SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

BACHELOR’S THESIS

论文题目：基于生成对抗网络的指纹图像增强算法研究

学生姓名: 黄悉偈

学生学号: 516021910433

专　　业: 测控技术与仪器

指导教师: 刘满华

学院(系): 电子信息与电气工程学院

基于生成对抗网络的指纹图像增强算法研究

摘要

指纹由于具有验证身份的唯一性，一直以来都是最重要的生物特征之一，指纹识别技术是刑侦领域侦破按键的重要手段之一，从犯罪现场采集到的现场指纹由于复杂的背景噪声，和不清晰的脊线纹理结构，导致指纹的图像质量低下，从而给现场指纹的识别带来了很大的挑战。 因此，图像增强是提高现场指纹识别准确率的关键步骤。在本文中，我们首先提出了一种基于生成式对抗网络（GAN）的指纹增强模型，并且设计了实验验证 GAN 模型在指纹增强任务上的有效性。然后，我们提出了一种改进的基于多任务学习的渐进式生成对抗网络的指纹增强方法。生成式对抗网络强大的表征能力能够很好地帮助完成从低质量指纹到高质量指纹图像的复杂变换。该方法分为两步：渐进式离线训练（POT）与迭代式在线测试（IOT）。渐进式训练使我们的模型不仅关注指纹特征点等局部特征，也保留了方向场等全局结构性特征。在迭代式在线测试阶段，原有的低质量现场指纹图像被送入网络进行迭代，直到输出的增强结果达到满意的图像质量。我们网络的训练数据基于 NIST DB 27 的背景噪声与 NIST DB 14 的指纹脊线纹理信息合成，并在 NIST DB 27 现场指纹图像集上对我们的模型进行了测试和验证。在方向场估计多任务学习与渐进式训练模式的帮助之下，我们的模型相比目前已有文献的结果,测试效率和识别准确率有了进一步提高。

关键词：指纹图像，生物特征，生成式对抗网络，多任务学习，图像增强

RESEARCH ON FINGERPRINT IMAGE ENHANCEMENT ALGORITHM BASED ON GENERATIVE ADVERSARIAL NETWORK

ABSTRACT

Fingerprints are one of the most important and reliable biometric modalities because of the characteristic that a person can be uniquely identified via it. Latent fingerprints, which are captured from the finger skin impressions at the crime scene, have been used to identify suspected criminals for a long time. However, poor latent fingerprint image quality caused by the overlapping patterns, unclear ridge structure, and various background noise has brought a great challenge to the recognition of latent fingerprints. Therefore, image enhancement is a crucial step for more accurate fingerprint recognition. In this paper, we first build a generative adversarial network-based model for latent fingerprint enhancement. We also design experiments to prove the effectiveness of GAN in latent fingerprint enhancement task. Then, a latent fingerprint enhancement method is proposed by combining the multi-task learning and progressive GAN. The powerful representative capacity of the GAN structure provides an efficient translation from the latent fingerprint image domain to the high-quality fingerprint image domain. Our multi-task progressive GAN method consists of two stages: Progressive Offline Training (POT) and Iterative Online Testing (IOT). Progressive offline training makes our model not only focus on the local features such as the minutiae but also preserve global structure features such as the orientation field. In the iterative online testing stage, the original low-quality latent fingerprint image is iteratively enhanced until the image quality is satisfied. To train our network, the training set is generated synthetically with noise patterns from NIST DB 27 dataset and high-quality fingerprints from NIST DB 14 dataset. We extensively evaluate our model on NIST DB 27 latent fingerprint dataset. With the help of the orientation estimation task and the progressive training scheme, our model achieves better recognition accuracy and efficiency than the previous state-of-the-art methods.

Key words: Fingerprint image, Biometric, Generative adversarial network (GAN), Multi-task learning, Image enhancement

Contents

Chapter One Introduction 1
1.1 Fingerprint and its characteristic 1
1.2 Automated Fingerprint Identification Systems (AFIS) 2
1.3 Deep Learning based Fingerprint Identification 4
1.4 Overview of proposed work 5
Chapter Two Related Work 8
2.1 Fingerprint Image Enhancement 8
2.2 Generative Adversarial Network (GAN) 13
2.3 Deep Learning based Image Translation 17
2.4 Chapter Summary 20
Chapter Three Fingerprint Enhancement based on Generated Adversarial Network 22
3.1 Fingerprint Image Enhancement: Problem Formulation 22
3.2 Generative Adversarial Network Architecture 23
3.3 Objective Function 25
3.4 Experimental Results 26
3.4.1 Dataset Preparation 26
3.4.2 Evaluation Metrics 29
3.4.3 Fingerprint Enhancement Results 30
3.5 Chapter Summary 33
Chapter Four Fingerprint Enhancement based on Multi-Task Progressive GAN 34
4.1 Multi-task Generative Adversarial Network Architecture 34
4.2 Objective Function 36
4.3 Progressive Training 37
4.4 Generation of Orientation Ground Truth 40
4.5 Experimental Results 42
4.5.1 Ablation Experiments 43
4.5.2 Comparisons with other methods 45
4.5.3 Efficiency 48
4.6 Chapter Summary 48
Chapter Five Conclusion and Future Work 50
REFERENCE 52
ACKNOWLEDGEMENTS 58
PUBLICATIONS 60

Chapter One Introduction

1.1 Fingerprint and its characteristic

How to identify a specific person accurately and quickly has always been an important problem and research topic. Traditional markers of human identity include knowledge-based methods such as password and objects-based methods such as driving license and ID card. In essence, these traditional human identification methods involve personal belongings identification rather than individual identification, which may result in significant drawbacks such as loss and potential theft of identification markers. As smart devices develop rapidly, these traditional methods have been gradually replaced by more reliable biometric identification.

Fingerprints are one of the most important and reliable biometric modalities because of the characteristic that a person can be uniquely identified via it. Over the whole life of a single person, the fingerprints are difficult to change dramatically, which makes it possible for long-term and stable human identification. As an instance, fingerprint recognition and identification have been widely adopted in smartphones, smart lock, and criminalistics.

The research of fingerprint identification has started from a very early stage. Edward et al. proposed the famous “Henry system” fingerprint classification system in 1899. In the system, the fingerprints are classified into 5 categories: arch, tented arch, right loop, left loop, and whorl. The establishment of the fingerprint classification system is a great progress of fingerprint research. The classification standard of the Henry System is still widely used in fingerprint recognition nowadays.

Minutiae point features are also an important feature of fingerprint ridges. Minutiae point can be categorized into delta, core, ridge ending, and bifurcation, etc. [1]. Minutiae feature based fingerprint matching algorithm is the most widely accepted algorithm. How to accurately extract minutiae features in the fingerprint image plays an important role in fingerprint identification. Missing and Spurious minutiae points can influence the matching and verification accuracy significantly.

However, the fingerprints acquired by sensors or other devices and methods are not always in good quality. Latent fingerprints, which are captured from the finger skin impressions unintentionally left at the crime scene by accident, are a representative type of low-quality fingerprint [2]. The latent fingerprint and its analysis are very useful in law and forensics applications. When they are compared to the plain and rolled fingerprints which are collected under controlled condition, latent fingerprints usually suffer from poor ridge structure and overlapping unstructured noise [3]. Fig.1.1 gives examples of latent, rolled, and plain fingerprints respectively. We can see from the examples shown here that the latent fingerprint suffers from severe corruption and noise. As a result, how to suppress the noise and recover the fingerprint pattern from corruption has been the core problem of fingerprint image enhancement.

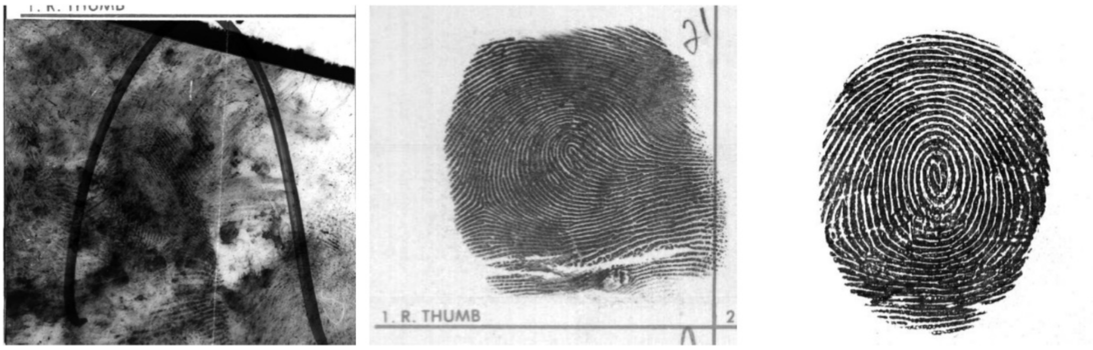


FIGURE 1.1 Examples of latent, rolled, and plain fingerprint images (from left to right), which are selected from NIST DB 27 dataset [4], NIST DB 14 dataset [5], and FVC dataset, respectively.

1.2 Automated Fingerprint Identification Systems (AFIS)

The Automated Fingerprint Identification Systems (AFIS) have been widely adopted for fingerprints identification. The AFIS usually consists of fingerprint image acquisition, fingerprint segmentation, orientation estimation, image enhancement, image binarization, minutiae extraction, fingerprint classification, and fingerprint matching [7]. The flowchart of the AFIS is shown in Fig.1.2. In a real application, some components of the AFIS may be removed or altered.

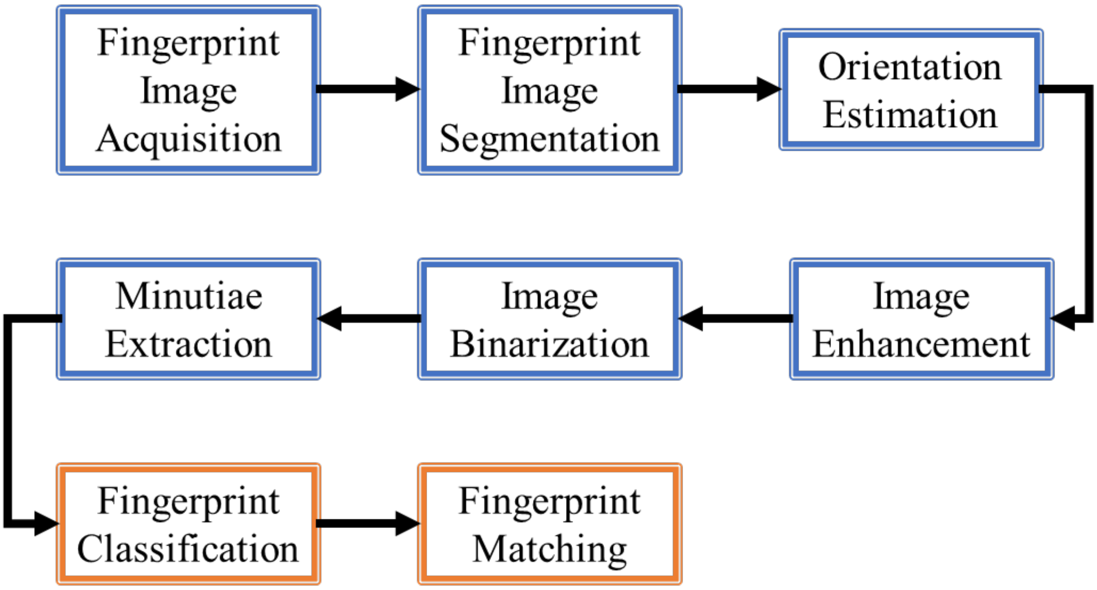


FIGURE 1.2 Flowchart of the Automated Fingerprint Identification Systems

Although AFIS achieves promising accuracy on plain and rolled fingerprints, the performance is still not satisfied for low-quality latent fingerprint images. As a result of the poor image quality of these fingerprint images, most of the common feature extraction techniques often fail to accurately extract useful features. To solve these problems, fingerprint image enhancement is the important processing step in the AFIS to remove the noise, recover the corrupted regions, and improve the clarity of the ridge structure. When low-quality fingerprint images are enhanced, a more efficient and accurate feature extraction algorithm can be facilitated for better fingerprint matching and identification performance.

Classic fingerprint image enhancement methods focus on how to filtering noise from the meaningful ridge pattern and remove it. Information in the frequency domain and orientation field is widely used because the orientation and frequency characteristics of noise are different from the fingerprint ridge pattern part, which makes it easier to enhance the useful region.

1.3 Deep Learning based Fingerprint Identification

During the past few decades, we have entered a new era and witnessed a huge revolution of artificial intelligence. Among all the breakthroughs of artificial intelligence techniques, Deep neural networks (DNNs) achieved state-of-the-art performance in many image recognition and understanding applications, such as object detection [8][9], self-driving cars [10][11] and face recognition [12][13]. Among all kinds of DNNs, Convolutional Neural Networks (CNNs) have been most widely used in image processing tasks, including image enhancement and image restoration. Since LeCun et al. [14] proposed the first CNN structure and applied it to zip code recognition, different CNNs have been emerging. After Deng et al. proposed ImageNet [15], the first large-scale image classification dataset, the research on CNN based image processing techniques has been growing at a tremendous speed. As an important breakthrough in image denoising and enhancement, Zhang et al. [16] proposed DnCNN model to handle blind Gaussian denoising with the residual learning strategy. Following the flourish of deep learning-based image denoising and image restoration, latent fingerprint enhancement based on the deep neural network has also been a hot research topic recently.

Fingerprint enhancement can be formulated as an image translation problem, where end-to-end deep learning methods can be applied. As the classic method fully utilize the information of the orientation field and the minutia, we can also develop a multi-task learning scheme to achieve better performance. In addition, fingerprint minutia extraction and fingerprint matching can also be tackled by deep learning-based methods. Graph neural network (GNN) [17] can be applied to fingerprint identification as the graph structure can represent the minutia and orientation field information well.

With the powerful capacity of the GNN based graph matching [18], the fingerprint identification performance can be boosted significantly.

1.4 Overview of proposed work

We will first build a model based on a generative adversarial network to enhance the low-quality fingerprint. As deep learning can be used to learn the mapping of image translation, GAN is a powerful tool for effective image enhancement. We will first formulate the latent fingerprint enhancement problem. To train our generative adversarial network with supervised learning, paired training data with original fingerprint images and high-quality fingerprint images should be given. This can be achieved by the synthetical generation. The enhancement effect can be evaluated by the performance of fingerprint matching. If the enhanced fingerprint can match precisely with the matched pair, we can draw a conclusion that the image quality improves. The evaluation metric we use to measure the performance of enhancement model is the Cumulative Match Characteristic (CMC).

In the next section, we will show the multi-task learning progressive GAN model for latent fingerprint enhancement. Motivated by the idea that orientation estimation and fingerprint enhancement can share representation in the deep neural network, we proposed a multi-task based progressive generative adversarial network (PGAN) model. The method we proposed can be decomposed into two stages: Progressive Offline Training (POT) and Iterative Online Testing (IOT). The deep GAN model is trained on the POT stage and then is applied to the IOT stage. The flowchart of our method is illustrated in Fig.1.3.

In the POT stage, our GAN is trained with paired data. To generate the paired data of latent fingerprint and corresponding high-quality fingerprint, similar to the data generation of our first model, we simulate the latent fingerprint synthetically combining the noise and high-quality fingerprint images. Besides, we also generate the ground truth orientation filed of the fingerprint image with classic methods. The orientation field label will be used in the training of the orientation estimation task. Another problem is that the generating performance of GAN is poor on high-resolution images and suffers from checkboard artifacts. This is because the generator of GAN follows an encoder-decoder structure, where deconvolution is applied in the decoder. To solve the problem, the progressive growth of GAN arises applied in the training process which can effectively boost the stability and performance of GAN.

In the IOT stage, the segmented latent fingerprint is input into the GAN trained in the POT stage. Generally, a fingerprint image after a single iteration of the image translation still suffers from unstructured noise. To solve this problem, we iteratively enhance the image until the fingerprint image quality achieves a satisfying result. In the latter iteration, the restored high-quality region can be a good guidance for the enhancement of remaining low-quality regions.

To show the effectiveness and efficiency of our method, we extensively evaluate our model on NIST DB 27 latent fingerprint dataset. From the Cumulative Match Characteristic (CMC) curve on NIST DB 27 and three subsets, we can see our progressive GAN based method outperforms the previous model by a great margin. The key contribution of this paper can be summarized as follows:

(1) We propose a generative adversarial network based latent fingerprint enhancement method, which consists of two stages: Progressive Offline Training (POT) and Iterative Online Testing (IOT).

(2) We propose a progressive training method to boost the stability and performance of GAN, and propose a multi-task training method to fully exploit the information of the orientation field and minutiae.

(3) We evaluate our model on NIST DB 27 latent fingerprint dataset and show the Cumulative Match Characteristic (CMC) curve on NIST DB 27 and its three subsets.

(4) We compare our model with state-of-the-art latent fingerprint enhancement methods. Our results are better than other enhancement methods in terms of both

effectiveness and efficiency.

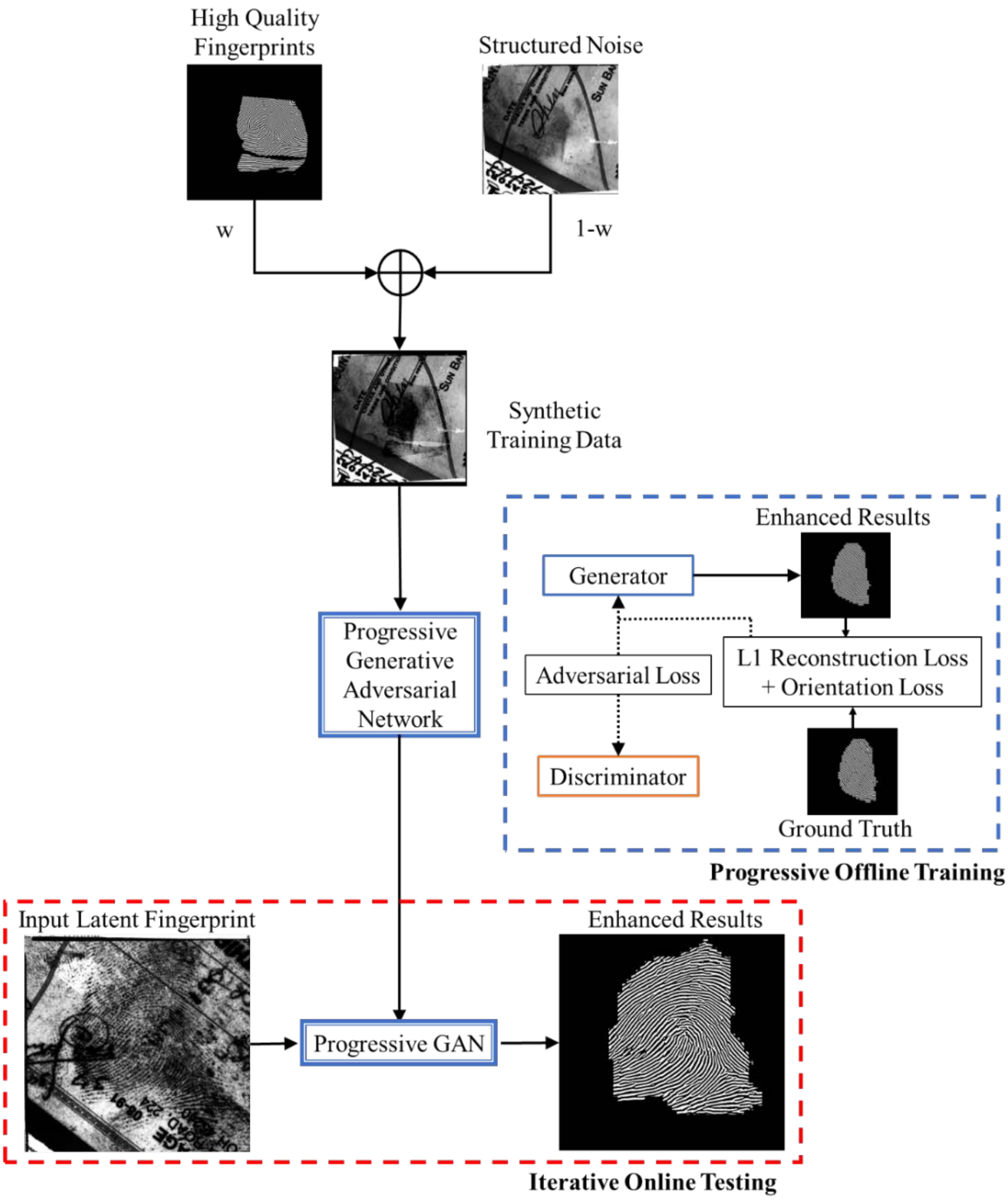


FIGURE 1.3 The flowchart of our latent fingerprint image enhancement method based on progressive generative adversarial network. Our method can be divided into two stages: Progressive Offline Training (POT) and Iterative Online Testing (IOT). Before the POT stage, paired training data are generated. In the POT stage, the dashed lines denote the loss backpropagation path in training our GAN progressively. In the IOT stage, the latent fingerprint images are enhanced by the GAN model iteratively . The number of enhanced iteration is fixed.

Chapter Two Related Work

In this chapter, we will introduce the related work of fingerprint image enhancement, generative adversarial network (GAN), and deep learning-based image translation, respectively. For fingerprint image enhancement methods, we will start from classic Gabor filter methods, then we introduce the total variation (TV) model and global dictionary based sparse representation model. In Generative adversarial network section, we show the core formula and architecture of basic GAN. We also mention some important improvements of GAN such as DCGAN, WGAN, LSGAN, and cGAN, which is the basis of GAN based image translation. Finally, we introduce some important topics of image translation including image denoising and image super-resolution. There are many deep learning based models proposed to address these problems. Latent fingerprint enhancement can be formulated as an image translating problem and these deep learning based method can spark our inspiration of effective fingerprint enhancement model.

2.1 Fingerprint Image Enhancement

To enhance the latent fingerprint images, Gabor filtering is proposed by Lin et al. [19]. As the fingerprint consists of interleaved parallel ridge and valley flow with well-defined frequency and orientation, Gabor filtering can make full use of this information. Gabor filter can be defined as a sinusoidal plane wave tapered by a Gaussian to capture the periodic and non-stationary nature of fingerprint regions, which achieve promising effects on the improvement of ridge clarity. The Gabor filter can be denoted as Equation 2-1.

$$g(x,y,\theta,f,\sigma_x,\sigma_y)=\exp\{-\frac{1}{2}[\frac{{x_\theta}^2}{{\sigma_x}^2}+\frac{{y_\theta}^2}{{\sigma_y}^2}]\}\cdot\cos(2\pi\cdot f\cdot x_\theta), \quad (2-1)$$

where

$$x_\theta = x\cdot\cos\theta + y\cdot\sin\theta, y_\theta = -x\cdot\sin\theta + y\cdot\cos\theta. \quad (2-2)$$

θ represent the rotation angle of the filter, f denotes the local frequency, σ_x, σ_y represents the standard deviation of the relative Gaussian function on x, y axis. We can see from Equation 2-1 and Equation 2-2 that the Gabor function is highly orientation and frequency selective. As a result, the performance of Gabor filtering is stable on fingerprint with homogenous local region orientation and frequency. Fig.2.1 gives an example of a Gabor filter with parameter $f=0.2, \theta=0$

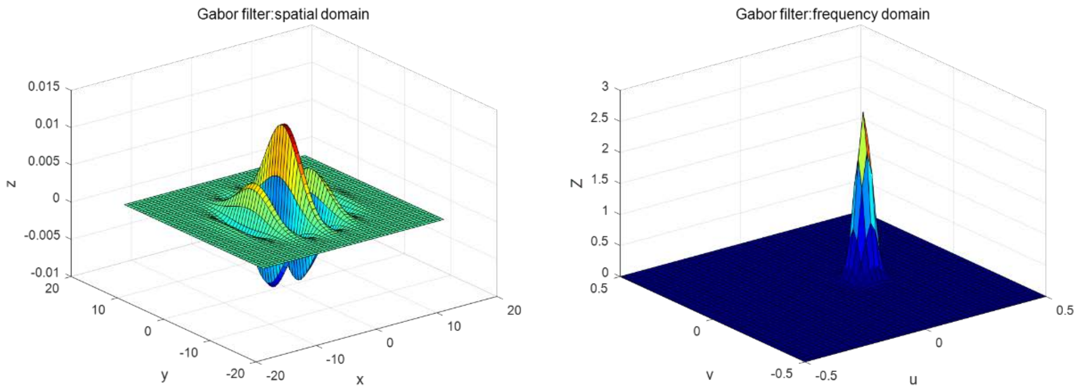


FIGURE 2.1 Spatial domain and frequency domain map of a given Gabor filter

An image can be decomposed into the cartoon part and texture part according to the total variation (TV) model. To exploit this cue, image decomposition based on the minimization of total variation has been researched to enhance the latent fingerprint images [22][23]. The core idea of the TV model is that the texture component represents a meaningful structure of the image while the cartoon component is characterized as non-repeated structured noise. Their characteristics in the frequency domain are different and easy to be divided. We can remove the cartoon component of the fingerprint image and keep the texture component as the enhanced result.

Zhang et al. proposed an adaptive directional total variation (ADTV) model [22] to improve the effectiveness of fingerprint segmentation and enhancement. Compared to the classical total variation model, this model fully exploits the unique features of the fingerprint image: scale and orientation. The model can achieve accurate segmentation and enhancement results which improve the fingerprint matching performance. The ADTV model can be formulated as

$$u^* = \arg\min_u \int |\nabla u \cdot \vec{a}(x)| dx + \frac{1}{2}\int \lambda(x)|u - f| dx, \tag{2-3}$$

Where $\lambda(x)$ represents the spatially varying feature scale parameter and $\vec{a}(x)$ denotes the orientation vector controlled by the local texture orientation. u is the cartoon layer along with all directions and the optimization target for the TV model. The Lagrangian method can be applied to solve the ADTV model formulated in Equation 2-3. Some experiment examples of the ADTV model are shown in Fig.2.2.

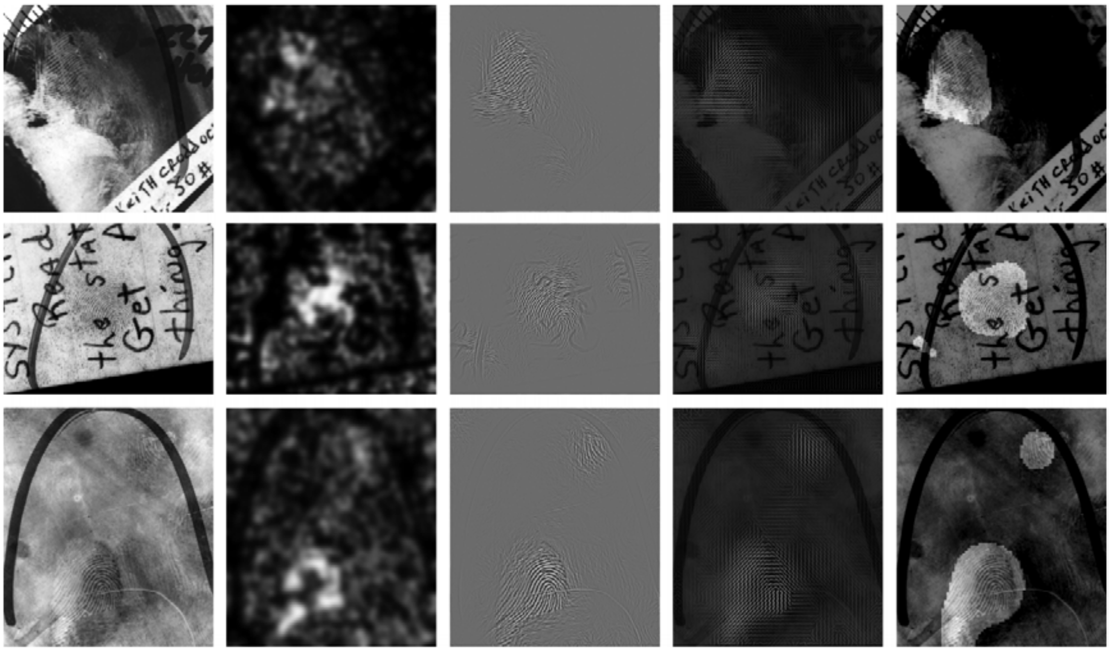


FIGURE 2.2 Samples of low-quality fingerprints (from left to right): original fingerprints , scale parameters , texture components , orientation vectors , and the segmentation results.

However, total variation-based methods are difficult to accurately restore the ridge pattern as cartoon component of the fingerprint image also contains some meaningful fingerprint pattern information. As a result, the enhanced fingerprint pattern is weak and suffers from missing information, which will lead to the poor performance of fingerprint matching and identification. Another performance is the testing efficiency, since solving the optimization is required in each test sample respectively.

Although the Gabor filtering and TV model can improve the ridge clarity to some extent, it fails to restore the ridge structure influenced by unstructured noise precisely. The Gabor filtering method is widely used in plain and rolled fingerprint images, while failing to perform well on latent fingerprint images. As we have mentioned above, the Gabor filter can remove untrusted noise. However, in latent fingerprint images, there are quite a lot of non-fingerprint components overlapping with the ridge pattern. These handwriting, stain, line, and arch are usually recognized as structured noise which shares similar frequency and orientation characteristics with the fingerprint ridge pattern.

To tackle this problem, the Global dictionary [20] was proposed to improve the accuracy of orientation field estimation. Liu et al. [21] proposed multi-scale patch based sparse representation with dictionaries. The essence of fingerprint enhancement is to remove the irrelevant patterns and highlight the ridge and valley pattern. This can be achieved in two stages including image decomposition and pattern reconstruction. In the pattern reconstruction part, spare representation method can be applied to enhance the useful information. More concretely, for each image signal y, it is made up of the linear combination of a set of basic components:

$$y = \sum_i a_i \psi_i + e_i = \Phi a + e \qquad (2-4)$$

where $\Phi = [\psi_1, \psi_2, \cdots, \psi_M]$ denotes a dictionary consists of basic components. For the 2D fingerprint image, the orientation and frequency of local patches can be seen as the oscillatory signal. The fingerprint enhancement problem can be transformed into a sparse representation problem. We can first build our dictionary Φ with methods such as Gabor functions. The Gabor filter is a proper choice because it can also extract the orientation and frequency features of the fingerprint image. As Equation 2-1 and Equation 2-2 denotes, Gabor filter is highly selective in orientation. For each angle, we can get a basic component for the dictionary. After the dictionary is established, we can get the sparse representation vector $a=[a_1,a_2,\cdots,a_M]$. Then we can reconstruct each local patch with formula $y=\Phi a$. The size of the local patch is an important parameter for the reconstruction step. If the patch size is small, more local details such as minutiae point can be preserved better. If the patch size is large, the noise can be removed to a greater extent. To fully exploit the capacity of sparse representation, we can change the size of the patch to achieve a multi-scale sparse representation. This can be formulated as a l_0 norm regularized minimization problem. This can be solved with a convex optimization:

$$\min \| a \|_1 \quad \text{s.t.} \ \| y_s - \Phi_s a \|_2^2 < \varepsilon_1; \| y_l - \Phi_l a \|_2^2 < \varepsilon_2, \tag{2-5}$$

where y_s, y_l denotes the patch for the smallest and largest size respectively, and Φ_s, Φ_l represents the corresponding dictionaries. The minimization problem can also be reformulated as a Lagrange multiplication form:

$$\hat{a} = \arg\min_a \| \tilde{y} - \tilde{\Phi} a \|_2^2 + \lambda \| a \|_1 \tag{2-6}$$

In the reconstruction process, the patch size usually increases from small size to large size. To evaluate the fingerprint quality in different scales, orientation coherence [62] can be applied to measure gradient direction coherence as:

$$Q = \frac{|\sum_W (G_{s,x}, G_{s,y})|}{\sum_W |(G_{s,x}, G_{s,y})|} = \frac{\sqrt{(G_{xx} - G_{yy})^2 + 4G_{xy}^2}}{G_{xx} + G_{yy}} \tag{2-7}$$

Where (G_x, G_y) denotes the gradient, which represents the orientation of the local patch. When the quality score Q is improved, it means that the fingerprint quality in this local patch is better. When we are using multi-scale local patches, we can use Q to choose if we will replace the specific area with our reconstructed results.

All these methods mentioned above, including Gabor filtering, Total Variation (TV) Model, and dictionary based sparse representation, are highly interpretable but perform badly when more and more fingerprint image with different noise is given.

Because the representative capacity of classical methods is relatively weak, learning-based method can be adopted.

2.2 Generative Adversarial Network (GAN)

Generative Adversarial Network (GAN) was proposed by Ian GoodFellow et al.[34]. It generates output through the adversarial game between the generative model and the discriminative model. Among them, the generation model G uses noise data to generate pictures, and the discriminative model D determines whether the pictures are generated from the real data set or G. G tries to keep D from being able to tell whether it is a generated picture or a real picture, and D tries to give himself the ability to distinguish the two pictures. The two parties constantly optimize themselves in the game, and ultimately can reach an optimal state. The optimization target can be formulated as

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]. \tag{2-8}$$

From Equation 2-8 we can see, in the training of GAN, the discriminative model D needs to make its output D (x) as close to 1 as possible for the data x from the real data set, and the output D (G (z)) for the data G (z) generated by the generation model G as close to 0 as possible. The greater the Equation 2-8 is, the stronger the discriminative ability of D. For generator G, it needs to make D think that the data it generates is real, so G needs D to output the discriminative results of data it generates as large as possible. The more similar the data generated by G is to the real data, the stronger the ability of G to generate samples.

In the max-min optimization shown in Equation 2-8, when the parameter of generator G is fixed, we can get the optimized discriminator D. To get the best D, we need to maximize V(G,D)

$$\begin{aligned} V(G,D) &= \int_x p_{data}(x)\log(D(x))dx + \int_z p_z(z)\log(1-D(G(z)))dz \\ &= \int_x p_{data}(x)\log(D(x)) + p_g(x)\log(1-D(x))dx \end{aligned} \tag{2-9}$$

If we want to maximize the Equation 2-5 for each given sample x, we should maximize the $\int_x p_{data}(x)\log(D(x)) + p_g(x)\log(1-D(x))dx$. It is easy to prove that the optimized D is

$$D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}. \qquad (2-10)$$

Similarly, the optimization goal of training discriminator D can also be formulated as maximizing the likelihood estimation of probability $P(Y = y \mid x)$. The Equation 2-8 can also be formulated as

$$\begin{aligned} C(G) &= \max_D V(G, D) \\ &= E_{x\sim p_{data}(x)}[\log D^*_G(x)] + E_{z\sim p_z(z)}[\log(1-D^*_G(G(z)))] \\ &= E_{x\sim p_{data}(x)}[\log D^*_G(x)] + E_{x\sim p_g}[\log(1-D^*_G(x))] \\ &= E_{x\sim p_{data}(x)}[\log \frac{p_{data}(x)}{p_{data}(x)+p_g(x)}] + E_{x\sim p_g}[\log \frac{p_g(x)}{p_{data}(x)+p_g(x)}] \end{aligned} \qquad (2-11)$$

We can also use JS (Jensen–Shannon) divergence and KL divergence to represents C(G)

$$\begin{aligned} C(G) &= \int_x p_{data}(x)\log\frac{p_{data}(x)}{p_{data}(x)+p_g(x)} + p_g(x)\log\frac{p_g(x)}{p_{data}(x)+p_g(x)}dx \\ &= -2\log 2 + \int_x p_{data}(x)\log\frac{p_{data}(x)}{(p_{data}(x)+p_g(x))/2} + p_g(x)\log\frac{p_g(x)}{(p_{data}(x)+p_g(x))/2}dx \\ &= -2\log 2 + KL(p_{data}(x) \| \frac{p_{data}(x)+p_g(x)}{2}) + KL(p_g(x) \| \frac{p_{data}(x)+p_g(x)}{2}) \\ &= -2\log 2 + 2JSD(p_{data}(x) \| p_g(x)) \end{aligned} \qquad (2-12)$$

Note that JS divergence is used to quantize the difference between two probability distribution $JSD(p(x) \| q(x)) = \frac{1}{2}KL(p(x) \| \frac{p(x)+q(x)}{2}) + \frac{1}{2}KL(q(x) \| \frac{p(x)+q(x)}{2})$. The JS divergence of two distribution is always non-negative. And it is easy to find that the optimization is achieved when $p_g = p_{data}$ and $D^*_G = 1/2$. This is also when the generated data are the same of the real data.

After the vanilla GAN is proposed, more and more different generative adversarial network came out. Combining the powerful convolutional structure and the adversarial training of generator and discriminator, DCGAN [35] is proposed. In classic GAN architecture, both the generator and discriminator are built on a multi-layer perceptron (MLP). However, the representative capacity of MLP is limited. After replacing the MLP with the convolutional layer, the performance of DCGAN is much better than previous vanilla GAN in much tasks, especially computer vision tasks.

Another problem of GAN is the mode collapse. The problem happens when the generated samples of G is far from real data and D give a positive evaluation of the generated samples. In this scenario, the G and D will converge in a local minimum and fail to generated high-quality data. According to the research proposed by Google [40], model collapse is the most severe problem in GAN training. Inception Score (IS) [41] and Frechet Inception Distance (FID) [42] can be used to evaluate the performance of GAN. These two metrics can be formulated as follow:

$$IS(G) = \exp(E_{x\sim G}[d_{KL}(p(y|x), p(y)])) \quad (2-13)$$

$$FID(x,g) = \|\mu_x - \mu_g\|_2^2 + Tr(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}) \quad (2-14)$$

For the Inception Score, it expects the entropy of output of discriminator $p(y|x)$ to be low and the entropy of the $p(y)$ to be high. Therefore, IS use KL divergence between these two entropies to describe the distance between the two different probability distribution. In the FID formula, $\mu_x, \mu_g, \Sigma_x, \Sigma_g$ are the average value and variance of real data and generated data. The core idea behind FID is to embed the generated data into the feature space of a specific layer of Inception Net. The output of the embedded layer can be seen as a continuous multivariate gaussian distribution.

To tackle the model collapse problem and improve the generated image quality, Wasserstein GAN (WGAN) [36] is proposed. In the original GAN, the loss function is built with cross-entropy to describe the distance between real data distribution and generated distribution. In WGAN, however, Wasserstein distance is applied to represent the distance between these two distributions.

$$W(P_r, P_g) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y)\sim\gamma}(\|x-y\|) \quad (2-15)$$

In Equation 2-15, $\prod$ represents the set of union distribution combined with these two distributions. Wasserstein distance is the lowest cost to approximate one distribution to another. Compared to KL divergence and JS divergence, Wasserstein distance is a better choice of the loss function, since Wasserstein distance is smoother in distribution. In the training process, Adam optimizer usually fails to achieve a satisfying performance. RMSProp and SGD are frequently used as the optimizer.

In addition, the generator architecture and the type of loss function can be altered to achieve stability in training and the diversity of generated samples. Applying the least square error to replace the loss function of GAN, LSGAN is put forward. Using the least squares method can close the gap between the decision boundary and the distribution of generated samples. This can be described as

$$\begin{cases} L_D{}^{LSGAN} = E[(D(x)-b)^2] + E[(D(G(z))-a)^2] \\ L_G{}^{LSGAN} = E[(D(G(z))-c)^2] \end{cases} \tag{2-16}$$

Auto-encoder architecture can also be an alternative choice for the discriminator. To exploit this cue, BEGAN [38] is proposed. BEGAN does not estimate the difference between the two distributions. However, BEGAN describe loss function to close the gap between them. It introduces a parameter to control the balance between diversity and sharpness. The final loss functions are

$$\begin{cases} L_D{}^{BEGAN} = L(x) - k_t L(G(z_D)) \\ L_G{}^{BEGAN} = L(G(z_G)) \\ k_{t+1} = k_t + \lambda_k(\gamma L(x) - L(G(z_G))) \end{cases} \tag{2-17}$$

All generative networks mentioned above follow an unsupervised training manner. There are training samples but the samples are not labeled. This is widely used in image generation without class labels. However, If we want to control the content of the generation with a label, or we want to generated samples based on a given sample in another domain, conditional GAN (cGAN) [39] is proposed. The optimization target of cGAN is

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x|y)] + E_{z\sim p_z(z)}[\log(1-D(G(z|y)))] \tag{2-18}$$

Compared the cGAN optimization target to the original GAN optimization target in Equation 2-8, the generated samples are restricted by the condition label y, which make it possible to generated specific type of samples with a given condition. cGAN also serves as a foundation of GAN based image translation. The image in the source domain can be the condition and the generated target is the image in the target domain. The structure comparisons of GAN and cGAN is depicted in Fig.2.3. As can been seen in the figure, the core improvement of cGAN is to add the condition label c, which will influence both the generation and discrimination process.

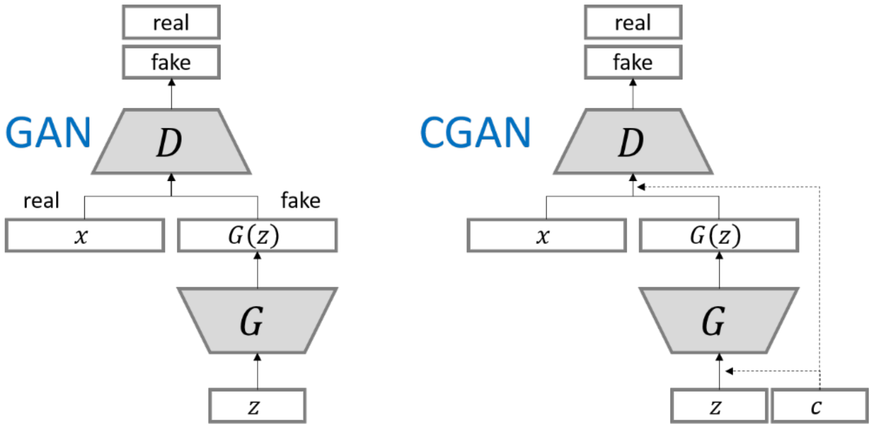


FIGURE 2.3 Structure comparison of GAN and cGAN.

2.3 Deep Learning based Image Translation

Image translation is of great importance in image processing and understanding. The essence of image translation is domain adaption. There are two different domains: source domain and target domain. The source domain and target domain usually belong to the same task, while their distribution is different. When the source domain and the target domain is the low resolution and high-resolution image respectively, the image translation task is image super-resolution. The task will be noise removing when source and target domain is the image with noise and high-quality image without noise. In our fingerprint enhancement task, the source domain is where low quality latent fingerprint images lie in and the target domain is where enhanced high-quality fingerprint images lie in.

When we want to translate an image into another type, the most straight forward idea is to design a CNN and build a mapping between input and output. The CNN can follow an auto-encoder structure. The original image in the source domain is input into the encoder, and the translated image is output from the decoder. There are lots of research on the CNN based image translation, including some representative work on image denoising such as DnCNN [16], Non-Local Recurrent Network [24], Adversarial Denoising [25], DAAM [26], and some representative work on image super-resolution such as SRCNN [27], DRRN [28], EDSR [29], IDN [30].

Our task in this paper focus on fingerprint enhancement, which is also a task of image translation. Moving to the era of deep learning, more learning-based latent fingerprint enhancement methods, especially CNN-based enhancement methods are much more widely adopted. Cao and Jain [31] put forward orientation estimation as a classification problem and used the CNN for fingerprint orientation estimation. Inspired by their work, FingerNet [32] is proposed by Li et al. Following the pixels-to-pixels and end-to-end learning manner, FingerNet [32] is a deep convolutional neural network-based method to enhance latent fingerprints and achieve state-of-the-art matching accuracy on various low-quality fingerprint datasets. The network structure of FingerNet is different from traditional auto encoder, as there are two decoders after the bottleneck. They also apply multi-task learning scheme in the training of convolutional neural network including an orientation estimation task and a fingerprint enhancement task. The network architecture of FingerNet is shown in Fig.2.4.

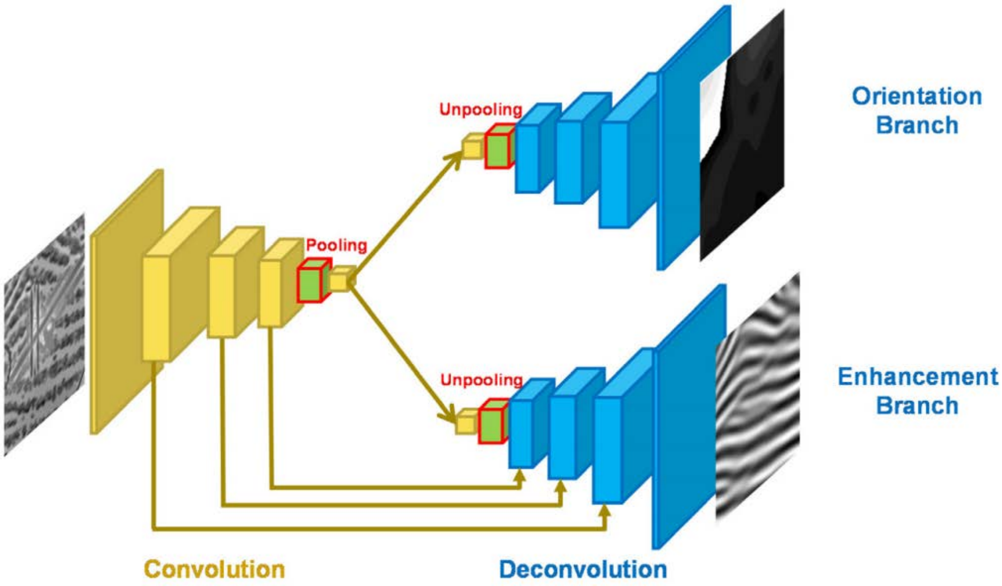


FIGURE 2.4 Convolutional neural network architecture of FingerNet [32]. As can be seen in the Figure, there are two branches of decoder corrresponding to two different tasks: orientation estimation and fingerprint enhancement.

More recently, Qian et al. proposed a DenseUNet based latent fingerprint enhancement model [26]. In their work, they generate the paired training data with the aid of a TV model to separate structured noise. Skip connections are used in DenseUNet [33] architecture which boosts the representation power of the network. Also, they use a quality control module as a switch of iterative testing, which effectively helps remove the noise in the latent fingerprint images. However, CNN based method still fails to enhance some examples of severely corrupted regions or heavy noises.

However, the translated image will be blurry. This is because the convolution operation and max-pooling fail to preserve some local information of the image. To solve this problem in the image translation problem, GAN based method has been proposed. The first GAN based image translation model is Pix2Pix [43]. In their work, GAN loss is added to penalize the model when it outputs a blurry image. The network architecture is the cGAN structure where the condition is the image in the source domain. The final loss function can be formulated as

$$G^{*} = \arg\min_{G}\max_{D} L_{cGAN}(G, D) + \lambda L_{L1}(G) \qquad (2-19)$$

where L1 loss is applied rather than L2 loss. This is because the L1 loss can preserve the details on edge better and can avoid the blurry output. For the architecture of the generator, U-Net [44] with skip connections is applied.

Follow the core idea of GAN, researchers have attempted to improve fingerprint enhancement by the generative adversarial network. Svoboda et al. [45] are the first to use the generative network to denoise minutiae and predict the missing parts of the ridge pattern. More recently, COOGAN [46] is proposed to utilize the supervision of orientation and quality and achieve state-of-the-art matching accuracy on the NIST SD27 dataset. In the COOGAN model, the orientation estimation module and quality evaluation module are applied to the multi-task learning scheme. These approaches bring a new direction for latent fingerprint enhancement.

2.4 Chapter Summary

To sum up, we give a detailed introduction of the related work of fingerprint image enhancement, generative adversarial network (GAN), and deep learning-based image translation problems, and corresponding classic methods. In fingerprint image enhancement method section, the formula and the characteristic of Gabor filter is introduced. We also illustrate the frequency and orientation selective characteristic of Gabor filter. The total variation (TV) model are proposed based on the idea that image can be decomposed into texture and cartoon component. We show some examples of TV model based latent fingerprint enhancement. As for sparse representation methods, we show how the dictionary is built and the final convex optimization of the problem. The generative adversarial network (GAN) is made up of generator and discriminator. The training of GAN is inspired by the adversarial game of two objects. Conditional GAN (cGAN) is an important improved GAN model. In cGAN, a condition is given to supervise the training. Image translation involves translating images in source domain to images in target domain. The source domain and target domain are different in various specific image translation tasks. We mentioned some important image denoising neural network and image super resolution neural network, including DnCNN, Non-Local Recurrent Network, Adversarial Denoising, DAAM, SRCNN, DRRN, EDSR, and IDN. These effective deep learning based methods inspire some CNN based fingerprint enhancement model. We give a detailed introduction of FingerNet and show the network structure. In addition, some deep learning based fingerprint enhancement methods are given, such as DenseUNet, COOGAN, and Svoboda et al.

Chapter Three Fingerprint Enhancement based on Generated Adversarial Network

In this section, we will first define the latent fingerprint image enhancement problem and describe how to apply conditional GAN (cGAN) to the enhancement task. Then we will show the network architecture and some important hyperparameters of our GAN model. Next, we will introduce how we generate the training samples and corresponding ground truth with the existing latent fingerprint dataset and high-quality fingerprint images. We will also briefly introduce the evaluation metrics we choose and show how we design our experiments to show the effectiveness and efficiency of our model. Finally, the results of latent fingerprint enhancement will be given.

3.1 Fingerprint Image Enhancement: Problem Formulation

Latent fingerprint enhancement involves translating the original latent fingerprint images into high-quality fingerprint with a clear ridge structure. So, it can be formulated as an image-to-image translation problem [43]. Our goal can be denoted as learning a mapping $f: D_L \rightarrow D_E$, where D_L represents the latent fingerprint domain and D_E represents the target domain where enhanced fingerprint images lie in.

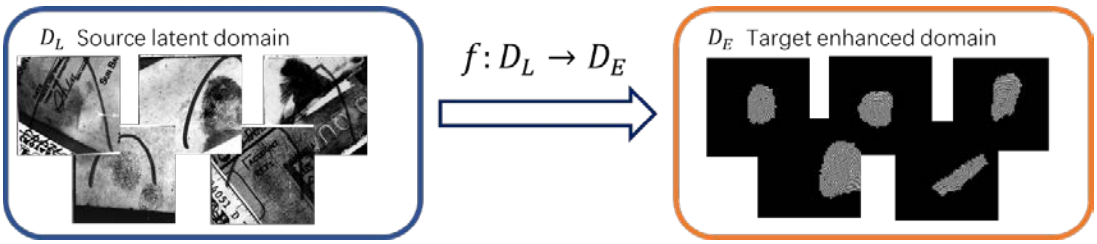


FIGURE 3.1 Illustration of image-to-image translation problem for latent fingerprint image enhancement.

Generative adversarial networks can map a sample from a random distribution p_{data} to the target domain. The generator G is supervised by reconstruction loss (i.e. L1 or L2 loss) with corresponding ground truth in the domain E. Meanwhile, the discriminator D learns to discriminate the generated sample from G and the ground truth from the target domain. The generator G tries to maximize the possibility of discriminator D making a mistake [34]. The objective function can be denoted as:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \qquad (3-1)$$

In our application of latent fingerprint enhancement, the source domain is the given latent fingerprint rather than a random distribution. The input of the generator is real samples of latent fingerprint images. Also, the input sample serves as a condition of the discriminator. As the discriminator takes both the enhanced fingerprint generated by G and the ground truth of the high-quality fingerprint, the G should learn to preserve the details and remove the noise to fool the discriminator D.

3.2 Generative Adversarial Network Architecture

Our proposed generative adversarial network is made up of generator and discriminator. An overview of the proposed GAN architecture is depicted in Fig.3.2. The input of the generator is the original latent fingerprint image, the output is the enhanced fingerprint image. Then the output of the generator and the corresponding ground truth (ground truth fingerprint image) is input into the discriminator. The output of the PatchGAN discriminator is a probability map of the same size as the number of patches.

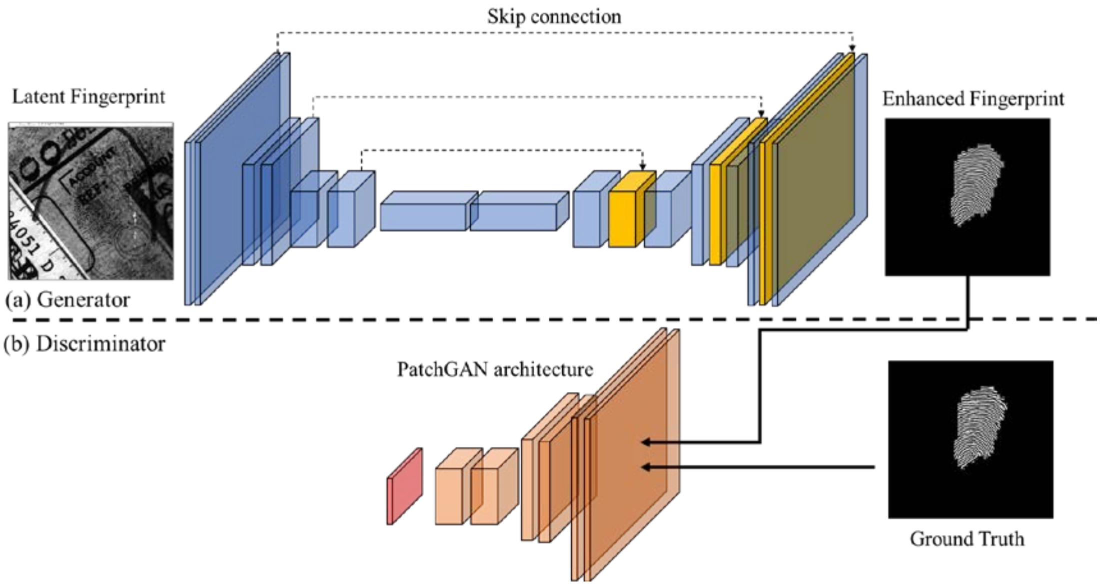


FIGURE 3.2 The architecture of our Generative Adversarial Network architecture. In the generator, each cuboid pair of the same size in the encoder part represents two consequent "Conv+Relu" layers. Each cuboid triple of the same size in the decoder represents two consequent "Deconv+Concat+Relu" layers, among which the concatenate layers are shown in yellow. The skip connection is between the i th Conv layer and $n-i$ th Deconv layer. The output of the generator is the enhanced fingerprint image. In the discriminator, we use the PatchGAN architecture the same as pix2pix [43].

The generator we use in our generative adversarial network is a "U-Net" structure [44]. In the "U-Net" generator, there are skip connections connecting i th Conv layer and $n-i$ th Deconv layer. The skip connection can pass the feature between the encoder and decoder which will help the preservation of the details such as ridge pattern. In the encoder-decoder architecture used in many image translation problems [47][48][49] , the original image in the source domain will firstly be down sampled in a series of layers in the encoder. Between the encoder and decoder there will be a bottleneck layer, after which the down sample process will turn to be up sample. However, there are a lot of common information in the representative of the input and output image. It would be desirable to pass the shared representative information directly from the encoder layer to the decoder layer. This is the reason why skip connection is important in the generator architecture.

In our work, the kernel size of the Conv and Deconv layer is set to be 5×5 and the strides are set to be 1×1. The activation we use is ReLu activation. To suppress the noise in the image, the feature extracted by the Conv layer should be abstract, which requires the receptive field to be large. We also replace the pooling and unpooling layers with the Conv and Deconv layers. The kernel size of these layers is set to be 2×2 and the stride is also set to be 2×2. These Conv and Deconv layers have the same function of pooling and unpooling layers while learning more abstract representation. The output of the generator is the enhanced fingerprint image and the estimated orientation field.

The discriminator we use is a "PatchGAN" [43] (also called Markovian discriminator) structure CNN classifier. Rather than output a single value of the possibility of true or fake, PatchGAN output a probability map. Each value in the map represents the probability of a single patch in the image. This architecture can help recover texture and preserve the details in the high-resolution image. No matter what the size of the generated image is, the image will be divided into patches with fixed size and input into D for training. There are many advantages for patch discriminator. Firstly, the input size of D will be relatively small, which indicate a lower computation cost and faster training speed. Secondly, the generator follows a full convolution manner and there is no limit for the scale of the image. Using the PatchGAN discriminator will also extend the expandability for the model. The patch size we use is 70×70, which is selected with grid search for the best training performance.

3.3 Objective Function

To train our generative adversarial network, the objective function should be given. In the proposed model, the loss function consists of two parts: adversarial loss and reconstruction loss.

Adversarial Loss: The objective function of GAN has been shown in Equ.3.1. In the backpropagation process during the training, both the generator G and discriminator D minimize the adversarial loss. The generator will be penalized if the generated image is correctly identified. In our latent fingerprint enhancement application, we are using a conditional generative adversarial network (cGAN). Thus, the loss function should be modified as

$$L_{cGAN}(G,D,x)=E_{(x,y)\sim p(x,y)}[\log(D(x,y))]+E_{x\sim p_x(x)}[\log(1-D(x,G(x)))]. \qquad (3-2)$$

Reconstruction Loss: The generator G can translate latent fingerprint x into enhanced high-quality fingerprint y, while the ground truth high-quality fingerprint is y^*. Reconstruction loss aims to preserve the similarities between the enhanced image and the corresponding ground truth. We use the L1 loss as the reconstruction loss, which can be formulated as:

$$L_{l_1}(G,x,y^*)=\| y^*-G(x)\|_1 . \qquad (3-3)$$

Total Loss: Combining these two loss functions, we can define our total loss as

$$L_{total}=L_{cGAN}+\lambda_1 L_1 \qquad (3-4)$$

λ_1 is the weighting coefficient used to balance between the different components. Therefore, our training goal is to optimize the objective function:

$$G_{opt}=\min_G \max_D L_{total}. \qquad (3-5)$$

3.4 Experimental Results

3.4.1 Dataset Preparation

The dataset used in this work is collected or generated from NIST DB 27 [4] and NIST DB 14 [5]. The training set contains 30000 synthetic fingerprint images generated by combining noise from NIST DB 27 and high-quality fingerprint from NIST DB 14. The testing set contains 258 latent fingerprint images from NIST DB 27. The size of each fingerprint is 816×816 pixels both in the training set and testing set.

In NIST DB 27 latent fingerprint dataset, there are 258 latent fingerprints and matched template fingerprints. According to the fingerprint quality level, each image is labeled with a quality level including "GOOD", "BAD", and "UGLY". The number of latent fingerprint cases is shown in Table 3-1.

Table 3-1 Number of latent fingerprints in NIST DB 27.

Quality Label	GOOD	BAD	UGLY	Total
Number of fingerprints	88	85	85	258

It requires a large amount of paired training data to train a generative adversarial network from scratch. In our work, the paired training data includes the latent fingerprints and the corresponding enhanced images and segmentation masks. However, there is no publicly available database consisting of pairs of low quality latent and high-quality enhanced fingerprint images for training. In addition, the number of available latent fingerprint images is limited comparing to rolled and plain fingerprint images. As mentioned above, in NIST DB 27 latent fingerprint dataset, there are merely 258 latent images and their corresponding template fingerprints. As a result, it is more widely accepted to use NIST DB 27 as the test images dataset rather than the training dataset. It is still challenging to get enough training data for training a deep neural network to enhance the latent fingerprints. To tackle this problem, a solution is to synthetically generate the latent fingerprints by simulating the conditions in which a latent fingerprint is typically acquired such as in different overlapping patterns and backgrounds. Fig.1.3 illustrates the process of generating paired training data for latent fingerprint enhancement. This procedure generally includes the generation of latent fingerprints and their corresponding enhancement ground truth, which are obtained with the good quality fingerprints from which the latent fingerprint is simulated. The samples in the training set simulate the structured noise and can represent the latent fingerprints acquired environment. Fig.3.3 gives some examples of generated training data.

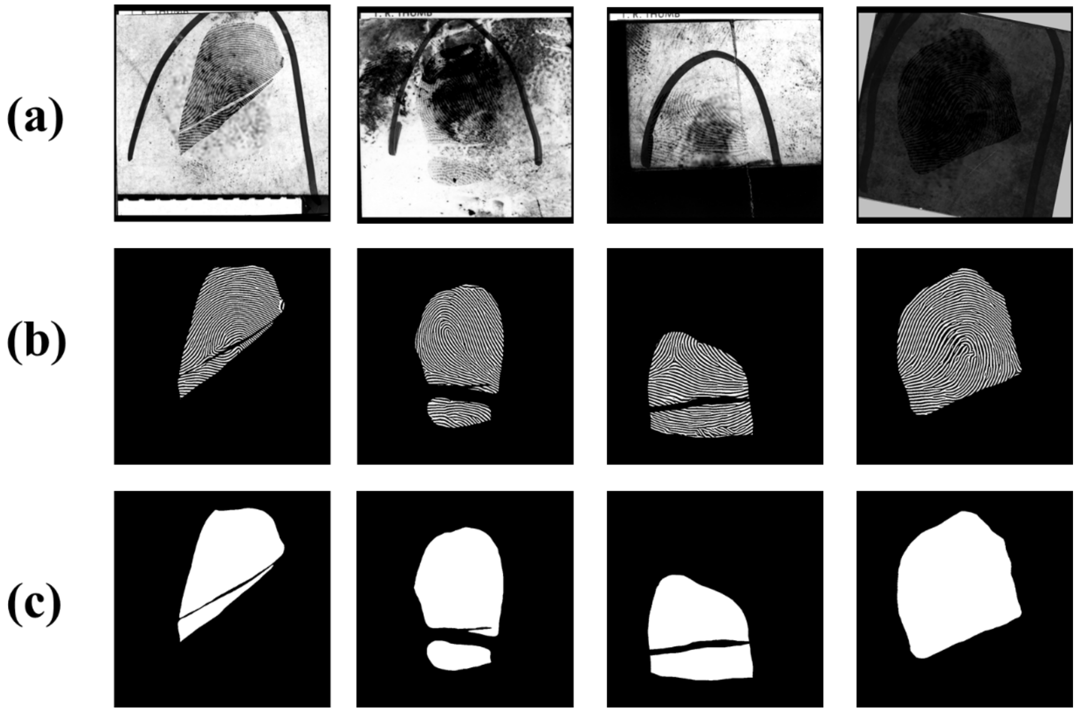


FIGURE 3.3 Examples of generated training data. (a) Generated Latent fingerprint images (b) Enhanced binary fingerprint image (c) Segmented fingerprint image mask.

To generate the latent fingerprints, we simulate the conditions under which the latent fingerprints are usually captured by adding different noises and backgrounds into good quality fingerprints. Specifically, we synthetically add the high-quality fingerprints with the structured noises to generate the latent fingerprints for training. Instead of NIST DB 27, we use NIST DB 14 to prepare the good quality fingerprints. NIST DB 14 consists of 27,000 pairs of rolled fingerprints with various fingerprint types. We manually check and choose 200 high quality fingerprints from NIST DB 14. Besides, we decompose the latent images of NIST DB 27 into texture and cartoon components using the TV method [60] to simulate the various noise of latent images. The structured noises are obtained from the cartoon components.

After we have the noise pattern and high-quality fingerprint, the simulated fingerprint I_l can be a linear combination as follows:

$$I_l = \omega \times I_g + (1-\omega) \times I_s \tag{3-6}$$

where I_g represents the good quality fingerprint and I_s is the decomposed noise pattern from the TV model. Weighting coefficient ω varies from 0.3 to 0.7 for whole patch noise. For data augmentation, we also add varying levels of Gaussian noise to the generated latent fingerprint and make some shifts and rotations of them. The final generated latent fingerprint images are of a size 816×816. Fig.3.3 (a) shows some examples of generated latent fingerprint images.

The size of each fingerprint images is 768×800 pixels and we use image padding to change the size into 816×816 which is more suitable for training. Finally, we generate 30,000 pairs of latent fingerprint images, their corresponding binarized enhancement ground truth, and their segmentation for the training.

3.4.2 Evaluation Metrics

The ultimate goal for latent fingerprint enhancement is to improve the clarity of useful details and suppress various noises from the image for better recognition accuracy. Therefore, we can evaluate the effectiveness of our model using the accuracy of fingerprint matching. The matching template for each image is 258 images from NIST DB 27 dataset and 27000 images from NIST DB 14 dataset. After the matching process we will obtain a matric of size 258×27258. Then we compute the Cumulative Match Characteristic (CMC) to show the results of fingerprint matching. In the fingerprint matching stage, we use commercial fingerprint matching software VeriFinger SDK 4.3 (https://www.neurotechnology.com/) for the feature extraction and matching. VeriFinger is commonly used by previous work and we can compare our model with the previous methods in the same environment. We use Cumulative Match Characteristic (CMC) curves to show the enhancement effectiveness of our model.

Similar to the Receiver Operating Characteristic (ROC) curve, the CMC curve is an important evaluation metric of the pattern recognition system, especially for human face, fingerprint, and iris recognition. The CMC curve is a precision curve that provides accuracy of recognition, verification, or detection for each rank. The core to acquire the ultimate CMC curve is to compute the top-k accuracy. Take a simple single-gallery-shot for example, each gallery identity has only one instance. For each query, the algorithm will rank all the samples according to the query. If the top-k ranked gallery samples contain the query identity, the accuracy for top-k in this query is 1, otherwise it will be 0. This is a shifted step function. The final CMC curve will take an average of the results of queries. CMC curve is a non-decreasing curve indicating identification rate or recognition accuracy in each rank, and the identification rate or recognition accuracy for the last rank is equal to 1.

3.4.3 Fingerprint Enhancement Results

After we generate our training data and choose the proper evaluation metric to show the effectiveness of our GAN model. We train our generative adversarial network for 10 epochs without any pre-trained model. The optimizer we use is Adam Optimizer with learning rate 1e-4, β_1=0.9, and β_2=0.999. The batch size is 4 and all training process is finished on a single NVIDIA 1080Ti GPU. After the training stage, we use image padding for our tested image from NIST DB 27 to resize their spatial resolution into 816×816. Then we input the resized image into the trained neural network. Some of the enhanced examples are shown in Fig.3.4. It can be seen from the enhancement result shown in Fig.3.4 that image quality improves a lot. The noise has been removed and the ridge pattern has been enhanced. Also, there are some additional failure cases. The original latent fingerprint image, gray and binary enhanced results is shown in Fig.3.5. In Fig.3.5, it can be seen from the difference between gray enhanced results and final binary result that the enhanced result tends to be blurry in severe corrupted regions. If the gray enhancement is blurry, the binary result will be influenced and minutiae features may be failed to be extracted.

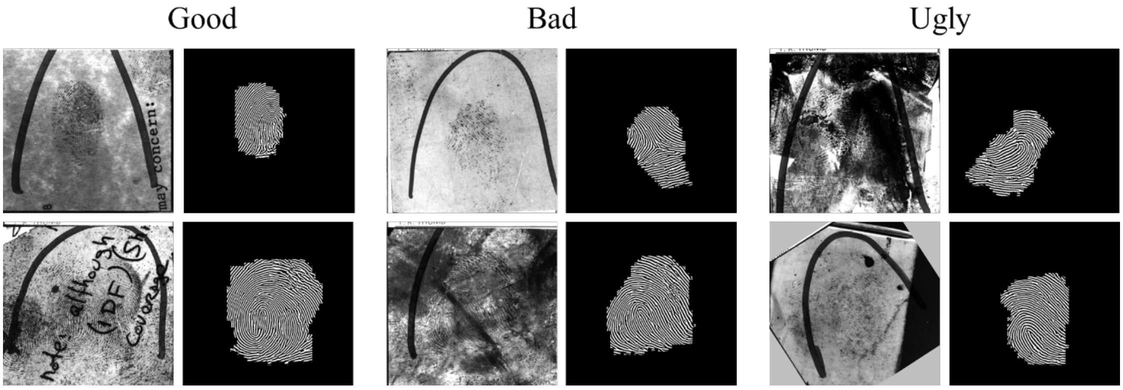


FIGURE 3.4 Examples of latent fingerprint enhancement results (left) as well as the original latent fingerprints (right) selected from three subsets of NIST DB 27.

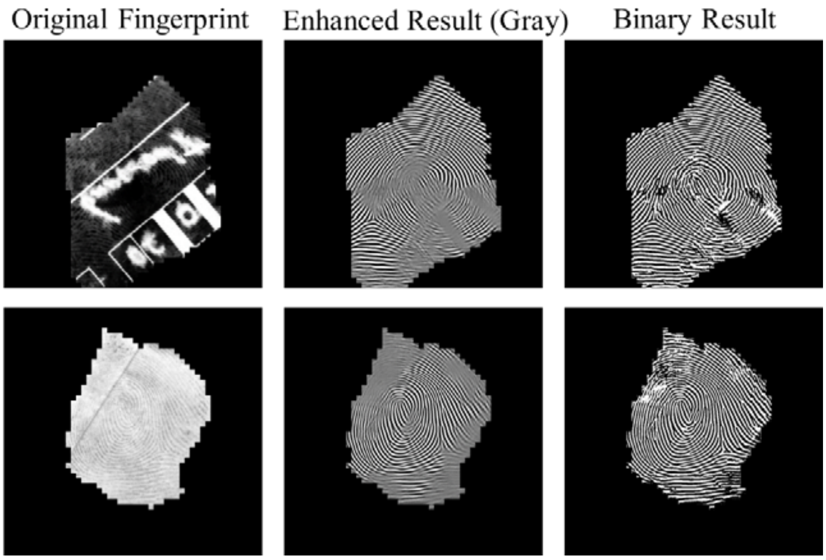
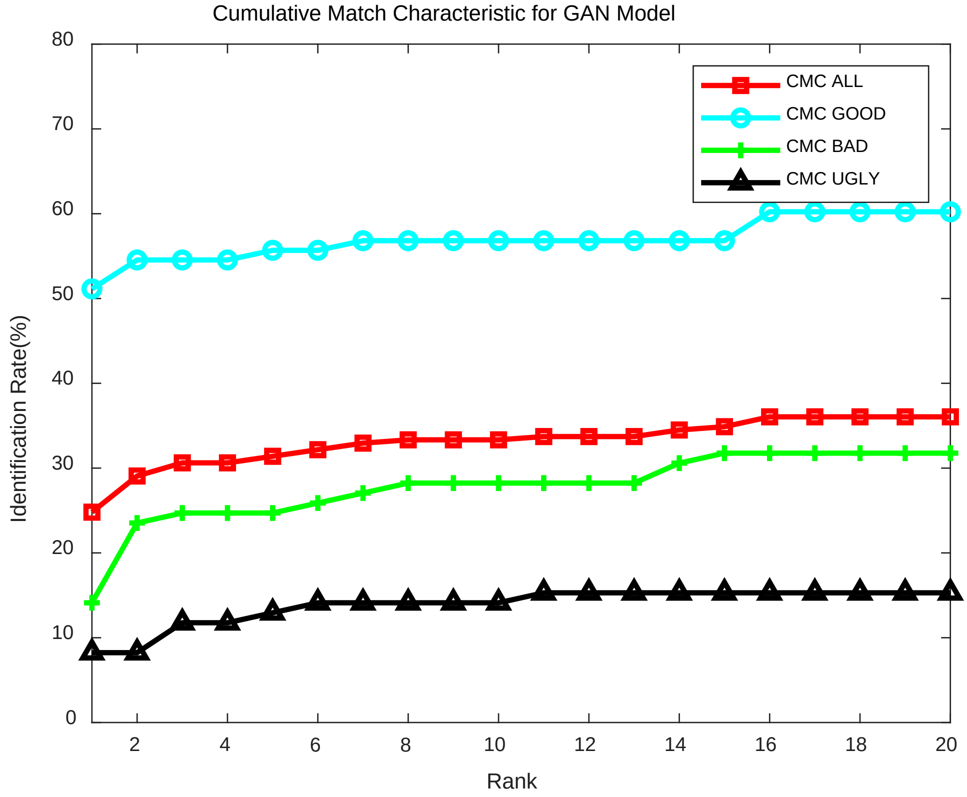


FIGURE 3.5 Selected fingerprint enhancement failure cases of our GAN model.

We then apply the enhanced fingerprint to the fingerprint recognition to evaluate the fingerprint quality. The matching results are shown in Fig.3.6. We can see that the identification rate for rank 1 has already achieved a satisfying result, which indicates that our GAN model does improve the image quality. Compared to the Cumulative Match Characteristic (CMC) of the original raw fingerprint image, the performance improvement is significant on all three subsets.



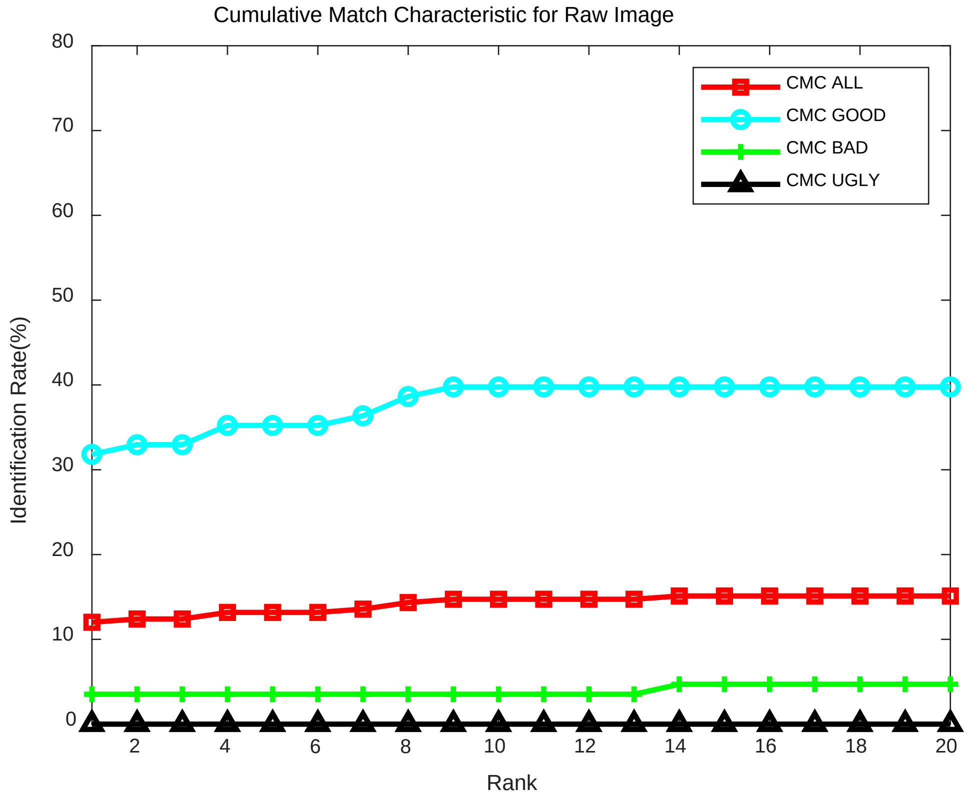


FIGURE 3.6 CMC curves of our GAN model and raw image for all samples in the NIST DB 27 and three subsets "GOOD", "BAD", and "UGLY".

3.5 Chapter Summary

In conclusion, we propose a generative adversarial network based latent fingerprint enhancement methods. We first formulate the fingerprint image enhancement problem into an image translation problem. As mentioned in section 2.3. Deep learning is a proper solution to learn the mapping from low-quality fingerprint image to high-quality one. We then define our generative adversarial network architecture and given detailed hyper parameter of the proposed network. The objective function is also given, which consists of two parts: adversarial loss and reconstruction loss. We also describe how to generate the paired training data concretely. The final synthetical image is made up of the noise derived from NIST DB 27 dataset using TV model, and the fingerprint pattern from high quality NIST DB 14 model. To evaluate our methods, we design fingerprint matching experiment to show the effectiveness of the fingerprint enhancement. Using VeriFinger software to match the enhanced results with fingerprint template, we can get the Cumulative Match Characteristic (CMC) results. We show some of the enhanced samples, from which we can see that noise are removed and ridge are enhanced. We compare the CMC result of enhanced images and original latent fingerprint image. The improvement of identification rate is significant. We can draw a conclusion that our GAN model possesses the ability to denoise and enhanced the low-quality fingerprint images.

Chapter Four Fingerprint Enhancement based on Multi-Task Progressive GAN

In this chapter, we will present our multi-task learning progressive generative adversarial network (PGAN) model for latent fingerprint enhancement. We will show how we design the objective function for multi-task learning. In addition, the progressive growing training method will be introduced. We also compare our method with existing deep learning based methods and some classic methods. In the ablation experiments, we will show why the multi-task learning scheme and progressive training strategy are useful. We will also compare some samples with iterative testing and without iterative testing to prove the effectiveness of our iterative online testing (IOT).

4.1 Multi-task Generative Adversarial Network Architecture

Our proposed generative adversarial network is also made up of generator and discriminator. However, to implement multi-task learning of our GAN, the architecture of generator is changed. To speak specifically, the input of the generator is the original latent fingerprint image and the corresponding manually labeled segmentation, the output is the enhanced fingerprint image and the orientation field estimation. The multi-input multi-output network structure is corresponding to our multi-task learning strategy. The output of the generator is also connected with the output of discriminator, which is the enhanced fingerprint and the corresponding ground truth. The output of the PacthGAN discriminator is a probability map of the same size as the number of patches. An overview of the proposed multi-task progressive GAN architecture is depicted in Fig.3.2.

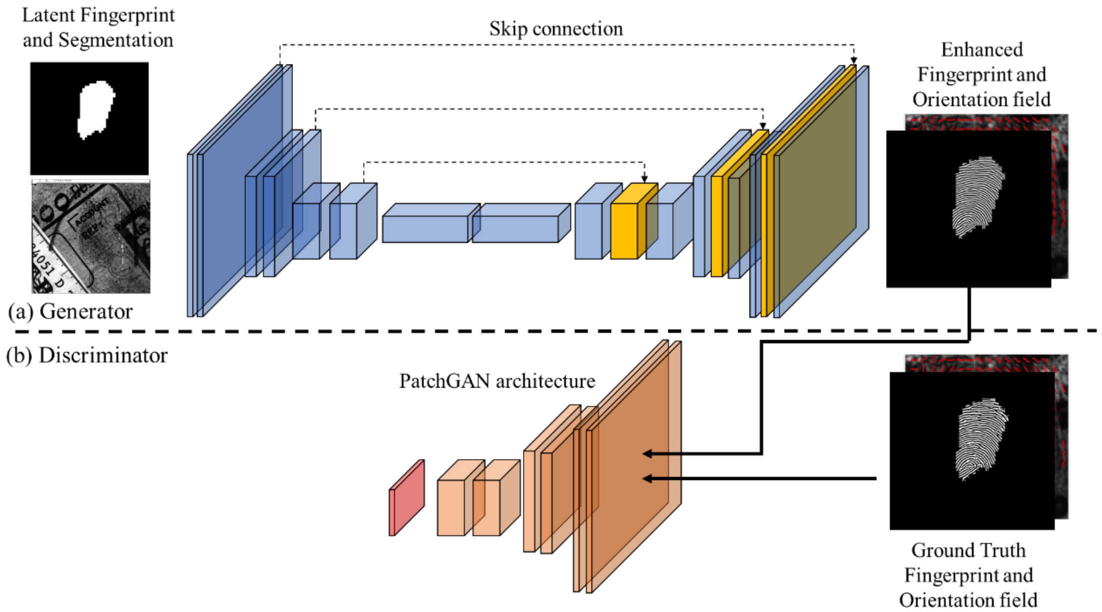


FIGURE 4.1 The architecture of multi-task Generative Adversarial Network. The core difference between our improved multi-task model and previous GAN is that the input and the output have been changed. The estimated orientation map is output as the orientation estimation task is added in the training.

Similar to our first model, the generator we use is a "U-Net" structure [44]. In the "U-Net" generator, there are skip connections connecting i th Conv layer and $n-i$ th Deconv layer. The skip connection can pass the feature between the encoder and decoder which will help the preservation of the details such as ridge pattern.

In our multi-task GAN model, the kernel size of each layer is the same of our first model, whose size of the Conv and Deconv layer is set to be 5×5 and the strides are set to be 1×1. Different from the previous model, the output of our multi-task learning generative adversarial network is the enhanced fingerprint image and the estimated orientation field.

The discriminator we use is also a "PatchGAN" [43] structure CNN classifier. Since we use a multi-task learning scheme, there are more than one feature map outputs of the generator. As can be seen in the objective function of our multi-task model, we only concatenate the enhancement fingerprint image output and the fingerprint ground truth. The loss function of the orientation estimation is computed in the output layer of the generator follow an auto encoder structure. The patch size output of our discriminator we use is 70×70.

4.2 Objective Function

To train our multi-task learning generative adversarial network, the objective function should be given. In the multi-task learning progressive GAN model, the loss function consists of three parts: adversarial loss, reconstruction loss, and orientation loss. Compare to our first model, the orientation loss is added to fully utilize the information of the orientation field.

Adversarial Loss: Similar to our first GAN mode, the objective function of GAN has been shown in Equation 4-1. In our latent fingerprint enhancement application, we are using a cGAN model. The loss function should be modified as

$$L_{cGAN}(G,D,x) = E_{(x,y)\sim p(x,y)}[\log(D(x,y))] + E_{x\sim p_x(x)}[\log(1-D(x,G(x)))]. \qquad (4-1)$$

Reconstruction Loss: Similar to our first model, reconstruction loss can preserve the similarities between the enhanced image and the corresponding ground truth. We adopt the L1 loss as the reconstruction loss:

$$L_{l_1}(G,x,y^*) = \| y^* - G(x) \|_1 . \qquad (4-2)$$

Orientation Loss: Our model follows a multi-task learning scheme: generating enhanced fingerprint and output the orientation field estimation. Since the orientation estimation and fingerprint enhancement share the same representation in the neural network, the multi-task learning from end to end can abstract more meaningful features than a single task. To reduce the dimension of the orientation feature, we down-sample the orientation field feature map and the corresponding ground truth generated by a gradient-based method the same as [32]. For each patch with a given size, the neural network can predict an angle value in the range of 180°, which makes up the estimation field o_e. To normalize and represent the orientation field for better training of the neural network, we use cosine and sine function. This can be formulated as:

$$o_x(i,j)=\cos(2\theta(i,j)), o_y(i,j)=\sin(2\theta(i,j)) \qquad (4-3)$$

Notice that in Equation 3-3, we can get two different orientation map representing different direction components. In the deep learning method, we use concatenate operation to combine these two feature maps of different direction. We can also calculate the combined orientation map with formula $O(i,j)=\frac{1}{2}\tan(\frac{\phi_y^{'}(i,j)}{\phi_x^{'}(i,j)})$ which can represent the synthetical orientation features.

If the ground truth orientation field is o_g, the orientation loss can be denoted as the cross-entropy (CE) between the estimated and ground-truth orientation field:

$$L_{ori}(o_e,o_g)=CE(o_e,o_g)=-\sum_i o_{g_i}\log(o_{e_i}) \qquad (4-4)$$

Total Loss: Combining these three loss functions, we can define our total loss as

$$L_{total}=L_{cGAN}+\lambda_1 L_1+\lambda_2 L_{ori} \qquad (4-5)$$

where λ_1 and λ_2 are weighing coefficient used to balance between the different components. Therefore, our training goal is to optimize the objective function:

$$G_{opt}=\min_G \max_D L_{total}. \qquad (4-6)$$

4.3 Progressive Training

The idea of the progressive growing training of GAN is proposed by Karras et al [50]. Before the progressive growing training of the generative adversarial network is proposed, many researchers notice the problem of the performance difference of deep neural network, especially GAN on different spatial resolutions. Wang et al. [51] first use multiple discriminators to solve the performance difference on input of different size, this is inspired by Durugkar et al. [52] who use multiple discriminators and a single generator to improve the generation ability and diversity. To achieve better performance, Ghosh et al. [53] proposed a GAN based model with multiple generators and a single discriminator. All of their work is limited to increase the number of the generator or discriminator, which will increase the number of parameters and the computational cost significantly. To tackle the problem and fully used the share information of different spatial resolution representation, a good solution is to progressively build our network.

The core idea of progressive training is to start training from a low-resolution and add new layers to the model increasingly. The generator fine details as training progress. This training method can significantly stabilize the training of GAN. In our work, the input image is 816×816, which possesses a relatively high resolution. When training an image translation GAN on such high resolution, the generated image quality is low and the loss is hard to converge. To tackle these problems, progressive training is introduced in our latent fingerprint application. The process of progressive training is shown in Fig.4.2.

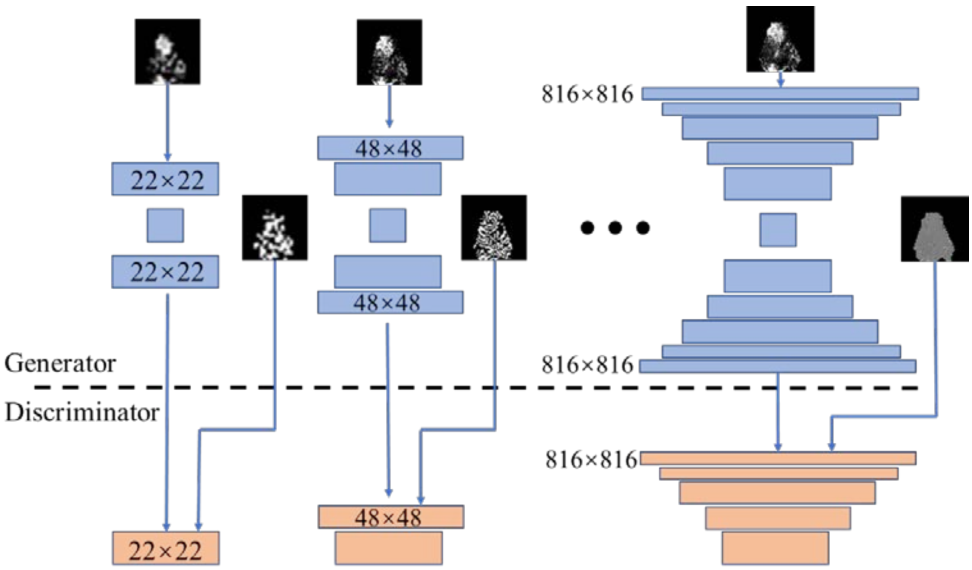


FIGURE 4.2 Progressive growing training of GAN. The blue rectangle represents the generator and the orange rectangle represents the discriminator. The training process is from left to right. The number N×N in the figure represent the spatial resolution of the input or the output.

As can be seen in Fig.4.2, we first start training our GAN at a small scale with only a few Conv and Deconv layers. The spatial resolution of the first step is 22×22. The original training data is down-sampled into this resolution. After some iteration, we add a new Conv and a new Deconv layer to the two ends of the generator, add a new Conv layer to the discriminator. In the second step, the spatial resolution is 48×48. After adding new layers for 5 times, the input resolution is 816×816, which is the spatial resolution of our generated data. In the progressive training stage, the number of convolutional layers before the bottleneck and the deconvolutional layers after the bottleneck are the same to ensure a mapping from input to output with the same size.

The model is trained on 30000 pairs of latent fingerprints and their corresponding ground truth. The optimizer we use is Adam Optimizer with learning rate 1e-4, β_1 =0.9, and β_2 =0.999. In the progressive growing stage, the GAN is trained for 1 epoch for each network scale and spatial resolution. After the growing training is completed, we then train the GAN for 10 epochs. The batch size is 8 and all training process is finished on a single NVIDIA 1080Ti GPU. After the progressive growing stage, we first train the orientation estimation task for 2 epochs and jointly train orientation task and enhancement task. The loss function and the model accuracy are shown in Fig.4.3. We can see from the loss curve that our model can converge smoothly.

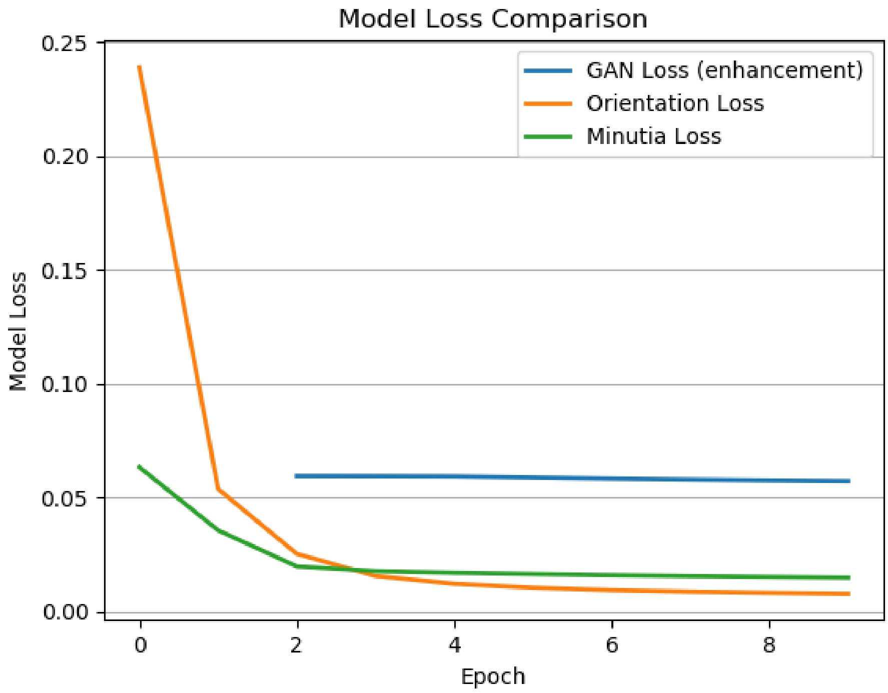


FIGURE 4.3 Model Loss function curve. Notice that minutia loss here denotes L1 loss of reconstruction. The enhancement task training is not started until the second epoch is finished, and pre-trained model of the generator is loaded.

4.4 Generation of Orientation Ground Truth

Orientation estimation has always been an important research topic because correctly estimation ridge orientation of fingerprint can help a lot in fingerprint image enhancement and fingerprint matching. In our model, we follow an end-to-end manner, which means the orientation estimation and enhancement share the same network. In fact, the enhanced fingerprint image and the estimated orientation map share the same representation. However, to train the neural network with supervised learning, we should also have the ground truth orientation as the label.

There is some work on orientation estimation already. Liu et al. [54] proposed Reliable Fingerprint Orientation Estimation Algorithm (RFOEA) and Fingerprint Orientation Restoration Algorithm (FORA). In their model, the orientation vector of the local patch is computed taking the hexagonal neighbors into consideration. Zhu et al. [55]proposed a systematic orientation estimation method which is still widely used today. In their model, the size of the estimated fingerprint image I is $m \times n$, $I(x, y)$ is the intensity of pixel (x, y). The image is first divided into patches with size $w \times w$. Therefore, we get $(m / w) \times (n / w)$ blocks. We can first compute the gradient vector of pixel (u, v) as Equation 4-7.

$$\begin{cases} G_x(u,v) = \dfrac{\partial I(u,v)}{\partial u} \\ G_y(u,v) = \dfrac{\partial I(u,v)}{\partial v} \end{cases} \tag{4-7}$$

After the gradient of each pixel is obtained, we can compute the orientation estimation within each segmented local patch with size $w \times w$ as follows:

$$\begin{cases} V_x = \sum_{u=i-w/2}^{i+w/2} \sum_{v=j-w/2}^{j+w/2} (G_x{}^2(u,v) - G_y{}^2(u,v)) \\ V_y = \sum_{u=i-w/2}^{i+w/2} \sum_{v=j-w/2}^{j+w/2} 2G_x(u,v)G_y(u,v) \end{cases} \tag{4-8}$$

The final orientation estimation can be computed with $\theta^I(x, y) = \tan^{-1}(V_y / V_x) / 2$. However, there are still some boundary conditions such as V_x or V_y equals to zero. Also, there should be no difference between θ and $\theta + \pi$. The final corrected orientation for each pixel is formulated in Equation 4-9.

$$O^I(i, j) = \begin{cases} \pi, & V_x = 0, V_y < 0 \\ 3\pi / 4, & V_x = 0, V_y \geq 0 \\ \theta^I(x, y) + \pi / 2, & V_x > 0 \\ \theta^I(x, y), & V_x < 0, V_y \leq 0 \\ \theta^I(x, y), & V_x < 0, V_y > 0 \end{cases} \tag{4-9}$$

Using this gradient based model, the estimated orientation field can be accurate when the noise and corruption are not severe. There are also some filter-based orientation estimation methods [57][58] to achieve a satisfying result in some specific scenarios. In our work, we follow the unsupervised orientation estimation method of sc_minutia [59] proposed by Joshua et al. Their orientation estimation model is similar to the gradient model above but they finish the fingerprint registration including orientation estimation and minutia extraction together, which will improve the orientation correctness. With the ground truth orientation map well prepared, we can begin the neural network training.

4.5 Experimental Results

To compare our multi-task progressive GAN with previous model, we use the same testing database and configuration. Some selected successfully enhanced latent fingerprint examples from NIST DB 27 are shown in Fig.4.4. It can be seen from the comparisons of gray enhanced result and the binary enhanced results that there is no blurry part in the enhancement. The results can also be compared to the results of our first GAN model which are shown in Fig.3.5. The enhanced ridge and valley structure are clearer and the noise is removed to a greater extent. The image quality achieves a significant improvement.

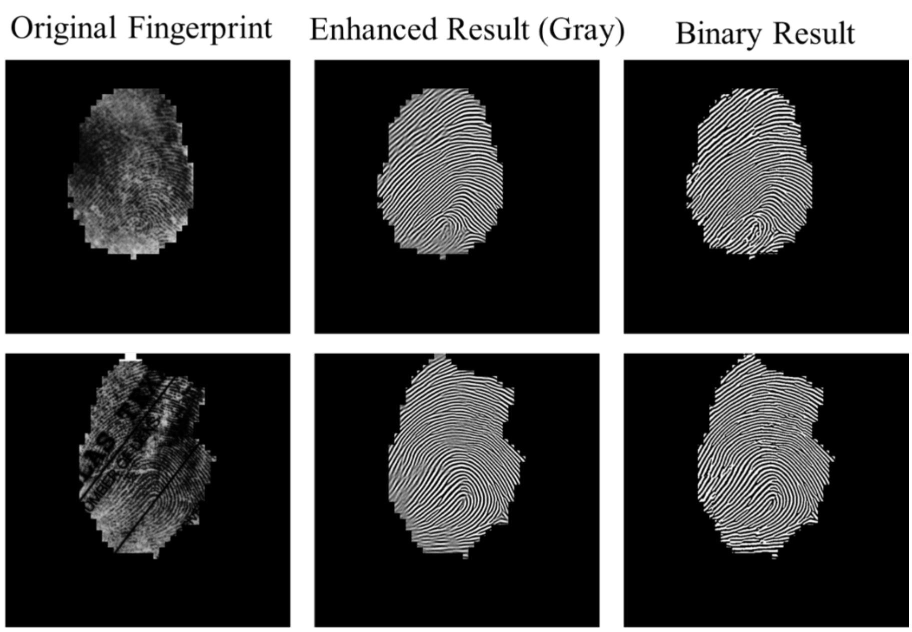


FIGURE 4.4 Selected successfully echanced example of NIST DB 27 latent fingerprint dataset for our multi-task progressive GAN model.

4.5.1 Ablation Experiments

Although from the enhancement results shown in Fig.4.4 we can see that the performance of our model improves, we still need to quantify the improvement and find out how each part of our model contributes to the final result. One way to achieve the evaluation is ablation experiments. In an ablation experiments, a specific part of our model is removed and other conditions keep the same. We can see from the change of results that if this part contributes to the result positively or negatively. Compare to our first GAN model, the major change in our multi-task progressive GAN is the multi-task learning scheme and the progressive growing training strategy. To show the effectiveness of these two parts of our method, we design some ablation experiments. The first and second experiment is to test the effectiveness of the multi-task learning scheme and progressive growing training. In these experiments, we compare the proposed multi-task learning scheme with those without orientation estimation and progressive training as well as the raw image. From the results shown in Fig.4.5 we can see that orientation estimation tasks and progressive growing training schemes play an important role in our model. We can conclude that orientation estimation shares the same representation with fingerprint enhancement and progressive growing training help stabilize our model and boost effectiveness.

In the IOT stage, the segmented latent fingerprint is input into the GAN trained in the POT stage. Generally, a fingerprint image after a single iteration of the image translation still suffers from unstructured noise. A solution to the problem is iteratively input the image into the neural network until the quality achieve a satisfying level. We adopt this iterative testing strategy in our method and achieve a better enhancement performance.

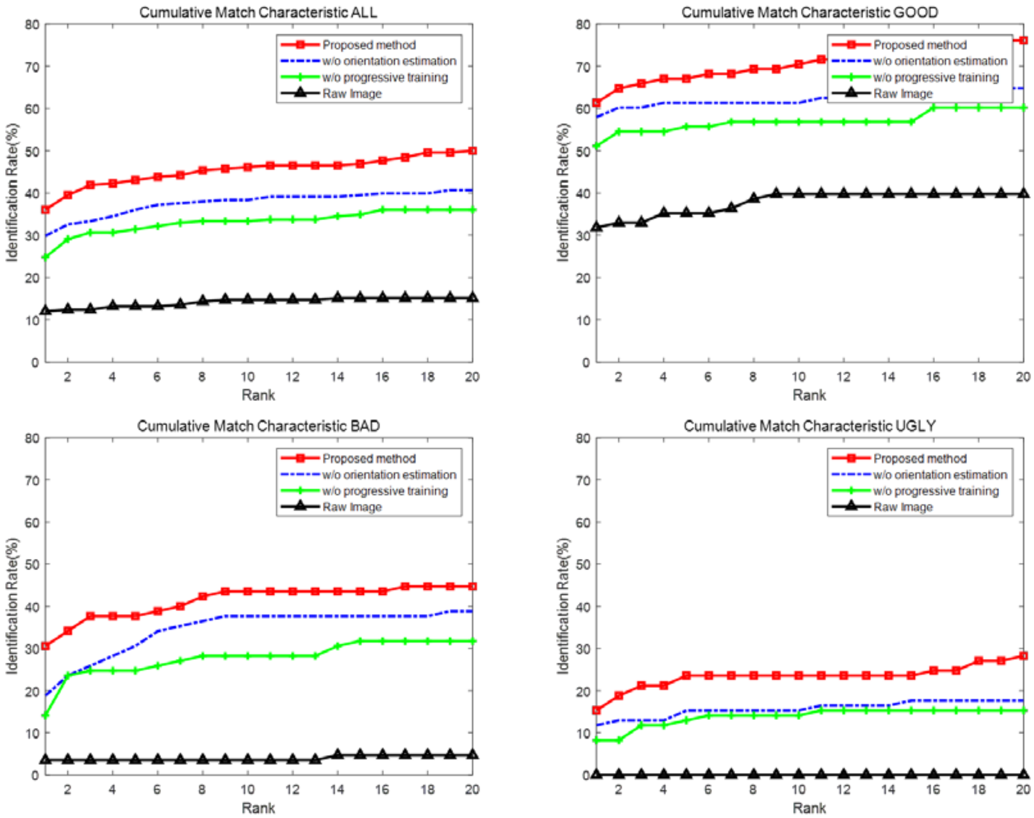


FIGURE 4.5 Ablation experiment results on Orientation estimation task and Progressive growing training scheme on all latent fingerprints and three subsets of different image quality. From the results, we can see that both the orientation estimation task and progressive growing training can improve the performance.

In addition, we also design the third experiment to compare the results with and without the iterative testing. Our iterative online testing (IOT) stage aims to iteratively enhance the latent fingerprint to make sure that the noise is removed as much as possible. From the enhancement results comparisons shown in Fig.4.6, we can see that the ridge pattern of an iteratively enhanced fingerprint is clearer and noise is removed to a greater extent.

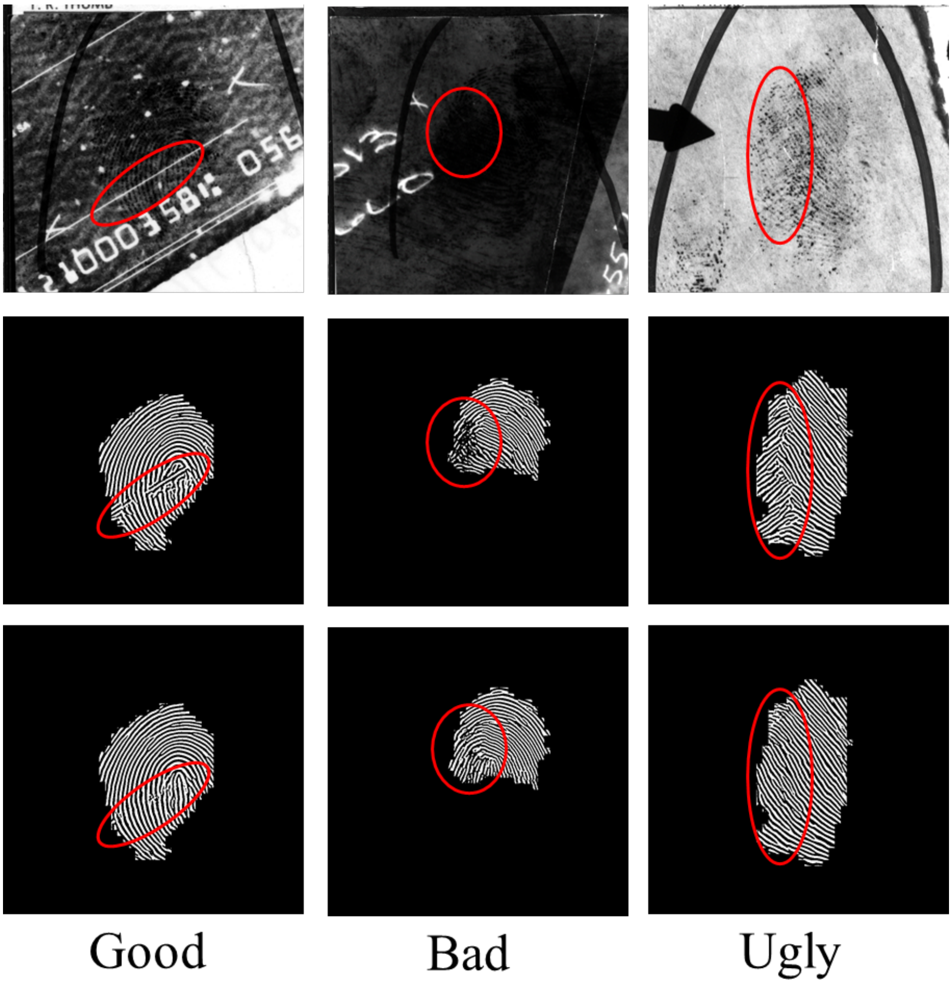


FIGURE 4.6 Ablation study on iterative online testing. Each column represents a sample image from the three subsets from NIST DB 27. The first, second, and third rows are the original latent fingerprint, the enhancement results without iteration, and the enhancement results with iteration, respectively. The red circle denotes the fingerprint region with significant improvement by iterative testing.

4.5.2 Comparisons with other methods

Finally, we compare our proposed method to the original raw image and other methods published in the literature[33] [32] [20]. These methods were also tested on the NIST 27 database. We directly used the results reported in the literatures for comparison. The Cumulative Match Characteristic (CMC) curves comparison on our test set are shown in Fig.4.7 and Fig.4.8. From the results we can see that our progressive GAN based method outperforms other enhancement algorithms significantly.

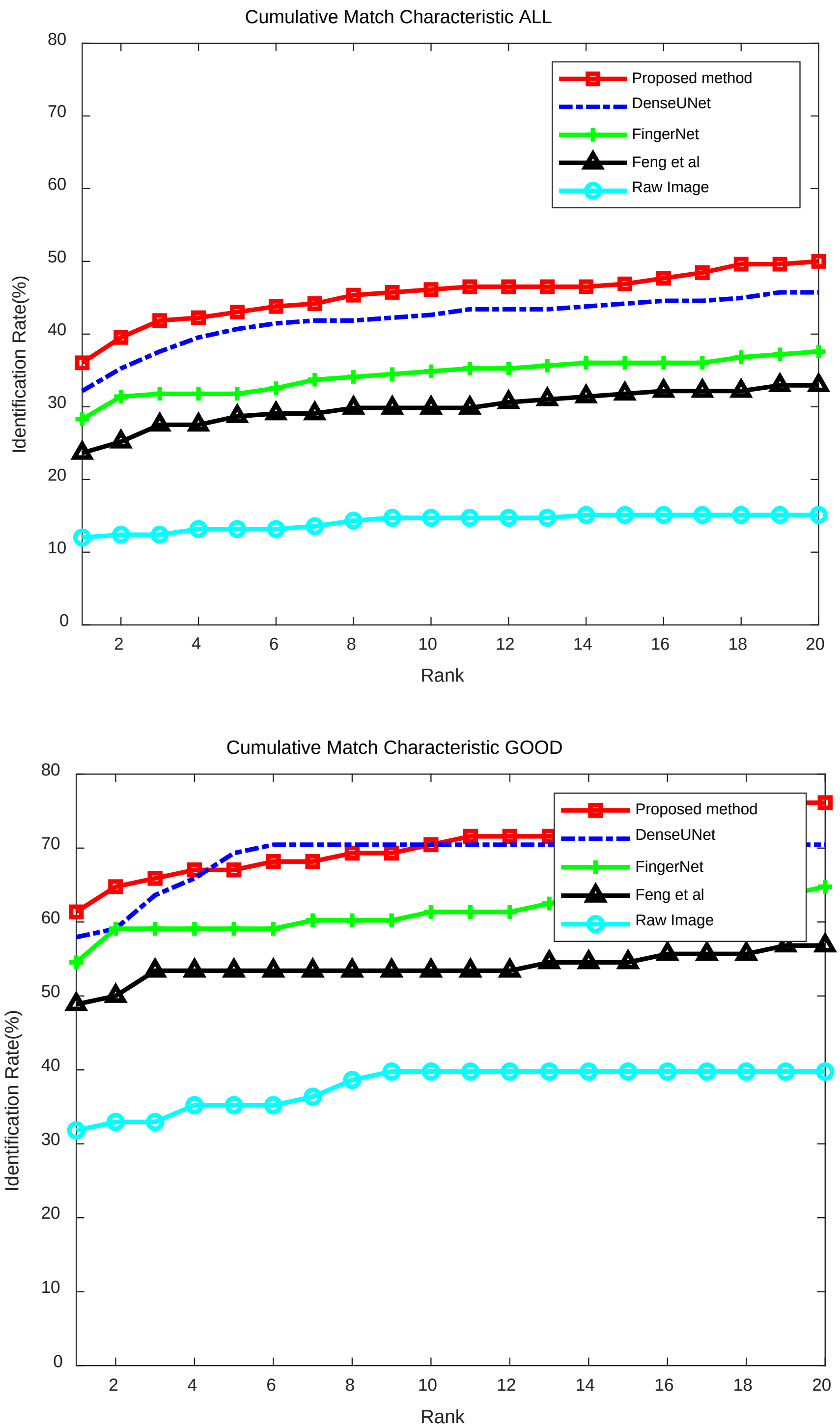
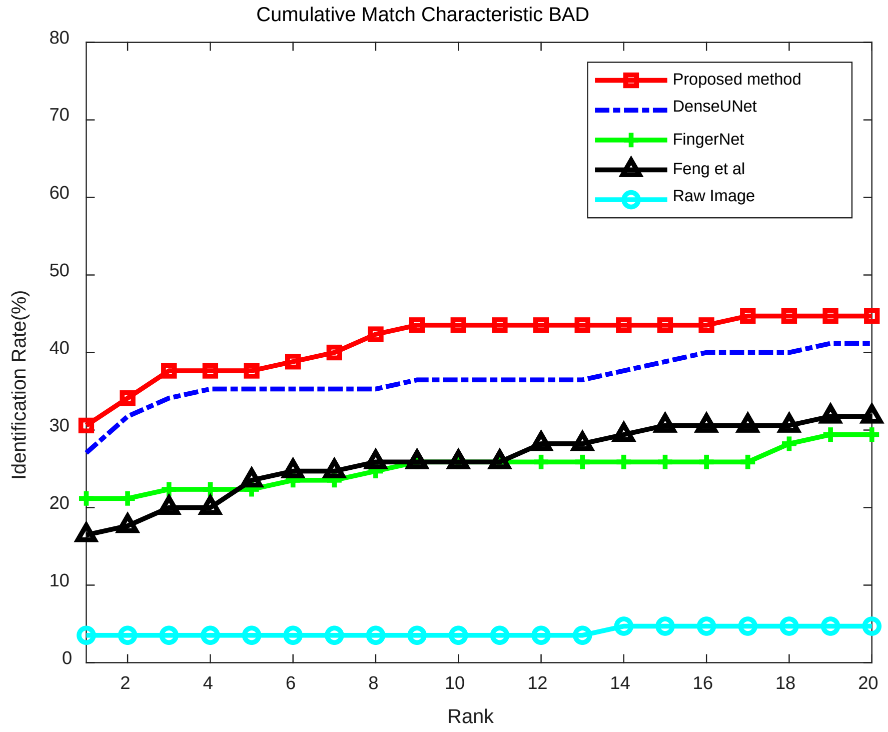


FIGURE 4.7 CMC curves on all images and "GOOD" subset comparing our method with previous latent fingerprint enhancement method DenseUNet [33], FingerNet [32], the method proposed by Feng et al. [6], and the raw image.



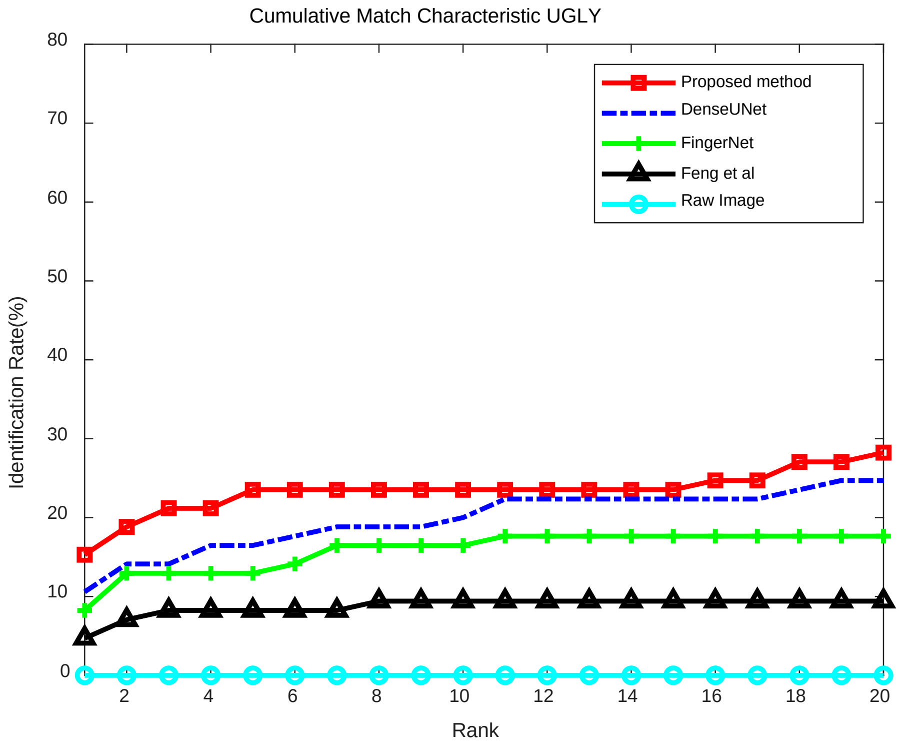


FIGURE 4.8 CMC curves on “BAD” and “UGLY” subset comparing our method with previous latent fingerprint enhancement method DenseUNet [33], FingerNet [32], method proposed by Feng et al. [6], and the raw image.

4.5.3 Efficiency

To look into the efficiency of our method, we compare the inference speed with other methods under the same configuration. As we can see in Tab.4.2, our progressive GAN based method has a significant boost in efficiency over the previous method. The high efficiency of our iterative online testing makes it possible for a more efficient real-time fingerprint identification and verification system.

Table 4-2 Inference time per image (sec.)

Method	ADTV [22]	Local Dict [61]	Ours
Test Speed	58.2	8.6	**0.034**

4.6 Chapter Summary

To draw a conclusion in the chapter summary, we improve the structure, objective function, training scheme of our vanilla generative adversarial network and proposed multi-task learning progressive generative adversarial network. We first introduce the core insight of our objective function design: the orientation estimation and fingerprint enhancement share the same representation in the neural network, which means that training orientation estimation at the same time with fingerprint enhancement will improve the final enhancement performance of our model. We illustrate the architecture and hyper parameter difference of our improved multi-task GAN model. The objective function is also given in detail: the loss function consists of three parts: adversarial loss, reconstruction loss, and orientation loss. To help stabilize the training of GAN and improve the performance on multi-scale patches, we adopt a progressive training scheme. We first train a few layers with relatively low-resolution training data down sampled from original one, and then increase the number of layers and the spatial resolution of training data accordingly. To show why multi-task learning and progressive training scheme is helpful, we design ablation experiments. The results show that model without orientation estimation task or progressive growing scheme will be outperformed by the multi-task progressive GAN significantly. We also compare our best results with previous published state-of-the-art results on NIST DB 27 dataset. Our model performs better than all other methods in term of both effectiveness (CMC curve) and efficiency (testing time per image).

Chapter Five　Conclusion and Future Work

In this thesis, we investigate the deep learning neural networks on the fingerprint image enhancement. The problem of fingerprint enhancement is formulated as an image-to-image translation problem. First, we develop a generative adversarial network (GAN) for latent fingerprint enhancement. Experiments are performed to show the effectiveness of the GAN on latent fingerprint enhancement. Second, we propose a multi-task learning GAN with progressive growing training for improving the latent fingerprint enhancement. The multi-task GAN network was built to fully exploit the shared representation between orientation estimation and enhancement and the progressive growing training scheme can stabilizes the training of GAN. Third, to train the proposed GAN network, we generate the training dataset synthetically including the latent fingerprints and the corresponding enhancement ground truths.

Finally, to show the effectiveness of our model, we evaluate our model on the NIST DB 27 dataset and apply the enhanced results to the fingerprint recognition application. We design ablation experiments to show the contribution of the orientation estimation task, progressive growing training scheme, and iterative testing. Our model outperforms the previous state-of-the-art method in terms of both recognition accuracy and efficiency.

The research on fingerprint enhancement and the generative adversarial network is a crucial research topic for biometric identification and deep learning. However, due to the limited time and computational resources for doing research, there are still some problems to be discussed.

Firstly, more types of low-quality fingerprint can be explored and tested. In this work, we used the latent fingerprints from NIST DB 27 to test the fingerprint enhancement algorithm. However, there are also different fingerprint datasets with different levels of noise and various extent of corruption. If the capacity of our progressive generative adversarial network model is strong enough, the performance is expected to achieve satisfying results on different samples.

Secondly, more tasks related to the fingerprint can be added into our multi-task learning scheme. Until now, we only apply the orientation estimation task in our network training. There are still some fingerprint-related tasks such as minutiae points detection, fingerprint segmentation, and quality estimation. Each of these tasks can be an optional component of the multi-task learning. All these fingerprint-related task share a similar representation of fingerprint including minutia and ridge pattern. The performance of our model is expected to be improved with these tasks.

REFERENCE

[1] Maltoni D, Maio D, Jain A K, et al. Handbook of fingerprint recognition[M]. Springer Science & Business Media, 2009.

[2] Chugh T, Cao K, Zhou J, et al. Latent fingerprint value prediction: Crowd-based learning[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(1): 20-34.

[3] Cole S A. Suspect identities: A history of fingerprinting and criminal identification[M]. Harvard University Press, 2009.

[4] Garris M D, Garris M D. NIST special database 27: Fingerprint minutiae from latent and matching tenprint images[M]. US Department of Commerce, National Institute of Standards and Technology, 2000.

[5] Watson C I. NIST Special Database 14. NIST Mated Fingerprint Card pairs 2 (MFCP2)[R]. 2016.

[6] Feng J, Shi Y, Zhou J. Robust and efficient algorithms for separating latent overlapped fingerprints[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(5): 1498-1510.

[7] Jain A K, Nandakumar K, Ross A. 50 years of biometric research: Accomplishments, challenges, and opportunities[J]. Pattern Recognition Letters, 2016, 79: 80-105.

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[9] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[10] Chen C, Seff A, Kornhauser A, et al. Deepdriving: Learning affordance for direct

perception in autonomous driving[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 2722-2730.

[11]Chen X, Ma H, Wan J, et al. Multi-view 3d object detection network for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1907-1915.

[12]Sun Y, Liang D, Wang X, et al. Deepid3: Face recognition with very deep neural networks[J]. arXiv preprint arXiv:1502.00873, 2015.

[13]Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[J]. 2015.

[14]LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

[15]Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[16]Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.

[17]Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.

[18]Zanfir A, Sminchisescu C. Deep learning of graph matching[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2684-2693.

[19]Hong L, Wan Y, Jain A. Fingerprint image enhancement: algorithm and performance evaluation[J]. IEEE transactions on pattern analysis and machine intelligence, 1998, 20(8): 777-789.

[20]Feng J, Zhou J, Jain A K. Orientation field estimation for latent fingerprint enhancement[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(4): 925-940.

[21]Liu M, Chen X, Wang X. Latent fingerprint enhancement via multi-scale patch based sparse representation[J]. IEEE Transactions on Information Forensics and Security, 2014, 10(1): 6-15.

[22] Zhang J, Lai R, Kuo C C J. Adaptive directional total-variation model for latent fingerprint segmentation[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(8): 1261-1273.

[23] Zhang J, Lai R, Kuo C C J. Latent fingerprint segmentation with adaptive total variation model[C]//2012 5th IAPR International Conference on Biometrics (ICB). IEEE, 2012: 189-195.

[24] Liu D, Wen B, Fan Y, et al. Non-local recurrent network for image restoration[C]//Advances in Neural Information Processing Systems. 2018: 1673-1682.

[25] Divakar N, Venkatesh Babu R. Image denoising via CNNs: an adversarial approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 80-87.

[26] Kim Y, Jung H, Min D, et al. Deeply aggregated alternating minimization for image restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6419-6427.

[27] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307.

[28] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3147-3155.

[29] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.

[30] Hui Z, Wang X, Gao X. Fast and accurate single image super-resolution via information distillation network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 723-731.

[31] Cao K, Jain A K. Latent orientation field estimation via convolutional neural network[C] //2015 International Conference on Biometrics (ICB). IEEE, 2015:

349-356.

[32] Li J, Feng J, Kuo C C J. Deep convolutional neural network for latent fingerprint enhancement[J]. Signal Processing: Image Communication, 2018, 60: 52-63.

[33] Qian P, Li A, Liu M. Latent Fingerprint Enhancement Based on DenseUNet[C]//2019 International Conference on Biometrics (ICB). IEEE, 2019: 1-6.

[34] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[35] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[36] Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks[C] //International Conference on Machine Learning. 2017: 214-223.

[37] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2794-2802.

[38] Berthelot D, Schumm T, Metz L. Began: Boundary equilibrium generative adversarial networks[J]. arXiv preprint arXiv:1703.10717, 2017.

[39] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[40] Lucic M, Kurach K, Michalski M, et al. Are gans created equal? a large-scale study[C] //Advances in neural information processing systems. 2018: 700-709.

[41] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans[C] //Advances in neural information processing systems. 2016: 2234-2242.

[42] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[C]//Advances in neural information processing systems. 2017: 6626-6637.

[43] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision

and pattern recognition. 2017: 1125-1134.

[44] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[45] Svoboda J, Monti F, Bronstein M M. Generative convolutional networks for latent fingerprint reconstruction[C]//2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017: 429-436.

[46] Liu Y, Tang Y, Li R, et al. Cooperative Orientation Generative Adversarial Network for Latent Fingerprint Enhancement[C]//2019 International Conference on Biometrics (ICB). IEEE, 2019: 1-8.

[47] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[48] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2536-2544.

[49] Wang X, Gupta A. Generative image modeling using style and structure adversarial networks[C]//European Conference on Computer Vision. Springer, Cham, 2016: 318-335.

[50] Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.

[51] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8798-8807.

[52] Durugkar I, Gemp I, Mahadevan S. Generative multi-adversarial networks[J]. arXiv preprint arXiv:1611.01673, 2016.

[53] Ghosh A, Kulharia V, Namboodiri V P, et al. Multi-agent diverse generative adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8513-8521.

[54] Liu L, Dai T S. A reliable fingerprint orientation estimation algorithm[J]. Journal of Information Science and Engineering, 2011, 27(1): 353-368.

[55] Zhu E, Yin J, Hu C, et al. A systematic method for fingerprint ridge orientation estimation and image segmentation[J]. Pattern recognition, 2006, 39(8): 1452-1472.

[56] Hou Z, Yau W Y, Wang Y. A review on fingerprint orientation estimation[J]. Security and Communication Networks, 2011, 4(5): 591-599.

[57] Kekre H B, Bharadi V A. Fingerprint orientation field estimation algorithm based on optimized neighborhood averaging[C]//2009 Second International Conference on Emerging Trends in Engineering & Technology. IEEE, 2009: 228-234.

[58] Wang Y, Hu J, Han F. Enhanced gradient-based algorithm for the estimation of fingerprint orientation fields[J]. Applied Mathematics and Computation, 2007, 185(2): 823-833.

[59] Abraham, Joshua, Paul Kwan, and Junbin Gao. "Fingerprint matching using a hybrid shape and orientation descriptor." State of the art in Biometrics (2011): 25-56.

[60] Yin, Wotao, Donald Goldfarb, and Stanley Osher. "A comparison of three total variation based texture extraction models." Journal of Visual Communication and Image Representation 18.3 (2007): 240-252.

[61] Yang X, Feng J, Zhou J. Localized dictionaries based orientation field estimation for latent fingerprints[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(5): 955-969.

[62] Bazen A M, Gerez S H. Segmentation of fingerprint images[C]//ProRISC 2001 Workshop on Circuits, Systems and Signal Processing. Veldhoven, The Netherlands, 2001: 276-280.

ACKNOWLEDGEMENTS

This study was supported in part by National Natural Science Foundation of China (NSFC) under grants No. 61773263. After a long period of hard work and diligence, I have finished my graduate thesis on fingerprint image enhancement, which is a challenging research topic. I have started my research in this field more than one year ago under the supervision of Prof. Manhua Liu. At the end of this thesis, I would like to show my gratitude to all the people who helped me in my undergraduate study and research life.

My deep gratitude goes first to the supervisor of my graduate thesis, Prof. Manhua Liu, for her constant guidance and instructive suggestion on my research. It was a milestone in my academic career when I first started working on fingerprint enhancement with Prof. Manhua Liu at SJTU. Although I have some experience in machine learning and image processing, I have no clear idea of what research area to choose as my undergraduate research topic. Prof. Manhua Liu provided me with treasured guidance when I felt confused. Also, her rigorous and insightful research style influence me a lot. I learn how to choose valuable ideas from all possible choice, how to think and write critically. I foster a lot of important scientific research skills, which will be my core competitiveness in my future PhD research life.

I would also express my great thanks to other research mentors of mine. Prof. Cewu Lu is a brilliant professor who achieves many accomplishments in the area of computer vision. I am really honored to be a member of Machine Vision and Intelligence Group (MVIG) where I started my first research project, submit and publish my first research paper. Prof. Mani Srivastava from University of California, Los Angeles is a great professor and can always give me useful suggestions. My research life in Los Angeles and UCLA is the most valuable experience in my life. All the professor in our department is helpful and warm-hearted, I would like to thank Prof. Yixin Ma, Yimei

Mao, Xuchu Mao, Zhengyu Huang…. Their courses and labs provide me solid math and electrical computer engineering knowledge foundation on research topics.

I am also fortunate to have excellent school mates, lab mates, and roommates. Peng Qian, Youglu Li, Liang Xu, Chao Xu from ZJU, Zhengbang Wu from Zju, and Xiao Liu from BNU always share their research ideas with me and serve as a role model in my research and study. My roommates Xiaoyu Zhang, Hongyuan Chen, and Chenzhang He accompanied me during the four years of my undergraduate life. I really enjoy the time we played and laughed together.

At last, I want to express my deepest gratitude to my beloved parents. In the difficult time of COVID-19 pandemic, my dad and mom provided me a great environment to conduct my research at home. They always support my choice and encourage me to pursue my dream. I want to say, Thanks, and I love you!

The finish of the thesis is a mark of the finish line of my undergraduate life at Shanghai Jiao Tong University. However, this is a new start for a new voyage to the new land of machine learning and artificial intelligence research. In my PhD studies, I will always remember what I have learnt in the pasting four years. At last, I want to say to myself: Stay humble and keep learning. We always keep minority spirits!

PUBLICATIONS

[1] **Huang X**, Qian P, Liu M H, et al. Latent Fingerprint Image Enhancement Based on Progressive Generative Adversarial Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Biometric Workshop. (CVPRW) 2020

[2] Li Y L, Zhou S, **Huang X**, et al. Transferable interactiveness knowledge for human-object interaction detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (CVPR) 2019

[3] Li Y L, Xu L, Liu X, **Huang X**, et al. PaStaNet: Toward Human Activity Knowledge Engine[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (CVPR) 2020

[4] Li Y L, Xu L, Liu X, **Huang X**, et al. Hake: Human activity knowledge engine[J]. arXiv preprint arXiv:1904.06539, 2019.

[5] **Huang X**, Alzantot M, Srivastava M. NeuronInspect: Detecting Backdoors in Neural Networks via Output Explanations[J]. arXiv preprint arXiv:1911.07399, 2019.