Assessing face image quality with LSTMs

Tommy Thorsen[*], Pankaj Wasnik, Christoph Busch,
R. Raghavendra, and Kiran Raja

Norwegian University of Science and Technology, Norway

Abstract

Biometric authentication using fingerprints or face recognition is making its way into the mainstream, and there is an urgent need to make these authentication methods as secure and reliable as possible. One way to achieve better performance with a biometric authentication method, is to introduce a quality estimation step early in the pipeline, so that unsuitable, or low-quality samples can be rejected.

While existing work predominantly focuses on algorithms for detecting specific properties of the face images, we investigate whether machine learning techniques can provide a general way to estimate overall face image quality.

We train a selection of neural network types, and discover that a type of Recurrent Neural Network (RNN) called *Long Short-Term Memory* (LSTM) can reliably estimate face image quality, with better performance than the bespoke algorithms.

1 Introduction

In all biometric authentication systems, there is a chance of false positives or false negatives. Some genuine login attempts will be denied, and some erroneous attempts will succeed. The chance of such errors increase if the biometric samples are of low quality. Some users will attempt to enter low-quality samples on purpose, to cheat the system, and others will do so unintentionally. For a face recognition system, a sample may be of low quality if for instance the illumination is bad, or the camera is out of focus, or if the user intentionally conceals parts of their face. Regardless of the users' motivations, it is important that we are able to detect and reject low quality samples.

Determining the quality of a face image is commonly done by running it through a set of algorithms that detect and measure specific properties of the image. A number of such algorithms, can be found in the specification ISO/IEC TR 29794-5. Sharpness and contrast was originally proposed by Werner et. al. [14]. How to use facial symmetry calculations to detect lighting and pose problems is described by Gao et. al. [2]. Raghavendra et. al. [9] use Co-occurrence matrices to detect rotation and yaw of the head. Wasnik et. al. [13] proposes another measure called *edge density*, and use it to determine pose and lighting.

Common for all of the methods above, is that they are algorithms for detecting specific face image properties that have been found to correlate well with face image quality. It would be desirable to develop an approach that would leave out the guesswork as to which properties to measure and how to measure them. Machine learning techniques might provide the tools that we need. LSTMs and other RNNs are commonly used to analyze and make predictions on time series, but it has been shown that they are also suitable for image processing tasks such as handwriting recognition [4] and natural scene labeling [1]. In this paper, we show that the LSTM's ability to learn spatial dependencies and analyze parts of images based on surrounding context, makes them well suited for the task of biometric quality estimation.

[*] The author presented this paper at the NISK 2018 conference.

Proceedings of the 11th Norwegian Information Security Conference, Sep. 19-20, 2018.
NISK Journal, e-ISSN 1894-7735. https://ojs.bibsys.no/index.php/NISK

1.1 Previous work

One of the earliest applications of biometric sample quality estimation was in the context of fingerprint recognition [10]. Ratha and Bolle's approach was to use wavelet compression to estimate fingerprint sample quality.

NIST's NFIQ is a well-known tool for estimating the quality of fingerprint images, but unfortunately the output of this tool is not as reliable as one could wish, and the small number of quality classes is quite limiting. Because of this, a lot of work has gone into the development of NFIQ 2.0, which outputs a score which complies to the international biometric sample quality standard ISO/IEC 29794-1:2016. This score is in the range [0-100], where 100 is the best possible quality score.

One paper that applies the ISO/IEC sample quality standard to face images is *Standardization of Face Image Sample Quality* [2], by Gao et. al. This paper looks at the challenge of estimating the score in a standardized and normalized way, and proposes some relevant measurements, such as facial symmetry.

The paper *Assessing Face Image Quality for Smartphone based Face Recognition System* [13], list an additional number of commonly measured properties, all of which are adopted from *ISO/IEC TR 29794-5*. These include *brightness*, *blur*, and *global contrast factor*.

Convolutional Neural Networks (CNNs), such as AlexNet, described by Krizhevsky et. al. [7], is well suited to processing images. Vizilter et. al. [12] use a CNN to do face image identification with good results. It seems likely that a CNN can also perform well at face image quality estimation.

LSTMs are defined by Hochreiter and Schmidhuber in their 1997 paper called *Long Short-Term Memory* [6]. An LSTM is a type of RNN, which can remember some data for a longer time. LSTMs contain something called *cell state* which can be written to according to the values of certain *gates* inside each LSTM cell. LSTMs typically learn faster, and perform better than regular RNNs.

2 Methods

2.1 Neural network architecture

We have implemented three different neural network architectures for comparison. As can be seen in figure 1, the bottom half of the models are the same for all architectures. This is to keep them as comparable as possible. The size of the models, as measured by the file-size of the fully trained graph, is also similar for all architectures.

2.1.1 LSTM

This model consists of LSTMs interleaved with CNN layers and max pooling layers, finally passing through two fully connected layers. There are two layers with an LSTM and a CNN in each layer. The output sizes for the layers (and for all of the components within the layers) are 32 for layer one and 64 for layer two. The model has been trained for 30 epochs with a batch size of 256 and a learning rate of 0.001.

2.1.2 Four-way LSTM

The construction of this model is the same as the regular LSTM model, except for one thing: Instead of using a single LSTM which analyzes the image data from the top-left towards the

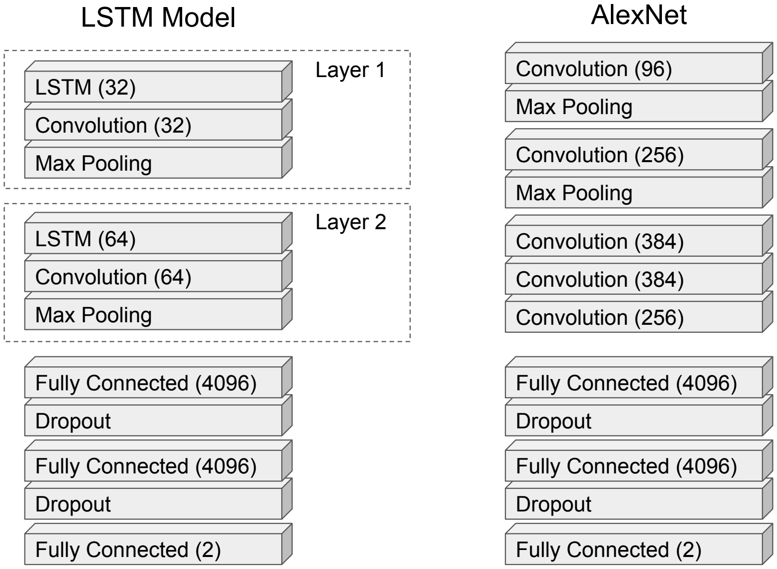

Figure 1: On the left is the LSTM based neural network model, which is used both for the simple LSTM architecture and the Four-way LSTM architecture. On the right is the AlexNet model.

bottom-right, we use four separate LSTMs that analyze the image in four different directions (see figure 2). This idea comes from the paper *Multi-dimensional Recurrent Neural Networks*[3], which makes use of four different 2D LSTMs to analyze images in all four directions. It has been shown that for one-dimensional data, running an RNN in both directions (a bidirectional recurrent neural network) can improve performance [11]. It seems likely that processing a two-dimensional image in all four directions might lead to a similar performance improvement.

To make the combined size of the four LSTMs in this model, the same as for the model with single LSTMs per layer, we divide the layer output size by four to get the individual LSTM output size. The first layer contains four LSTMs with output size 8, and the second layer contains four LSTMs without output size 16. The model has been trained for 30 epochs with a batch size of 256 and a learning rate of 0.001.

2.1.3 AlexNet

This architecture is the AlexNet model as described by Krizhevsky et. al. [7]. This network represents a pure CNN architecture, and is a good reference for us to measure how our LSTM architectures compare to other neural network architectures. The model has been trained for 30 epochs with a batch size of 32 and a learning rate of 0.001.

2.2 Datasets

Our neural networks have been trained and tested on a dataset containing of face image from a variety of sources. Among them are: *AR Face database*, *FRGC database*, *NCKU Face database*, *Yale Face database* and the *CASIA Face V5 database*. The training dataset consists of roughly

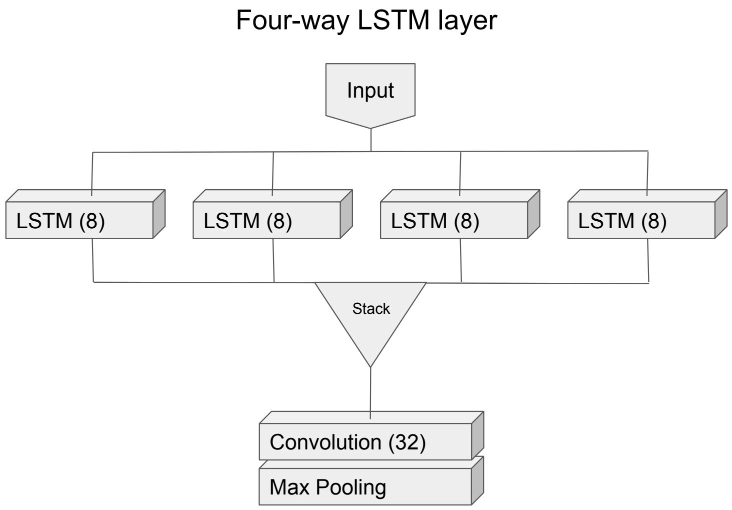

Figure 2: The four-way LSTM layer consists of four regular LSTMs which read the image in four different directions. The results from each LSTM are stacked to provide a single 32 unit output.

30000 face images, labeled as either `good` or `bad`. The neural networks have been trained for binary classification between these classes. The classification confidence (the estimated probability that a given image belongs to the `good` class) is used to derive an ISO compliant score in the range [0, 100].

The datasets used to test the performance of the trained networks contain face images captured with the front cameras of an iPhone 6 Plus and a Samsung Galaxy S7, making them quite realistic, at least in the context of face recognition for mobile phones. Each of these databases contain images of 101 different subjects, with 10 images of varying quality for each subject.

3 Results

To measure and visualize the performance of our face image quality assessors, we will use Error Reject Curves (ERCs), as recommended by Grother and Tabassi [5]. ERCs were originally designed to measure the performance of fingerprint image quality assessors, but are suitable for any kind of biometric system. In addition to our three neural network based systems, we also provide the results for a commercial solution.

To calculate an ERC, it is necessary to couple the quality assessor with a face recognition system. The curve shows how the performance of the face recognition system is affected by the quality assessment algorithm operating at different rejection rates. As such, the ERC is a good indicator of the predictive performance of a quality algorithm.

Figure 3 shows the ERCs for both test databases. The curves show that AlexNet does not perform as well as one would have thought, for this task. The regular LSTM performs really

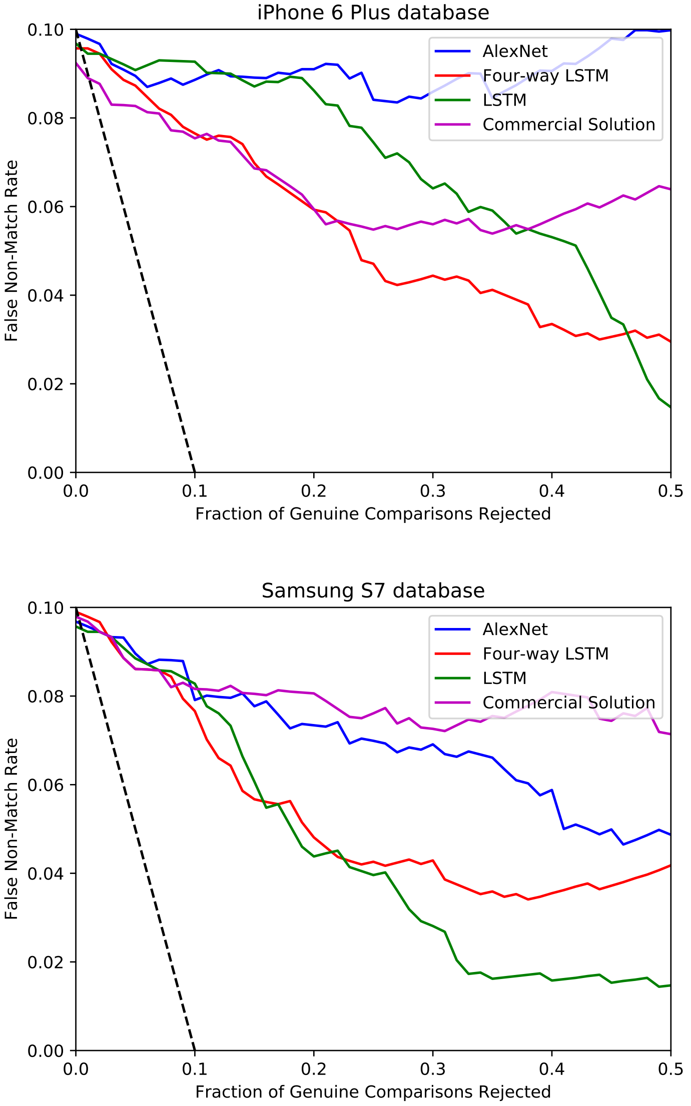

Figure 3: Top: ERCs for the iPhone 6 Plus database. Bottom: ERCs for the Samsung S7 database

well for the Samsung dataset, but for the iPhone dataset, the performance at low rejection rates is not so competitive. The difference between these datasets is mainly the due to the quality of the cameras of the two different smartphones. The Samsung frontal camera is better than the one on the iPhone, and produces higher quality images.

The Four-way LSTM model seems to give the most stable and predictable results. Especially at low rejection rates, it performs really well. The good performance is likely to be due to the four-way architecture's ability to consider each pixel given its surrounding context on all sides.

For a more quantitative analysis of the ERCs, we turn to Olsen et. al. [8] who propose the

	iPhone 6 Plus Database		Samsung S7 Database	
	η^{erc}_{auc}	η^{erc}_{pauc20}	η^{erc}_{auc}	η^{erc}_{pauc20}
LSTM	**0.0316**	0.0132	**0.0206**	0.0102
Four-way LSTM	0.0328	0.0106	0.0333	**0.0098**
AlexNet	0.0810	0.0131	0.0470	0.0118
Commercial Solution	0.0495	**0.0102**	0.0739	0.0120

Table 1: AUC and PAUC for ERC plots for all quality assessors on both databases.

following metrics:

$$\eta^{erc}_{auc} = \int_0^1 ERC - area\ under\ theoretical\ best$$

$$\eta^{erc}_{pauc20} = \int_0^{0.2} ERC - area\ under\ theoretical\ best$$

The first metric is simply the integral of the ERC, giving us the area under the curve. We subtract the area under theoretical best, which corresponds to the area under the black dashed line in the ERCs. The second metric is the same as the first, except we only look at the first 20% of the ERC.

Table 1 shows these metrics. We can see that the regular LSTM performs best for the whole curve, while the Four-way LSTM gives better results for the first 20%.

4 Conclusion

We have shown that some types of deep neural networks can work well for estimating biometric sample quality for face images. The results indicate that the performance is as good as, if not better than that of traditional quality estimation methods, as implemented in a commercial face recognition product. We have also shown that RNNs (here represented by an LSTM) performs better at this task than a pure CNN of similar size.

When we consider the relative simplicity of development of a neural network quality assessor, compared to the development complexity of a set of quality measurement algorithms, we believe that this will be a very attractive approach for many prospective implementers.

References

[1] Wonmin Byeon, Thomas M Breuel, Federico Raue, and Marcus Liwicki. Scene labeling with lstm recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3547–3555, 2015.

[2] Xiufeng Gao, Stan Z Li, Rong Liu, and Peiren Zhang. Standardization of face image sample quality. In *International Conference on Biometrics*, pages 242–251. Springer, 2007.

[3] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Multi-dimensional recurrent neural networks. In *ICANN (1)*, pages 549–558, 2007.

[4] Alex Graves and Jürgen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in neural information processing systems*, pages 545–552, 2009.

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

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[8] Martin Aastrup Olsen, Vladimír Šmida, and Christoph Busch. Finger image quality assessment features–definitions and evaluation. *IET Biometrics*, 5(2):47–64, 2016.

[9] Ramachandra Raghavendra, Kiran B Raja, Bian Yang, and Christoph Busch. Automatic face quality assessment from video using gray level co-occurrence matrix: An empirical study on automatic border control system. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 438–443. IEEE, 2014.

[10] Nalini K Ratha and Ruud Bolle. Fingerprint image quality estimation. 1999.

[11] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[12] Yuri Vizilter, Vladimir Gorbatsevich, Andrey Vorotnikov, and Nikita Kostromov. Real-time face identification via cnn and boosted hashing forest. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 78–86, 2016.

[13] Pankaj Wasnik, Kiran B Raja, Raghavendra Ramachandra, and Christoph Busch. Assessing face image quality for smartphone based face recognition system. In *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pages 1–6. IEEE, 2017.

[14] Martin Werner and Michael Brauckmann. Quality values for face recognition. In *NIST Biometric Quality Workshop*, volume 3, 2006.