconstruction. Minutiae are taken from Anguli (top) and URU (bottom) fingerprints. Note that for Anguli, the reconstruction examples are with **aug39k_DL_BN_60+60ep** and for URU with **aug39k_DL_IN_15ep**.



Figure 9. Distributions of NFIQ2 scores for our synthetic datasets AMSL SynFP P2P v1 and AMSL SynFP P2P v2 in comparison to the 4000 Anguli generated fingerprints taken as a source of minutiae.

In comparison to the NFIQ2 scores of Anguli fingerprints, which have a narrow symmetric distribution around the mean of approx. 65, the distributions of synthetic datasets are negatively skewed and wider spread, having significantly higher standard deviations. The mean value of the AMSL SynFP P2P v1 is approximately 60, which is lower than that of Anguli, and the mean value of the AMSL SynFP P2P v2 is approximately 75, which is higher than that of Anguli. The NFIQ2 scores validate that the AMSL SynFP P2P v1 dataset includes a small portion of fingerprints that might be challenging for minutiae extractors, leading to verification errors. In contrast, almost all of the fingerprints in the AMSL SynFP P2P v1 dataset are of high utility, which is useful for checking that an advanced fingerprint matching algorithm produces almost no verification errors with this dataset.

4.2.2. Estimation of the Verification Performance

Finally, we take a look at verification scores produced by the Verifinger SDK with our compiled synthetic datasets. Our evaluation protocol is defined as follows. Bearing in mind that we have in total 500 persons with 8 fingers each and 10 impressions per finger, we build the set of mated fingerprint pairs by pairing all impressions of a single finger, resulting in $10 \times 9/2 = 45$ pairs. In total, we have $4000 \times 45 = 180,000$ mated fingerprint pairs spread between different basic patterns, as shown in Table 4. The sets of non-mated fingerprint pairs are built by combining only the first impressions of each finger.

Basic Pattern	Ulnar Loop (UL)	Radial Loop (RL)	Whorl (WH)	Arch (AR)	Tented Arch (TA)
Relative number	55%	5%	20%	10%	10%
Impostor trials	99,000	9000	36,000	18,000	18,000
Genuine trials	3996	3996	3995	3990	3990

Table 4. Distribution of basic patterns our synthetic datasets.

Although we disregard pairs that contain fingerprints with different basic patterns and, therefore, significantly decrease the total number of possible non-mated pairs, the number of pairs is still enormous. Hence, we decided to limit the number of non-mated pairs by randomly selecting approximately 4000 pairs for each basic pattern. For exact numbers, see Table 4.

By calculating the Verifinger scores for mated and non-mated pairs, we obtain genuine and impostor score distributions, respectively. Figure 10 demonstrates the violin plots of score distributions for different fingerprint basic patterns taken separately. The genuine and impostor distributions are clearly separated for the **AMSL SynFP P2P v2** dataset (right plot). The majority of genuine scores range from 100 to 300, indicating the high number of very confident matches, while the majority of impostor scores are below 80 with a significantly smaller spread. For the case of the **AMSL SynFP P2P v1** dataset (left plot), there is are huge overlap between genuine and impostor distributions in the range from 25 to 60, indicating the possibility for many verification errors at all standard decision thresholds: 36, 48, and 60. The spread of impostor scores is higher in comparison to that in the **AMSL SynFP P2P v2** dataset, while the spread of genuine scores is lower, indicating the low number of very confident matches.



Figure 10. Distributions of Verifinger matching scores with the **AMSL SynFP P2P v1** (left) and **AMSL SynFP P2P v2** (right) datasets for different fingerprint basic patterns considered separately.

5. Results and Discussions

Our experiments have demonstrated that a pix2pix network in conjunction with "pointing minutiae" or "directed line" encoding is a viable solution to reconstruct fingerprints from minutiae. The network has a scalable architecture, allowing for training with 512×512 pixel images. The ridge patterns are realistic, which can be seen in Figure 8 for Anguli and URU fingerprints, respectively. The images demonstrate that our models perform a style transfer, meaning that the resulting fingerprints bear a resemblance to those obtained using a CrossMatch Verifier 300 sensor. The training of pix2pix models with fingerprint data from other sensors would allow the style transfer to other sensor domains. Note that our pix2pix models have been trained on 2168 original fingerprint images only. The only limitation of our fingerprints. Although the ridge patterns in reconstructed fingerprints may not precisely match those in target fingerprints, the models reproduce the minutiae co-allocation accurately enough to facilitate matching with the source of minutiae.

Our experiments confirm that the AMSL SynFP P2P v1 dataset is a challenging dataset for the Verifinger SDK. This dataset can be utilized for benchmarking matching performances of different fingerprint matchers. The AMSL SynFP P2P v2 dataset can be seen as a means for confirming the power of a fingerprint matcher. For instance, Verifinger SDK v12.0 makes almost no verification errors with this dataset. Note that the higher diversity of Verfinger scores for mated fingerprints in the AMSL SynFP P2P v1 dataset can be caused by an inaccurate fingerprint reconstruction, implying that the transformations applied to minutiae templates may lead to a vanishing identity in a reconstructed fingerprint. Hence, a systematic analysis of the influence of a particular minutiae template transformation to the reconstruction process is required. However, this analysis is beyond the scope of this paper and will be conducted in our future studies. Note that the numbers of virtual subjects (currently 500) and fingerprint impressions (currently 10) are taken as an example to demonstrate the feasibility of the approach and can be easily up-scaled. The manipulation of minutiae and the consequent reconstruction of fingerprints from minutiae for fingerprint identity control and the production of mated impressions is a completely novel technique that seems to be equally or even more effective than affine transformations in an image domain or the thin-plate spline modeling. Comparing our synthetic datasets to the few existing synthetic datasets, we claim that PrintsGAN [17] is the only dataset with a mechanism for controlling fingerprint identity that surpasses our datasets in terms of the number of fingerprints.

6. Conclusions

Bearing in mind the objective of getting around privacy concerns arising when evaluating biometric systems with real biometric samples, the study reported in this paper aims to compile large-scale synthetic fingerprint datasets that are suitable for assessing the performance of fingerprint matching algorithms. It is demonstrated how model-based and data-driven approaches can be combined to generate realistic, anonymous, and sufficiently diverse fingerprints in a fully controlled environment, enabling the synthesis of not only non-mated but also mated fingerprints. The applied technique is a reconstruction of realistic fingerprints from minutiae extracted from pseudo-randomly synthesized non-realistic fingerprints. Mated fingerprints are generated by reconstruction from slightly modified minutiae templates. In particular, we train pix2pix generative models with high-quality fingerprint images at a fingerprint-native resolution from public datasets. The generative models reconstruct fingerprints from so-called minutiae maps-images in which minutiae are encoded either by *directed lines* or by *pointing minutiae*. The pix2pix architecture is extended to handle 512×512 pixels images. The evaluation results indicate that the application of the pix2pix network to the reconstruction problem is a viable solution and validate the utility of the compiled datasets for evaluating fingerprint matching algorithms. Future work will be devoted to experimenting with a broad variety of alternative fingerprint matching systems relying not only on minutiae but also on DCNN-based techniques and to analyzing the influence of a particular minutiae template transformation on the fingerprint reconstruction process.

Author Contributions: Conceptualization, A.M. and J.D.; methodology, A.M.; software, A.M. and V.S.M.; validation, A.M. and V.S.M.; formal analysis, A.M.; investigation, A.M. and V.S.M.; resources, J.D.; data curation, A.M.; writing—original draft preparation, A.M. and V.S.M.; writing—review and editing, A.M., V.S.M. and J.D.; visualization, A.M. and V.S.M.; supervision, J.D.; project administration, J.D.; and funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded in part by the Deutsche Forschungsgemeinschaft (DFG) through the research project GENSYNTH under the number 421860227.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets applied for training of our generative models are available at http://bias.csr.unibo.it/fvc2002/databases.asp (accessed on 4 September 2023), http://bias.csr.unibo.it/fvc2004/databases.asp (accessed on 4 September 2023), and https://www.neurotechnology.com/download.html (accessed on 4 September 2023). The links to our synthetic datasets and generative models are available at https://gitti.cs.uni-magdeburg.de/Andrey/gensynth-pix2pix (accessed on 4 September 2023). The download password will be shared after signing the corresponding license agreements at https://omen.cs.uni-magdeburg.de/disclaimer_gensynth_models/index.php (accessed on 4 September 2023).

Acknowledgments: We would like to thank Christian Kraetzer and Stefan Seidlitz for their advice and technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Carlini, N.; Chien, S.; Nasr, M.; Song, S.; Terzis, A.; Tramer, F. Membership Inference Attacks From First Principles. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–26 May 2022.
- Seidlitz, S.; Jürgens, K.; Makrushin, A.; Kraetzer, C.; Dittmann, J. Generation of Privacy-friendly Datasets of Latent Fingerprint Images using generative adversarial networks. In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP'21), Virtual, 8–10 February 2021; VISAPP; Farinella, G.M., Radeva, P., Braz, J., Bouatouch, K., Eds.; 2021; Volume 4, pp. 345–352. [CrossRef]
- Bahmani, K.; Plesh, R.; Johnson, P.; Schuckers, S.; Swyka, T. High Fidelity Fingerprint Generation: Quality, Uniqueness, And Privacy. In Proceedings of the IEEE International Conference on Image Processing (ICIP'21), Anchorage, AK, USA, 19–22 September 2021; pp. 3018–3022. [CrossRef]
- Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In Proceedings of the 6th International Conference on Learning Representations (ICLR'18), Vancouver, BC, Canada, 30 April–3 May 2018. [CrossRef]
- Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 43, 4217–4228. [CrossRef]
- Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. *arXiv* 2019, arXiv:1912.04958.
- Makrushin, A.; Mannam, V.S.; Dittmann, J. Data-Driven Fingerprint Reconstruction from Minutiae Based on Real and Synthetic Training Data. In Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications—Volume 4: VISAPP, SCITEPRESS, Lisbon, Portugal, 19–21 February 2023; pp. 229–237.
- Makrushin, A.; Kauba, C.; Kirchgasser, S.; Seidlitz, S.; Kraetzer, C.; Uhl, A.; Dittmann, J. General Requirements on Synthetic Fingerprint Images for Biometric Authentication and Forensic Investigations. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'21), Virtual, 22–25 June 2021; pp. 93–104.
- 9. Kücken, M. Models for fingerprint pattern formation. Forensic Sci. Int. 2007, 171, 85–96. [CrossRef] [PubMed]
- 10. Ram, S.; Bischof, H.; Birchbauer, J. Modelling fingerprint ridge orientation using Legendre polynomials. *Pattern Recognit.* **2010**, 43, 342–357. [CrossRef]
- 11. Zinoun, F. Can a Fingerprint be Modelled by a Differential Equation? arXiv 2018, arXiv:1802.05671.
- Zhao, Q.; Jain, A.K.; Paulter, N.G.; Taylor, M. Fingerprint image synthesis based on statistical feature models. In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; pp. 23–30. [CrossRef]
- 13. Cappelli, R.; Maio, D.; Maltoni, D. Synthetic fingerprint-database generation. In Proceedings of the 2002 International Conference on Pattern Recognition (ICPR), Quebec, QC, Canada, 11–15 August 2002; Volume 3, pp. 744–747. [CrossRef]
- Johnson, P.; Hua, F.; Schuckers, S. Texture Modeling for Synthetic Fingerprint Generation. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 154–159. [CrossRef]
- 15. Cappelli, R. SFinGe. In *Encyclopedia of Biometrics;* Li, S.Z., Jain, A., Eds.; Springer: Boston, MA, USA, 2009; pp. 1169–1176. [CrossRef]
- 16. Ansari, A.H. Generation and Storage of Large Synthetic Fingerprint Database. Master's Thesis, Indian Institute of Science Bangalore, Karnataka, Indian, 2011.
- 17. Engelsma, J.J.; Grosz, S.; Jain, A.K. PrintsGAN: Synthetic Fingerprint Generator. *IEEE Trans. Pattern Anal. Mach. Intell.* 2023, 45, 6111–6124. [CrossRef]
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; Curran Associates, Inc.: New York, NY, USA, 2014; Volume 27.
- 19. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 2642–2651.
- 20. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* 2015, arXiv:1511.06434.
- Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; Aila, T. Alias-Free generative adversarial networks. In Proceedings of the NIPS, Online, 6–14 December 2021.
- 22. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* 2014, arXiv:1411.1784. https://doi.org/10.48550/ARXIV. 1411.1784.
- Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
- Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
- 25. Bouzaglo, R.; Keller, Y. Synthesis and Reconstruction of Fingerprints using Generative Adversarial Networks. *arXiv* 2022, arXiv:2201.06164.

- Minaee, S.; Abdolrashidi, A. Finger-GAN: Generating Realistic Fingerprint Images Using Connectivity Imposed GAN. *arXiv* 2018, arXiv:1812.10482.
- Cappelli, R.; Ferrara, M.; Franco, A.; Maltoni, D. Fingerprint Verification Competition 2006. *Biom. Technol. Today* 2007, 15, 7–9. [CrossRef]
- Zhao, Q.; Zhang, D.; Zhang, L.; Luo, N. High resolution partial fingerprint alignment using pore-valley descriptors. *Pattern Recognit.* 2010, 43, 1050–1061. [CrossRef]
- Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA,USA, 4–9 December 2017; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.
- 30. Fahim, M.A.N.I.; Jung, H.Y. A Lightweight GAN Network for Large Scale Fingerprint Generation. *IEEE Access* 2020, *8*, 92918–92928. [CrossRef]
- Cao, K.; Jain, A. Fingerprint Synthesis: Evaluating Fingerprint Search at Scale. In Proceedings of the 2018 International Conference on Biometrics (ICB), Gold Coast, Australia, 20–23 February 2018; pp. 31–38. [CrossRef]
- Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training of Wasserstein GANs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
- 33. Mistry, V.; Engelsma, J.J.; Jain, A.K. Fingerprint Synthesis: Search with 100 Million Prints. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–10. [CrossRef]
- 34. Wyzykowski, A.B.V.; Segundo, M.P.; de Paula Lemes, R. Level Three Synthetic Fingerprint Generation. *arXiv* 2020, arXiv:2002.03809.
- 35. Hill, C.J. Risk of Masquerade Arising from the Storage of Biometrics. Bachelor's Thesis, Australian National University, Canberra, Australia, 2001.
- Cappelli, R.; Maio, D.; Lumini, A.; Maltoni, D. Fingerprint Image Reconstruction from Standard Templates. *IEEE Trans. Pattern* Anal. Mach. Intell. 2007, 29, 1489–1503. [CrossRef]
- Ross, A.; Shah, J.; Jain, A.K. From Template to Image: Reconstructing Fingerprints from Minutiae Points. *IEEE Trans. Pattern Anal. Mach. Intell.* 2007, 29, 544–560. [CrossRef]
- Vizcaya, P.R.; Gerhardt, L.A. A nonlinear orientation model for global description of fingerprints. *Pattern Recognit.* 1996, 29, 1221–1231. [CrossRef]
- Feng, J.; Jain, A.K. Fingerprint Reconstruction: From Minutiae to Phase. *IEEE Trans. Pattern Anal. Mach. Intell.* 2011, 33, 209–223. [CrossRef]
- 40. Cao, K.; Jain, A.K. Learning Fingerprint Reconstruction: From Minutiae to Image. *IEEE Trans. Inf. Forensics Secur.* 2015, 10, 104–117. [CrossRef]
- 41. Watson, C.I. NIST Special Database 14. NIST Mated Fingerprint Card Pairs 2 (MFCP2); NIST: Gaithersburg, MD, USA, 2008.
- Wijewardena, K.P.; Grosz, S.A.; Cao, K.; Jain, A.K. Fingerprint Template Invertibility: Minutiae vs. Deep Templates. arXiv 2022, arXiv:2205.03809. https://doi.org/10.48550/ARXIV.2205.03809.
- Kim, H.; Cui, X.; Kim, M.G.; Nguyen, T.H.B. Reconstruction of Fingerprints from Minutiae Using Conditional Adversarial Networks. In Proceedings of the IWDW'18, Chengdu, China, 2–4 November 2019; pp. 353–362.
- 44. Makrushin, A.; Mannam, V.S.; Rao, B.N.M.; Dittmann, J. Data-driven Reconstruction of Fingerprints from Minutiae Maps. In Proceedings of the IEEE 24th Int. Workshop on Multimedia Signal Processing (MMSP'22), Shanghai, China, 26–28 September 2022.
- Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
- FVC2000. The First Fingerprint Verification Competition. Available online: http://bias.csr.unibo.it/fvc2000/databases.asp (accessed on 4 September 2023).
- FVC2002. The Second Fingerprint Verification Competition. Available online: http://bias.csr.unibo.it/fvc2002/databases.asp (accessed on 4 September 2023).
- FVC2004. The Third Fingerprint Verification Competition. Available online: http://bias.csr.unibo.it/fvc2004/databases.asp (accessed on 4 September 2023).
- NIST. NIST Fingerprint Image Quality (NFIQ) 2. Available online: https://www.nist.gov/services-resources/software/nfiq-2 (accessed on 4 September 2023).
- 50. Neurotechnology. VeriFinger SDK. Available online: https://www.neurotechnology.com/verifinger.html (accessed on 4 September 2023).
- 51. Neurotechnology. Neurotechnology CrossMatch dataset. Available online: https://www.neurotechnology.com/download.html (accessed on 4 September 2023).
- 52. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* 2015, arXiv:1505.04597.
- 53. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. arXiv 2014, arXiv:1411.4038.
- 54. Zhou, Y.; Berg, T.L. Learning Temporal Transformations From Time-Lapse Videos. arXiv 2016, arXiv:1608.07724.
- 55. Wang, X.; Gupta, A. Generative Image Modeling using Style and Structure Adversarial Networks. *arXiv* **2016**, arXiv:1603.05631.

- 56. Neurotechnology. Cross Match Verifier 300 Classic. Available online: https://www.neurotechnology.com/fingerprint-scannercross-match-verifier-300-classic.html (accessed on 4 September 2023).
- 57. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv* 2016, arXiv:1607.08022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.