applied sciences

MDPI

Article

A Realistic Hand Image Composition Method for Palmprint ROI Embedding Attack

Licheng Yan [1,2], Lu Leng [1,2,*], Andrew Beng Jin Teoh [3] and Cheonshik Kim [4,*]

[1] School of Software, Nanchang Hangkong University, 696 Fenghe Nan Avenue, Nanchang 330063, China; 2116085400002@stu.nchu.edu.cn
[2] Key Laboratory of Jiangxi Province for Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China
[3] School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120749, Republic of Korea; bjteoh@yonsei.ac.kr
[4] Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea
* Correspondence: leng@nchu.edu.cn (L.L.); mipsan@sejong.ac.kr (C.K.)

Abstract: Palmprint recognition (PPR) has recently garnered attention due to its robustness and accuracy. Many PPR methods rely on preprocessing the region of interest (ROI). However, the emergence of ROI attacks capable of generating synthetic ROI images poses a significant threat to PPR systems. Despite this, ROI attacks are less practical since PPR systems typically take hand images as input rather than just the ROI. Therefore, there is a pressing need for a method that specifically targets the system by composing hand images. The intuitive approach involves embedding an ROI into a hand image, a comparatively simpler process requiring less data than generating entirely synthetic images. However, embedding faces challenges, as the composited hand image must maintain a consistent color and texture. To overcome these challenges, we propose a training-free, end-to-end hand image composition method incorporating ROI harmonization and palm blending. The ROI harmonization process iteratively adjusts the ROI to seamlessly integrate with the hand using a modified style transfer method. Simultaneously, palm blending employs a pretrained inpainting model to composite a hand image with a continuous transition. Our results demonstrate that the proposed method achieves a high attack performance on the IITD and Tongji datasets, with the composited hand images exhibiting realistic visual quality.

Keywords: palmprint recognition; palmprint ROI extraction; hand composition; hand attack; ROI harmonization

check for updates

Citation: Yan, L.; Leng, L.; Teoh, A.B.J.; Kim, C. A Realistic Hand Image Composition Method for Palmprint ROI Embedding Attack. *Appl. Sci.* **2024**, *14*, 1369. https://doi.org/10.3390/app14041369

Academic Editor: José Salvador Sánchez Garreta

Received: 26 December 2023
Revised: 27 January 2024
Accepted: 30 January 2024
Published: 7 February 2024

1. Introduction

Palmprint recognition (PPR) has recently been widely studied due to its robustness and accuracy. A typical PPR is composed of two stages. Stage one is the preprocessing stage, in which the palmprint region of interest (ROI) is located and extracted; stage two is the recognition stage, which entails feature extraction, matching, and the final decision. ROI localization is indispensable for PPR, which determines which part of the palmprint will be used for recognition, and it usually requires an input hand satisfying certain conditions.

Given the vital role of ROI in PPR systems, ROI attacks aimed at spoofing and evading PPR systems [1–3] have emerged. While these attacks primarily target the palmprint ROI, their practicality may be limited, as PPR systems typically utilize hand images as input rather than the ROI. Consequently, synthesized hand images are required for impersonation attacks. Generally, two distinct approaches are employed to achieve hand image synthesis: generating a hand image based on the ROI and embedding the ROI into an existing hand image. The merit of the embedding method over the generating method stems from its reduced dependency on training data for a generator to yield realistic outcomes, rendering it a more practical choice.

Specifically, the embedding approach employs a hand image as a carrier and integrates the ROI into it, a procedure dubbed ROI embedding. After acquiring composited images through ROI embedding, these images can be employed to launch impersonation attacks. In this paper, we aim to formulate an ROI embedding method tailored to substantiate the impersonation attack through composited hands, namely a hand attack instead of synthesized ROI.

As illustrated in Figure 1, the components delineated by the black dashed line represent the palmprint recognition flow, constituting a PPR system that exclusively accepts the entire hand as a valid input. The ROI attack flow attempts to compromise the system by employing an ROI image deemed less practical. Conversely, the hand attack flow poses a more substantial threat, enabling attacks through legitimate inputs.

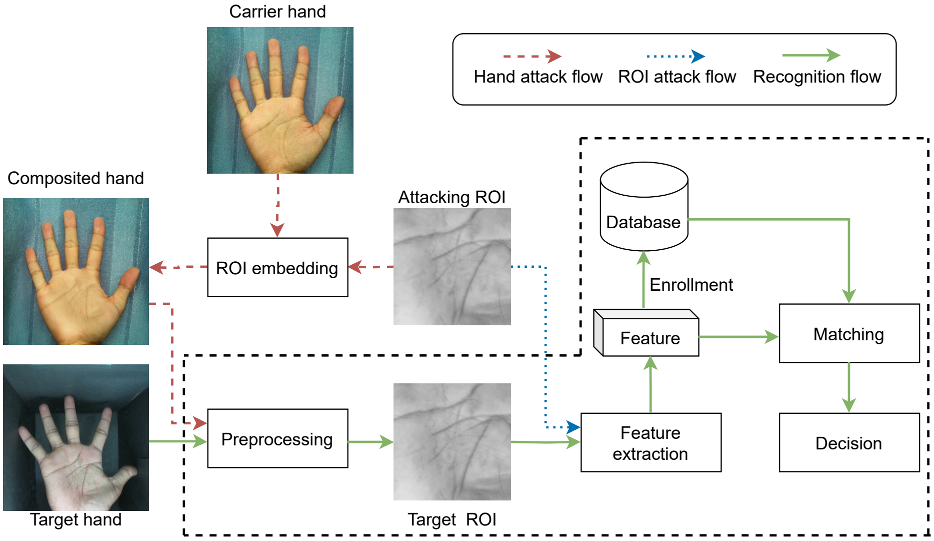


Figure 1. Hand attack, ROI attack, and palmprint recognition flow.

The hand attack flow consists of a series of steps. Initially, the attacking ROI is acquired with the goal of compromising the target ROI. Subsequently, an arbitrary full-hand image is selected as the carrier hand. The third step involves utilizing an ROI embedding method to seamlessly integrate the attacking ROI into the chosen carrier hand. Ultimately, this process creates a composite hand, enabling the execution of the intended attack.

The ROI embedding is intuitive yet non-trivial, posing two principal challenges. The first challenge involves embedding localization, necessitating the precise placement of the ROI on the carrier hand at an appropriate scale. However, a singular ROI position is not universally applicable to a hand, given the diverse preprocessing methods in PPR systems. The second challenge pertains to appearance consistency, wherein the attacking ROI's hue, luminance, and skin color must seamlessly integrate with the carrier hand. Moreover, the region around the embedding position in the carrier hand must exhibit continuous texture and natural transitions. Appearance consistency is an intuitive metric for ROI embedding, ensuring that the composited hand image attains desirable visual quality and fidelity, as illustrated in Figure 2. In this study, we focus on the second challenge, assuming the availability of a decent preprocessing method employed by the target PPS system.

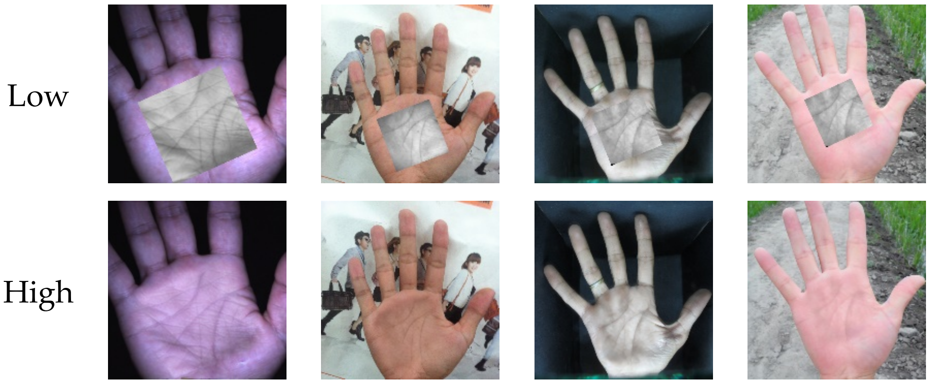


Figure 2. The illustration of low and high appearance consistency in ROI embedding. The second row depicts a seamless integration of carrier and ROI images.

This paper introduces an end-to-end, training-free approach for ROI embedding to achieve a realistic composition of hand images. Our methodology comprises two key components: ROI harmonization and palm blending. In the ROI harmonization phase, we employ a modified style transfer method to closely transform the attacking ROI's appearance to resemble that of the carrier hand. This adaptation leverages an optimization method based on the pretrained Contrastive Arbitrary Style Transfer (CAST) network [4], which can achieve desirable results within a few iterations.

In the palm blending phase, the harmonized ROI image warps to the embedding position on the carrier hand. Concurrently, the area surrounding the embedding position on the carrier hand is masked and then regenerated using a pretrained inpainting network to facilitate a natural transition. The ROI image of the carrier hand serves as the reference image for inpainting, capitalizing on the ability of the inpainting method [5] to generate the masked area with semantic information from a reference image without requiring additional training. Ultimately, this process yields a composited hand image with heightened appearance consistency.

In conclusion, our contributions are as follows.

(1) We present the first end-to-end, training-free approach for synthesizing hand images, extending the scope of ROI attacks to encompass hand attacks.

(2) To enhance the visual coherence of the composited hand images, we introduce an efficient and effective optimization method based on CAST and a pretrained reference-based inpainting technique.

(3) The experimental results underscore the efficacy of our proposed method. The resulting composited hand images exhibit a noteworthy attack success rate and demonstrate superior appearance consistency in several datasets.

The subsequent sections of this paper are structured as follows: Section 2 presents an overview of the relevant prior research. Section 3 delineates the designed hand image composition methods. Section 4 presents comprehensive experiments on the composition methods and the composited hand images. Lastly, Section 5 encompasses the conclusions drawn from the study and discusses potential avenues for future research.

2. Related Works

2.1. Palmprint Recognition

Zhang first introduced PalmCode [6] for PPR, which is characterized by four distinct steps: preprocessing, feature extraction, matching (or enrollment), and decision. Subsequent developments in PPR have predominantly focused on enhancing feature extraction, matching, and decision-making. These advancements include approaches based on handcrafted methodologies [7–10] as well as those rooted in deep learning [11–14].

A competitive coding scheme [7] has been introduced to enhance the discriminative features, selecting the highest responsive direction order as a feature from six-directional

Gabor features. Building upon this competition scheme, the work [8] employs a Modified Finite Radon Transform (MFRAT) to replace Gabor, extracting more robust features. Another paper [9] enhances the discriminative nature of the competitive feature by incorporating an additional adjacent direction matrix. To leverage more comprehensive information, ref. [10] employs six-directional Gabor features and utilizes the average match score of these features to inform the decision-making process.

In the realm of deep-learning-based approaches, several studies [11,12] have integrated the competition mechanism with multi-scale learnable Gabor kernels to enhance feature extraction effectively. Furthermore, the work [13] employs a hashing method for extracting concise features, while [14] introduces a single convolution layer to extract multi-directional features. These PPR methods are dedicated to exploring discriminative features within the palmprint ROI.

Conversely, other studies have focused on refining preprocessing methods [15–20], directing their efforts toward devising universal and effective techniques for accurately locating palmprint ROIs. The majority of ROI localization methods are based on identifying finger valley points, exemplified by methodologies such as that described in [15], which employs Convolutional Neural Networks (CNN) for hand classification and segmentation [16], which employs an improved active shape model [17], which utilizes a hand landmark detector model, and [18], which integrates a segmentation network and two regression networks. Some methodologies directly predict the ROI bounding box, including [19], which incorporates a finger classifier and transfer learning, and [20], which employs a regression network to simultaneously detect palmprint and palm vein ROIs.

A plethora of diverse methods for localizing ROI lead to variations in the obtained ROI outcomes. Recognition results, derived from matching and decision processes, are intricately dependent on preprocessing and feature extraction outcomes. Therefore, accurate knowledge of the ROI position is crucial to the efficacy of hand attacks. The successful execution of a hand attack appears unattainable without prior knowledge of the target preprocessing method, given the diverse nature of preprocessing techniques. Consequently, our research is grounded in the assumption that knowledge of the target preprocessing method is available.

2.2. Attacks on Palmprint

Given the popularity of palmprint ROI recognition methods, there has been an emergence of ROI attacks aimed at spoofing and evading these recognition systems. One such approach is the False Acceptance Attack proposed by Wang [1]. This method leverages Generative Adversarial Networks (GANs) to generate a substantial volume of synthetic palmprint ROI images. Subsequently, a k-means algorithm is employed to eliminate indistinguishable images, constituting an attack strategy against the recognition system.

Sun has introduced two reinforcement strategies for PPR attacks [2], employing iterative algorithms to modify a given ROI image. The objective is to enhance the similarity between the modified ROI and target features, thus subverting the PPR system. On the other hand, Yang utilizes a style-transfer method to generate a realistic appearance based on ROI features to attack the system [3]. Notably, these attack methodologies primarily target the palmprint ROI, but their efficacy may be limited in practical environments as they typically do not successfully navigate through the preprocessing stage within a PPR system.

2.3. Image Composition

Image composition, a classical task in image processing, encompasses subtasks such as object placement, image harmonization, and image blending [21]. Object placement focuses on placing the foreground to the background with a suitable position, size, and orientation, as shown in [22], in which the authors use two generative modules and a spatial transformation network to decide the position for placement, and [23], which applies the bounding box calculated by the features from the masked convolutions to locate the position. Considering the differences between the color and hue of the foreground and background, image harmonization is used to decrease these differences, as shown in [24], in which the authors designed an image harmonization network with collaborative dual transformation for pixel and RGB transformation to harmonization, and [25], which applied the self-consistent style contrastive learning scheme with background-attentional adaptive instance normalization to harmonization. For seamless composition, image blending is applied to diminish the unnatural edge around the foreground in the composited image, as described in [26], which optimizes the style and content loss with the Poisson blending loss and iteratively updates the pixel of the boundary region for blending, and [27], which proposes using a densely connected multi-stream fusion network to fuse the foreground and the background image. Generally, these tasks have been geared towards achieving realistic visual quality by situating independent objects within a background without explicit precision requirements.

However, the challenge intensifies when dealing with hand and palmprint ROI, as they constitute an integrated entity with continuous texture connection rather than two distinct objects. Consequently, hand image composition poses a more substantial challenge, demanding a higher level of precision in seamlessly placing an ROI into a palm and ensuring a cohesive appearance of the palmprint as a unified entity. Moreover, the foreground palmprint must preserve most of texture information under harmonization. Conventional image composition methods fall short of meeting the specific requirements of hand composition. Figure 3 illustrates the distinction between image composition and the more intricate task of hand composition.

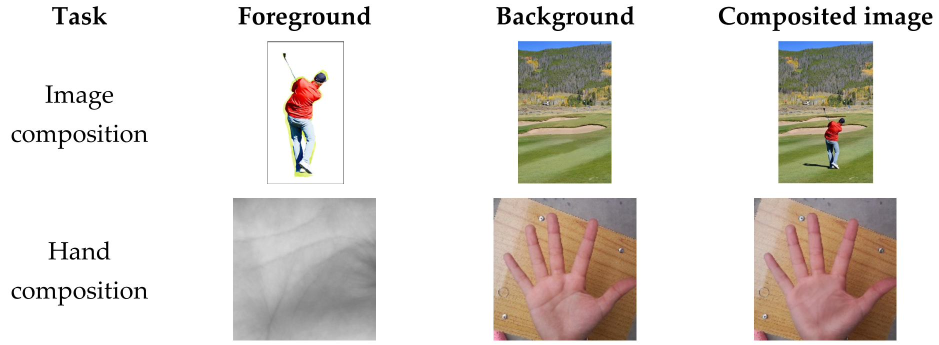


Figure 3. The comparison of image composition [21] and hand composition.

3. Methodology

3.1. Threat Model for Hand Attacks

Let $\mathcal{P}(\cdot)$ represent the preprocessing procedure of the PPR system, where H denotes the full hand image. The ROI is obtained through $\mathcal{P}(H)$ and is subsequently utilized for palmprint recognition.

Suppose R_{Atk} and C denote an attacking ROI and an arbitrary carrier hand image, respectively, where R_{Atk} is assumed to be an ROI image designed to match the target ROI in the feature domain. The R_{Atk} can be sourced from a compromised dataset or synthesized using the methods detailed in Section 2.2. We assume that an adversary possesses knowledge of the $\mathcal{P}(\cdot)$ of the target system, implying awareness of the precise position P of ROI within C; this is referred to as the embedding position.

Designating $\mathcal{E}(\cdot,\cdot,\cdot)$ as our proposed ROI embedding technique, a highly realistic composited hand image H_{Atk} can be generated through $\mathcal{E}(R_{Atk}, C, P)$ in an end-to-end and training-free manner. By inputting H_{Atk} into the target PPR system, R_{Atk} can be accurately extracted through $\mathcal{P}(\cdot)$, achieving the objective of the attack.

3.2. Overview

$\mathcal{E}(\cdot,\cdot,\cdot)$ comprises two crucial components: ROI harmonization and palm blending. Let R_{Ref} denote the reference ROI—an image extracted from C at position P, utilized to facilitate both ROI harmonization and palm blending. During the ROI harmonization, an iterative optimization process is implemented to transform the R_{Atk} style to align with R_{Ref}. Subsequently, the harmonized R_{Atk} is seamlessly integrated into the carrier image C during the palm blending phase. In this phase, the area surrounding position P is masked to facilitate inpainting. The inpainting operation employs R_{Ref} as a reference image, utilizing its semantic information to regenerate the masked area. The resulting composited hand image represents the culmination of the entire process and is referred to as ROI embedding. The schematic representation of the ROI embedding process is displayed in Figure 4.

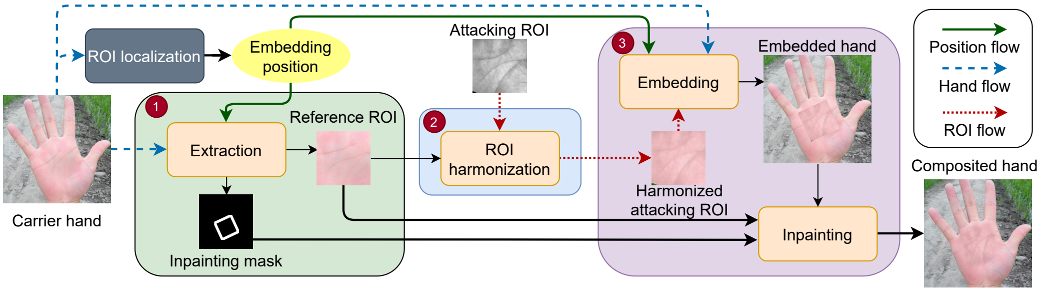


Figure 4. The pipeline of ROI embedding.

The ROI embedding pipeline, shown in Figure 4, comprises three distinct components. The first component is the localization, which is responsible for furnishing essential R_{Ref} and an inpainting mask required for subsequent ROI harmonization and palm blending. The inputs for this stage are carrier image C and the specified embedding position P. The ROI localization is determined based on the known target preprocessing.

The second component involves ROI harmonization, wherein R_{Atk} transforms harmonized R_{Atk} aligned with the style of R_{Ref} to mitigate visual disparities between R_{Atk} and C. ROI harmonization adopts a modified style transfer method, detailed in Section 3.3. The last component constitutes the palm blending, which results in generation of the composited hand image. The harmonized R_{Atk} is initially incorporated into C based on the predetermined P. Subsequently, the embedded hand image is input into the inpainting network with R_{Ref}, facilitating the regeneration of the masked inpainting area. The composited hand image is obtained, with further details discussed in Section 3.4.

3.3. ROI Harmonization

Given potential hue, illumination, and skin color disparities between R_{Atk} and carrier hand image C, ROI harmonization becomes imperative to mitigate these differences. Moreover, for computational efficiency in matching and aligning with the grayscale nature of ROI images, R_{Atk} is converted to a grayscale image.

Specifically, we leverage the pretrained Contrastive Arbitrary Style Transfer (CAST) proposed in [4] to achieve ROI harmonization. As shown in Figure 5, CAST employs cycle generation to train encoders and decoders, adeptly separating and merging style and content. However, training a model that accommodates the significant diversity within hand images across various datasets proves challenging. We devise an iterative style transfer method based on pretrained CAST to initialize and optimize image quality for individual palms. This iterative process resembles the training process of CAST but focuses on optimizing the image quality rather than the model ability, which enables the synthesis of realistic images through limited iterations.

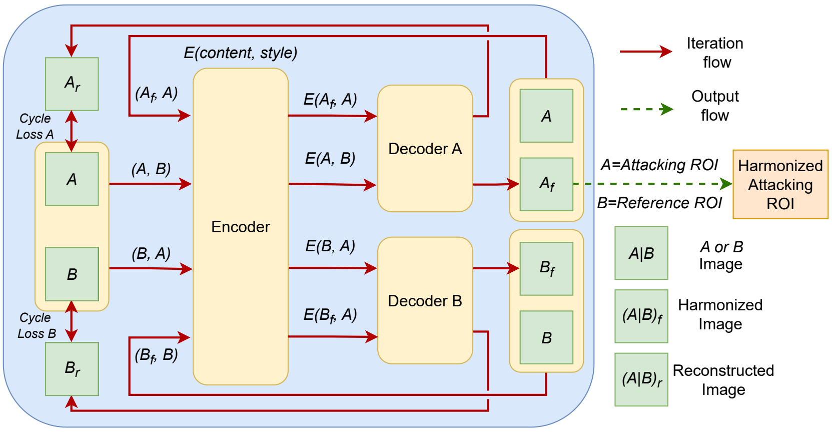


Figure 5. The pretrained CAST with modified loss functions is utilized to realize ROI harmonization.

Specifically, let A and B symbolize R_{Atk} and R_{Ref}, respectively. The notation $A|B$ and $B|A$ denote A or B and B or A, respectively. The ROI harmonization process commences by employing the encoder $E(\cdot,\cdot)$ to amalgamate the content of $A|B$ and the style of $B|A$ from the two input images, A and B. Subsequently, the decoder operates on $E(A|B, B|A)$ to generate the harmonized image $(A|B)_f$ based on the content of $A|B$ and the style of $B|A$. Once $(A|B)_f$ is obtained, it is combined with $A|B$ to yield a reconstructed image $(A|B)_r$. The final harmonized image is refined by minimizing the difference between A and A_r and B and B_r.

Within our proposed ROI harmonization framework, the loss function is essential in enhancing both image quality and harmonization effectiveness. Notably, the loss function employed in CAST solely incorporates the $\mathcal{L}_1$ loss and assigns equal weights to $\mathcal{L}_{cycA}$ and $\mathcal{L}_{cycB}$. However, this approach proves insufficient for attaining optimal image quality and harmonization effects. Furthermore, preserving texture information in R_{Atk} is critical, necessitating measures to minimize interference from R_{Ref}. Consequently, essential and imperative modifications must be made to the loss function to address these considerations.

Building upon the existing $\mathcal{L}_1$ loss, the cycle loss is improved by introducing two loss functions, namely $\mathcal{L}_2$ and $\mathcal{L}_{lpips}$ [28], to enhance the preservation and separation of content and style information, in which the overall image visual quality can be improved. To further preserve texture information in synthetic images, a texture loss $\mathcal{L}_T$ is devised. This involves utilizing a separate texture feature extractor, $f_t()$ which is composed of two Gabor convolution layers and a leakyReLU function for feature extraction. For a comprehensive feature extraction, a six-direction Gabor kernel is employed. Following the acquisition of texture features, $\mathcal{L}_1$ is employed to calculate the differences between input images and reconstructed images. The $\mathcal{L}_T$ is the average of the six Gabor feature losses, as depicted in Equation (2).

$$f_t\ (x, g_i) = \varphi_{LR}\ (Conv(Conv(x, g_i\ ), g_i)) \tag{1}$$

$$\mathcal{L}_T(x, y) = \frac{\sum_{i=1}^{6} \mathcal{L}_1(f_t(x, g_i),\ \ f_t(y, g_i))}{6} \tag{2}$$

where $Conv$ denotes convolution, φ_{LR} is the LeakyReLU function, and g_i is the Gabor convolution kernel toward $(i-1) \times 30$ degrees.

The final modified cycle loss is Equation (4), which constitutes $\mathcal{L}_{cycA}$ and $\mathcal{L}_{cycB}$. The two losses, $\mathcal{L}_{cycA}$ and $\mathcal{L}_{cycB}$, that are specified by Equation (3), are almost identical except for their input and loss coefficients.

$$\mathcal{L}_{cycA|B}(x, y) = \lambda_{L1}\mathcal{L}_1(x, y) + \lambda_{L2}\mathcal{L}_2(x, y) + \lambda_{LP}\mathcal{L}_{lpips}(x, y) + \lambda_{LT}\mathcal{L}_T(x, y) \tag{3}$$

$$\mathcal{L}_{cyc}(A, B, A_r, B_r) = \mathcal{L}_{cycA}(A, A_r) + \mathcal{L}_{cycB}(B, B_r) \tag{4}$$

where x and y represent the original image and the reconstructed image, respectively. The λ_{L1}, λ_{L2}, λ_{LP}, and λ_{LT} denote the weight coefficients of $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_{lpips}$ and $\mathcal{L}_T$ loss, respectively.

Since we need to preserve more detail in A and more appearance in B, the loss coefficients of $\mathcal{L}_{cycA}$ are experientially set as 0.5, 0.3, 0.3, and 0.7 for λ_{L1}, λ_{L2}, λ_{LP}, and λ_{LT}, respectively. On the other hand, the loss coefficients of $\mathcal{L}_{cycB}$ are set to 0.5, 0.7, 0.7, and 0.3. Moreover, since our goal is to improve the image quality, not the network performance, and the adversarial loss produced by the discriminators of CAST will slow and disturb the image optimization process, the discriminators and their loss are disabled to speed up and stabilize the optimization process.

3.4. Palm Blending

Each palm exhibits a distinct texture characterized by unique and specific patterns. Directly overlaying attacking ROI R_{Atk} onto carrier hand C can introduce noticeable inconsistencies in the palm area surrounding R_{Atk}. To address this challenge, we have devised an intuitive solution that uses a pretrained inpainting network for region regeneration. Specifically, we utilize the Paint by Example (PbyE) inpainting network [5], which employs a reference image to generate a mask area. The generated area, while not identical to the reference image, mirrors the semantic and visual characteristics of the reference. Moreover, the mask's surrounding area is also considered a reference for the inpainting process, making this network suitable for generating a seamless and continuous palm area without additional training.

The process begins by obtaining the embedded hand by warping the harmonized R_{Atk} into the specified embedding position P on carrier hand C. Subsequently, reference ROI R_{Ref} is employed as the reference image to guide the inpainting process. Finally, an inpainting mask is generated based on the embedding position P of carrier hand C, as illustrated in Figure 6.

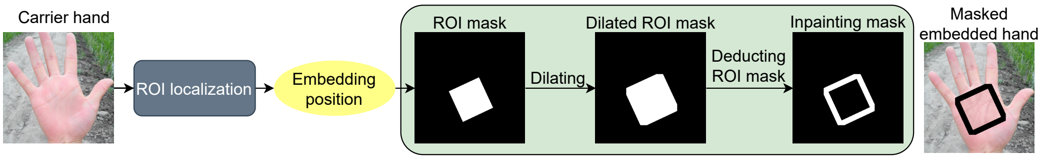


Figure 6. The pipeline of making inpainting masks.

As illustrated in Figure 6, ROI localization is employed to extract the embedding position P from the carrier image C. Subsequently, the inpainting mask is deduced based on P through a three-step procedure. Firstly, an ROI mask is delineated using P, where white represents the mask and black signifies the background to be disregarded. Secondly, employing a dilation kernel, the ROI mask from the first step is expanded to generate a dilated ROI mask. The final inpainting mask is derived by subtracting the ROI mask from the first step from the dilated ROI mask obtained in the second step. The impact of this inpainting mask on the masked embedded hand is depicted in Figure 6.

Nevertheless, the size of the dilation kernel requires careful consideration. A larger kernel size leads to a more extensive inpainting area, potentially introducing significant changes in C and consequently diminishing the realism of the composited hand image. Conversely, a smaller kernel size results in a more confined inpainting area, making the transition harder than a larger kernel. To address this, we devised an adaptive kernel size strategy, selecting one of the five side lengths of the ROI as the dilation kernel size.

Once the embedded hand, inpainting mask, and reference image R_{Ref} have been obtained, the final composited hand image is synthesized using the Paint by Example (PbyE) method. The palm blending pipeline is depicted in Figure 7.

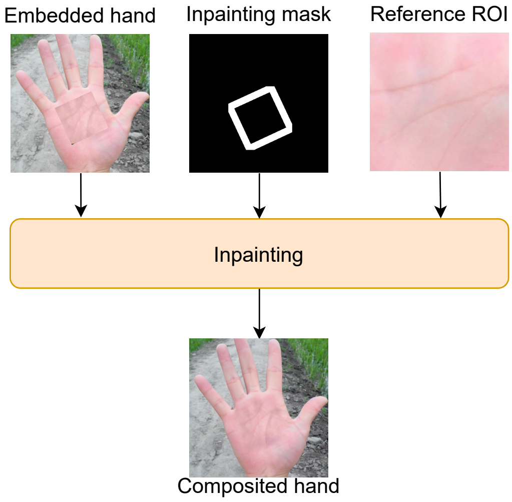


Figure 7. The illustration of Palm Blending.

4. Experiments

The experimental setup is as follows: an Intel(R) Xeon(R) Platinum 8222CL CPU @3.00 GHz from Chengdu in China, equipped with 64 GB of internal storage, an NVIDIA 3090Ti GPU from Jiangsu in China, and an Ubuntu 20.04.3 LTS 64-bit operating system. Four distinct datasets were employed in the ensuing experiments, with comprehensive details available in Table 1, and a few image samples of each dataset are shown in Figure 8.

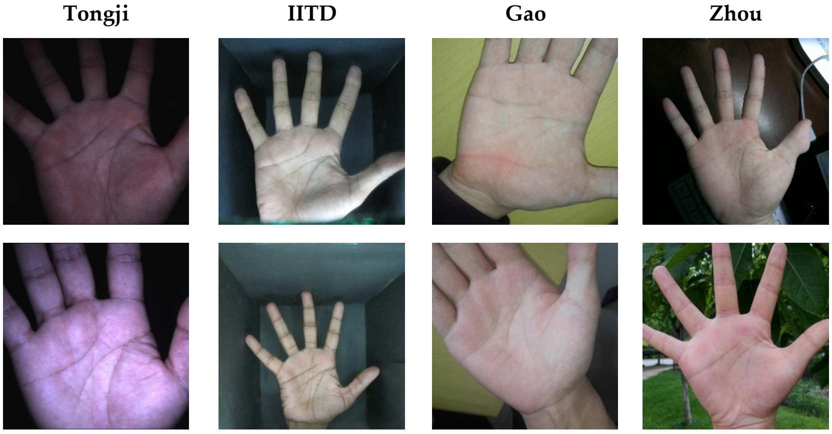


Figure 8. The dataset samples of Tongji, IITD, Gao and Zhou.

Table 1. Datasets Explanation.

Datasets	All Data		Used Data		Resolution	Condition
	Hands	Images	Hands	Images		
Tongji [29]	600	12,000	200	200	800 × 600	Simplex
IITD [30]	460	2601	200	200	1600 × 1200	Simplex
Gao [31]	204	816	200	200	720 × 1184	Indoor
Zhou [32]	160	918	160	200	Multi-resolution	Indoor and outdoor

A judicious data selection was undertaken given the subtle variations within hand images of single datasets and the imperative to conduct experiments under cross-dataset conditions. Specifically, 200 images were randomly sampled from each dataset for experimentation purposes, and all 200 images sampled from each dataset were from different hands, except for the 200 images sourced from Zhou [32], which were from 160 hands.

All images from the four datasets underwent cropping and resizing to facilitate uniformity, resulting in a standardized resolution 512 × 512. In our experimental setup, the Palm

Keypoint Localization Neural Network (PKLNet) [18], renowned for its exceptional performance in ROI localization, was employed.

4.1. ROI Harmonization

4.1.1. Cross-Dataset Harmonization

In a real-world scenario, there can be substantial discrepancies between H and R_{Atk}. Cross-dataset harmonization was executed to emulate this environment, requiring only 30 iterations to harmonize an individual image. The content images are intentionally rendered in grayscale to simulate potential variations in R_{Atk}. The assessment of harmonized images involves two metrics: $\mathcal{L}_{lpips}$ for evaluating visual quality and texture loss $\mathcal{L}_T$ for assessing the preservation of textures. The visual outcomes of the harmonization process are illustrated in Figure 9, while the corresponding quantitative results for $\mathcal{L}_{lpips}$ and $\mathcal{L}_T$ are provided in Table 2 and Table 3, respectively.

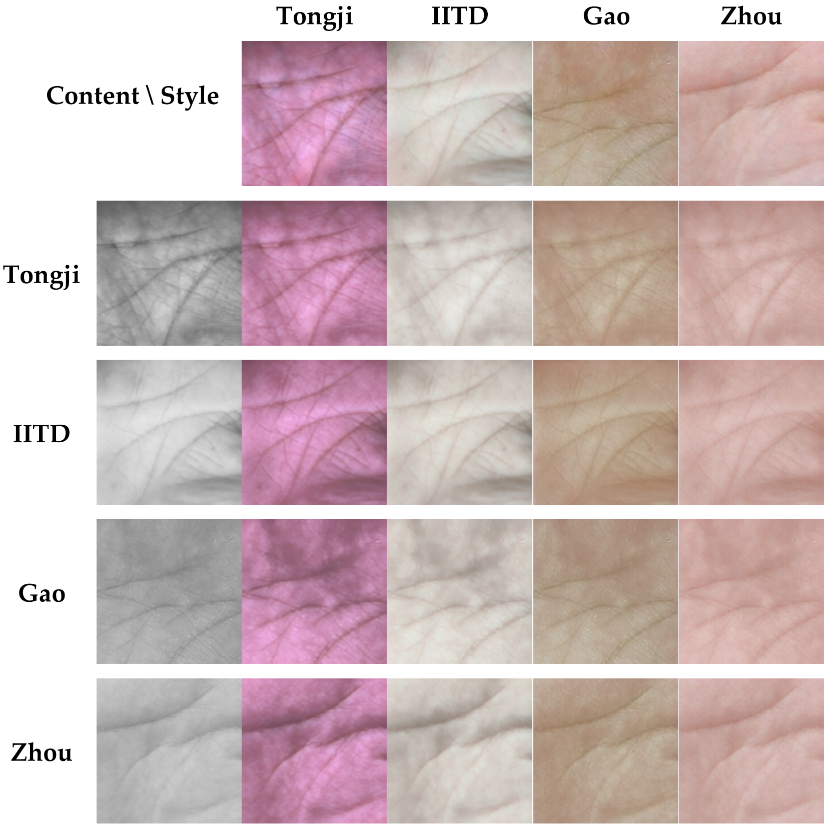


Figure 9. Cross-dataset harmonized images result.

Table 2. Average Lpips score between style images and harmonized images, the lower the better.

Content\Style	Tongji	IITD	Gao	Zhou
Tongji	0.117 ± 0.060	0.355 ± 0.066	0.336 ± 0.075	0.326 ± 0.097
IITD	0.273 ± 0.045	0.124 ± 0.084	0.363 ± 0.081	0.330 ± 0.097
Gao	0.292 ± 0.052	0.400 ± 0.071	0.203 ± 0.093	0.342 ± 0.103
Zhou	0.282 ± 0.057	0.367 ± 0.068	0.345 ± 0.081	0.142 ± 0.081

Table 3. Average texture loss between content images and harmonized images, the lower the better.

Content\Style	Tongji	IITD	Gao	Zhou
Tongji	0.029 ± 0.011	0.055 ± 0.017	0.062 ± 0.017	0.055 ± 0.019
IITD	0.044 ± 0.015	0.020 ± 0.011	0.040 ± 0.010	0.042 ± 0.015
Gao	0.052 ± 0.016	0.037 ± 0.009	0.028 ± 0.010	0.044 ± 0.016
Zhou	0.047 ± 0.017	0.040 ± 0.016	0.042 ± 0.015	0.024 ± 0.012

As depicted in Figure 9, our ROI harmonization method consistently produces remarkable visual results irrespective of the content or style employed. The outcomes presented in Table 2 indicate that harmonized images adopting the Tongji style outperform other styles, while those harmonized with the IITD style exhibit slightly inferior results. This discrepancy can be attributed to the robust color profile of Tongji images, facilitating more effective transfer, whereas the lighter color of IITD images poses challenges for seamless transfer. Moreover, as detailed in Table 3, all harmonized images exhibit lower $\mathcal{L}_T$, signifying effective preservation of texture across the harmonization process.

4.1.2. Ablation Study

Experiments have been devised to elucidate the impact of iteration and losses. The Zhou dataset, characterized by diverse image conditions, was selected for the ensuing ablation study to provide comprehensive insights. The visual results across different iterations are depicted in Figure 10, and the corresponding evaluation results are shown in Figure 11.

In Figure 10, images from iteration 0 are deemed unsatisfactory, while those from iteration 30 demonstrate a significant enhancement, approaching the quality of the original images. Subsequent iterations yield only marginal improvements. Figure 11 reveals that all losses' elbow points are reached before iteration 30, which means that the mean and std of all losses are relatively lower, and there is not much room for dropping after this point. Considering effectiveness and efficiency, we designate 30 iterations as the optimal parameter for all experiments.

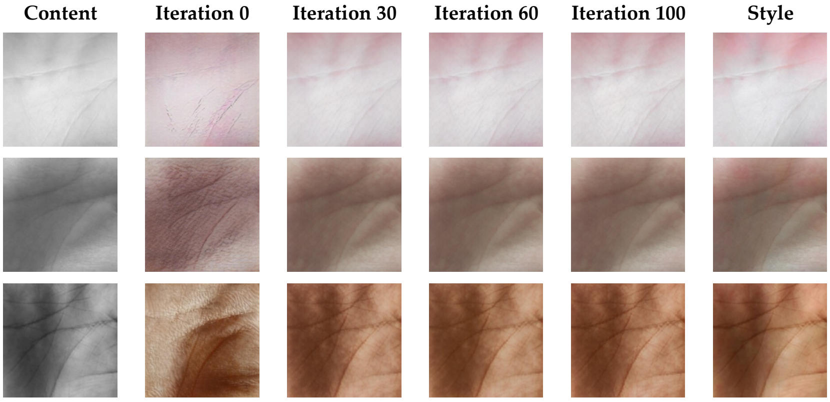
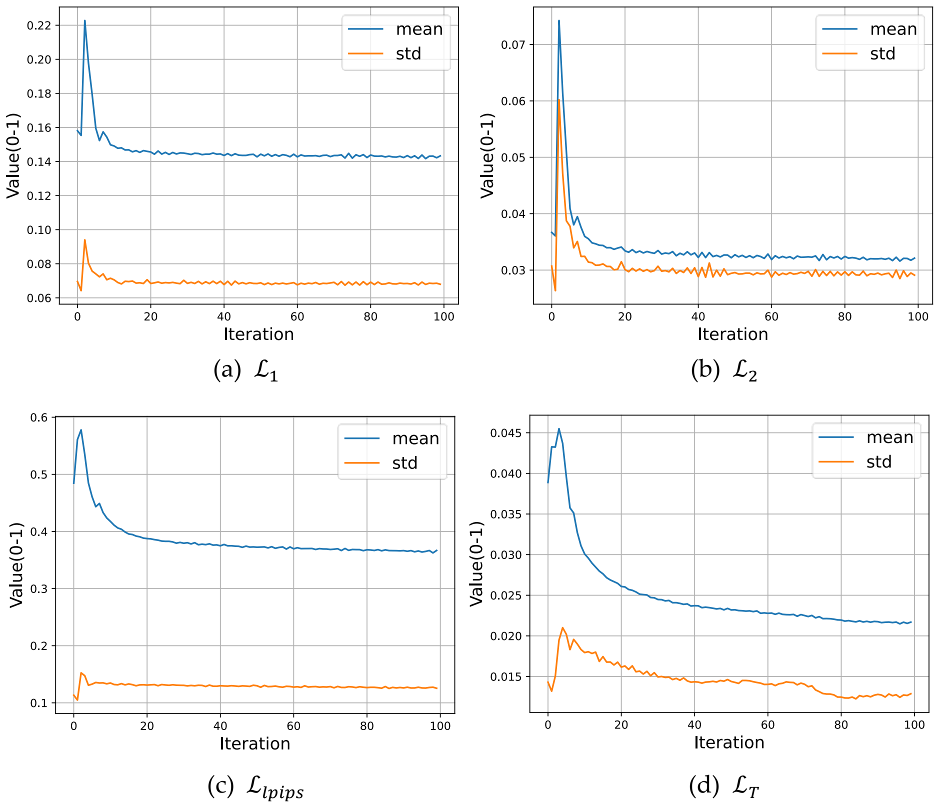


Figure 10. Harmonized images in different iterations.

Building upon the findings at iteration 30, an exploration into the impact of $\mathcal{L}_{cycle}$ was conducted. The visual outcomes are presented in Figure 12, while a comprehensive evaluation is enumerated in Table 4. The results indicate that the inclusion of additional $\mathcal{L}_{lpips}$ and $\mathcal{L}_2$ significantly enhances both visual quality and realism compared to the original cycle loss. Moreover, our devised texture loss exhibits the ability to preserve more texture while imparting slight alterations to visual quality.



(a) $\mathcal{L}_1$

(b) $\mathcal{L}_2$

(c) $\mathcal{L}_{lpips}$

(d) $\mathcal{L}_T$

Figure 11. The L1, L2, Lpips, and texture loss in different iterations.

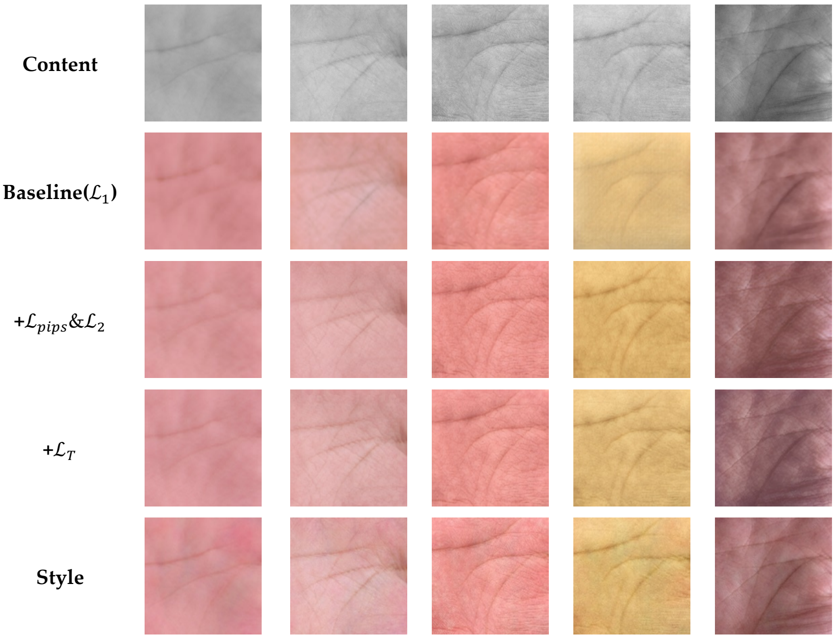

Content
Baseline($\mathcal{L}_1$)
+$\mathcal{L}_{pips}$&$\mathcal{L}_2$
+$\mathcal{L}_T$
Style

Figure 12. The images of ablation study of cycle losses.

Table 4. The evaluation results of ablation study of cycle losses; the bolded values are the best.

Setting	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_{lpips}$	$\mathcal{L}_T$
Baseline($\mathcal{L}_1$)	0.145 ± 0.070	0.034 ± 0.030	0.496 ± 0.140	0.026 ± 0.013
$+\mathcal{L}_2 + \mathcal{L}_{lpips}$	0.145 ± 0.066	**0.033 ± 0.028**	**0.381 ± 0.120**	0.027 ± 0.016
$+\mathcal{L}_T$	**0.144 ± 0.070**	0.033 ± 0.030	0.384 ± 0.128	**0.024 ± 0.013**

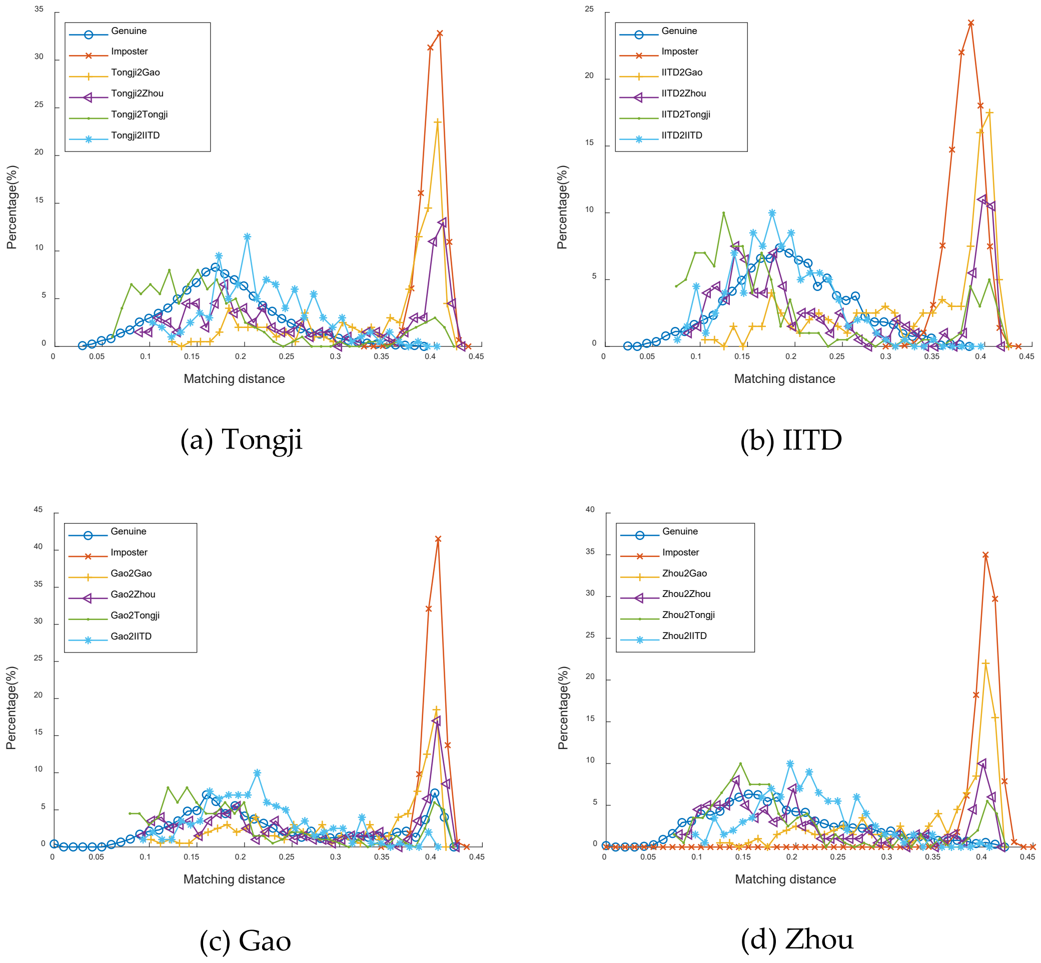
4.2. ROI Embedding Attack

To assess the performance of our ROI embedding attack, we employ the Binary Orientation Co-occurrence Vector (BOCV) [10] as the recognition method, and we utilize PKLNet [18] to extract all ROI. The success threshold for the attack is established based on the Equal Error Rate (EER) calculated using BOCV for each dataset. EER is determined by utilizing all data from the original dataset to calculate the EER specifically for their ROIs.

The attack's results are presented in Table 5, displaying the attack success rate for each scenario. Additionally, Figure 13 illustrates the distribution of matching distances. In Table 5, the first column represents the attacking ROI, and the last column denotes the corresponding dataset's EER. In Figure 13, datasetA**2**datasetB signifies using ROI from datasetA as the attacking ROI while employing datasetB as the carrier hand. The grayscale ROIs from each dataset serve as both the attacking and target ROIs.

Table 5. The attack success rate of ROI embedding attack (%).

ROI	Tongji	IITD	Gao	Zhou	EER
Tongji	88	98.5	36	63.5	0.29
IITD	82.5	99	41.5	68	1.09
Gao	88	97.5	65	69.5	14.5
Zhou	85.5	99	44	77.5	2.14



(a) Tongji

(b) IITD

(c) Gao

(d) Zhou

Figure 13. The distribution of ROI embedding attack in four datasets (ROI**2**Hand).

The efficacy of the attack diminishes notably when Gao and Zhou serve as the carrier hands. This can be attributed to the diverse and intricate environments associated with Gao and Zhou, which may introduce complexities affecting the accuracy of PKLNet ROI localization. In contrast, the attack significantly improves performance when Tongji and IITD act as the carrier hands.

Figure 14 presents the composited hand images, visually representing the attack scenarios. The quality of these composited hand images is quantitatively assessed by measuring the disparity between the composited hands and the carrier hands. Evaluation metrics such as the peak signal-to-noise ratio (PSNR) and $\mathcal{L}_{lpips}$ are employed, and the corresponding results are presented in Tables 6 and 7.

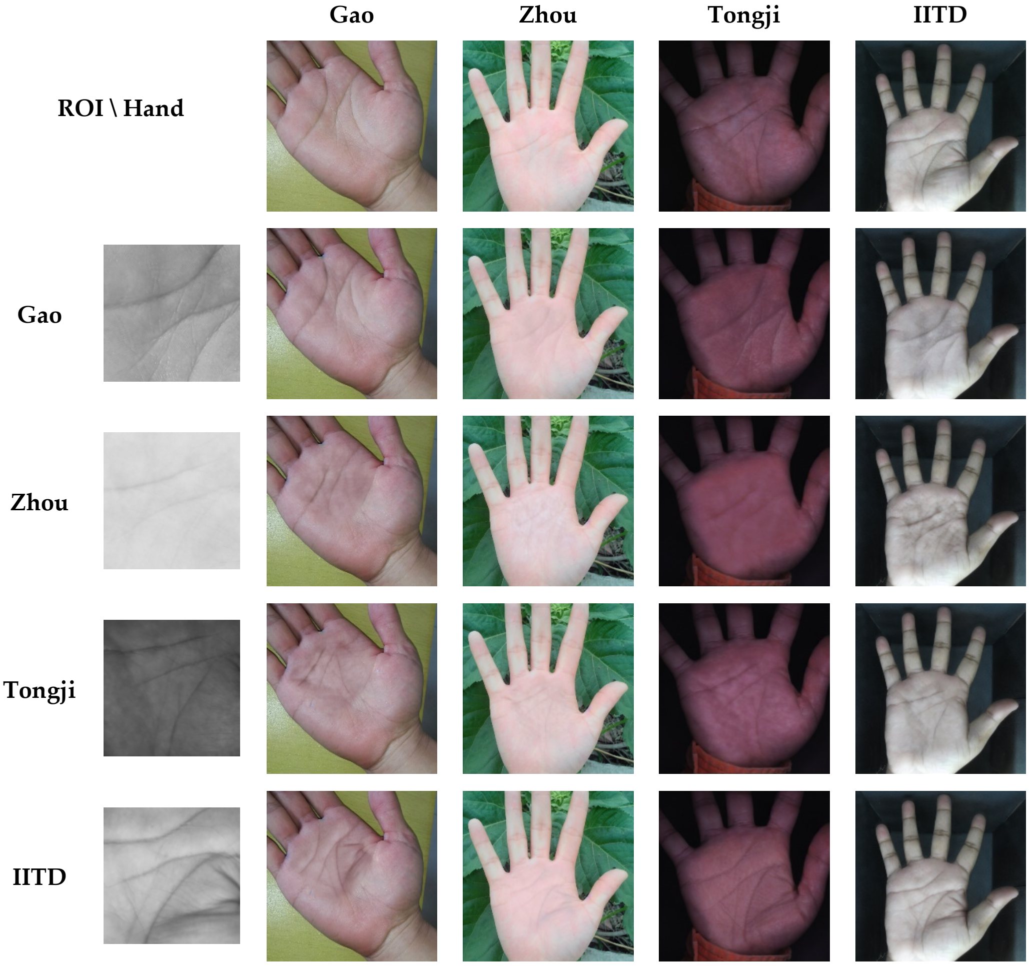


Figure 14. The composited hand images of four datasets.

Table 6. PSNR of composited hand images, the higher the better.

ROI\Hand	Tongji	IITD	Gao	Zhou
Tongji	30.96 ± 2.78	28.42 ± 1.96	29.13 ± 1.61	25.74 ± 3.47
IITD	25.38 ± 2.73	32.09 ± 1.53	28.83 ± 1.61	25.97 ± 3.26
Gao	25.70 ± 2.66	27.92 ± 2.07	29.83 ± 1.69	26.10 ± 3.41
Zhou	24.38 ± 2.83	27.89 ± 2.21	28.69 ± 1.72	29.34 ± 3.06

Table 7. Lpips of composited hand images, the lower the better.

ROI\Hand	Tongji	IITD	Gao	Zhou
Tongji	0.085 ± 0.059	0.057 ± 0.014	0.067 ± 0.030	0.110 ± 0.067
IITD	0.129 ± 0.066	0.037 ± 0.011	0.070 ± 0.028	0.108 ± 0.064
Gao	0.137 ± 0.061	0.062 ± 0.017	0.054 ± 0.022	0.112 ± 0.066
Zhou	0.140 ± 0.061	0.061 ± 0.017	0.073 ± 0.032	0.073 ± 0.060

Regarding image quality, Figure 14 visually presents realistic outcomes wherein an attacking ROI seamlessly integrates into a distinct carrier hand, producing a natural visual effect. Furthermore, the evaluation results in Tables 6 and 7 corroborate these findings, indicating a favorable and satisfactory outcome.

5. Conclusions and Future Works

This paper introduced an end-to-end, training-free method for hand image composition designed for ROI embedding attacks. This method leverages a modified style transfer approach, demonstrating notable success in achieving harmonized ROI images within a few iterations. The intuitive and effective utilization of a pretrained inpainting model to regenerate the area surrounding the embedding position contributes to the robustness of our framework. Notably, our approach requires only an ROI image and a hand image for execution, eliminating the need for training and enhancing practicality. In contrast with earlier approaches in palmprint attacks, our research introduces a novel full-hand attack strategy that builds upon an existing palmprint ROI. This innovative approach extends the applicability of attacks to more realistic environments, addressing a previously overlooked aspect. Furthermore, our method presents an intuitive and effective means of synthesizing a realistic composite hand image. Notably, this idea exhibits universality across different styles and achieves enhanced precision in texture preservation compared to conventional image composition methods. It is essential to highlight a key observation derived from our methodology: the palmprint recognition system is susceptible to significant threats even when exposed to minor and incomplete information leaks, such as an incomplete ROI. Moving forward, we will focus on devising a more efficient method that eliminates the need for Region of Interest (ROI) localization information from the target system. Additionally, our approach can be extended to a data augmentation strategy, synthesizing highly realistic full-hand images to enhance the performance of full-hand recognition methods. Nevertheless, we acknowledge that there is still ample room for improvement in our work. ROI harmonization is not so sensitive to the light color style and is susceptible to strong stains. In addition, the attack result should have a higher performance. However, we will aim to explore ways to enhance its efficacy in the future.

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

Funding: This research was supported by the National Natural Science Foundation of China (61866028), the Technology Innovation Guidance Program Project (Special Project of Technology Cooperation, Science and Technology Department of Jiangxi Province) (20212BDH81003), and the Innovation Foundation for Postgraduate Students of Nanchang Hangkong University (YC2022-S765).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy issues.

Acknowledgments: Thank you to the reviewers who reviewed this paper and the MDPI editor who edited it professionally.

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

Abbreviations

The following abbreviations are used in this manuscript:

PPR	Palmprint recognition
ROI	Region of interest
CAST	Contrastive arbitrary style transfer
CNN	Convolutional neural networks
GANs	Generative adversarial networks
PKLNet	Palm keypoint localization neural network
BOCV	Binary orientation co-occurrence vector
EER	Equal error rate
PSNR	Peak signal-to-noise ratio
$\mathcal{P}(\cdot)$	Preprocessing procedure of the PPR system
H	Full hand image
C	Carrier hand image
P	Embedding position
$\mathcal{E}(\cdot,\cdot,\cdot)$	Proposed ROI embedding technique
R_{Atk}	Attacking ROI
R_{Ref}	Reference ROI
H_{Atk}	Composited hand image

References

1. Wang, F.; Leng, L.; Teoh, A.B.J.; Chu, J. Palmprint false acceptance attack with a generative adversarial network (GAN). *Appl. Sci.* **2020**, *10*, 8547. [CrossRef]
2. Sun, Y.; Leng, L.; Jin, Z.; Kim, B.G. Reinforced palmprint reconstruction attacks in biometric systems. *Sensors* **2022**, *22*, 591. [CrossRef]
3. Yang, Z.; Leng, L.; Zhang, B.; Li, M.; Chu, J. Two novel style-transfer palmprint reconstruction attacks. *Appl. Intell.* **2023**, *53*, 6354–6371. [CrossRef]
4. Zhang, Y.; Tang, F.; Dong, W.; Huang, H.; Ma, C.; Lee, T.Y.; Xu, C. Domain enhanced arbitrary image style transfer via contrastive learning. In Proceedings of the ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada, 7–11 August 2022; pp. 1–8.
5. Yang, B.; Gu, S.; Zhang, B.; Zhang, T.; Chen, X.; Sun, X.; Chen, D.; Wen, F. Paint by example: Exemplar-based image editing with diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18381–18391.
6. Zhang, D.; Kong, W.K.; You, J.; Wong, M. Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1041–1050. [CrossRef]
7. Kong, A.K.; Zhang, D. Competitive coding scheme for palmprint verification. In Proceedings of the 17th International Conference on Pattern Recognition—ICPR 2004, Cambridge, UK, 26 August 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 1, pp. 520–523.
8. Jia, W.; Huang, D.S.; Zhang, D. Palmprint verification based on robust line orientation code. *Pattern Recognit.* **2008**, *41*, 1504–1513. [CrossRef]
9. Xu, Y.; Fei, L.; Wen, J.; Zhang, D. Discriminative and robust competitive code for palmprint recognition. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *48*, 232–241. [CrossRef]
10. Guo, Z.; Zhang, D.; Zhang, L.; Zuo, W. Palmprint verification using binary orientation co-occurrence vector. *Pattern Recognit. Lett.* **2009**, *30*, 1219–1227. [CrossRef]
11. Liang, X.; Yang, J.; Lu, G.; Zhang, D. Compnet: Competitive neural network for palmprint recognition using learnable Gabor kernels. *IEEE Signal Process. Lett.* **2021**, *28*, 1739–1743. [CrossRef]
12. Yang, Z.; Huangfu, H.; Leng, L.; Zhang, B.; Teoh, A.B.J.; Zhang, Y. Comprehensive competition mechanism in palmprint recognition. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 5160–5170. [CrossRef]
13. Wu, T.; Leng, L.; Khan, M.K. A multi-spectral palmprint fuzzy commitment based on deep hashing code with discriminative bit selection. *Artif. Intell. Rev.* **2023**, *56*, 6169–6186. [CrossRef]
14. Fei, L.; Zhao, S.; Jia, W.; Zhang, B.; Wen, J.; Xu, Y. Toward efficient palmprint feature extraction by learning a single-layer convolution network. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *32*, 9783–9794. [CrossRef] [PubMed]
15. Bao, X.; Guo, Z. Extracting region of interest for palmprint by convolutional neural networks. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
16. Gao, F.; Cao, K.; Leng, L.; Yuan, Y. Mobile palmprint segmentation based on improved active shape model. *J. Multimed. Inf. Syst.* **2018**, *5*, 221–228.

17. Matkowski, W.M.; Chai, T.; Kong, A.W.K. Palmprint recognition in uncontrolled and uncooperative environment. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 1601–1615. [CrossRef]
18. Liang, X.; Fan, D.; Yang, J.; Jia, W.; Lu, G.; Zhang, D. PKLNet: Keypoint localization neural network for touchless palmprint recognition based on edge-aware regression. *IEEE J. Sel. Top. Signal Process.* **2023**, *13*, 662–676. [CrossRef]
19. Izadpanahkakhk, M.; Razavi, S.M.; Taghipour-Gorjikolaie, M.; Zahiri, S.H.; Uncini, A. Deep region of interest and feature extraction models for palmprint verification using convolutional neural networks transfer learning. *Appl. Sci.* **2018**, *8*, 1210. [CrossRef]
20. Li, Z.; Liang, X.; Fan, D.; Li, J.; Zhang, D. BPFNet: A unified framework for bimodal palmprint alignment and fusion. In Proceedings of the Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, 8–12 December 2021; Proceedings, Part VI 28; Springer: Berlin/Heidelberg, Germany, 2021; pp. 28–36.
21. Niu, L.; Cong, W.; Liu, L.; Hong, Y.; Zhang, B.; Liang, J.; Zhang, L. Making images real again: A comprehensive survey on deep image composition. *arXiv* **2021**, arXiv:2106.14490.
22. Lee, D.; Liu, S.; Gu, J.; Liu, M.Y.; Yang, M.H.; Kautz, J. Context-aware synthesis and placement of object instances. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 10414–10424.
23. Volokitin, A.; Susmelj, I.; Agustsson, E.; Van Gool, L.; Timofte, R. Efficiently detecting plausible locations for object placement using masked convolutions. In Proceedings of the Computer Vision–ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; Proceedings, Part IV 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 252–266.
24. Cong, W.; Tao, X.; Niu, L.; Liang, J.; Gao, X.; Sun, Q.; Zhang, L. High-resolution image harmonization via collaborative dual transformations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18470–18479.
25. Hang, Y.; Xia, B.; Yang, W.; Liao, Q. Scs-co: Self-consistent style contrastive learning for image harmonization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 19710–19719.
26. Zhang, L.; Wen, T.; Shi, J. Deep image blending. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Online, 2–5 March 2020; pp. 231–240.
27. Zhang, H.; Zhang, J.; Perazzi, F.; Lin, Z.; Patel, V.M. Deep image compositing. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikola, HI, USA, 3–8 January 2021; pp. 365–374.
28. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
29. Zhang, L.; Li, L.; Yang, A.; Shen, Y.; Yang, M. Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach. *Pattern Recognit.* **2017**, *69*, 199–212. [CrossRef]
30. Kumar. Iit delhi Touchless Palmprint Database Version 1.0. 2009. Available online: https://www4.comp.polyu.edu.hk/csajaykr/IITD/Database_Palm.htm (accessed on 14 July 2023).
31. Gao, F.M. Research on the Palmprint Authentication Algorithm of Mobile Terminal Assisted by Two Lines and One Point. Master's Thesis, Nanchang Hangkong University, Nanchang, China, 2019.
32. Zhou, Z.; Chen, Q.; Leng, L. Key point localization based on intersecting circle for palmprint preprocessing in public security. *J. Def. Acquis. Technol.* **2019**, *1*, 24–31. [CrossRef]

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.