Performance Evaluation of Latent Fingerprint Enhancement and its perspectives

Ning Fang, Haowei Ren, Fangsheng Ye, Xinwei Liu* Zhejiang Wanli University, Ningbo, Zhejiang, China Zhejiang Wanli University (ZWU) Ningbo, Zhejiang, China Corresponding author: xinwei_liu@yeah.net

*--***ABSTRACT***--*

This paper analyzes the impact of existing image enhancement techniques on latent fingerprint recognition, scores different datasets and their enhanced recognition results with an evaluation tool, summarizes and reviews existing datasets and related techniques, and analyzes the reasons for their mixed results. It evaluates the effects of different enhancement models, such as FingerGAN, on both public fingerprint datasets and newly compiled latent datasets, namely the MUST and LFIW databases. Using metrics like GMean, GSTD, AUC, and EER, the paper compares the recognition results before and after enhancement to determine the effectiveness of these techniques. The findings suggest that FingerGAN significantly improves recognition rates for latent fingerprints of poorer quality, while it has a mixed or negative impact on higher quality datasets. The analysis highlights the potential and challenges of enhancing latent fingerprints, especially in complex real-world scenarios.

Keywords - **fingerprint recognition, latent in the wild, performance evaluation**

Date of Submission: October 13, 2024 Date of Acceptance: December 04, 2024

I. INTRODUCTION

 $\mathbf F$ ingerprints have long been a cornerstone of human identity verification and matching. For common use, fingerprints are generally obtained from optical and capacitive capture devices, both of which have more accurate fingerprint capture quality. Normal fingerprint-matching technology has become more mature and has been widely used in various industries for security and authentication. However, latent fingerprints have always been a challenge.

In the most widespread fingerprint detection method, the algorithm determines the fingerprint minutiae by detecting the endpoints of the fingerprint ridges and the bifurcation points of the fingerprint ridges. The fingerprint minutiae is saved as a template after noise reduction, de-duplication, and merging of proximity points. After alignment, the fingerprint minutiae is saved as a template by comparing the direction of the minutiae with the position of the minutiae in the template. After the alignment, the direction and position of the minutiae in the template can be compared to determine whether the fingerprints match or not according to the minutiae.

In criminal investigations, fingerprints are often lifted from surfaces such as objects and walls, where their quality is typically inferior to that of fingerprints obtained through specialized equipment. Compared to normal fingerprints, latent fingerprints have less usable area, more background noise, carry item texture, have discontinuous recognizable areas, and possess unclear structures. Due to the above factors, the recognition rate of latent fingerprints will be much lower than that of ordinary fingerprints, and this undoubtedly slows down the speed of identity verification.

Latent fingerprint detection is common in many contexts, yet most available models focus primarily on exact fingerprints as used in industry or research settings. To address this gap, we have compiled and evaluated several existing models using a widely recognized benchmark. Our evaluation was conducted on both public fingerprint datasets and a new set of latent fingerprint datasets. Additionally, we employed an existing latent fingerprint enhancement model to improve the quality of latent fingerprints and compared the enhanced results to the original scores. The primary focus of this paper is on the enhancement methods and the subsequent results they yield.

The main contributions of this paper are summarized as follows.

- 1. **Evaluation of Latent Fingerprint Enhancement Techniques**: it provides a comprehensive evaluation of various latent fingerprint enhancement techniques, comparing their effectiveness across different image quality levels. It systematically analyzes methods based on their ability to improve fingerprint clarity, ridge detail, and minutiae extraction under various conditions.
- 2. **Benchmarking of Performance Metrics**: it introduces and applies a set of standardized performance metrics (e.g. *FingerGAN*[1]-[4]) to assess the quality of enhanced fingerprints. These metrics likely include factors like image sharpness, contrast, accuracy of minutiae detection, and the ability to recover fingerprint details from noisy or partial prints.
- 3. **Comparison with State-of-the-Art Methods**: it compares current enhancement techniques against the most advanced state-of-the-art methods. By doing so, it highlights the strengths and weaknesses of each approach, providing a clear picture of the technological landscape and the practical limitations of current fingerprint enhancement technologies.
- 4. **Perspectives on Future Research Directions**: Beyond evaluation, the paper offers perspectives on future research needs in the field. This could involve identifying gaps in current methods, suggesting new techniques or technologies (e.g., deep learning-based approaches), and proposing ways to address challenges like handling low-quality or compromised latent prints.

5. **Practical Implications for Forensic Applications**: The study's findings have direct implications for forensic science, particularly in improving the reliability of latent fingerprint matching in criminal investigations. By enhancing fingerprint quality, the paper contributes to advancing the field of biometric identification, particularly in the context of real-world forensic challenges.

The rest of this paper is organized as follows. Section II provides background information on related techniques. Section III describes the experimental setups in detail. Section IV presents and discusses the experimental results. Finally, the paper is concluded in Section V.

II. BACKGROUND

Evaluating potential fingerprint enhancements will utilize many existing techniques and databases. This cannot be accomplished without the support of these techniques.

A. Latent Fingerprint Recognition Techniques

Neurotechnology[5] is a long history company and make a fast, accurate fingerprint evaluate model. This model has implemented a "latent mode" to match latent fingerprint that is the more usable latent fingerprint detection method. We evaluate this model in two datasets and their enhanced copy.

B. Latent Fingerprint Enhance Techniques

FingerGAN[1]-[4] is a project that utilizes adversarial generative neural networks for latent fingerprint enhancement. The detail information is optimized by means of regularization and weighting to improve the recognition rate of latent fingerprints.

Generative Adversarial Networks (GAN)[6] is a type of machine learning model. Ian Goodfellow and his team introduced GANs in 2014. A GAN consists of two neural networks: a Generator and a Discriminator. The Generator creates data that mimics real examples. The Discriminator checks if the data is real or fake.

In image processing, GANs are powerful. They learn complex patterns in data and can create high-quality images. They are used in image translation, style transfer, and image enhancement. For fingerprint enhancement, GANs improve the details of latent fingerprints and help increase recognition accuracy.

GANs face challenges like mode collapse, where they produce limited varieties of images, and instability in training. Recent techniques have improved their stability and output quality.

C. Latent Fingerprint Databases

1) Multi-Surface Multi-Technique (MUST) Latent Fingerprint Database

Multi-Surface Multi-Technique (MUST) Latent Fingerprint Database[7] is a massive Multi-Surface latent fingerprint dataset that consists of more than 16,000 latent fingerprint

impressions from 120 unique classes (120 fingers from 12 participants). Through its fingerprint sampling of participating samples on up to 35 different surfaces, a highquality dataset of complex latent fingerprints was constructed for new latent fingerprint identification studies. This paper primarily evaluates the performance of the enhancement model by assessing the recognition accuracy of its database before and after enhancement.

2) Latent Fingerprint in the Wild (LFIW) Database Latent Fingerprint in the Wild (LFIW) Database[8] is a brand-new latent fingerprint dataset that complements the MUST[7] database. It provides a large quantity of fingerprint data from different samples in specific contexts, making up for the deficiency of the MUST[7] database in terms of sample size. The LFIW[8] database consists of 13,180 fingerprint samples from 132 subjects, collected under various real-world conditions. It includes the largest number of unique fingerprint instances among existing databases, offering a unique resource for assessing the performance of fingerprint recognition systems in challenging and diverse scenarios. This paper primarily evaluates the performance of the enhancement model by assessing the recognition accuracy of its database before and after enhancement.

D. Evaluate performance Techniques

PyEER[9]-[11] is a python package intended for biometric systems performance evaluation. It has been developed with the idea of providing researchers and the scientific community in general with a tool to correctly evaluate and report the performance of their systems [12].

III. EXPERIMENTAL SETUPS

A. General

Set up FingerGAN[1]-[4] and use the pre-trained model it provides. Compile its corresponding matcher. Then in order for the image to be read correctly by the model, we need to convert the image to 8-bit grayscale format so that the preparation is complete. And we use PyEER[9]-[11] to evaluate the performance and matching rate of the models and generate reports that have been compared.

We used the MUST[7] dataset and the LFIW[8] dataset and their augmented versions, considering that the augmented LFIW[8] format color is different from MUST[7], we introduced a version with the same base color as the MUST[7] dataset after inversion of the colors, so that there are five datasets in total.

Then, we followed the following formula to invert the image to get a consistent background color in enhanced images.

$$
I_{\text{new}}(x,y) = 255 - I(x,y)
$$

Finally, we identified them using the evaluation library provided by Neurotechnology, obtained the finalize outputs and analyzed them using PyEER[9]-[11].

B. Intruduce Datasets

The types of LFIW[8] images employed in this study include the following:

- Opt-N1/N2: Reference fingerprints obtained from optical
- sensors across two distinct sessions.
- Capa-N1/N2: Reference fingerprints obtained from capacitive sensors across two distinct sessions.
- Wall: Latent fingerprints captured from wall surfaces.
- iPad: Latent fingerprints captured from iPad surfaces.
- Metal: Latent fingerprints captured from aluminum foil surfaces.

The types of MUST[7] images employed in this study include the following:

- Black dustbin bag $+$ Super glue: Latent fingerprints on black dustbin bag developed with super glue, photographed and scanned.
- Black tape + White wetwop: Latent fingerprints on black tape developed with white wetwop, photographed and scanned.
- Brown cardboard + Ninhydrin: Latent fingerprints on brown cardboard developed with ninhydrin, photographed under normal and UV light, and scanned.
- Ceramic Plate + Black powder: Latent fingerprints on ceramic plate developed with black powder, photographed, lifted with clear and frosted tape, and scanned.
- Ceramic Plate + Magnetic powder: Latent fingerprints on ceramic plate developed with magnetic powder, photographed and lifted with clear tape.
- Clear duct tape + Black wetwop: Latent fingerprints on clear duct tape developed with black wetwop, photographed and scanned with white background.
- Glass bottle + Powder: Latent fingerprints on glass bottle developed with white and orange fluorescent powder, photographed.

IV. EXPERIMENTS RESULTS

A. For MUST

The results from the MUST[7] dataset demonstrate significant improvements across various performance metrics following enhancement, underscoring the effectiveness of the applied techniques. The analysis reveals prominent changes, especially in metrics related to genuine and impostor score distributions, classification accuracy, and error rates.

The GMean and GSTD saw substantial increases postenhancement. GMean rose from 84.33 to 733.00, and GSTD increased from 150.34 to 728.04. This indicates that the enhancement markedly shifted the genuine score distribution in terms of central tendency and variability. The significant increase in both mean and standard deviation suggests that genuine scores became more distinguishable from impostor scores. This contributed to improved model performance. By expanding the distribution of genuine scores, the distinction between genuine and impostor matches became clearer. This is crucial for enhancing accuracy.

The Impostor Scores Distribution Mean (IMean) increased moderately from 8.92 to 13.63, while the Standard Deviation (ISTD) slightly decreased from 12.30 to 11.65.This reduced overlap with genuine scores. A lower standard deviation among impostor scores indicates greater consistency in assessing impostor matches. This helps the model better distinguish between true and false matches.

The Sensitivity Index (d') increased from 0.71 to 1.40, reflecting a significant improvement in the model's ability to differentiate genuine from impostor scores. A higher d' value indicates better separation, leading to more reliable identity matches. The Area Under the Curve (AUC) improved from 0.77 to 0.84. This shows an increased ability of the classifier to discriminate between genuine and impostor matches. This metric highlights improved overall classification capacity, demonstrating better true positive identification while minimizing false positives.

The J Index rose from 0.50 to 0.65, while the Matthews Correlation Coefficient (MCC) increased from 0.61 to 0.74. These metrics indicate an improved balance between sensitivity and specificity. Higher values signify better identification of correct matches and rejections, with fewer misclassifications. MCC, known for its robustness against class imbalance, further underscores the enhancement's effectiveness in improving model robustness.

The Equal Error Rate (EER) and related metrics (EERlow and EERhigh) all showed reductions. This points to fewer instances where the false acceptance rate (FAR) and false rejection rate (FRR) were equal. EER decreased from 0.30 to 0.24, with EERlow dropping from 0.30 to 0.22 and EERhigh decreasing from 0.31 to 0.26. Reduced EER values indicate a more accurate balance between FAR and FRR. This suggests the enhancement procedure improved the model's reliability by lowering false positive and negative rates.

FMR and FNMR metrics also declined. This reinforces that the enhancement led to fewer incorrect matches and fewer falsely rejected genuine matches. For instance, FMR1000 decreased from 0.58 to 0.37 and FMR100 from 0.51 to 0.34. This demonstrates improved precision in separating genuine and impostor matches. Moreover, ZeroFMR and ZeroFNMR metrics either remained high or showed slight improvement. This indicates enhanced model performance even under stringent zero-false thresholds.

Overall, the enhancements applied to the MUST[7] dataset substantially improved key performance metrics. These include the ability to differentiate between genuine and impostor scores, classification accuracy, and error rates. The increased GMean, decreased impostor score variability, enhanced Sensitivity Index (d'), and improved AUC collectively illustrate the enhancement's reinforcement of the model's robustness and accuracy. Moreover, lower EER and FMR values suggest improved threshold determination, further minimizing false decisions. These improvements indicate the enhancement methodology was highly effective, boosting overall performance and reliability for practical identity verification tasks.

Fig.3 Original DET

Fig.4 Enhanced DET

B. For LFIW

The results from the LFIW[8] dataset provide a nuanced view of the impact of different enhancement techniques on model performance. The analysis focuses on the original data, the enhanced version, and the enhanced (inverted) version, highlighting how each modification affected key performance metrics related to score distributions, classification accuracy, and error rates. These changes help illustrate both the strengths and the potential limitations of the applied enhancement techniques.

The Genuine Scores Distribution Mean (GMean) increased from 251.67 to 286.14 after enhancement, while the enhanced (inverted) version showed a slight increase to 254.88. The Standard Deviation (GSTD) of the genuine scores also rose, from 309.98 to 458.84 for the enhanced version and to 393.19 for the inverted version. These changes indicate that the enhancement process led to a broader spread of genuine scores, making them slightly more distinguishable from impostor scores. However, the inverted enhancement showed a smaller increase in variability, suggesting a more conservative effect compared to the regular enhancement.

The Impostor Scores Distribution Mean (IMean) decreased from 11.36 to 10.54 after enhancement, while the enhanced (inverted) version showed a moderate decrease to 10.37. The Standard Deviation (ISTD) for impostor scores also decreased from 9.38 to 9.64 for both enhanced versions. This indicates that both types of enhancement reduced the central tendency and variability of impostor scores, helping to improve model performance by tightening the impostor score distribution and making it more distinct from genuine scores.

The Sensitivity Index (d') dropped from 1.10 to 0.85 after enhancement, but slightly increased to 0.88 for the enhanced (inverted) version. This suggests that the enhancement reduced the ability to distinguish genuine matches from impostor matches, while the inverted enhancement mitigated some of this negative impact. Lower d' values typically indicate reduced discrimination power, which highlights the trade-offs involved in these enhancement techniques.

Fig. 7: Enhanced (Invert)

The Area Under the Curve (AUC) decreased from 0.82 to 0.77 for the enhanced version and to 0.78 for the enhanced (inverted) version. This reduction suggests that the classifier's overall ability to distinguish between genuine and impostor matches was slightly compromised. The \textit{AUC} metric is a comprehensive indicator of model classification performance, and this decrease implies that enhancement had some unintended effects on the model's discrimination ability.

The Youden's J Index (J-Index) decreased from 0.63 to 0.51 for the enhanced version and to 0.52 for the enhanced (inverted) version. Similarly, the Matthews Correlation Coefficient (MCC) dropped from 0.73 to 0.66 for the enhanced version and to 0.67 for the inverted enhancement. These metrics are both measures of classification performance, with higher values indicating a better balance between sensitivity and specificity. The observed declines indicate that enhancement negatively impacted the balance between correct and incorrect matches.

The Equal Error Rate (EER) slightly increased, with EER values going from 0.30 to 0.32 after enhancement. This increase reflects a higher point of balance between false acceptance rate (FAR) and false rejection rate (FRR), which means the enhancement led to a slight decline in overall classification performance. EERlow and EERhigh also showed increases, pointing to a general trend of reduced precision post-enhancement.

FMR and FNMR values increased for various thresholds. For example, FMR1000 increased from 0.41 to 0.51 for the enhanced version and to 0.49 for the inverted version. FMR100 also rose from 0.37 to 0.48 for the enhanced version and to 0.47 for the inverted version. This demonstrates that the enhanced versions had a higher tendency to incorrectly accept impostor matches. However, metrics like ZeroFMR and ZeroFNMR remained high, indicating that the model still performed well under strict thresholds despite the increase in false match rates.

The enhancements applied to the LFIW[8] dataset showed mixed results. On one hand, the genuine score distribution was broadened, which could theoretically make it easier to

distinguish genuine matches from impostor matches. On the other hand, decreases in key metrics like AUC, J-Index, MCC, and increases in EER suggest that the enhancement techniques may have introduced new challenges for classification accuracy. The enhanced (inverted) version appeared to mitigate some negative impacts but did not fully address the decline in performance. These findings suggest that while enhancement can improve certain aspects of score distribution, careful consideration MUST[7] be given to the specific effects on classification accuracy and the balance between false matches and false non-matches. A more tailored enhancement approach may be needed to fully optimize performance for the LFIW[8] dataset.

Fig. 8: Original DET

Fig. 9: Enhanced DET

10: Enhanced (Invert) Fig. **DET**

C. Subjective Overview

LFIW Fig. 12: LFIW En- Fig. 13: LFIW En- $11:$ $Fig.$ Original hanced hanced (Invert)

By visually inspecting the enhanced images, it is clear that the enhancement process had a different impact on the dataset. In the MUST[7] sample (Fig.14), the enhanced background

was significantly cleaner, which helped to improve feature extraction and clarity. This process effectively reduces noise and removes extraneous background elements, thus improving the quality of the fingerprint features.

In contrast, the results for the LFIW[8] sample (Fig.11) are more complex. The intricate background of the original LFIW[8] image resulted in some background elements being incorrectly enhanced as fingerprint features, causing the problem of over-enhancement, i.e., irrelevant regions are enhanced along with the real fingerprint features, thus interfering with the recognition. The inverted version of the enhanced LFIW[8] sample (Fig.13) solves some of these problems by mitigating the enhancement effect so that the enhancement of the background is less pronounced. However, the complexity of the original background remains a challenge, and interference with fingerprint features remains.

Overall, the MUST[7] samples benefited more from the enhancement process due to the simpler backgrounds, while the LFIW samples showed over-enhancement in areas with complex backgrounds. This emphasizes the importance of considering the original image quality and background features when applying enhancement techniques.

V. CONCLUSION

A more rough conclusion is that, for better quality datasets, FingerGAN[1]-[4] rather deteriorates the recognition rate. But for poorer quality datasets, meaning more "latent", FingerGAN[1]-[4] can improve the recognition rate.

Therefore, when using enhancement techniques to improve recognition rates, it is important to combine them with a fingerprint scoring system that avoids enhancing higherquality fingerprints. This approach ensures more consistent and superior results. Additionally, it provides guidance for data capture: using a cleaner background plate enhances outcomes.

REFERENCES

- [1] Yanming Zhu, Xuefei Yin, and Jiankun Hu, "FingerGAN: A Constrained Fingerprint Generation Scheme for Latent Fingerprint Enhancement," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
- [2] Yanming Zhu, Xuefei Yin, Xiuping Jia, and Jiankun Hu, "Latent Fingerprint Segmentation Based on Convolutional Neural Networks," in 2017 IEEE Workshop on Information Forensics and Security (WIFS), IEEE, 2017, pp. 1-6.
- [3] Yanming Zhu, Jiankun Hu, and Jinwei Xu, "A Robust Multi-Constrained Model for Fingerprint Orientation Field Construction," in 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2016, pp. 156-160.
- [4] Yi Wang, Jiankun Hu, and Damien Phillips, "A Fingerprint Orientation Model Based on 2D Fourier Expansion (FOMFE) and Its Application to Singular-Point Detection and Fingerprint Indexing," IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), 2007, pp. 573-585.
- [5] Neurotechnology, "Fingerprint Biometrics," Neurotechnology, accessed October 13, 2024. Available: https://neurotechnology.com/fingerprint biometrics.html.
- [6] Ian Goodfellow et al., "Generative Adversarial Nets," Advances in Neural Information Processing Systems, 27, 2014.
- [7] Aakarsh Malhotra, Mayank Vatsa, Richa Singh, Keith B. Morris, and Afzel Noore, "Multi -Surface Multi - Technique (MUST) Latent Fingerprint Database," IEEE Transactions on Information Forensics and Security, 19, 2024, pp. $1041-1055$. Available: https://doi.org/10.1109/TIFS.2023.3280742.
- [8] Xinwei Liu, Kiran Raja, Renfang Wang, Hong Qiu, Hucheng Wu, Dechao Sun, Qiguang Zheng, et al., "A Latent Fingerprint in the Wild Database," IEEE Transactions on Information Forensics and Security, 19, 2024, pp. 3703 Available: https://doi.org/10.1109/TIFS.2024.3368892.
- [9] D. Maltoni et al., Handbook of Fingerprint Recognition, Springer -Verlag London Limited, 2009.
- [10] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain, "FVC2000: Fingerprint Verification Competition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3), 2002, pp. 402 -412.
- [11] J. Hernandez -Palancar, A. Munoz -Briseno, and A. Gago -Alonso, "Using a triangular matching approach for latent fingerprint and palmprint identification," International Journal of Pattern Recognition and Artificial Intelligence, 28, 2014, 1460004.
- [12] Liu, Xinwei, Yimin Song, Guoyun Ye, Min Fu, and Wen Liu. "Advanced Obstacle Detection and Distance Estimation for Forklift Operations through Integrated Deep Learning Networks." International Journal of Advanced Networking and Applications 16, no. 1 (2024): 6239 -6245.