Latent Fingerprint Image Quality Assessment Using Deep Learning

Jude Ezeobiejesi and Bir Bhanu
Center for Research in Intelligent Systems
University of California at Riverside, Riverside, CA 92521, USA
e-mail: jezeobie@cs.ucr.edu, bhanu@cris.ucr.edu

Abstract

Latent fingerprints are fingerprint impressions unintentionally left on surfaces at a crime scene. They are crucial in crime scene investigations for making identifications or exclusions of suspects. Determining the quality of latent fingerprint images is crucial to the effectiveness and reliability of matching algorithms. To alleviate the inconsistency and subjectivity inherent in feature markups by latent fingerprint examiners, automatic processing of latent fingerprints is imperative. We propose a deep neural network that predicts the quality of image patches extracted from a latent fingerprint and knits them together to predict the quality of a given latent fingerprint. The proposed approach eliminates the need for manual ROI markup and manual feature markup by latent examiners. Experimental results on NIST SD27 show the effectiveness of our technique in latent fingerprint quality prediction.

1. Introduction

The accuracy of latent fingerprint identification by latent fingerprint forensic examiners has been the subject of increased study, scrutiny, and commentary in the legal system and the forensic science literature. Errors in latent fingerprint matching can be devastating, resulting in missed opportunities to apprehend criminals or wrongful convictions of innocent people. Latent fingerprint image quality assessment provides an indication as to whether the latent fingerprint is a good candidate for further analysis and feature annotations. Figure 1 shows latent fingerprints of different qualities. Currently, latent fingerprint examiners assign one of the following values to a given latent fingerprint image: value for individualization (VID), value for exclusion only (VEO), and no value (NV). Latent fingerprints marked as VID have sufficient salient information for matching. Latent fingerprints identified by latent examiners as VEO and NV are generally considered to be valuable and are subject to further processing [5]. As reported by Yoon et. al. [15], 63% of VEO latents in NIST SD27 [3] and WVU [2] la-

tent fingerprint databases can be identified at rank 100 while 40% can be identified at rank 1. Incorrect NV determination for a latent fingerprint could result in missed opportunity to identify a crime suspect.

To process latent fingerprints, latent experts manually mark the regions-of-interest (*ROIs*) in latent fingerprints and use the *ROIs* to search large databases of reference fingerprints and identify a small number of potential matches for manual examination. Given the large size of law enforcement databases containing rolled and plain fingerprints and the inconsistency and subjectivity inherent in feature markups by latent fingerprint examiners, it is very desirable to perform latent fingerprint processing in a fully automated way. The determination of the quality of latent fingerprint images is an essential step in automatic processing of latent fingerprints.

This paper proposes a deep learning model for latent fingerprint quality assessment that eliminates the need for manual feature markup. The first stage in our model uses deep learning to segment a latent fingerprint. Feature vectors computed from the segmented latent fingerprint are used as input to a multi-class perceptron that predicts the quality of the fingerprint. Experimental results on NIST SD27 fingerprint database show the promise of the proposed approach. NIST SD27 database is the most suitable database for this work because all the latent fingerprint images in it have quality labels assigned by latent experts. *To the best of our knowledge, no previous work [5, 14, 15] on latent fingerprint image quality assessment performs latent fingerprint region-of-interest segmentation and quality assessment in a lights-out mode (minimal involvement of latent examiners). This work requires no manual ROI and feature markups by latent examiners in the segmentation and quality assessment steps.*

The rest of this paper is organized as follows. Section 2.1 presents a review of existing algorithms for latent fingerprint image quality estimation. Section 2.2 describes the contributions of this paper while Section 3 highlights our technical approach and framework. Section 3.1 presents an overview of Restricted boltzman machine (RBM) used to
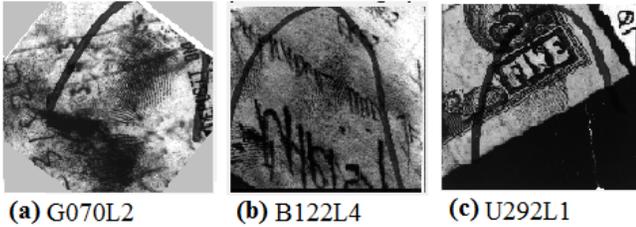
Figure 1. NIST SD27: Latent fingerprints images of different qualities: (a) good, (b) bad, and (c) ugly.

build the deep learning model and as well as a brief description of the segmentation stage of our framework. The quality assessment stage is discussed in Section 3.3. Section 3.3.1 discusses the features used to train the quality estimation neural network layer. Experimental results and performance evaluation are presented in Section 4, while Section 5 contains the conclusions and future work.

2. Related Work and Contributions

2.1. Related Work

Fingerprint quality assessment has received considerable attention in the literature [4]. Some recent studies on latent fingerprint quality assessment used local image features for quality assessment while others used global image features. The work presented in [15] used average ridge clarity, number of manually annotated minutiae, ridge connectivity, minutiae reliability, and finger position to define the quality of latent fingerprint. The authors used a semi-automated quality assessment algorithm and achieved 80% quality prediction accuracy. However, their use of manually annotated minutiae makes their quality assessment results fraught with subjectivity. The method presented in [5] used number of minutiae, ridge clarity, core and delta, and ridge flow features for automatic latent value determination. Although their value determination algorithm required no manual feature markups, it still relied on manually marked ROI for segmentation . In [14], the authors used ridge clarity and ridge quality features to assess the quality of latent fingerprints. Their approach required manually annotated minutiae and manually marked ROI. Chugh et. al [7] used a crowdsourcing based framework and multidimensional scaling to identify and understand how fingerprint experts assigned values to fingerprint images. They trained a prediction model that automatically assigned quantitative values to query latent fingerprints.

In our work, we use local features consisting of Gabor features, orientation certainty level, local ridge clarity, ridge frequency, ridge thickness, ridge-to-valley thickness, and spatial coherence to assess the quality of latent fingerprints. Unlike most of the other approaches that rely on manually

segmented ROIs in the quality estimation process, our approach performs latent fingerprint quality assessment in a fully automated way. In the first stage of our approach, we segment the latent fingerprint ROIs using deep learning as described in Section 3.1. The segmented ROIs are split into 32x32 patches and local features are computed from the patches to build feature vectors used to train a multi-class perceptron classifier as detailed in Section 3.3. The classification results are used to assess the quality of the latent fingerprint. Note that this work does not consider overlapped latent fingerprints.

2.2. Contributions

The paper makes the following contributions:

1. Poses latent fingerprint image quality assessment as a classification problem and solves it by using a deep neural network built by stacking RBMs. The depth chosen for our network was the one that gave the best performance and was found via experimentation. The depth is optimal for the problem being solved since going deeper did not yield appreciable performance gains and took longer to converge.

2. Unlike previous approaches, this work provides a region-of-interest based latent quality assessment strategy that requires no human intervention in latent fingerprint quality determination. The segmentation of the latent fingerprint and its quality assessment are done with no manual intervention or feature markups.

3. Technical Approach

Our latent fingerprint quality assessment architecture has two main stages. In the first stage, we use deep learning to segment the latent fingerprint. This stage involves feature learning, feature extraction and classification of the fingerprint patches into fingerprint and non-fingerprint classes. The segmented latent fingerprint referred to as the regions-of-interest (ROIs) consists of patches classified as fingerprint. In the second stage, we use a multi-class perceptron classifier to classify the fingerprint patches into three bins labelled 1 (good), 2 (bad) and 3 (ugly). The quality of the latent fingerprint is indicated by the label of the bin that contains the greatest number of patches. Ties are broken optimistically as explained in section 3.3. The block diagram of our proposed approach is shown in Figure 2.

3.1. Segmentation using Deep Learning

Restricted Boltzmann Machines (RBMs) are the building blocks for the proposed deep learning model. RBM is a stochastic neural network in which the nodes form an undirected bipartite graph. With RBM, a k-dimensional input can be mapped to a j-dimensional or m-dimensional feature
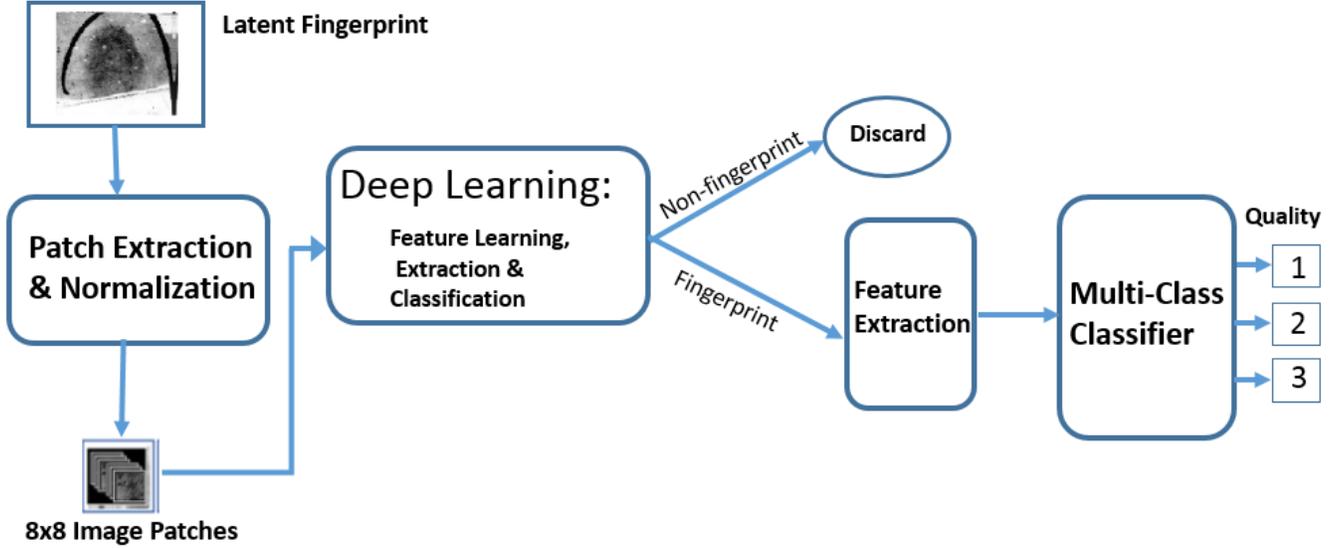
Figure 2. Proposed framework for latent fingerprint quality assessment. The first stage uses a deep learning architecture similar to that in [9], for feature learning, extraction and classification. In the second stage, features are extracted from the segmented fingerprint (ROI) and fed to a multi-class perceptron classifier. The target values from the classifier are 1 (Good), 2 (Bad) and 3 (Ugly).

space, where $j < k < m$. RBM has no intra-layer connections and given the visible unit activations, the hidden unit activations are mutually independent. Also the visible unit activations are mutually independent given the hidden unit activations [6]. These characteristics of RBMs make them ideal for identity mapping. From experiments, neural networks built with RBMs are suitable for learning input representations that can be used to reconstruct the inputs with minimal reconstruction error. This makes such networks attractive for patch based latent fingerprint segmentation.

The segmentation stage of the proposed model is similar to that in [9]. In this stage, latent fingerprint image is partitioned into 8x8 non overlapping patches. Stochastic features that model a distribution over image patches are learnt using a generative multi-layer feature extractor. The features are used to train a single layer perceptron classifier that classifies the patches into fingerprint and non-fingerprint classes. The fingerprint patches are used to reconstruct the latent fingerprint image and the non-fingerprint patches which contain the structured noise in the original latent fingerprint are discarded. The segmented latent fingerprints from this stage are used as inputs to the quality assessment stage. The choice of patch size of 8x8 for the segmentation stage is based on its optimality [8].

3.2. Network Hyper-Parameters

The values of the parameters used in the proposed segmentation and quality assessment networks are shown in Table 1 and Table 2, respectively. The values were selected through experiments.

Segmentation Network							
Parameter	L_1	L_2	L_3	L_4	L_5	L_6	L_7
Number of Neurons	64	800	1000	1200	1024	1024	2
Batch Size	-	100	100	100	100	100	-
Epochs	-	50	50	50	50	-	-
Learning Rate	-	1e-3	5e-4	5e-4	5e-4	5e-4	-
Momentum	-	0.70	0.70	0.70	0.70	0.70	-
Iteration	-	-	-	-	-	50	-

Table 1. Parameters and values for segmentation network. L_i refers to layer i. L_1 is the input layer. Layers 2, 3, 4 and 5 are RBM layers. L_6 is the perceptron layer and L_7 is the output layer

Quality Assessment Network			
Parameter	$InputLayer$	$HiddenLayer$	$OutputLayer$
Number of Neurons	64	450	3
Batch Size	-	32	-
Epochs	-	10	-
Transfer function	-	logsig	tansig

Table 2. Parameters and values for the quality assessment network.

3.3. Quality Assessment

In the quality assessment stage, 32x32 patches are extracted from the segmented (only fingerprint segments) ROIs and features computed from them are used as the quality assessment training dataset. The choice of 32x32 is based on the fact that for 500 pixels per inch (ppi) images, the width of a pair of ridge and valley is 8 to 12 pixels wide [13]. This implies that a patch size of at least 24x24 pixels is required to cover two ridges with a valley in between. Given a segmented latent fingerprint image L, let g, b, u be the number of its 32x32 patches classified into bins B_1, B_2, B_3, respectively. Let $val = max\{g, b, u\}$. The
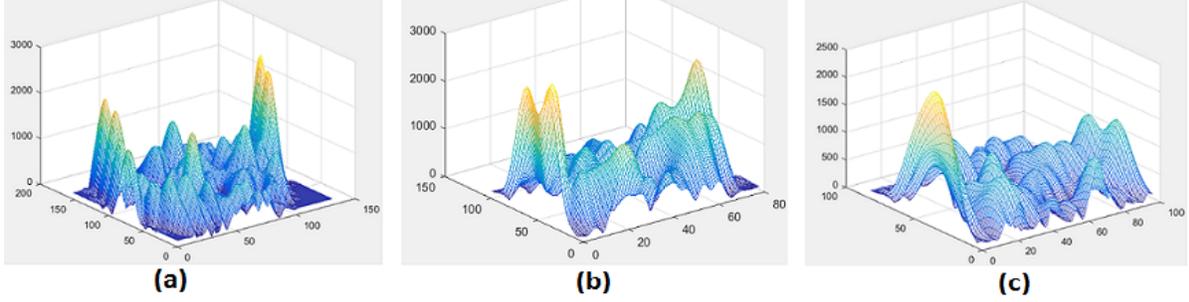
Figure 3. Gabor magnitude responses to sample segmented fingerprints : (a) Good (b) Bad, and (c) Ugly). As can be seen from the figures, good quality patches have more well-defined peaks than the bad and ugly patches. Also the peaks in (b) are more distinctive than in (c).

Features	Description
Peak Kurtosis	Kurtosis of image patch magnitude and phase response
Mean Kurtosis	
Peak Skewness	Skewness of image patch magnitude and phase response
Mean Skewness	
Ridge frequency [12]	Values computed from a sinusoidal model of ridges and valleys in the image patch
Ridge thickness	
Ridge-to-valley thickness	
Orientation certainty level	Measure of orientation strength
Spatial coherence	Computed from the gradient of image patch

Table 3. Local features used for latent quality assessment.

quality of L is defined as:

$$Q(L) = \begin{cases} 1, & \text{if val = g;} \\ 2, & \text{if val = b;} \\ 3, & \text{if val = u.} \end{cases} \quad (1)$$

Ties are broken in an optimistic manner. For example, if $g = b$ and $b > u$, then $Q(L) = 1$.

3.3.1 Features used for Quality Assessment

The local features used for latent fingerprint quality estimation are shown in Table 3.

We use kurtosis and skewness of the magnitude and phase of Gabor filter response to measure the local quality of an image patch. Skewness is defined as a measure of symmetry [1]. A distribution is symmetric if the left and right sides of its central point are similar. Kurtosis is a normalized form of the fourth central moment of a distribution and a measure of how heavy-tailed or light-tailed a given distribution is relative to a normal distribution. Given a vector $V = \{v_1, v_2, \cdots, v_k\}$, the skewness S and kurtosis K are defined as:

$$S = \frac{\frac{1}{|V|} \sum_{j=1}^{|V|} (v_i - \hat{v})^3}{\sigma_3}, \quad (2)$$

$$K = \frac{\frac{1}{|V|} \sum_{j=1}^{|V|} (v_i - \hat{v})^4}{\sigma_4}, \quad (3)$$
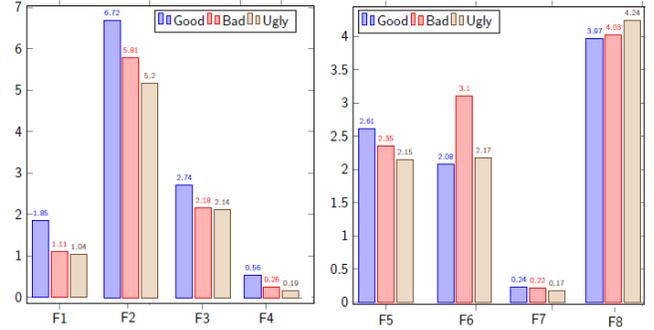

Figure 4. Gabor features used to train the multi-class perceptron classifier for image patch quality assessment. The features were computed from the kurtosis and skewness of the Gabor filter responses of the image patches. F1 and F4 are the peak and mean skewness of the magnitude response, respectively. F2 and F5 are the peak and mean kurtosis of the phase response, respectively. F3 is the mean kurtosis of the magnitude response. F6 and F7 are the peak and mean skewness of phase response, and F8 is the peak of the magnitude response. F7 was scaled up by 0.2 for visibility. The charts show that together, the features exhibit discriminative potential for classifying patches into good, bad and ugly bins.

where σ, and $\hat{v}$ are the standard deviation, and mean, respectively. From experiments, we found that the areas of a fingerprint image with a regular ridge-valley patterns tend to have a high Gabor filter magnitude responses while those with unclear ridge-valley patterns have low and sometimes constant Gabor magnitude filter responses. Figure 3 shows the Gabor filter magnitude responses for sample good, bad and ugly segmented latent fingerprints from our deep learning model and shows the discriminative potential of the selected Gabor features for classifying patches into good, bad and ugly bins.

4. Experiments and Results

We implemented our algorithms in Matlab R2014a running on Intel Core i7 CPU with 8GB RAM and 750GB hard drive. Our implementation relied on NNBox (a Mat-

lab toolbox for neural networks, and multi-class perceptron with Levenberg-Marquardt optimization. We evaluated our model on NIST SD27 [3] latent fingerprint database. The 258 latent fingerprint images in NIST SD27 consists of 88 Good, 85 Bad and 85 Ugly quality latent fingerprint images. The quality assigned to each image was based on the condition of the image in the location in which the minutia was positioned and on how clearly identifiable the type of the minutia was in the image [3]. The results of the segmentation and quality assessment stages of our network were evaluated using the NIST SD27 ground-truth quality datasets from [8]. We generated the ground-truth dataset used in evaluating the results of the quality assessment stage. The details are provided in section 4.2.3.

4.1. Performance Evaluation Metrics

We used the following metrics to evaluate the performance the segmentation and quality assessment stages of our network.

- **Missed Detection Rate (*MDR*):** This is the percentage of class $\mathcal{C}_1$ patches classified as class $\mathcal{C}_2$ patches and is defined as.

$$MDR = \frac{FN}{TP + FN} \qquad (4)$$

where *FN* is the number of false negatives and *TP* is the number of true positives.

- **False Detection Rate (*FDR*):** This is the percentage of class $\mathcal{C}_2$ patches classified as class $\mathcal{C}_1$ patches. It is defined as:

$$FDR = \frac{FP}{TN + FP} \qquad (5)$$

where *FP* is the number of false positives and *TN* is the number of true negatives.

- **Segmentation Accuracy (SA):** It gives a good indication of the segmentation reliability.

$$SA = \frac{TP + TN}{TP + FN + TN + FP} \qquad (6)$$

For the segmentation stage, $\mathcal{C}_1$ = fingerprint, $\mathcal{C}_2$ = non-fingerprint, and for the quality assessment stage, $\mathcal{C}_1 \in \{\text{Good, Bad, Ugly}\}$ and $\mathcal{C}_2 \in \{\text{Good, Bad, Ugly}\}$. We also used **precision** and **recall** to evaluate the performance of classifier used for quality assessment.

- **Precision:** Precision is the percentage of examples that truly belong to class *k* among all examples that the classifier predicted as belonging to class *k*.

- **Recall:** Recall is the percentage of examples correctly predicted as belonging to class *k* among all examples that truly belong to class *k*.

		Predicted Patch Class (Training)	
		Fingerprint	Non-Fingerprint
Actual Patch Class	Fingerprint	23,665	11
	Non-Fingerprint	0	108,324

		Predicted Patch Class (Validation)	
		Fingerprint	Non-Fingerprint
Actual Patch Class	Fingerprint	13,637	163
	Non-Fingerprint	2	36,198

		Predicted Patch Class (Testing)	
		Fingerprint	Non-Fingerprint
Actual Patch Class	Fingerprint	13,914	188
	Non-Fingerprint	5	35,893

Table 4. NIST SD27 - Confusion matrix for training, validation and testing for the segmentation stage.

4.2. Latent Fingerprint Database

The ROI segmentation and quality assessment stages of our model were trained, validated and tested on NIST SD27 latent fingerprint databases. This database contains images of **258** latent crime scene fingerprints and their matching rolled tenprints. The images are grouped into good, bad or ugly categories. The grouping is based on the quality of the image determined by latent examiners. NIST SD27 has **88** Good, **85** Bad and **85** ugly quality latents. The latent prints and rolled prints are at **500** ppi.

4.2.1 Segmentation: Training, Validation and Testing

The training, validation and testing of the segmentation part of the model was done with 232,000 8x8 patches (132,000 for training, 50,000 for validation and 50,000 for testing) from the NIST SD27 database with 40% from good 30% from bad, and 30% from ugly NIST image categories. Table 4 shows the confusion matrix reflecting the results of training, validation and testing. We did not notice any appreciable performance gain when the model was trained with more than 132,000 patches.

Figure 5 shows the segmentation results of our proposed method on sample good, bad and ugly quality images from the NIST SD27 database. It shows the original latent fingerprint images and the segmented fingerprints constructed using patches classified as fingerprints.

4.2.2 Stability of the Segmentation Network

The stability of the segmentation network was investigated by selecting 40 images at random from each (Good, Bad and Ugly) category of NIST SD27 database and extracting 50,000 8x8 patches from each category for a total of 150,000 8x8 patches. We performed 5 runs of network training and for each of the 5 runs, we used 20,000 patches randomly sampled from the 150,000 patches. All the model parameters (number of epochs, number of iterations etc.)
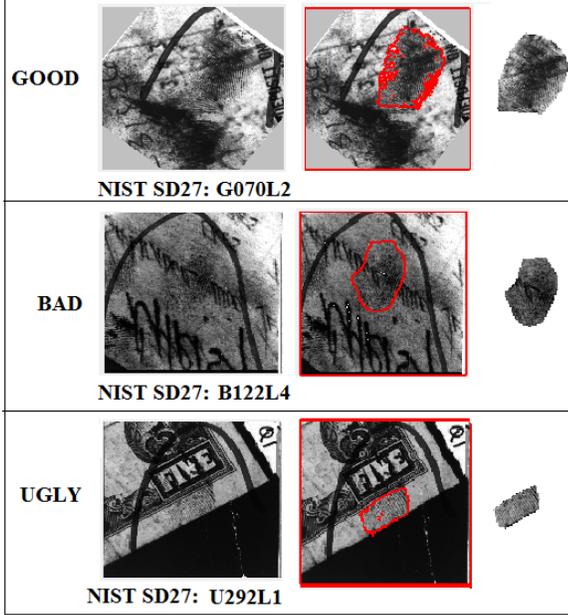
Figure 5. NIST SD27: Segmentation results without post classification processing for Good (row 1), Bad (row 2) and Ugly (row 3) latents. Each row shows the original image followed by an outline of the segmented fingerprint superimposed on the original image, and the segmented fingerprint only part. The segmented fingerprint part was constructed with patches classified as fingerprint.
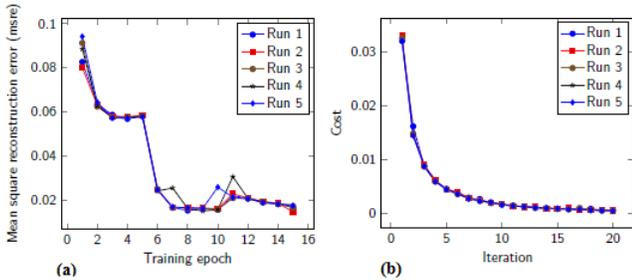

Figure 6. Segmentation Network Stability: (a) shows that the mean square reconstruction error (MSRE) during the pre-training phase for the 5 runs followed similar trajectories. Similarly, (b) shows that the error during the fine-tuning phase for the 5 runs were close. These results are indicative of the stability of the network.

shown in Table **??** remained unchanged across the runs. The mean square reconstruction error (msre) and mean error cost at convergence, as well as the standard deviation for the 5 runs are shown in Table 6. Plots of the error during training for each run are shown in Figure 6. These results indicate that the proposed segmentation model is stable.

4.2.3 Quality Assessment: Training, Validation and Testing

There are 258 latent fingerprint images in NIST SD27 database with 88, 85 and 85 in the Good, Bad and Ugly

Table 5. Segmentation Network Stability:: The mean square reconstruction error (MSRE) at convergence during the pre-training phase, cost at convergence during the fine-tuning phase, MDR, and FDR for the 5 different runs are close. The mean and standard deviation indicate stability across the 5 runs.

Run #	MSRE	Cost	MDR	FDR
1	0.0169	5.769E-04	3.220E-04	1.10E-05
2	0.0159	5.406E-04	3.950E-04	0.00
3	0.0148	5.420E-04	1.720E-04	0.00
4	0.0167	5.562E-04	3.310E-04	1.20E-05
5	0.0175	5.145E-04	2.091E-04	0.00
Mean	0.01636	0.00055	2.8E-04	4.6E-06
Standard Deviation	0.00104	2.289E-05	9.235E-05	6.309E-06

categories [11]. The 258 latent fingerprint images were segmented using our trained segmentation model. We extracted 7,000 32x32 patches from 50 Good, 50 Bad and 50 Ugly ROIs for training, 1,500 32x32 patches from 20 Good, 20 Bad and 20 Ugly ROIs for validation, and 1,500 32x32 patches from 18 Good, 15 Bad and 15 Ugly ROIs for testing. The multi-class perceptron classifier (MPC) in the quality assessment stage of our model was trained, validated and tested with the training, validation and testing datasets that were independently drawn from the fingerprint only segments from the 258 latent fingerprint images in NIST SD27 database.

To label the patches in the 32x32 patch datasets, we computed the average fractal dimension (FD_{av}) and fractal dimension spatial frequency (FD_{sf}) for each patch. The label L_p for each patch p was determined using equation 7. The thresholds used in equation 7 were empirically determined. Figure 7 shows sample patches and their FD_{av} and FD_{sf}.

$$
L_p = \begin{cases} 1, & \text{if } \tau > 1.75 \text{ and } \kappa < 0.65; \\ 2, & \text{if } 1.65 < \tau < 1.75 \text{ and } 0.65 < \kappa < 0.70; \\ 3, & \text{if } \tau < 1.70 \text{ and } \kappa > 0.70. \end{cases}
$$
(7)

where τ and κ are the FD_{av} and FD_{sf} of patch p, respectively. Figure 8 shows the confusion matrix for MPC training, validation and testing, as well as the validation performance and error histogram on NIST SD27.

The quality assessment model achieved a quality assessment accuracy of 96.1% for Good, 91.1% for Bad and 96.7% for ugly latent fingerprints on the testing dataset, and 97.9% for Good, 92.7% for Bad and 97.5% for ugly latent fingerprints on the validation dataset.

4.2.4 Stability of the Quality Assessment Network

To investigate the stability of the quality assessment network, we performed 5 runs of training, validation and testing of the network using the dataset created in 4.2.3. All the model parameters (number of epochs, number of iterations etc.) remained unchanged across the runs. The overall precision for training, validation and testing, as well as the

1	2	3	4	5	6	7	8	9	10	11
Image Patch										
FD_{av}	1.773	1.804	1.767	1.675	1.737	1.727	1.590	1.633	1.645	1.512
FD_{sf}	0.611	0.617	0.649	0.669	0.675	0.659	0.713	0.772	0.768	0.825

Figure 7. Images patches from NIST SD27 with their computed average fractal dimension (FD_{av}) and fractal dimension spatial frequency (FD_{sf}). Patches with visible fingerprint patterns (columns 2, 3, & 4) have higher average FD and lower FDsf, than those with little visible fingerprint patterns(columns 5, 6 & 7) or no visible fingerprint patterns (columns 8, 9, 10 & 11). The higher the FD_{av} and the lower the FD_{sf}, the better the quality of the patch, and conversely.
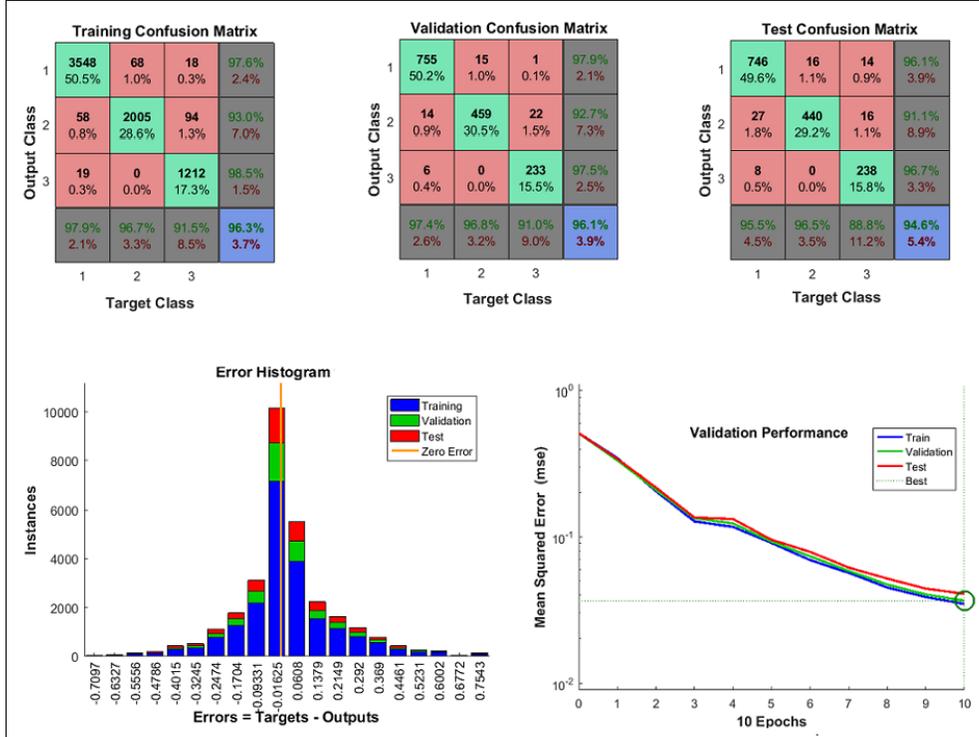

Figure 8. NIST SD27 - Confusion matrix for training, validation and testing, error histogram, and validation performance for the quality assessment neural network. Class 1 = Good, Class 2 = Bad, and Class 3 = Ugly. 7,000 patches were used for training, 1,500 patches for validation and 1,500 patches for testing. The training, validation and testing samples were independently drawn from the dataset. Output Class is the predicted class while target class is the ground-truth class. The fourth row contains Recall values while the fourth column contain the Precision. In the testing confusion matrix, Precision=96.1% and Recall=95.5% for Class 1 means that out of the times Class 1 was predicted, the classifier was correct 96.1% of the time, and out of all the times Class 1 should have been predicted 95.5% of the predictions were correct. The small numbers on all cells but the diagonal (that contains the true positives for the respective classes), as well as the error histogram, and validation performance plots, indicate good classifier performance.

mean accuracy, and standard deviation for the five runs are shown in Table 6. The precision and recall for the three classes in the five runs are provided in Table 7. These results indicate the stability and reliability of the network.

4.3. Evaluation of Latent Quality Prediction

We compare the latent quality predictions of the proposed model with the VID, VEO, and NV value determination by latent examiners [10] as well as the quality value

predictions by Expert Crowd [7]. Note that NIST SD27 database is the only latent database with available latent value determinations by latent examiners. As reported in [10], there are 210 VID, 41 VEO , and 7 NV latents in NIST SD27. A total of 166 latents (155 VID , and 11 VEO) out of the 256 latents in NIST SD27 are retrieved at Rank-1 using state-of-the-art latent AFIS [7]. To ensure a fair comparison between the three results being compared, we follow the protocol used in [7]. The 258 latents are sorted in as-

Run #	Training	Validation	Testing
1	93.9%	93.1%	93.6%
2	96.3%	96.1%	94.6%
3	95.5%	94.0%	94.0%
4	97.6%	96.4%	95.8%
5	94.9%	94.1%	93.8%
Mean (%)	95.46	94.74	94.36
Standard Deviation (%)	1.12	1.44	0.89

Table 6. Network Stability: The precision values (computed with the true positives along the diagonal of the confusion matrix) in each column are close. The mean and standard deviation indicate stability across the five runs.

Validation: Run #	Class	Precision	Recall
1	1	91.9%	97.3%
	2	93.8%	88.1%
	3	95.7%	89.7%
2	1	97.9%	97.4%
	2	92.7%	96.8%
	3	97.5%	91%
3	1	92.1%	99.2%
	2	94.4%	88.7%
	3	100%	88.7%
4	1	98.7%	96.7%
	2	92%	98.3%
	3	97.1%	92.8%
5	1	96.8%	95.7%
	2	90.6%	95.3%
	3	95.2%	92.8%

Testing: Run #	Class	Precision	Recall
1	1	93.0%	98.0%
	2	93.2%	89.6%
	3	96.0%	87.3%
2	1	96.1%	95.5%
	2	91.1%	96.5%
	3	96.7%	88.8%
3	1	92.1%	99.1%
	2	94.2%	88.0%
	3	100%	88.9%
4	1	98.4%	97.6%
	2	93.7%	98.8%
	3	97.0%	92.4%
5	1	96.7%	96.5%
	2	88.2%	95.3%
	3	95.2%	85.4%

Table 7. Network Stability: Precision and Recall for the three classes in the five runs. For both the validation and testing samples, there is no marked difference between the Precision for a given class from one run to the next. The same applies to the Recall. This indicates that the network is reliable.

cending order of the quality [1-3] predicted by our model, and then partitioned into three, P_1, P_2, P_3. Partition P_1 contains the first 210 latents, P_2 contains the next 41, while P_3 contains the last 7 latents. Following [10], the latents in P_1, P_2 and P_3 are considered as VID, VEO and NV, respectively. Table 8 shows a comparison of the number of latents retrieved at rank-1 using value determination by latent examiners [10], the predicted latent value from [7], and the predicted latent quality from our quality prediction model. A reference dataset containing 2,257 rolled prints created

	VID	VEO	NV
Latent Examiners [10]	155/210	**11**/41	0/7
Expert Crowd [7]	161/210	5/41	0/7
This Work	**164**/210	4/41	0/7

Table 8. NIST SD27 latent fingerprints retrieved at Rank-1 using a state-of-the-art latent AFIS. The results show that the proposed quality assessment model performs slightly better than Expert Crowd [7] in predicting latent AFIS performance for VID latents (164 vs 161). However, both Latent examiners and Expert Crowd are better than the proposed model in predicting latent AFIS performance for VEO latents.

from 2,000 fingerprint images in NIST SD4 database, and the 257 rolled images in NIST SD27 database was used for this performance comparison. In terms of predicting latent AFIS performance, the quality prediction by our model is better than the value determination latent examiners and value prediction by Expert Crowd. 164 latents predicted by our model as VID latents were identified at Rank-1 compared to 161 identified at Rank-1 based on value determination by Expert Crowd.

5. Conclusions and Future Work

We proposed automatic region-of-interest based latent fingerprint quality assessment technique using deep learning. The first stage of our proposed method uses feature learning, extraction and classification to segment the latent fingerprint image. In the second stage, 32x32 patches are extracted from the segmented ROI image and features computed from them are fed to a multi-class perceptron that classifies each fingerprint patch into Good, Bad or Ugly bins. The quality of a latent fingerprint is indicated by the label of the bin that contains the greatest number of patches, with ties broken optimistically (if the number of patches in the Good bin is equal to that in the Bad bin and greater the number in the Ugly bin, the quality of the fingerprint is set to Good). We demonstrated the performance of our model on the NIST SD27 latent fingerprint database. We presented a comparative analysis showing that in terms of predicting latent AFIS performance, the quality prediction by our model performs better than the state-of-the-art latent fingerprint value determination model. Part of our future work involves using NIST Finger Image Quality (NFIQ 2.0) as a baseline for mapping latent fingerprint quality assessment to recognition performance.

Acknowledgment

References

[1] In *https://goo.gl/5wmG3H*. 4

[2] Integrated Pattern Recognition and Biometrics Lab, West Virginia University. *http://www.csee.wvu.edu/*. 1

[3] NIST Special Database 27. *Fingerprint Minutiae from Latent and Matching Ten-print Images.* *http://www.nist.gov/srd/nistsd27.htm*. 1, 5

[4] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A comparative study of fingerprint image-quality estimation methods. *IEEE Transactions on Information Forensics and Security*, 2(4):734–743, Dec 2007. 2

[5] K. Cao, T. Chugh, J. Zhou, E. Tabassi, and A. K. Jain. Automatic latent value determination. In *International Conference on Biometrics (ICB)*, pages 1–8, June 2016. 1, 2

[6] M. A. Carreira-Perpinan and G. Hinton. On contrastive divergence learning. In *AISTATS*, volume 10, pages 33–40. Citeseer, 2005. 3

[7] T. Chugh, K. Cao, J. Zhou, E. Tabassi, and A. K. Jain. Latent fingerprint value prediction: Crowd-based learning. *IEEE Transactions on Information Forensics and Security*, 13(1):20–34, Jan 2018. 2, 7, 8

[8] J. Ezeobiejesi and B. Bhanu. Latent fingerprint image segmentation using fractal dimension features and weighted extreme learning machine ensemble. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016. 3, 5

[9] J. Ezeobiejesi and B. Bhanu. Latent fingerprint image segmentation using deep neural network. In *Deep Learning for Biometrics*, pages 83–107, 2017. 3

[10] R. Hicklin, J. Buscaglia, M. A. Roberts, S. B. Meagher, W. Fellner, M. J. Burge, M. Monaco, D. Vera, L. R. Pantzer, C. C. Yeung, and T. N. Unnikumaran. Latent fingerprint quality: a survey of examiners. *Journal of Forensic Identification*, 61:385–419, 07 2011. 7, 8

[11] M. Indovina, R. A. Hicklin, and G. I. Kiebuzinski. Elft efs: Evaluation of latent fingerprint technologies: Extended feature sets [evaluation no. 1]. NISTIR 7775,2011, 2011. 6

[12] E. Lim, X. Jiang, and W. Yau. Fingerprint quality and validity analysis. In *Proceedings. International Conference on Image Processing*, volume 1, pages I–469–I–472 vol.1, 2002. 4

[13] D. Maltoni, D. Maio, A. K. Jain, , and S. Prabhakar. Lfiq: Latent fingerprint image quality. In *Handbook of Fingerprint Recognition*. Springer Publishing Company, Incorporated, 2009. 3

[14] A. Sankaran, M. Vatsa, and R. Singh. Automated clarity and quality assessment for latent fingerprints. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2013. 1, 2

[15] S. Yoon, K. Cao, E. Liu, and A. K. Jain. Lfiq: Latent fingerprint image quality. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8, Sept 2013. 1, 2