Analysis of Master Vein Attacks on Finger Vein Recognition Systems

Huy H. Nguyen[1], Trung-Nghia Le[1,3,4], Junichi Yamagishi[1], and Isao Echizen[1,2]
[1]National Institute of Informatics, Tokyo, Japan [2]The University of Tokyo, Tokyo, Japan
[3]University of Science, VNU-HCM, Vietnam [4]Vietnam National University, Ho Chi Minh City, Vietnam
{nhhuy,jyamagis,iechizen}@nii.ac.jp

Abstract

Finger vein recognition (FVR) systems have been commercially used, especially in ATMs, for customer verification. Thus, it is essential to measure their robustness against various attack methods, especially when a hand-crafted FVR system is used without any countermeasure methods. In this paper, we are the first in the literature to introduce master vein attacks in which we craft a vein-looking image so that it can falsely match with as many identities as possible by the FVR systems. We present two methods for generating master veins for use in attacking these systems. The first uses an adaptation of the latent variable evolution algorithm with a proposed generative model (a multi-stage combination of β-VAE and WGAN-GP models). The second uses an adversarial machine learning attack method to attack a strong surrogate CNN-based recognition system. The two methods can be easily combined to boost their attack ability. Experimental results demonstrated that the proposed methods alone and together achieved false acceptance rates up to 73.29% and 88.79%, respectively, against Miura's hand-crafted FVR system. We also point out that Miura's system is easily compromised by non-vein-looking samples generated by a WGAN-GP model with false acceptance rates up to 94.21%. The results raise the alarm about the robustness of such systems and suggest that master vein attacks should be considered an important security measure.

1. Introduction

Finger vein authentication (using a FVR system [24]) was first commercially implemented in Japan in 1997 and has become well-recognized because of its application in ATMs to authenticate users [28]. Its usage frees users from remembering and regularly changing passwords to maintain security. Due to their convenience, biometric authentication methods (including finger vein ones) have become widely used. Therefore, it is essential to evaluate their robustness and identify potential harms. A presentation attack is a common way to attack biometric recognition systems [17]. Besides presenting a captured biometric trait of the victim, the attacker can use a wolf sample [27], which can match enrolled models of multiple identities. Master prints [2] and master faces [21, 25, 20] are examples of wolf samples generated using generative models. In reality, not all FVR systems have countermeasure methods deployed properly, allowing master vein attacks to compromise them.

Besides presentation attacks, there are several other ways that an attacker can compromise a biometric recognition system [23], as shown in Fig. 1. Moreover, in theory, it is possible to craft a physical object (known as a presentation attack instrument, or PAI) from a synthetic master vein (an image clearly showing the center lines of the veins) and use it to perform a *presentation attack* (attack 1 in Fig. 1) [26]. It is also possible to translate the synthetic master vein into a captured vein sample using a convolutional neural network (CNN) [22] and use it to perform a *logical attack* (attack 2 in Fig. 1). Due to these reasons, we focus on a logical attack with a clear vein image (attack 4 in Fig. 1) in this work. We craft a **master vein** image and then inject it as a probe to attack FVR systems built based on Miura *et al.*'s design [18, 19]. *A master vein image is a probe vein-looking image that can be falsely accepted as a match with enrolled models of multiple identities by a FVR system.* Although non-vein-looking images may have better attack ability, it is harder for them to generalize on other systems and other attack scenarios. There are two possible solutions to craft such master veins: using the latent variable evolution (LVE) algorithm [2] and using an adversarial machine learning (AdvML) attack [11].

The LVE algorithm is a common way to generate master biometric samples [2, 21, 25, 20]: a pre-trained generative model is used along with an evolutionary algorithm. The original work on generating master prints used a traditional generative adversarial network (GAN) model [5] while the work on generating master faces [21, 25, 20] used an advanced GAN model trained on a large facial database [12]. With the original variational autoencoder (VAE) [13] and β-VAE [8, 3] generative models, there is a trade-off between

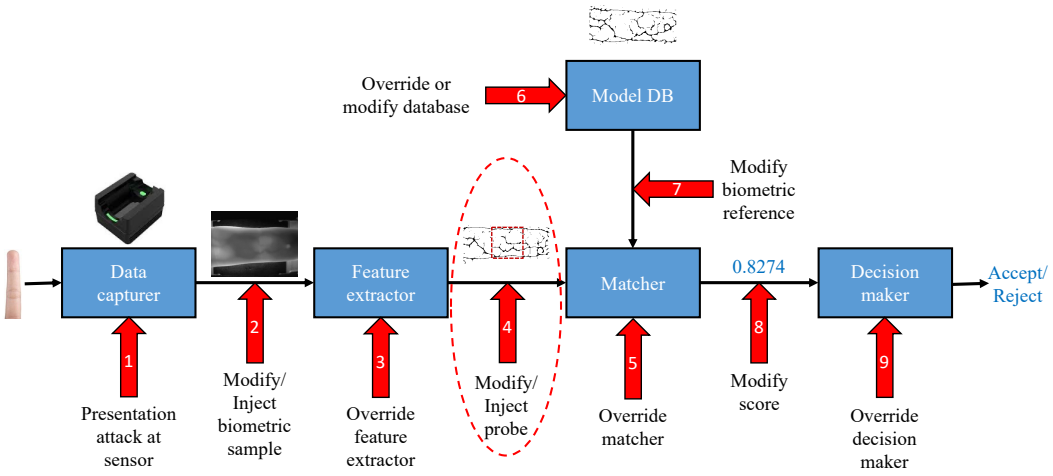


Figure 1. Overview of FVR system and possible attacks on it, inspired by Ratha *et al.* [23]. This paper focuses on attack 4 by injecting a master vein probe image.

image quality and the ability to learn disentangled representations. The traditional VAE and GAN [1, 5] models have trouble generating large images (320×240 pixels in our case), while advanced models are data-hungry. The ability of the LVE algorithm to achieve good results depends on the disentanglement ability of the generative model. Moreover, compatibility with modern vein capturing devices depends on the model's ability to generate high-resolution vein images.

An AdvML attack using an adversarial example can be used to change the output of a CNN [11]. It is assumed that an attacker using a master vein attack does not know the identities of the enrolled models. The attacker thus attempts to generate a master vein that can match as many enrolled models as possible. State-of-the-art CNN-based FVR systems use different approaches between training and testing. For example, for a system [14] that uses the additive angular margin loss [4] in training, the task is to minimize the cross-entropy loss or focal loss [15] using the provided labels. In evaluation, cosine similarity is used to calculate the distance between the two embedded features of the probe and the model veins. Therefore, traditional adversarial attacks could not be applied in this case. Moreover, adversarial attacks are exclusive to machine-learning-based recognition systems. They are unlikely to generalize well to handcrafted recognition systems.

This work aims to solve to two above problems and then combine the two newly proposed solutions to generate master vein images that can attack both hand-crafted and deep-learning-based vein recognition systems. For the generative model used by the LVE algorithm, we proposed a method to combine the β-VAE model and the Wasserstein GAN with a gradient penalty (WGAN-GP) [5] model. The combination model can effectively learn disentanglement latent representations essential for the LVE algorithm and is capable of generating images with higher quality than the single models. Using this setting, we can successfully attack a hand-crafted system with about 70% of false acceptance rates (FARs). However, this LVE-based method could not work on the CNN-based FVR systems, leading to the development of the adversarial-attack-based one. Unlike traditional adversarial attack methods, we propose using k labels as targets. Since the target system uses cosine distance between the two embedded features in inference mode, we attack its training configuration, which uses an advanced addictive angular margin loss function. To make the attack more general, we combine these two proposed methods. By performing an adversarial attack on the master vein generated by the LVE-based method, the crafted master vein can fool both hand-crafted and CNN-based recognition systems with higher FARs (up to 88.79% for the hand-crafted system) than those of master veins created by single methods.

In summary, the contributions of this work are four-fold:

- We point out that a hand-crafted vein recognition system without any countermeasure methods can be easily compromised by **non-vein-looking** images generated by a WGAN-GP model. We are also the first in the literature to investigate synthesized **vein-looking** images to perform master vein attacks.
- We introduce a way to combine a β-VAE model and a WGAN-GP model to generate large, good-quality **vein-looking** images with better disentanglement. The trained β-VAE decoder extracted from this combination is then used in the LVE algorithm.
- We present a k-label targeted adversarial attack for use in attacking a CNN-based FVR system. This target CNN-based system was trained using an advanced

loss function (additive angular margin), outperforming a hand-crafted system.

- We describe a highly successful attack that combines a latent LVE-based attack with an adversarial attack on a hand-crafted FVR system. We show that robustness against master vein attacks is an important measure for FVR systems.

2. Related Work

2.1. Finger Vein Recognition Systems

A typical vein recognition system usually has four modules (visualized in Fig. 1): a data capturer, a feature extractor, a matcher, and a decision maker [23]. Pre-processing operations may be applied before feature extraction. In the original work of Miura *et al.* [18, 19], the maximum curvature method and the repeated line tracking method were used for the feature extractor module and the cross-correlation method was used for the matcher module. The maximum curvature method was designed to be robust against varying vein widths and non-uniform brightness. We used it in a baseline handcrafted FVR system, which we refer to as "Miura's system."

Besides fully handcrafted systems as introduced above, machine learning methods have been used to build the feature extractor and/or the matcher [24]. For instance, Kuzu *et al.* [14] used a modified version of the DenseNet-161 model [10] to build a feature extractor and used cosine distance as the metric for matching. The modified DenseNet-161 model was trained using the additive angular margin loss [4] on the provided ground-truth labels. When testing, it was used to calculate the embedded features of the probe and the model finger vein. An overview of this kind of system is shown in Fig. 2. To build a conceptual attackable CNN-based model for evaluation, we used this kind of CNN as an additional feature extractor to take the output of

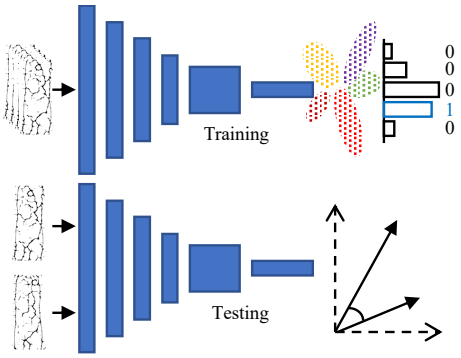


Figure 2. Illustration of training and testing phases of CNN-based FVR system. The training phase uses the additive angular margin loss with ground-truth labels while the testing phase uses cosine distance between the two embedded features.

the maximum curvature-based feature extractor. We combined the additional CNN-based feature extractor with the cosine similarity-based matcher to form a new matcher. In our experiments, we use two modified versions of ResNet-18 [7] and a modified version of MobileNetV3-Large [9] (representing small networks) and ResNeXt-50 [29] (representing a large network) as the additional CNN feature extractors.

2.2. Attacks on Biometric Recognition Systems

A sample is considered a "wolf" if it can be falsely accepted as a match with models from multiple identities in a biometric recognition system [27]. A wolf sample can be either biometric or non-biometric. Wolf attacks using wolf samples were initially used to target fingerprint recognition systems [23]. A master biometric attack is a particular case of a "wolf attack" in which the wolf sample looks similar to a biometric trait. A non-biometric wolf sample does not have any constraint on it, hence can be in any appearance. Therefore, non-biometric wolf samples are easier to craft and may have better attack ability than master biometric samples. However, a spoofing detector or a quality assessor integrated into a biometric recognition system can quickly reject them before being recognized. Moreover, since most non-biometric wolf samples mainly focus on a specific flaw in a particular system, they may not generalize well. Therefore, in this paper, we choose to investigate a master finger vein attack.

Master biometric attacks using master biometric samples have been recently used to attack partial fingerprint recognition systems [2] and face recognition systems [21, 25, 20]. A latent variable evolution algorithm [2] is used to generate such master biometric traits by combining an evolutionary algorithm with a pre-trained generative model. The covariance matrix adaptation evolution strategy (CMA-ES) [6] is a popular choice for the evolution algorithm due to its novel design for non-linear and non-convex functions. It is sufficient for a low-resolution biometric trait (like partial fingerprints) generator to use a traditional generative model like WGAN-GP [5]. For high-resolution biometric traits like faces, it requires a more complex generative model like StyleGAN [12]. In this work, vein images do not need very high-resolution like facial ones, but low-resolution images like partial fingerprints are not sufficient. Thus, we could not simply use WGAN-GP as the generative model.

3. Proposed Methods

We first discuss the attack strategy used in this paper. We then introduce two methods for generating master veins, one using the LVE algorithm and one using an adversarial attack, and describe a way to combine them. We assume that the target FVR systems **do not use any spoofing detector or quality assessor**.

3.1. Attack Strategy Analysis

There are several positions where an attack can be carried out on a FVR system, as shown in Fig. 1. We aim to maximize the scope and effectiveness of master vein attacks given limited resources. It is crucial to ensure that the crafted master veins can be used to attack various systems under various conditions. The captured finger vein images are sensor-dependent, and the structure of the veins is unclear because of noise and lack of pre-processing. If we generate coarse master veins to carry out attack 2, the generative model cannot effectively learn the vein representations. The generated master veins also do not work well with other data capture devices. The vein images used to perform attack 4 are more precise, which is more suitable for training the generative model, performing attacks, and analysis. It is possible to translate a master vein into a corresponding captured image using a CNN [22] to perform attack 2. Moreover, an attacker can craft a corresponding PAI given an image of finger veins [26] that can be used to carry out a presentation attack (attack 1). In summary, *in theory, it is possible to carry out attacks 1 and 2 if we can carry out attack 4*.

Miura's system can perform both symmetric matching (or full matching) and asymmetric matching (or partial matching). For partial matching, the probe is a randomly cropped image of the complete one. This is similar to the scenario in the work of Bontrager *et al.* [2]. Before performing random cropping on an input vein image, the system uses an algorithm to calculate a mask to extract the vein-only region first. *For simplicity, we assume that this region is provided. In reality, because of this algorithm, non-vein-looking master vein images may not be cropped appropriately, reducing their attack ability.* For CNN-based systems, the networks can only perform full matching. Therefore, to ensure generalizability, we focus on generating full master vein images. Furthermore, the full master vein images can be cropped for partial matching in Miura's system.

3.2. Attack Using LVE-Based Method

3.2.1 Method's Description

The work of Bontrager *et al.* [2] used a WGAN-GP [5] to generate partial fingerprints (hereafter **LVE^1**). However, a WGAN-GP is hard to train, especially with limited training data. A β-VAE is easier to train and could learn better disentangled representations (hereafter **LVE^2**). However, its generated images have low quality. Therefore, we fuse their strengths in our proposed generator and use the LVE algorithm [2, 21] to generate master veins (hereafter **LVE^3**).

An overview of our proposed LVE^3 method is shown in Fig. 3. To achieve a better generative model with better learned disentanglement representations and high-resolution output, we first train a β-VAE model [8, 3] by

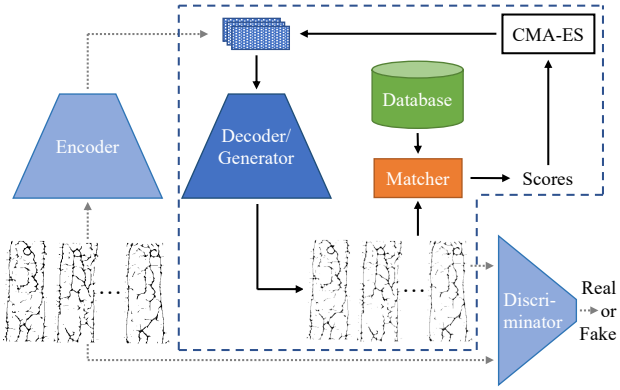
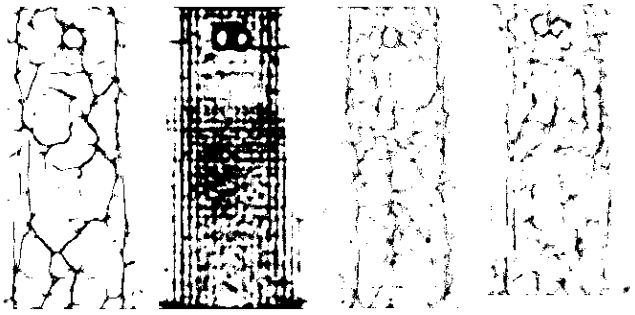


Figure 3. Proposed LVE-based method. Only modules in the dashed polygon are used when running the LVE algorithm.



a. Original image | b. WGAN-GP (LVE^1) | c. β-VAE (LVE^2) | d. Our method (LVE^3)

Figure 4. Original image and images generated using WGAN-GP method, β-VAE method, and our proposed method (best viewed in the digital version with zoom-in). More samples are shown in the Supplementary Material. The WGAN-GP method failed to generate a vein-looking image while the β-VAE method generated a blurry image. Our proposed method generated a clearer image than the other two methods.

minimizing Eq. 1.

$$\mathcal{L}^{\text{β-VAE}}(\theta, \phi; \mathbf{x}, \mathbf{z}, C) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] \\ -\gamma|D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p_\theta(\mathbf{z})) - C|, \quad (1)$$

where ϕ and θ parameterize the distributions of the encoder q_ϕ and decoder p_θ, respectively, and $D_{KL}(\ \parallel\ )$ represents Kullback-Leibler divergence.

Then we fine-tune the decoder by using the WGAN-GP discriminator (minimizing Eq. 2). Using this discriminator improves the quality of the generated images. To ensure stability, we freeze the parameters of q_ϕ and most of p_θ except for the last three convolutional layers of p_θ when minimizing $\mathcal{L}^{\text{GAN}}$. Finger vein images generated using the WGAN-GP method, the β-VAE method, and our proposed method are shown in Fig. 4 in this paper and in Fig. 1 in the Supplementary Material. The WGAN-GP method failed to generate a realistic vein image while our method generated

a clearer image than β-VAE.

$$\mathcal{L}^{\text{GAN}} = \mathbb{E}_{\tilde{\mathbf{x}}\sim\mathbb{P}_g}[D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x}\sim\mathbb{P}_r}[D(\mathbf{x})] + \lambda\mathbb{E}_{\hat{\mathbf{x}}\sim\mathbb{P}_{\hat{\mathbf{x}}}}[(\| \nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}}) \|_2 - 1)^2], \quad (2)$$

where

- $\tilde{\mathbf{x}} = p_\theta(\mathbf{x}|\mathbf{z}) = p_\theta(\mathbf{x}|q_\phi(\mathbf{z}|\mathbf{x}))$.
- $\mathbb{P}_r$ and $\mathbb{P}_g$ are the real and generated data distributions, respectively.
- $\mathbb{P}_{\hat{\mathbf{x}}}$ is sampled uniformly along straight lines between pairs of points sampled from $\mathbb{P}_r$ and $\mathbb{P}_g$.

The LVE algorithm is described in Alg. 1 in the Supplementary Material. For simplicity, we use the CMA-ES [6] for the evolutionary algorithm. Only the decoder p_θ of the β-VAE is used when running the LVE algorithm. It plays the role of the generator to generate vein images.

3.2.2 Preliminary Analysis

The false acceptance rate (FAR - the rate at which unauthorized or illegitimate users are verified) calculated when running the LVE algorithm are plotted in Fig. 5. The master vein generated using the LVE^3 method on Miura's system is shown in Fig. 6.b. Surprisingly, random non-vein-looking finger veins generated by the LVE^1 method (using WGAN-GP) easily fooled Miura's system with the FARs higher than 90%, even without the help of the LVE algorithm. Its cross-correlation-based matcher module failed to work correctly with these wolf samples. This finding raises **an urgent alarm on the reliability of the Miura's system without a spoofing detector or a quality assessor integrated**.

Besides the above irregular case, the proposed LVE^3 method worked better than the LVE^2 method (using β-VAE) on Miura's system (with the FARs about 70% and 50%, respectively). This result confirms the effectiveness of our multi-stage combination of β-VAE and WGAN-GP for the generator. Although they are not perfect-looking, finger veins generated by the LVE^3 method are more natural than those generated by the LVE^1 and the LVE^2 methods, reducing the possibility of being detected by the spoofing detectors or being rejected by the quality assessors. For a CNN-based recognition system, the LVE^1 and LVE^3 methods failed to work on the ResNeXt-50-based system with near-zero FARs. This failure may be due to the ResNeXt-50-based system being a large network trained on a well-designed loss function [4], preventing the formation of dense clusters in its embedding space (discussed in Nguyen *et al.*'s work [20]). We thus investigated another method to attack the CNN-based system, as described in the next section.

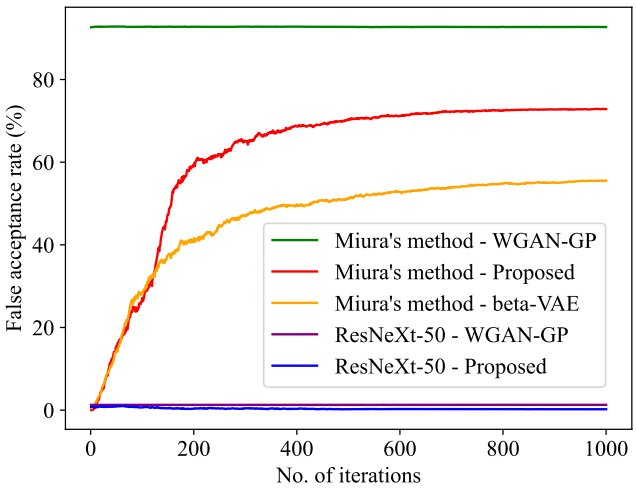


Figure 5. FARs when running the LVE algorithm on Miura's and ResNeXt-50-based systems. When the number of iterations increases, only the FARs of Miura's system increase, implying that the LVE-based method only works on this system.

3.3. Attack Using Adversarial Machine Learning Method

We propose using a modified version of the l_∞ projected gradient descent attack [16] described by Eq. 3. We use a filter K to control the shape of the perturbations, a mask M to limit the area of the perturbations to the area containing the veins, and a soft-label vector $\mathbf{y}$ to control the target identities of the attack. Since the synthesized veins lay on fingers with similar shapes and locations, the mask M can be easily approximated and crafted beforehand by hand or using a gap-filling algorithm.

$$\mathbf{x}^{t+1} = \text{Clip}_{\mathbf{x},\epsilon}(\mathbf{x}^t + \alpha(\zeta * K) \odot M) \quad \text{with} \quad \zeta = \nabla_{\mathbf{x}}\mathcal{L}(\theta, \mathbf{x}^t, \mathbf{y}), \quad (3)$$

where $\mathbf{x}$ is the input image, $\mathbf{y}$ is a target soft-label vector, θ is the set of target model parameters, K is the filter kernel, M is the finger vein mask, $*$ is the convolutional operator, and $\odot$ is the element-wise multiplication operator.

Unlike traditional adversarial targeted attacks, we target multiple labels instead of a single label. We call this **k-label targeted attack**. In more detail, we choose $1 < k < N$ of the total N labels as target labels and set their probabilities to $1/k$, with k as a hyper-parameter. For example, if we choose 3/5 of all labels, the target vector $\mathbf{y}$ is set to $[0.33, 0, 0.33, 0.33, 0]$. Since the FVR system does not calculate class probabilities during its testing phase but embeddings (see Fig. 2), we need to attack the configuration of the training phase. In addition to randomly selecting k target labels, we can choose the top-k labels with the highest predicted probabilities. Doing so can make the optimizer process converge faster with fewer perturbations. Examples of randomly selected k labels and top-k labels are shown

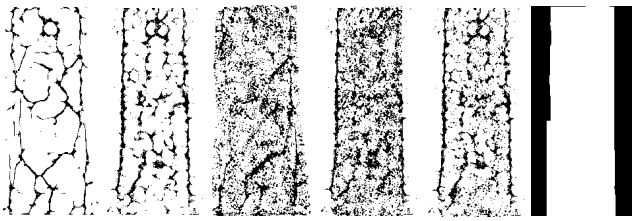


Figure 6. Master veins generated using LVE method, adversarial machine learning method, their combination, and the corresponding mask used for adversarial attacks (best viewed in the digital version with zoom-in).

in Fig. 6.d and Fig. 6.e, respectively. If k is close to 1, the attack is meaningless. Otherwise, it is hard to optimize the perturbations. In our experiments, the target CNN-based FVR system was powerful, with 0% in both false acceptance rate and false rejection rate on the training set. Therefore, we used a small top-k $= 5\%$ in our experiments so that the optimization could successfully converge.

Besides the well-known hyper-parameter ϵ, the number of iterations and the kind of filter K are also important hyper-parameters. If the number of iterations is too small, the optimization will not converge. If it is too large, the perturbations will be too strong, damaging the original image. A damaged master vein image may be rejected by the quality assessor if implemented. Furthermore, it could not be generalized to other systems. Fig. 2 in the Supplementary Material shows such an effect. When the number of iterations is around 200, the image is noisy, and when it is 500 or 1000, it is almost impossible to perceive the veins. We used 100 iterations in our experiments.

Regarding filter K, it is used to control the shape of the perturbations. Ideally, the perturbations should look like veins rather than random patterns or noise. In reality, it is very challenging to achieve this goal. We initially used a (differentiable) CNN-based vein/non-vein classifier as a loss function in optimizing the AdvML attack. However, it was easily fooled. On the other hand, a non-differentiable hand-crafted one was not useful for optimization with back-propagation. Therefore, we simply use a filter kernel for K to regularize the perturbations. The goal is to avoid tiny-dot noise, which can be easily destroyed and is harder to adapt to other attack scenarios. We evaluated Gaussian blur, low-pass, high-pass, and Laplacian kernel. They have different effects on the rate of convergence of the optimization process and the quality of the master vein images. Examples are shown in Fig. 3 in the Supplementary Material. The Gaussian blur kernel helped the optimization process converge the fastest and produced large-size perturbations with the least amount of tiny-dot noise (except the low-pass kernel), so it is the most suitable candidate kernel. The low-pass kernel destroyed almost all adversarial perturbations, preventing the adversarial attack. The high-pass and Laplacian kernels allowed too much tiny-dot noise. When a filter was not used, the optimizer process took a long time to converge, and the crafted perturbations were also tiny-dot noise.

Instead of using a bona fide image, we can use a master vein image as the input image $\mathbf{x}$. If we use a master vein image crafted using the LVE-based method, the corresponding adversarial image can work on both handcrafted recognition systems like Miura's one and CNN-based recognition systems, resulting in better generalizability. An example result of this combination attack is shown in Fig. 6.

4. Experiments

We investigated the attack abilities of master veins crafted using the LVE-based method, the adversarial machine learning method, and their combination in white-box, gray-box, and black-box scenarios.

4.1. Settings

4.1.1 Databases

We used two finger vein databases:

- The SDUMLA-HMT Database [30] contains images of six fingers per subject (six images per finger). We divided it into a training set containing the images for 80 subjects and a test set containing the images for 26 subjects. The training set was used to train the CNN-based recognition systems, set the recognition systems' matching thresholds, train the generative models, and generate master veins. We used both the training and test sets for evaluation.

- The VERA FingerVein Database [26] contains bona fide images of two fingers of 110 subjects (two images per finger). We used the entire database to evaluate black-box attacks (in terms of database). Since its distribution is different from that of the SDUMLA-HMT Database, the recognition systems' matching thresholds calculated on the SDUMLA-HMT Database could not be used and needed to be re-set for this database.

4.1.2 Finger Vein Recognition Systems

We used three FVR systems: one hand-crafted system (Miura's system) and three CNN-based systems (one based on ResNet-18, one based on ResNeXt-50, and one based on MobileNetV3-Large). We chose MobileNetV3-Large (MobileNetV3-L) from the MobileNet family since it performed the best on FVR. For Miura's system (customized

from Idiap bob's implementation), we evaluated both partial matching and full matching. We are aware of other hand-crafted recognition systems [24] using other features extracted from the local binary patterns, principal component analysis, or Gabor filters. This paper focuses on attack 4 in Fig. 1, however, these systems have different feature extractors. Attacking them requires building a mapping network to convert the master veins to their raw forms and perform attack number 2. We treated it as future work.

To generate master veins, we used Miura's system and the ResNeXt-50-based system as surrogate FVR systems. Since the LVE-based method failed to generate master veins on the ResNeXt-50-based recognition system, we ignored this case in all of our experiments. Hereafter, the LVE-based method is assumed to be the one running on Miura's system.

To evaluate performance, we used Miura's system and the ResNeXt-50-based system for evaluating white-box attacks and the ResNet-18- and MobileNetV3-L-based ones for evaluating black-box attacks in terms of FVR systems.

4.2. Evaluation Methodology

We use FAR as the primary metric to define the effectiveness of master vein attacks. We compare the FARs of a FVR system on a normal dataset (without master veins) and a master vein dataset where the zero-effort imposter's probes were replaced by the master vein probes. If the FAR on the master vein dataset is moderately higher than the FAR on the normal dataset, the master vein attack is considered successful.

For the partial matching mode of Miura's system, we first randomly cropped the full vein images to the size of 128×128 pixels and then used them as the probes. Modal images are always full-size ones (320×240 pixels).

4.3. Results and Discussion

4.3.1 Attacks on Known Databases and Systems

We first evaluated master vein attacks on the same and similar configurations we used to craft the master vein. In terms of databases, we attacked the SDUMLA-HMT one, resulting in a white-box attack. In terms of FVR systems, we attacked two types of systems, resulting in both white-box and black-box attacks. In more detail, attacks using master veins generated using Miura's system on CNN-based systems are black-box ones and vice versa, while attacking the same system are white-box ones. Merging both terms, we have white-box and gray-box attacks.

The FARs on bona fide veins and master veins are shown in Table 1. Miura's system is extremely vulnerable to **non-vein-looking** wolf attack generated by the LVE^1 method, with about 69% for partial matching and 93.5% for full matching, as mentioned earlier in section 3.2.2. The possible reason is that when developing this system, wolf attacks had not been introduced. Therefore, the designation of its matching algorithm (based on cross-correlation) only considered vein-looking probes.

Miura's system is also vulnerable to **vein-looking** master vein attacks using the $\text{LVE}^{2,3}$ methods, AdvML method, and their combination. The LVE^3 method outperformed the LVE^2 one and achieved the FARs of about 70% on both train and test sets. It means that nearly two-thirds of the identities were falsely matched with the master veins. It happened because these generators can generate non-exist random finger vein images, and Miura's matcher algorithm has flaws. The LVE algorithm then guided them to generate finger vein images with wolf characteristics with large-enough iterations.

Although it is a black-box attack in terms of FVR systems, the AdvML method achieved the FARs of about 12% for partial matching and about 40% for full matching by Miura's system. The plausible explanation is that the AdvML master veins were optimized with full matching mode (CNN-based systems' only mode), their performance on partial matching mode is suboptimal.

A combination of the AdvML method and the LVE^3 method sustainably raised the FARs of Miura's system in the full matching mode, which is about 85%. However, it was not effective for the partial matching mode. Using the top probabilities label (denoted as (Top) in the table) helped increase the FARs on Miura's system in this partial matching mode and on the CNN-based systems. Its combination with the LVE^1 method has the reverse tendency (possibly because LVE^1 and LVE^3 have different characteristics). Regarding robustness, CNN-based systems resist master vein attacks better than Miura's system. Their FARs only slightly increased (about 1 to 3%) when attacks occurred. Regarding generalization, master vein attacks could not work well on unseen CNN-based systems (ResNet-18 and MobileNetV3-L).

4.3.2 Cross-Database and Cross-System Attacks

Next, we evaluated master vein attacks on more challenging scenarios. In terms of database, we attack a different database - the VERA FingerVein Database. In terms of FVR systems, attacks on Miura's system and the ResNeXt-50 system are white-box while attacks on ResNet-18 and MobileNetV3-L are black-box. Table 2 shows the FARs on bona fide and master veins.

Miura's system continued to be vulnerable to wolf attacks and master veins attacks. For master vein attacks, the FARs were around 20% and could reach 47.73% when we used the combination method to attack the full matching mode. Using top labels helped increase the FAR of the attack on partial matching mode to 22.25%. On the

Table 1. FARs (in %) of three FVR systems on SDUMLA-HMT Database with bona fide and master veins. Gray cells indicate gray-box attacks; white cells indicate white-box attacks. Bold numbers indicate that the master vein attacks have FARs higher than those of the corresponding bona fide cases by at least 1%.

Matcher	**Miura's system (Partial matching)**		**Miura's system (Full matching)**		**ResNeXt-50**		**ResNet-18**		**MobileNetV3-L**	
Attack \ Dataset	**Train set**	**Test set**	**Train set**	**Test set**	**Train set**	**Test set**	**Train set**	**Test set**	**Train set**	**Test set**
Bona fide	07.57	08.02	08.46	08.98	0.00	2.25	0.00	3.37	0.00	1.31
LVE^1 (WGAN-GP)	**68.24**	**70.41**	**92.46**	**94.21**	**1.85**	1.92	**1.51**	2.25	0.67	1.50
LVE^2 (β-VAE)	**59.63**	**59.27**	**54.75**	**43.89**	0.10	1.44	0.90	2.42	0.33	0.33
LVE^3 (Combination)	**70.47**	**69.85**	**73.29**	**71.84**	**1.46**	**6.07**	0.96	**5.86**	0.53	2.03
AdvML	**11.34**	**13.11**	**32.02**	**49.52**	**1.88**	**3.69**	**1.44**	2.24	0.61	1.46
LVE^3 + AdvML	**48.20**	**50.00**	**82.36**	**88.79**	**1.82**	**3.35**	**1.15**	1.93	0.48	0.64
LVE^3 + AdvML (Top)	**62.73**	**62.52**	**77.82**	**80.41**	**2.37**	**5.32**	**1.60**	4.00	**1.03**	**3.47**
LVE^1 + AdvML (Top)	**76.60**	**76.95**	**91.86**	**93.81**	**1.68**	1.85	**1.52**	2.09	0.55	0.40

Table 2. FARs (in %) of three FVR systems on VERA FingerVein Database with bona fide and master veins. Bold numbers indicate that the master vein attacks have FARs higher than those of the corresponding bona fide cases by at least 1%.

Matcher / **Attack**	**Miura's system (Partial matching)**	**Miura's system (Full matching)**	**ResNeXt 50**	**ResNet 18**	**Mobile NetV3-L**
Bona fide	04.07	03.13	8.22	7.28	8.10
LVE^1 (WGAN)	**38.84**	**43.86**	0.18	0.10	0.18
LVE^2 (β-VAE)	**15.08**	02.92	0.00	0.00	0.00
LVE^3 (Comb.)	**20.84**	**19.54**	0.54	0.00	0.01
AdvML (A)	03.12	03.57	0.20	0.04	0.18
LVE^3+A	**16.37**	**47.73**	0.42	0.01	0.18
LVE^3+A (Top)	**22.25**	**26.34**	0.82	0.52	0.21
LVE^1+A (Top)	**39.28**	**44.49**	0.18	0.01	0.17

other hand, the CNN-based recognition systems were robust against master vein attacks. However, it is important to note that the CNN-based recognition systems could not generalize well on the VERA FingerVein Database, resulting in higher FARs for bona fide vein attacks. It implies that if we want to use the CNN-based recognition systems effectively (so that bona fide users are not falsely rejected), we need to train them on the current dataset. However, doing so also opens a chance for master vein attacks.

4.4. Summary

Miura's system in partial matching and full-matching modes was vulnerable to non-vein-looking wolf attacks and vein-looking master vein attacks in white-box and gray-box scenarios. Both attacks substantially increased the FARs for Miura's system while barely increasing them for the CNN-based systems. A combination of the LVE^3 method and the AdvML method can reach 88.79% FAR on the full matching mode of Miura's system. Small increments of FARs on CNN-based systems indicate that they are more robust on master vein attacks.

It was challenging to perform master vein black-box attacks when both the target recognition system and the database were unknown. However, in reality, handcrafted FVR systems have already been deployed in ATMs, and the replacement cost is high. Since their variety is limited, attackers can narrow the scope of their attacks to gray-box or even white-box. Attackers can also prepare a set of potentially effective master veins in advance. Due to these reasons, master vein attacks still be a viable threat.

4.5. Social Impacts

To avoid possible harm, we use academic freely-accessed finger vein databases and open-source FVR systems. We believe our findings are necessary to raise awareness and promote improving the robustness of such systems. Besides robustness, we suggest using a fake finger vein detector to detect master veins.

5. Conclusion and Future Work

We have demonstrated that non-vein-looking wolf samples (generated by WGAN-GP) and vein-looking master veins generated using our proposed methods (LVE-based method, adversarial machine learning attack, and their combination) can successfully perform while-box and gray-box attacks on FVR systems. Miura's handcrafted system is fragile against such attacks, while CNN-based methods are more robust. Since not all commercial FVR systems are deep learning-based, their variety is limited, and the countermeasure methods are not always available, the threat of master vein attacks should not be underestimated.

Future work will focus on performing adversarial attacks to minimize cosine distance instead of maximizing label probabilities, improving the shape of adversarial perturbations to make them more vein-like, and evaluating more FVR systems and databases.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grants JP16H06302, JP18H04120, JP20K23355, JP21H04907, and JP21K18023, and by JST CREST Grants JPMJCR18A6 and JPMJCR20D3, including the AIP challenge program, Japan.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *ICML*, pages 214–223, 2017.

[2] Philip Bontrager, Aditi Roy, Julian Togelius, Nasir Memon, and Arun Ross. Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution. In *BTAS*, pages 1–9. IEEE, 2018.

[3] Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β-VAE. In *NIPSW*, 2017.

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699, 2019.

[5] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. In *NIPS*, pages 5767–5777, 2017.

[6] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[8] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *ICLR*, 2017.

[9] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for MobileNetV3. In *ICCV*, pages 1314–1324, 2019.

[10] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, pages 4700–4708, 2017.

[11] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin IP Rubinstein, and J Doug Tygar. Adversarial machine learning. In *Workshop on Security and Artificial Intelligence*, pages 43–58. ACM, 2011.

[12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410. IEEE, 2019.

[13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.

[14] Rıdvan Salih Kuzu, Emanuele Maiorana, and Patrizio Campisi. Loss functions for cnn-based biometric vein recognition. In *EUSIPCO*, pages 750–754. IEEE, 2021.

[15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017.

[16] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *ICLR*, 2018.

[17] Sébastien Marcel, Mark S Nixon, Julian Fierrez, and Nicholas Evans. *Handbook of biometric anti-spoofing: Presentation attack detection*, volume 2. Springer, 2019.

[18] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Machine Vision and Applications*, 15(4):194–203, 2004.

[19] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE Transactions on Information and Systems*, 90(8):1185–1194, 2007.

[20] Huy H. Nguyen, Sébastien Marcel, Junichi Yamagishi, and Isao Echizen. Master face attacks on face recognition systems. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2022.

[21] Huy H Nguyen, Junichi Yamagishi, Isao Echizen, and Sébastien Marcel. Generating master faces for use in performing wolf attacks on face recognition systems. In *IJCB*, 2020.

[22] Yingxue Pang, Jianxin Lin, Tao Qin, and Zhibo Chen. Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, 2021.

[23] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634, 2001.

[24] Kashif Shaheed, Hangang Liu, Gongping Yang, Imran Qureshi, Jie Gou, and Yilong Yin. A systematic review of finger vein recognition techniques. *Information*, 9(9):213, 2018.

[25] Ron Shmelkin, Liar Wolf, and Tomer Friedlander. Generating master faces for dictionary attacks with a network-assisted latent space evolution. In *FG 2021*, pages 01–08. IEEE, 2021.

[26] Pedro Tome, Ramachandra Raghavendra, Christoph Busch, Santosh Tirunagari, Norman Poh, BH Shekar, Diego Gragnaniello, Carlo Sansone, Luisa Verdoliva, and Sébastien Marcel. The 1st competition on counter measures to finger vein spoofing attacks. In *ICB*, pages 513–518. IEEE, 2015.

[27] Masashi Une, Akira Otsuka, and Hideki Imai. Wolf attack probability: A new security measure in biometric authentication systems. In *ICB*, pages 396–406. Springer, 2007.

[28] Gerik Alexander von Graevenitz. Biometric authentication in relation to payment systems and atms. *Datenschutz und Datensicherheit-DuD*, 31(9):681–683, 2007.

[29] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *CVPR*, pages 1492–1500, 2017.

[30] Yilong Yin, Lili Liu, and Xiwei Sun. Sdumla-hmt: a multimodal biometric database. In *Chinese Conference on Biometric Recognition*, pages 260–268. Springer, 2011.