Advancing Fingerprint Recognition Quality Assessment: Introducing the FRBQ Metric for Enhanced Fingerprint Recognition

Prateek Jaiswal*
prateek.jaiswal@research.iiit.ac.in
CVIT, IIIT Hyderabad
India

Arka Koner
arka.koner@uidai.net.in
UIDAI (Aadhaar)
India

Anoop M. Namboodiri
anoop@iiit.ac.in
CVIT, IIIT Hyderabad
India

Abstract

In the field of biometric security, the quality assessment of fingerprint images is paramount for boosting the accuracy of fingerprint recognition systems. These systems are fundamental for the secure and efficient authentication and identification of individuals. Our research presents FRBQ (Fingerprint Recognition-Based Quality), an innovative quality metric designed to navigate the limitations of the NFIQ2 model. FRBQ exploits deep learning algorithms in a weakly supervised setting and utilizes matching scores from DeepPrint, a Fixed-Length Fingerprint Representation Model. Each score is paired with labels indicating the robustness of fingerprint image matches. However, in a fully referenced setting, these labels can be subjective, lacking a clear definition of what "image quality" inherently means. This weakly labeled approach strives to capture diverse perspectives on image quality, potentially making it a more encompassing metric. In comparison to NFIQ2, our research showcases the superior performance of the FRBQ model. It not only correlates better with recognition scores but also effectively evaluates challenging images that NFIQ2 struggles with. Validated by the esteemed FVC 2004 dataset, FRBQ proves its efficacy in fingerprint image quality assessment. This study underscores the transformative potential of AI in biometrics, emphasizing its capability to capture details that traditional methods might overlook. Our work stresses the critical role of precise quality assessment in the evolution of fingerprint recognition systems.

CCS Concepts

• **Computing methodologies** → **Computer Vision, Biometrics**.

Keywords

Fingerprint Image Quality, Fingerprint Recognition System, Image Quality Assessment, Weakly Supervised Learning

ACM Reference Format:
Prateek Jaiswal, Arka Koner, and Anoop M. Namboodiri. 2023. Advancing Fingerprint Recognition Quality Assessment: Introducing the FRBQ Metric for Enhanced Fingerprint Recognition. In *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '23), December 15–17, 2023, Rupnagar, India.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3627631.3627649

*Corresponding author

ICVGIP '23, December 15–17, 2023, Rupnagar, India

ACM ISBN 979-8-4007-1625-6/23/12.
https://doi.org/10.1145/3627631.3627649

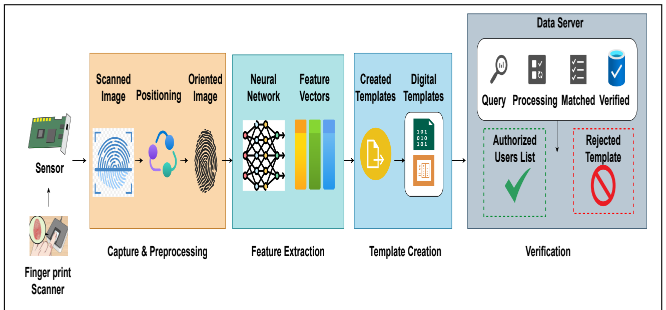

Figure 1: AFIS(Automated Fingerprint Identification System): The system scans a user's fingerprint, processes it for quality features using a neural network, and generates a feature vector. This vector creates templates which are compared to a database for matching. Based on this, the system decides to accept or reject the input.

1 Introduction

Deep-learning models have revolutionized the field of biometrics in recent years. They have shown state-of-the-art performance in face recognition [7, 21, 27] tasks, including image quality assessment, and have several advantages over traditional methods [32, 33], such as greater accuracy and the ability to handle large amounts of data. These models have also been applied to other biometric modalities, such as fingerprint and iris recognition, with promising results. The integration of deep learning into biometrics has enabled the development of more robust and reliable biometric systems that can be used in various applications, including security, authentication, and identification. Fingerprint recognition is widely used for biometric authentication, but the accuracy of fingerprint recognition and verification [12] systems depends heavily on the quality of the fingerprint images used [10]. The diagram in Figure 1 illustrates an AFIS system for fingerprint recognition and verification in a biometrics system. Traditional methods [32, 33] of assessing image quality rely on handcrafted features and necessitate the use of reference images to compare the quality of the test image. However, these methods have limitations in dealing with different image distortions, such as blur, noise, and compression artifacts.

No-Reference Fingerprint Image Quality Assessment: Fingerprint recognition [15, 23] relies heavily on quality features and traditional methods use handcrafted ISO NFIQ [32, 33] features to set quality parameters. This paper introduces a deep learning-based approach called FRBQ for no-reference image quality assessment of fingerprint images. The objective of this work is to assess the recognition performance of fingerprint images across all quality score thresholds, regardless of whether the score is high or low. As far as we know, this is the first study that can determine the recognition ability of a fingerprint image based solely on its quality score. FRBQ employs a CNN architecture [1, 20] to predict the quality scores of fingerprint images, enabling the assessment of recognition performance across all quality score thresholds.

In this paper, we present our contribution in the form of FRBQ, a simple quality network for fingerprint image quality assessment. The key features of this model are as follows:

- FRBQ generates quality scores for fingerprint images that reflect their recognition ability.
- The model is fine-tuned on ResNet18 deep learning architecture and does not require ground truth labels for training. Instead, the matching score is used as a proxy ground truth in a weakly supervised manner.
- FRBQ perform accurately even on low-resolution images where state-of-the-art models fail to perform.

The effectiveness of FRBQ was evaluated on the FVC 2004 [14] fingerprint image dataset, and it outperformed the state-of-the-art NFIQ2 [32, 33] in accuracy and robustness to various image distortions. This method can serve as a preliminary step for fingerprint recognition systems, ensuring that the biometric authentication process is accurate. The structure of the paper is as follows: Section 2 provides a concise overview of the current state of research on no-reference image quality assessment. Section 3, discussed the rationale behind FRBQ and introduces the proposed method. Section 4 outlines the experimental methodology, presents the results, and includes additional study or ablation studies. Finally, in Section 5, summarize the key findings and conclude the paper.

2 Related Works

In scenarios where explicit label information is not available, No Reference Image Quality Assessment (NR-IQA) plays a crucial role [2, 13, 16, 34]. To address this challenge, adopting a weakly supervised learning approach has been proposed [19, 38]. Additionally, Remy et al. have explored fingervein quality assessment [24], and Oblak et al. have conducted a comprehensive survey on deep learning ensemble models for fingerprint image quality assessment [18]. These studies offer valuable insights and alternative methods to enhance recognition accuracy in scenarios with limited label information.

Deep Learning in Biometric Applications: DL models have shown significant improvement in accuracy and robustness in various biometric applications, including face recognition, fingerprint recognition, and iris recognition. These models have been shown to outperform traditional feature extraction and classification methods by learning complex representations directly from the raw biometric data. This has led to a paradigm shift in biometric recognition, where deep learning models are becoming the go-to solution for a wide range of biometric recognition tasks. However, as with any technology, there are still challenges to be addressed, such as ensuring the privacy and security of biometric data, addressing potential biases and discrimination, and improving the interpretability and transparency of deep learning models.

In recent years, deep learning models [11, 25, 29, 31] have displayed immense potential in resolving various computer vision tasks, including image quality assessment. Specifically, no-reference image quality assessment, [2, 13, 16, 26] which does not require a reference image, has gained significant attention.

Deep learning models such as DeepPrint [8] have been utilized for fingerprint recognition to improve accuracy. DeepPrint extracts important information from fingerprint images without extracting varying feature information through a fixed-length fingerprint representation. It combines deep features extracted by deep networks with minutiae handcrafted features to get the critical feature information. Four different networks are used to extract fingerprint features as well as deeply learned features to prevent overfitting and extract interpretable deep features. The architecture of DeepPrint uses multitask branches to first extract the fingerprint image representation and subsequently learn how to classify it. The information from both branches is then fused to get a recognition score. Prior to DeepPrint, other deep networks were used to improve specific sub-modules of fingerprint recognition systems such as segmentation [5, 9, 30, 39], orientation field estimation [3, 22, 28], and minutiae extraction [4, 6, 17, 35, 37].

MiDeCon [36] is one of the few works that have utilized an approach incorporating minutiae information for quality score generation in fingerprint recognition. However, this approach has not been widely implemented, and DeepPrint provides a more advanced method for learning quality features. By incorporating minutiae maps during training, DeepPrint's feature-guided deep network has shown superiority over traditional methods in quality-based recognition tasks. As a result, DeepPrint is used in this study to generate matching scores and create proxy ground truth labels for accurate quality score prediction.

3 Methodology

This section provides an overview of the main methodology proposed in the paper and the architecture details of the deep learning models utilized in the approach. Subsection 3.1 presents the rationale behind the proposed approach, followed by a detailed description of the proposed method and loss function used to fine-tune the model in Subsection 3.2. We discuss the importance of using label information to improve the accuracy of the model in predicting the quality score in Subsection 3.3. Finally, in Subsection 3.4, we explore various approaches to calculate match scores and strategies to manage non-mated pairs for recognition.

3.1 Rationale

The assessment of fingerprint image quality holds paramount significance in biometric recognition systems. The proposed methodology leverages deep learning networks for learning fingerprint image features in a weakly supervised setting. This initiative roots in the concept of matching, training the network to yield a quality score simultaneously valid for fingerprint matching. Utilizing simulated ground truth information from labeled data, the network

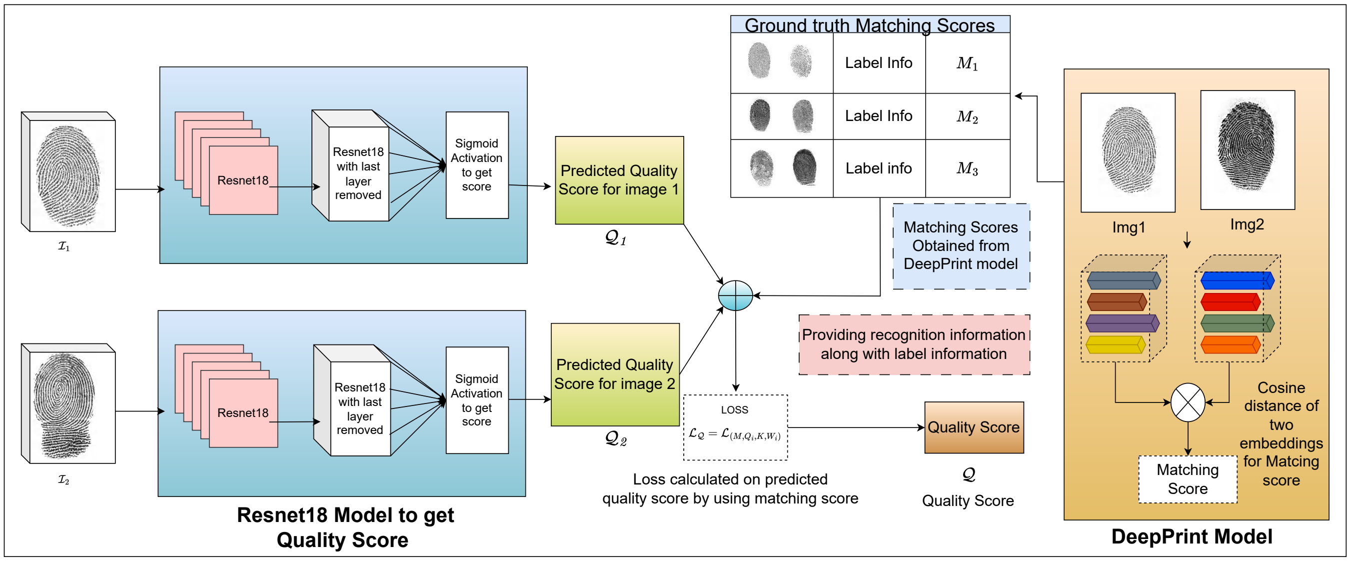

Figure 2: In this proposed method, the ResNet neural network takes in the input images, ($\mathcal{I}^1$) and ($\mathcal{I}^2$), and produces predicted quality scores, ($\mathcal{P}_I^1$) and ($\mathcal{P}_I^2$). These predicted scores are used in a loss function, along with proxy ground truth quality scores that are matching scores ($\mathcal{M}$) with their label info, generated by the DeepPrint model, in order to train the ResNet to produce scores that are as close as possible to the ground truth. The output of this process is a fingerprint quality score, (Q).

adeptly learns features specific to fingerprint images, enhancing the accuracy and reliability of the assessment.

3.2 FRBQ Method

The application of this innovative approach involves training a deep learning network on the FVC dataset of fingerprint images employing a weakly supervised learning algorithm, illustrated in Figure 2. The network, designed to extract valuable features from raw fingerprint image data, is further refined with ground truth information on the matching score and label information from DeepPrint [8]. Post-training, the network is adept at generating quality scores for new fingerprint images, offering a robust assessment of their matching suitability.

The distinctive methodology of DeepPrint [8] is exhibited in Figure 3, explicating the calculation of the matching score. Utilizing DeepPrint for creating matching scores and proxy ground truth labels enhances the learning of quality features as it incorporates a minutiae map during training. This inclusion embeds crucial minutiae information into the network, leading to a feature-guided deep network with superior performance in quality-based recognition tasks.

/subsectionCalculation of Quality Score

The network processes two distinct input images, $\mathcal{I}^1$ and $\mathcal{I}^2$, predicting a quality score for each, denoted as $\mathcal{P}^1$ and $\mathcal{P}^2$. Utilizing the matching score $\mathcal{M}$ from DeepPrint, a proxy ground truth for quality is obtained. This process necessitates both images to ensure the precise calculation of the quality score. Alongside the matching score, label information $\mathcal{K}$ augments the accuracy of the quality score calculation. The harmonic mean of the quality score, rather than the average matching score, is utilized for calculating the matching score. This use of harmonic average emphasizes the lower values in the set, capturing the impact of low scores more prominently, and consequently offering a more precise quality score for images with limited matching ability.

Loss Function Overview: The employed network, as illustrated in Figure 2, incorporates a specially designed loss function denoted as $\mathcal{L}_{(M,Q_i,K,W_i)}$ or $\mathcal{L}_Q$ to enhance its fine-tuning. In this notation:

- Q_i specifies the two input images.
- M specifies the matching score of these images.
- K indicates whether the two images are from the same class or different classes.
- W_i weights for balancing.

The primary objective of this loss function is to ensure robust performance under imbalanced data conditions. It aims to minimize the discrepancy between the predicted quality scores and the true quality of the input images, particularly in scenarios where:

- The label is 0 (indicating high quality) with a low matching score.
- The label is 1 (indicating low quality) and the matching score is low.

Conversely, the loss should increase when:

- There's a low matching score combined with a label of 1.
- There's a high matching score paired with a label of 0.

This loss function is applied as binary cross-entropy loss for labels 0 and 1.

Expressed as $\mathcal{L}_Q(\mathcal{M}, Q_1, Q_2, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1)$,:

$$\mathcal{L}_i(\mathcal{M}, Q_i, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1) = \log Q_i \left(-\mathcal{W}_1 \mathcal{M} + (1-\mathcal{K})\mathcal{W}_0(1-\mathcal{M})\right), \ \forall i = 1, 2 \tag{1}$$

$$\mathcal{L}_Q(\mathcal{M}, Q_1, Q_2, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1) = \mathcal{L}_1 + \mathcal{L}_2 \tag{2}$$

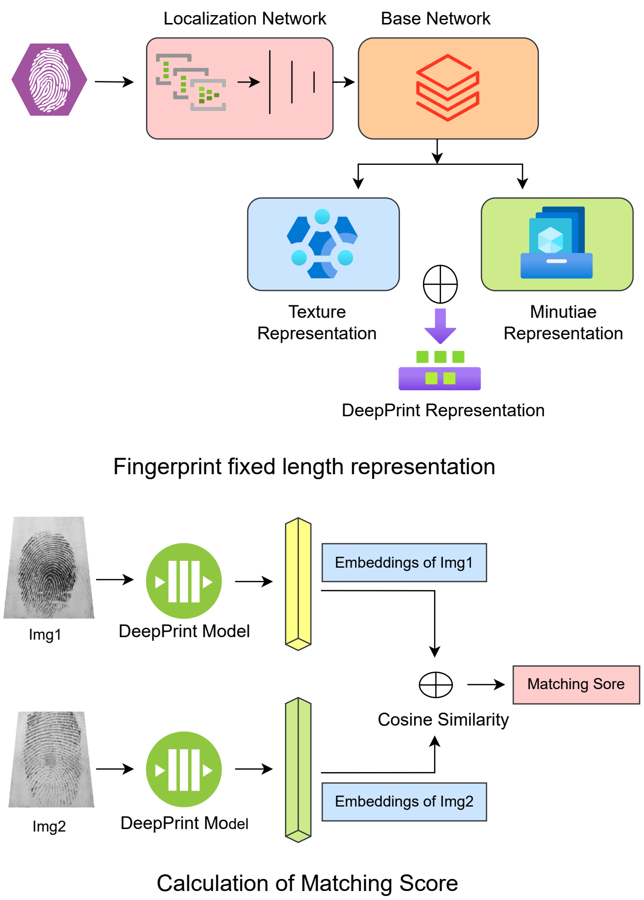

Figure 3: Block 1: The input image, $\mathcal{I}$, undergoes processing to generate a 192-dimensional vector. Texture and minutiae details are extracted and combined to form a DeepPrint representation $\mathcal{F}_d$. Block 2: Embedding vectors from DeepPrint are compared using cosine similarity to produce a matching score for fingerprint comparison.

$\mathcal{L}_Q$ is a combination of the above two losses where $\mathcal{L}_1$ is calculated for Q_1 the predicted quality of *image-1* and the $\mathcal{L}_2$ is calculated on Q_2 the predicted quality of *image-2* where $\mathcal{M}$ represents the Matching score obtained on a pair of images that are used as a proxy ground truth to finetune the network. Q_1 and Q_2 represent the Quality Score of *image-1* and *image-2* respectively and $\mathcal{K}$ represents the Label information that is 0 for different images and 1 for same images. When $\mathcal{K}$ is 1 loss functions try to reduce the loss directly proportional to $\mathcal{M}$ in predicting the Q_1 and Q_2 and when $\mathcal{K}$ is 0 loss function learns the loss in inversely proportional and try to reduce loss incorporates in predicting the Q_1 and Q_2 based on the inverse of $\mathcal{M}$.

Weight Calculation: To address the unbalanced nature of the training data, weights $\mathcal{W}_0$ and $\mathcal{W}_1$ are assigned to balance the loss function. The weights are determined based on the relative proportions of the samples in each class. Specifically, they are calculated as:

$$\mathcal{W}_k = \frac{1}{N_k}\left(\frac{N}{2}\right) \tag{3}$$

Where:

- k represents the class label (either 0 or 1).
- N is the total number of samples in the dataset.
- N_k denotes the count of samples in class k.

The weights $\mathcal{W}_0$ and $\mathcal{W}_1$ effectively balance the contribution of each class by inversely scaling with their prevalence, ensuring that neither class dominates the loss due to its abundance or scarcity.

3.3 Label Information effect on Quality Scores

In the context of fingerprint image quality assessment in our research, we employ proxy ground truths obtained from a DeepPrint matcher. The DeepPrint matcher leverages label information to enhance recognition accuracy concerning quality scores.

Labelling of Pairs: Table 1 presents an overview of the labeling process. Our approach involves training a deep learning model using labeled data that consists of pairs of fingerprint images. Each pair can either be from the same finger or from different fingers. The DeepPrint matcher utilizes this label information to distinguish between genuine pairs and impostors.

Training with Labels: During model training, we incorporate these fake ground truth labels, which are derived from matching scores obtained from the DeepPrint matcher. These scores serve as indicators of the likelihood of a pair of fingerprint images matching, with high scores signifying a good match and low scores indicating a poor match. This approach allows us to harness recognition-based information that cannot be obtained through manual image quality annotation, enabling our model to predict the quality score of fingerprint images accurately.

3.4 Analytical Study on Match Score:

We have explored alternative approaches that show promise for future applications, especially in the context of large fingerprint image datasets. These methods not only open up new avenues for further research but also provide valuable insights into the potential enhancements of fingerprint image analysis on a larger scale. In this paper, we have introduced an approach that involves utilizing a pair of images during model training. However, for scenarios where only a single image is available, it becomes essential to derive a representative score that indicates its recognizability. This score guides the model in accurately assessing image quality.

To compute a match score for a single image, it is imperative to gather a collection of match samples encompassing a wide range of both good and bad matches.

Match Score of Single Image To derive a single representative score from multiple match scores, statistical methods such as mean, harmonic mean, and median percentile prove valuable. The selection between these methods depends on the specific dataset characteristics and requirements. Here's a concise overview of both methods

- **Harmonic Mean:**
  - The harmonic mean is a type of average that gives more weight to lower values in the set. It is particularly effective at handling situations where extreme outliers or very low scores need to be taken into account.
  - The harmonic mean is suitable when the impact of the worst or lowest scores in the dataset. It can help in scenarios where you want to ensure that the overall quality is not solely driven by a few high scores.
  - It is important to note that the harmonic mean is sensitive to extremely low scores, and a single low score can significantly affect the resulting average.
- **Percentile:**
  - The percentile represents the relative position of a particular score within a distribution. It indicates the percentage of scores that are equal to or below a given value.
  - Using percentiles allows you to determine where a specific score lies within the distribution of all scores. It helps capture the overall quality of a matched image compared to others in the dataset.
  - Percentiles are useful when you want to establish a threshold or cutoff point to classify images as poor or good quality based on their relative position in the score distribution.

The choice between the harmonic mean and percentile depends on the specific requirements and characteristics of the dataset, as well as the significance that we want to place on different scores.

Dealing with non-mated pairs: In this paper, we have conducted a comprehensive study encompassing the utilization of both mated and non-mated pairs, which are commonly referred to as genuine and impostor pairs, respectively. This inclusion allows us to thoroughly assess the performance of our proposed method across diverse scenarios, reflecting real-world fingerprint recognition challenges.

- **Exclusive Mated Pair Analysis**
  - Looking ahead, we envision conducting further experiments focused on exclusively using mated pairs. This aligns with real-world scenarios where the primary objective is to verify the matching of images for authorized individuals.
  - On the other hand, non-mated pairs present a distinct challenge, where the emphasis on recognizability may not be as critical.

  For these cases, we consider the possibility of assigning lower or even negative recognizability scores, given that these pairs do not belong to authorized individuals or are not among the designated matches.
- **Non-Mated Image Quality Assessment:**
  - One key aspect is the ability to predict scores for non-mated pairs, potentially indicating zero or negative recognizability. It would greatly impact the quality score, as it would enable us to identify images that do not match any individual.
  - Although, we recognize that solving this complex problem for a single image without using additional references presents significant challenges. Notwithstanding, we maintain a positive outlook on future developments that could potentially empower us to forecast such results.

Label Info	Matching Score	Quality Score
0	High	Low
1	High	High
0	Low	High
1	Low	Low

Table 1: The table shows that when the matching score is high and the label information is positive, the quality score is high. Conversely, when the matching score is low and the label information is negative, the quality score is also low.The matching scores are used along with label information to accurately predict the quality score of fingerprint images.

Database	Genuine Pair	Impostor Pairs	Total Pairs
DB_A	2800	4950	7750
DB_B	280	45	325

Table 2: FVC dataset information

By conducting these additional experiments, we aim to gain a deeper understanding of the effectiveness of our method across different use cases, and refine its performance to suit specific fingerprint recognition scenarios. This research contributes to the broader field of fingerprint image quality assessment and holds promise for advancing the reliability and accuracy of fingerprint recognition systems in practical applications.

4 Experiments

This section begins by introducing the datasets that were used for the experiments in 4.1, followed by a detailed description of the implementation process and the workflow in 4.2. 4.3 discusses about the details of experimental setup. The findings of the experiments conducted with this approach are also discussed in 4.4. Finally, the limitations of the proposed approach are discussed in 4.5.

4.1 Dataset

The experiments were conducted on the widely used FVC 2004[14] dataset, which is commonly used for evaluating fingerprint images. Table 2 provides information on the FVC 2004 dataset, which comprises four different databases of various sensor types. The dataset includes two databases, DB_A, and DB_B. DB_A was used in the study and provided 7751 genuine and impostor image pairs, while DB_B provided 326 image pairs. All images in the dataset had a resolution of over 500 DPI, which is recommended by NFIQ for high-quality capture.

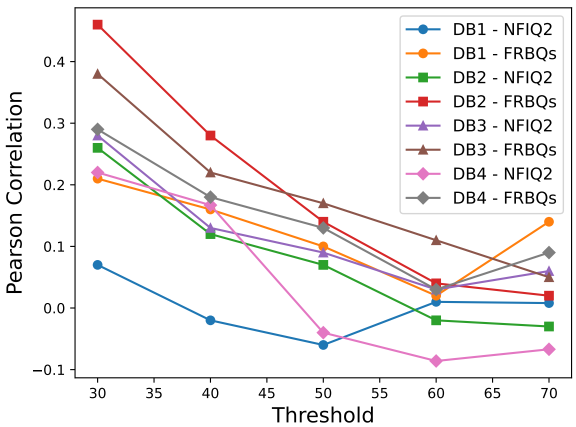

Figure 4: The performance of two quality score methods, NFIQ2, and FRBQ, were evaluated using matching scores generated from DeepPrint on four databases in the FVC Dataset A. The correlation between the quality scores and matching scores was analyzed at different quality thresholds. Results showed that FRBQ had higher correlation scores than NFIQ2 at varying quality thresholds for all the datasets.

4.2 Workflow and Implementation details

To finetune the ResNet18 model for fingerprint recognition, a substitute ground truth is given for the quality score. The matching score generated by the DeepPrint model is used to accomplish this requirement. By applying the DeepPrint model, the matching score have been calculated and then used with label information that determines if the image pair was genuine or an impostor. If the label was 1, the pair belonged to the same image set and was, therefore, genuine. Conversely, if the label was 0, the pair was from different image sets and was, hence, an impostor pair.

Utilizing DeepPrint-Based Substitute Ground Truth: It utilizes this data to establish a dependable proxy ground truth for quality score, which can be used to train the ResNet18 model. This methodology helped us surmount the limitations of the NFIQ2 baseline, which proved insufficient in providing accurate quality scores for fingerprint recognition.

The DeepPrint-based substitute ground truth was used to evaluate the performance of FRBQ, which proved to be a promising approach in accurately assessing the quality of fingerprint images. The matching scores for sample image pairs along with their label information are presented in Table 3. Furthermore, the table displays the NFIQ2 and FRBQ scores for the same image pairs.

4.3 Experiment Setup

In the experimental setup, we utilized a ResNet model with pretrained weights. Fine-tuning was conducted using a subset of 7750 pairs from Database A. For comprehensive details regarding the dataset employed in our experiments, please refer Table 2. This table provides the information about the FVC 2004 dataset, encompassing the count of genuine and impostor image pairs, sensor types, and recommended image resolution.

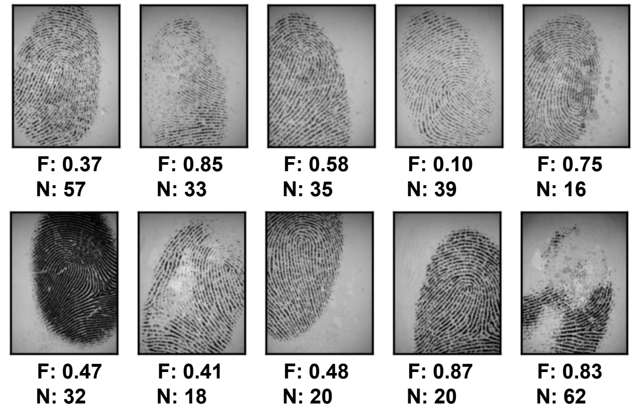

Figure 5: The quality scores for fingerprint images are obtained using two different methods, namely NFIQ2 and FRBQ, referenced as N and F, respectively.

Learning Representation: ResNet models are widely used for image classification tasks. In these models, the initial layers learn general features such as edges and textures, while the later layers focus on more specific features like object parts and shapes. By freezing the initial layers, the model can leverage these general features, which have shown to be beneficial for predicting image quality.

Finetuning of model: Image pairs are processed individually through the network. The predicted scores, matching scores, and label information are combined in a loss function to learn the quality score for recognition. The DeepPrint model is initially used to create a proxy ground truth for fine-tuning the ResNet model. All matching scores for the image pairs are collected. Subsequently, the ResNet model is modified by replacing the last layer with a 1-dimensional fully connected layer. To improve prediction accuracy, only the output from this last fully connected layer is utilized while freezing the initial layers.

Prediction of Score: The quality score is determined by passing the image through the ResNet model, which predicts its probability score for quality. With the trained FRBQ (Fingerprint Recognition Based Quality) model, accurate predictions of quality scores are made, reflecting the recognition performance of the images.

Setup: *The code was implemented using the PyTorch library and run on an NVIDIA GPU. Fine-tuning was done using Adam optimization. The hyperparameters were set as follows: Number of epochs=100 epochs, a learning rate=*10^{-5}*, and a batch size=8.*

4.4 Findings

The images were compared in pairs by taking the average of their scores to ensure fairness. A score above a pre-defined threshold was considered to be good quality and therefore a good match. DeepPrint matching scores were also obtained, and their correlation with the quality scores was evaluated using Pearson correlation coefficients. The results are presented in the correlation graphs

Img1					
Img2					
Matching Score	0.25	0.66	0.42	0.05	0.27
FRBQs Img1	0.24	0.53	0.41	0.51	0.53
NFIQ2 Img1	56	39	71	36	26
FRBQs Img2	0.51	0.61	0.53	0.57	0.37
NFIQ2 Img2	72	19	61	23	5
Label	1	1	1	0	0

Table 3: The table displays data on image pairs including their matching scores, FRBQs, and NFIQ2 scores, as well as label information for each pair. NFIQ2 scores range from 0 to 100, while FRBQs scores range from 0 to 1 and indicate the recognition ability of the quality score. When matching scores are low for different images, high-quality pairs of images are identified through FRBQs scores. Additionally, FRBQs scores cover all cases of lower matching scores for the same images.

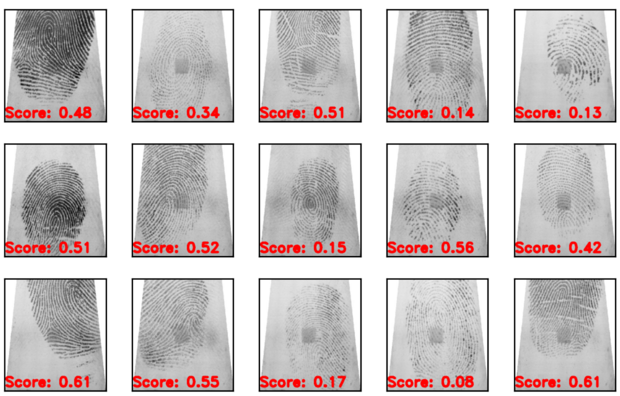

Figure 6: FRBQ was able to provide quality scores for images that NFIQ2 was unable to predict scores on, even though these images were highly distorted and noisy.

in Figure 4. These graphs, which also include data from Figure 5 (showing the quality scores for fingerprint images obtained using NFIQ2 and FRBQ, referenced as N and F, respectively), were generated by varying the quality threshold for different datasets to compare FRBQ performance against NFIQ2.

Reliability of Score: NFIQ2 scores are only reliable indicators of recognition accuracy when the quality of the fingerprint image is high, and even then, they do not work for all cases. When the quality score is less than 40, NFIQ2 scores fail to predict the recognition performance of the image.

The results of the study showed that the model was able to accurately identify images with good matching scores as good quality. The model has a better correlation score with the DeepPrint matcher, for all the different quality thresholds (in the same scale of 0 to 100) when compared to NFIQ2. This indicates that the proposed approach has the potential for accurately assessing the quality of good matches and their potential for use in fingerprint-matching systems.

Table 3 displays quality scores generated by FRBQ, outperforming NFIQ2 scores.

- The FRBQ model's optimal threshold of 0.4 was determined by filtering FVC 2004 databases and assessing performance at different levels.
- FRBQ performed better than NFIQ2 in predicting recognition performance for all different quality thresholds, making it a more useful approach for fingerprint matching suitability. Additionally, the quality score generated by FRBQ effectively predicts the matching score, demonstrating the usefulness of combining quality and matching information for fingerprint recognition tasks.

The methodology unfolds by utilizing label information coupled with matching scores to derive accurate quality scores. Nevertheless, navigating through the assessment of quality becomes increasingly complex when confronted with low matching scores and image quality.

Two Distinct Scenarios Unfold:

(1) Low quality may be ascribed to either or both images within a pair, potentially leading to an inaccurate high matching score due to noise. In a high-score scenario, both images might rightfully share comparable quality or, if of low quality, may falsely show high matching scores.
(2) To tackle these challenges, the deployment of a comprehensive and varied dataset is essential. This dataset should encompass all possible scenarios of image quality and matching scores, fortified with a sufficient number of image pairs, allowing the model to thoroughly evaluate each facet of label information alongside matching scores.

The model must be robust to efficiently counteract noise, ensuring a clear distinction between genuine and noise-induced matches. This enhancement ensures accurate quality assessments even under tough conditions.

The depicted histograms clearly present the average matching scores, ranging from 0 to 100, alongside quality scores from 0 to 1. As shown in Figure 7, the distribution highlights an important feature: quality scores fill the range where matching scores are absent, showcasing their strong predictive ability even with limited matching information.

4.5 Discussion and ablation study

This study employed NFIQ2 as a benchmark, highlighting its limitations in delivering accurate fingerprint image quality scores.

- The usage of a pre-established version, NFIQ2.2.0, further underscored these constraints. The study unveiled NFIQ2's inability to assign quality scores to certain images, specifically those of low resolution or poor capture quality, consequently assigning a score of zero.
- In stark contrast, FRBQ emerged as a robust alternative, providing substantial scores for these otherwise ungraded images, as visually represented in Figure 6. This aspect accentuates FRBQ's superiority over NFIQ2 in meticulously assessing fingerprint image quality, thereby potentially enhancing the precision of fingerprint recognition systems.
- The robust and consistent performance of FRBQ, even for images that NFIQ2 could not adequately assess, underscores its reliability and precision. Preliminary findings posit FRBQ as a potential leader in fingerprint recognition, especially in situations involving low-quality or improperly captured images.

This research significantly contributes to forging more reliable and precise fingerprint recognition systems, finding extensive application in diverse domains such as law enforcement, security, and access control.

Model Training: Figure 8 present the loss and training curves of the FRBQ model, offering crucial insights into its training performance. The loss curve illustrates the progression of optimization, whereas the training accuracy curve reveals the model's learning capacity from the training data. Concurrently, the validation accuracy curve indicates the model's generalization capabilities to unseen data, providing a comprehensive understanding of the model's convergence and effectiveness in ensuring elevated accuracy across both training and validation datasets.

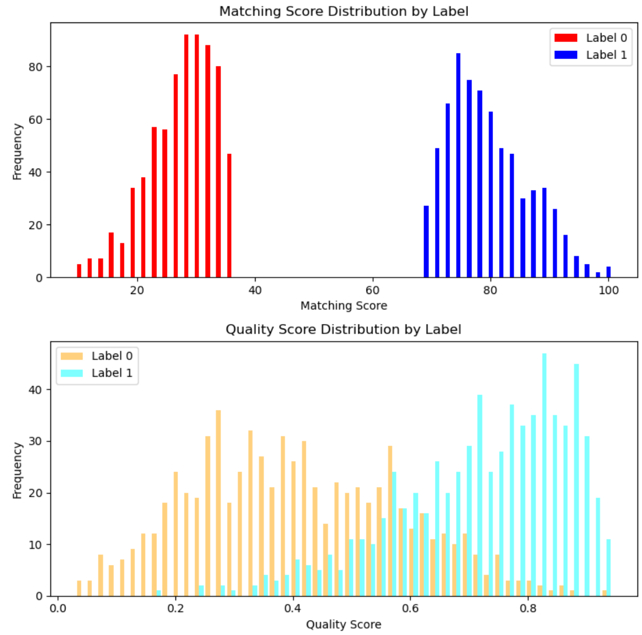

Figure 7: The histogram illustrates the match scores derived from the DeepPrint model and the corresponding quality scores obtained from our model, demonstrating improved separation between label 0 scores and label 1 scores.

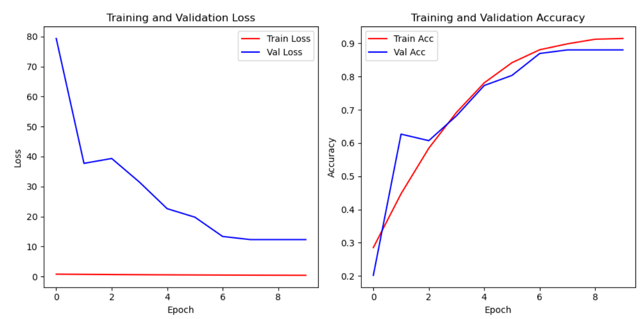

Figure 8: Training loss and accuracy results.

5 Conclusion

In conclusion, the proposed FRBQ model for assessing fingerprint image quality with deep learning networks shows substantial promise. It successfully generates dependable quality scores for new fingerprint images and surpasses established models. This approach offers numerous advantages, such as enhancing fingerprint matching systems, diminishing manual inspection and labeling, and seamlessly integrating with current recognition systems. Including label information and matching scores enriches the quality assessments' comprehensiveness. Nevertheless, assessing low-quality images remains a notable challenge. It requires an intricate understanding of the tie between image quality and matching scores. The answer to these challenges lies in a sophisticated algorithm, capable of handling data complexity and integrating additional information, like feature data. Moreover, an extensive training dataset could offer deeper insights into issues related to low-quality fingerprints, enabling the overcoming of proxy ground truth limitations. In essence, the FRBQ model holds the potential to refine the accuracy of fingerprint recognition systems and smoothen the quality assessment process. However, surmounting the challenges posed by low-quality images necessitates a deeper exploration into the correlation between image quality and matching scores. Future research efforts should focus on developing more robust and efficient algorithms to enhance the efficacy of FRBQ in fingerprint recognition systems.

References

[1] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. 2017. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*. Ieee, 1–6.

[2] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. 2017. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing* 27, 1 (2017), 206–219.

[3] Kai Cao and Anil K Jain. 2015. Latent orientation field estimation via convolutional neural network. In *2015 International Conference on Biometrics (ICB)*. IEEE, 349–356.

[4] Kai Cao and Anil K Jain. 2018. Automated latent fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence* 41, 4 (2018), 788–800.

[5] Xiaowei Dai, Jie Liang, Qijun Zhao, and Feng Liu. 2017. Fingerprint segmentation via convolutional neural networks. In *Biometric Recognition: 12th Chinese Conference, CCBR 2017, Shenzhen, China, October 28-29, 2017, Proceedings 12*. Springer, 324–333.

[6] Luke Nicholas Darlow and Benjamin Rosman. 2017. Fingerprint minutiae extraction using deep learning. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 22–30.

[7] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4690–4699.

[8] Joshua J Engelsma, Kai Cao, and Anil K Jain. 2019. Learning a fixed-length fingerprint representation. *IEEE transactions on pattern analysis and machine intelligence* 43, 6 (2019), 1981–1997.

[9] Jude Ezeobiejesi and Bir Bhanu. 2017. Latent fingerprint image segmentation using deep neural network. *Deep Learning for Biometrics* (2017), 83–107.

[10] Patrick Grother and Elham Tabassi. 2007. Performance of biometric quality measures. *IEEE transactions on pattern analysis and machine intelligence* 29, 4 (2007), 531–543.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[12] Anil K Jain, Arun Ross, and Salil Prabhakar. 2004. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology* 14, 1 (2004), 4–20.

[13] Le Kang, Peng Ye, Yi Li, and David Doermann. 2014. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1733–1740.

[14] Dario Maio, Davide Maltoni, Raffaele Cappelli, Jim L Wayman, and Anil K Jain. 2004. FVC2004: Third fingerprint verification competition. In *Biometric Authentication: First International Conference, ICBA 2004, Hong Kong, China, July 15-17, 2004. Proceedings*. Springer, 1–7.

[15] Davide Maltoni, Dario Maio, Anil K Jain, Salil Prabhakar, et al. 2009. *Handbook of fingerprint recognition*. Vol. 2. Springer.

[16] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* 21, 12 (2012), 4695–4708.

[17] Dinh-Luan Nguyen, Kai Cao, and Anil K Jain. 2018. Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. In *2018 International Conference on Biometrics (ICB)*. IEEE, 9–16.

[18] Tim Oblak, Rudolf Haraksim, Peter Peer, and Laurent Beslay. 2022. Fingermark quality assessment framework with classic and deep learning ensemble models. *Knowledge-Based Systems* 250 (2022), 109148.

[19] Maxime Oquab, Léon Bottou, Ivan Laptev, and Josef Sivic. 2015. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 685–694.

[20] Keiron O'Shea and Ryan Nash. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458* (2015).

[21] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep face recognition. (2015).

[22] Zhenshen Qu, Junyu Liu, Yang Liu, Qiuyu Guan, Ruikun Li, and Yuxin Zhang. 2018. A novel system for fingerprint orientation estimation. In *Image and Graphics Technologies and Applications: 13th Conference on Image and Graphics Technologies and Applications, IGTA 2018, Beijing, China, April 8–10, 2018, Revised Selected Papers 13*. Springer, 281–291.

[23] Nalini Ratha and Ruud Bolle. 2003. *Automatic fingerprint recognition systems*. Springer Science & Business Media.

[24] Oliver Remy, Jutta Hämmerle-Uhl, and Andreas Uhl. 2022. Fingervein Sample Image Quality Assessment using Natural Scene Statistics. In *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 1–6.

[25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115 (2015), 211–252.

[26] Michele A Saad, Alan C Bovik, and Christophe Charrier. 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE transactions on Image Processing* 21, 8 (2012), 3339–3352.

[27] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.

[28] Patrick Schuch, Simon-Daniel Schulz, and Christoph Busch. 2017. Deep expectation for estimation of fingerprint orientation fields. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 185–190.

[29] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[30] Nancy Singla, Manvjeet Kaur, and Sanjeev Sofat. 2020. Automated latent fingerprint identification system: A review. *Forensic science international* 309 (2020), 110187.

[31] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[32] Elham Tabassi. 2014. Towards NFIQ II Lite: Self-organizing maps for fingerprint image quality assessment. (2014).

[33] Elham Tabassi and Charles L Wilson. 2005. A novel approach to fingerprint image quality. In *IEEE International Conference on Image Processing 2005*, Vol. 2. IEEE, II–37.

[34] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE transactions on image processing* 27, 8 (2018), 3998–4011.

[35] Yao Tang, Fei Gao, Jufu Feng, and Yuhang Liu. 2017. FingerNet: An unified deep network for fingerprint minutiae extraction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 108–116.

[36] Philipp Terhörst, André Boller, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. 2021. Midecon: Unsupervised and accurate fingerprint and minutia quality assessment based on minutia detection confidence. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 1–8.

[37] Song Wu, Bing Liu, Zheng Wang, Zexi Jia, and Jufu Feng. 2022. Minutiae-awarely Learning Fingerprint Representation for Fingerprint Indexing. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 1–8.

[38] Zhi-Hua Zhou. 2018. A brief introduction to weakly supervised learning. *National science review* 5, 1 (2018), 44–53.

[39] Yanming Zhu, Xuefei Yin, Xiuping Jia, and Jiankun Hu. 2017. Latent fingerprint segmentation based on convolutional neural networks. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE, 1–6.