# Fingerprint Spoof Detection: Temporal Analysis of Image Sequence

Tarang Chugh and Anil K. Jain
Department of Computer Science and Engineering
Michigan State University, East Lansing, Michigan 48824
{chughtar, jain}@cse.msu.edu

## Abstract

*We utilize the dynamics involved in the imaging of a fingerprint on a touch-based fingerprint reader, such as perspiration, changes in skin color (blanching), and skin distortion, to differentiate real fingers from spoof (fake) fingers. Specifically, we utilize a deep learning-based architecture (CNN-LSTM) trained end-to-end using sequences of minutiae-centered local patches extracted from ten color frames captured on a COTS fingerprint reader. A time-distributed CNN (MobileNet-v1) extracts spatial features from each local patch, while a bi-directional LSTM layer learns the temporal relationship between the patches in the sequence. Experimental results on a database of* $26,650$ *live frames from* $685$ *subjects (*$1,333$ *unique fingers), and* $32,910$ *spoof frames of* $7$ *spoof materials (with a total of* $14$ *material variants), show that the proposed approach exceeds the state-of-the-art performance in both known-material and cross-material (generalization) scenarios. For instance, the proposed approach improves the state-of-the-art cross-material performance from TDR of* $81.65\%$ *to* $86.20\%$ *@ FDR =* $0.2\%$.

## 1. Introduction

Fingerprint recognition technology is now widely adopted across the globe for a plethora of applications, including international border crossing[1], forensics[2], unlocking smartphones[3], and national ID[4] programs. However, one of the most critical premise for this wide acceptance is that users have trust in the security of a fingerprint recognition system, namely protection of enrolled fingerprints (templates) and detection of fingerprint spoofs [18]. In this paper, we focus on fingerprint spoof detection.

978-1-7281-9186-7/20/$31.00 ©2020 IEEE

[1]https://www.dhs.gov/obim
[2]https://www.fbi.gov/services/cjis
[3]https://support.apple.com/en-us/HT201371
[4]https://uidai.gov.in/

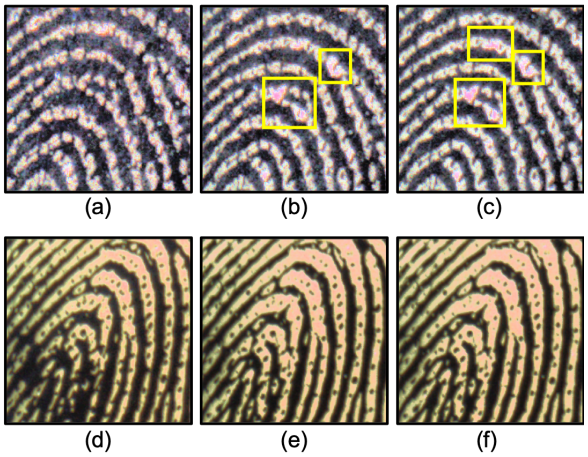


Figure 1. A sequence of ten color frames are captured by a SilkID SLK20R fingerprint reader in quick succession (8 fps). The first, fifth, and tenth frames from a live ((a) - (c)), and spoof (tan pigmented Third Degree silicone) ((d) - (f)) finger are shown here. Unlike spoofs, in the case of live fingers, appearance of sweat pores (highlighted in yellow boxes) and changes in skin color (pinkish red to pale yellow) across frames can be observed.

Fingerprint spoof attacks[5] refer to finger-like artifacts with an accurate imitation of one's fingerprint fabricated for the purpose of stealing their identity. Techniques ranging from simple molding and casting to sophisticated 3D printing have been utilized to create spoofs with high fidelity [19, 3, 8]. Various readily available and inexpensive materials (e.g. gelatin, wood glue) can be used to create spoofs that are capable of circumventing a fingerprint recognition system. For instance, in March 2018, a gang in Rajasthan (India) bypassed the biometric attendance system, using wood glue spoofs casted in wax molds, to provide proxies for police academy entrance exams[6]. In April

[5]Fingerprint spoofs are one of the most common forms of presentation attacks (PA). The ISO standard IEC 30107-1:2016(E) defines presentation attacks as the *"presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system"*. Other forms of PAs include use of altered fingerprints and cadaver fingers.

[6]https://www.medianama.com/2018/03/223-cloned-thumb-prints-used-to-spoof-biometrics-and-allow-

Table 1. Studies primarily focused on fingerprint spoof detection using temporal analysis.

| Study | Approach | Database | Performance |
|---|---|---|---|
| Parthasaradhi et al. [24] | Temporal analysis of perspiration pattern along friction ridges | 1,840 live from 33 subjects and 1800 spoof images from 2 materials, and 700 cadaver from 14 fingers | Avg. Classification Accuracy = 90% |
| Kolberg et al. [15] | Blood flow detection using a sequence of 40 Laser Speckle Contrast Images | 1,635 live from 163 subjects and 675 spoof images of 8 spoof materials (32 variants) | TDR = 90.99% @ FDR = 0.05% |
| Plesh et al. [25] | Fusion of static (LBP and CNN) and dynamic (changes in color ratio) features using a sequence of 2 color frames | 14,892 live and 21,700 spoof images of 10 materials | TDR = 96.45% (known-material) @ FDR = 0.2% |
| **Proposed Approach** | Temporal analysis of minutiae-based local patch sequences from 10 color frames using CNN + LSTM model | 26,650 live from 685 subjects and 32,910 spoof images of 7 materials (14 variants) | TDR = **99.15%** (**99.45%**) (known-material) and TDR = **86.20%** (**96.83%**) (cross-material) @ FDR = 0.2% (1.0%) |

2019, a Galaxy S10 owner with a 3D printer and a photo of his own fingerprint spoofed the ultrasonic in-display fingerprint sensor on his smartphone[7]. Other similar successful spoof attacks have been reported showing the vulnerabilities of fingerprint biometric systems[8,9]. It is likely that a large number of these attacks are never detected and hence not reported. Moreover, in response to this growing threat, a series of fingerprint Liveness Detection (LivDet) competitions [33] have been held since 2009 to benchmark various spoof detection solutions. Other initiatives include U.S. Government's IARPA ODIN Program [23] and European Commission's TABULA RASA program[10], with the goal of developing robust spoof detection systems for fingerprints, face, and iris biometric modalities.

With the goal to detect such spoof attacks, various hardware and software-based spoof detection approaches have been proposed in literature [18]. The hardware-based approaches typically utilize specialized sensors to detect the signs of vitality (blood flow, heartbeat, etc.) and/or sensing technologies for sub-dermal imaging [15, 31, 7, 20]. On the other hand, software-based approaches extract salient cues, related to anatomical (pores) [27] and texture-based features [32], from the captured fingerprint image(s). Chugh et al. [6] utilized minutiae-based local patches to train deep neural networks that achieves state-of-the-art spoof detection performance. Zhang et al. [35] utilized center of gravity (CoG)-based local patches to train a light version of the residual network, called Slim-ResCNN, to achieve the best performance in LivDet 2017 competition [21]. Gonzalez-Soler et al. [10] proposed fusion of feature encodings of dense-SIFT features for robust spoof detection.

Software-based approaches can be further classified into static and dynamic approaches based on the input. A static approach extracts discriminative spatial features from a single fingerprint image, while a dynamic approach utilizes an image sequence to extract spatial and/or temporal features for spoof detection. For a comprehensive review on the published static approaches, readers are referred to [18, 9].

In the case of dynamic approaches, published studies utilize temporal analysis to capture the physiological features, such as perspiration [24, 17], blood flow [34, 15], skin distortion [1], and color change [34, 25]. Table 1 summarizes the dynamic approaches for fingerprint spoof detection reported in the literature. Some of the limitations of these studies include long capture time (2-5 seconds), expensive hardware, and/or small number of frames in the sequence. Moreover, it is likely that some live fingers may not exhibit any of these dynamic phenomenons to separate them from spoofs. For instance, some dry fingers may not exhibit signs of perspiration during the finger presentation or a spoof may produce similar distortion characteristics as that of some live fingers.

We posit that a learning-based approach, as opposed to hand-crafted features, of the dynamics involved in the presentation of a finger can provide more robust and highly discriminating cues to distinguish live fingerprints from spoof fingerprints. In this study, we propose to use a CNN-LSTM architecture to learn the spatio-temporal features across different frames in a fingerprint image sequence. We utilize a sequence of minutiae-centered local patches extracted from ten colored frames captured by a COTS fingerprint reader, SilkID SLK20R[11], at 8 fps to train the network in an end-to-end manner. The use of minutiae-based local patches has been shown to achieve state-of-the-art spoof detection performance compared to randomly selected local patches in static images as well compared to the whole images [5]. Additionally, using minutiae-based local patches provides a large amount of training data, 71,530 minutiae-based patch sequences, compared to 5,956 whole-frame sequences.

proxies-to-answer-online-rajasthan-police-exam/

[7] https://imgur.com/gallery/8aGqsSu

[8] http://fortune.com/2016/04/07/guy-unlocked-iphone-play-doh/

[9] https://srlabs.de/bites/spoofing-fingerprints/

[10] https://ec.europa.eu/digital-single-market/en/content/tabula-rasa-protecting-biometric-recognition-external-attacks

[11] https://www.zkteco.com/en/product_detail/SLK20R.html

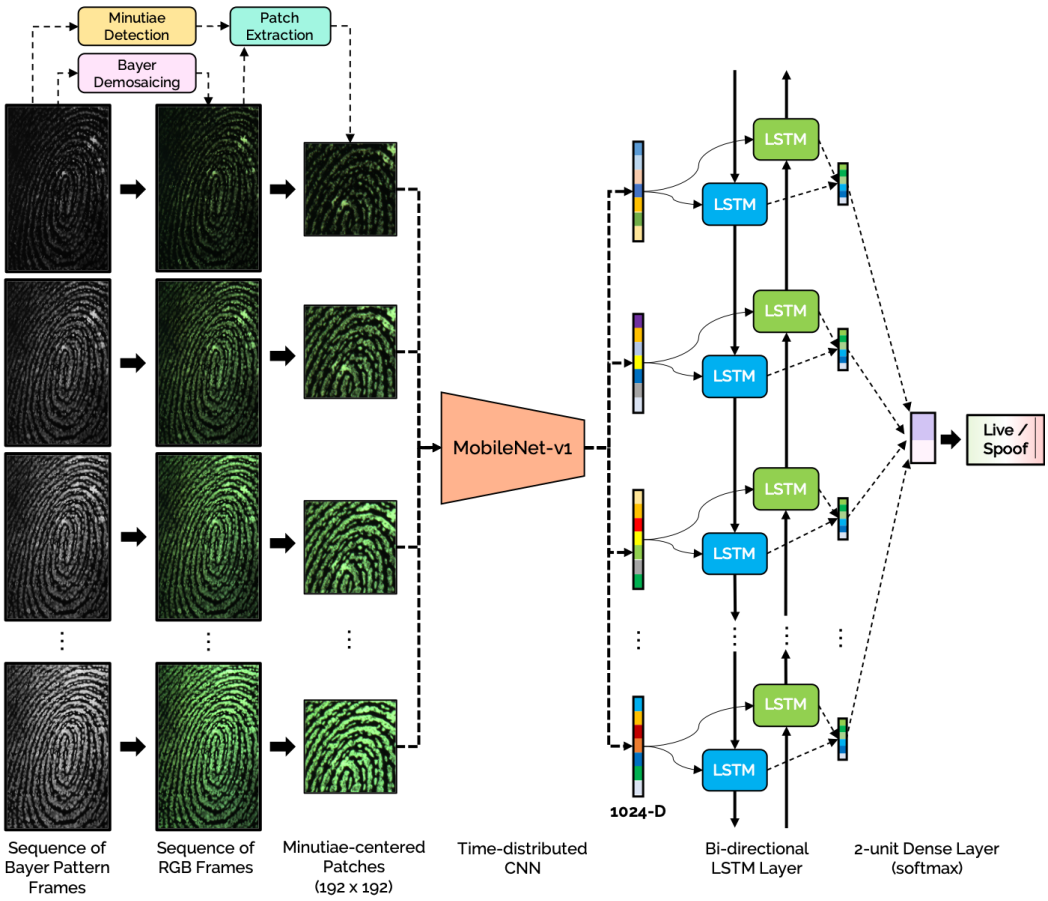


Figure 2. An overview of the proposed approach utilizing a CNN-LSTM model trained end-to-end on sequences of minutiae-centered local patches for fingerprint spoof detection.

The main contributions of this study are:

- Utilized sequences of minutiae-based local patches to train a CNN-LSTM architecture, with the goal of learning discriminative spatio-temporal (dynamic) features for fingerprint spoof detection. The local patches are extracted from a sequence of ten colored frames captured in quick succession (8 fps) using a COTS fingerprint reader, SilkID SLK20R.

- Experimental results on a database of $26,650$ live frames from $685$ subjects ($1,333$ unique fingers) and $32,930$ spoof frames of $7$ spoof materials (with a total of $14$ material variants) show that the proposed approach improves the state-of-the-art cross-material performance from TDR of $81.65\%$ ($91.00\%$) to $86.20\%$ ($96.83\%$) @ FDR = $0.2\%$ ($1.0\%$).

## 2. Proposed Approach

The proposed approach consists of: (a) detecting minutiae from each of the frames and selecting the frame with the highest number of minutiae as the reference frame, (b) preprocessing the sequence of frames to convert them from Bayer pattern grayscale images to RGB images (see section 2.2), (c) extracting local patches ($192 \times 192$) from all ten RGB frames based on the location of detected minutiae in the reference frame, and (d) end-to-end training of a CNN-LSTM architecture using the sequences of minutiae-centered patches extracted from the ten frames. While a time-distributed CNN network (MobileNet-v1) with shared weights extracts deep features from the local patches, a bidirectional LSTM layer learns the temporal relationship between the features extracted from the sequence. An overview of the proposed approach is presented in Figure 2.

### 2.1. Minutia Detection

When a live (or spoof) finger is presented to the SilkID SLK20R fingerprint reader, a sequence of ten color frames is captured, $F = \{f_1, f_2, ..., f_{10}\}$, at 8 frames per second[12] (fps) and a resolution of 1000 ppi. The complete sensing region ($h \times w$) in a SilkID fingerprint reader is $800 \times 600$ pixels. However, each of the ten colored frames are captured

[12] It takes an average of 1.25 seconds to capture a sequence of ten frames.

from a smaller central region of $630 \times 390$ pixels to ensure the fast frame rate of $8$ fps. It is observed that the starting and ending frames in the sequence may have little or no friction ridge details if the finger is not yet completely placed or quickly removed from the reader, respectively. Therefore, we extract minutiae information[13] from all ten frames using the algorithm proposed by Cao et al. [4]. The frame with the maximum number of detected minutiae is selected as the reference frame ($f^{ref}$) and the corresponding minutiae set as the reference minutiae set ($M^{ref}$).

## 2.2. Bayer Demosaicing

A digital sensor, containing a large array of photosensitive sites (pixels), is typically used in conjunction with a color filter array to permit only particular colors of light at each pixel. SilkID fingerprint reader employs one of the most common filter arrays, called *Bayer filter array*, consisting of alternating rows of red-green (RG) and green-blue (GB) filters. With the goal to recover the complete spatial color information in all three channels, the Bayer pattern grayscale image is converted to a three-channel RGB image using a process, called *Bayer Demosaicing*.

Bayer Demosaicing [16] (or debayering) is the process of converting a bayer pattern image to an image with complete RGB color information at each pixel. It utilizes bilinear interpolation technique [30] to estimate the missing pixels in the three color planes as shown in Figure 3. The original sequence of grayscale Bayer pattern frames ($10 \times 630 \times 390$) is converted to RGB colorspace[14]. SilkID readers are calibrated with strong gains on green pixels for generating high quality FTIR images. Therefore, after debayering, the RGB frames have high pixel intensity values in the green channel (see Figure 2). We utilize these raw images for our experiments. For visualization purposes, we reduce the green channel intensity values by a factor of $0.58$ and perform histogram equalization on intensity value in the HSV colorspace[15] (see Figures 1 and 4).

## 2.3. Local Patch Extraction

For each of the detected minutiae from the reference frame, $m_i \in M^{ref}$, we extract a sequence of ten local patches, $P_i = \{p_i^{f_1}, p_i^{f_2}, ..., p_i^{f_{10}}\}$, of size $192 \times 192$, from the ten frames $(F)$, centered at the minutiae location[16], *i.e.* $m_i = \{x_i, y_i\}$. This results in a total of $k$ patch sequences,

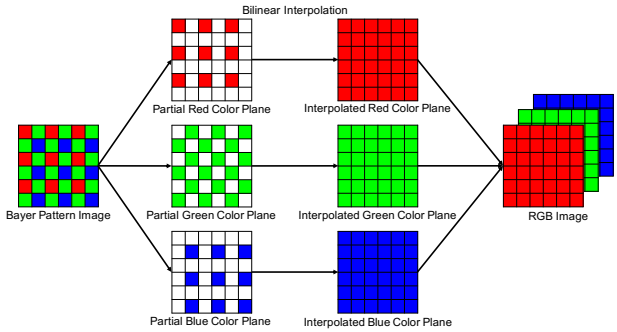


Figure 3. A Bayer color filter array consists of alternating rows of red-green and green-blue filters. Bilinear interpolation of each channel is utilized to construct the RGB image.

where $k$ is equal to the number of detected minutiae in the reference frame. Chugh et al. [5] reported that for $500$ ppi fingerprint images, the minutiae-based patches of size $96 \times 96$ pixels achieved the best performance compared to patch sizes of $64 \times 64$ pixels and $128 \times 128$ pixels. Therefore, for $1000$ ppi images in our case, we select a patch size of $192 \times 192$ pixels to ensure similar amount of friction ridge area in each patch, as contained in $96 \times 96$ pixels patch size for $500$ ppi fingerprint images. Each local patch from the reference frame is centered around the minutiae. However, this might not hold true for non-reference frames where the minutiae may shift due to non-linear distortion of human skin and non-rigid spoof materials. We hypothesize that the proposed approach can utilize the differences in thex non-linear shift along the sequence of local patches as a salient cue to distinguish between live and spoof presentations.

Table 2. Summary of the fingerprint database utilized in this study.

| Spoof Material (Pigment) | Mold Type | # Static Images | # Frames |
|---|---|---|---|
| **Ecoflex silicone** | | | |
| • Ecoflex 00-35 (flesh tone) | Dental | 757 | 7,570 |
| • Ecoflex 00-50 (flesh tone) | 3D Printed | 138 | 1,380 |
| • Ecoflex 00-50 (tan) | 3D Printed | 130 | 1,300 |
| **Gelatin** | | | |
| • Ballistic gelatin (flesh tone) | 3D Printed | 50 | 500 |
| • Knox gelatin (clear) | 3D Printed | 84 | 840 |
| **Third degree silicone** | | | |
| • Light flesh tone | Dental | 131 | 1,310 |
| • Tan | Dental | 98 | 980 |
| • Beige suede powder | Dental | 43 | 430 |
| • Medium flesh tone | Dental | 36 | 360 |
| **Crayola Model Magic** | | | |
| • White color | Dental | 910 | 9,100 |
| • Red color | Dental | 308 | 3,080 |
| **Dragon Skin (flesh tone)** | Dental | 452 | 4,520 |
| **Conductive Silicone** | 3D Printed | 87 | 870 |
| **Unknown Spoof (JHU-APL)** | 3D Printed | 67 | 670 |
| **Total Spoofs** | | **3,291** | **32,910** |
| **Total Lives** (685 subjects) | | **2,665** | **26,650** |

[13]Since the minutiae detector proposed in [4] is optimized for 500 ppi fingerprint images, all frames are resized before extracting the minutiae.

[14]We utilize an OpenCV [2] function, *cv2.cvtColor()*, with the parameter $flag = cv2.COLOR\_BAYER\_BG2RGB$

[15]Reducing gain in green channel and histogram equalization achieved similar or lower performance compared to using raw color images. Therefore, raw images were used for all experiments.

[16]Minutiae coordinates extracted from the resized 500 ppi frames are doubled to correspond to minutiae coordinates in the original 1000 ppi frames.

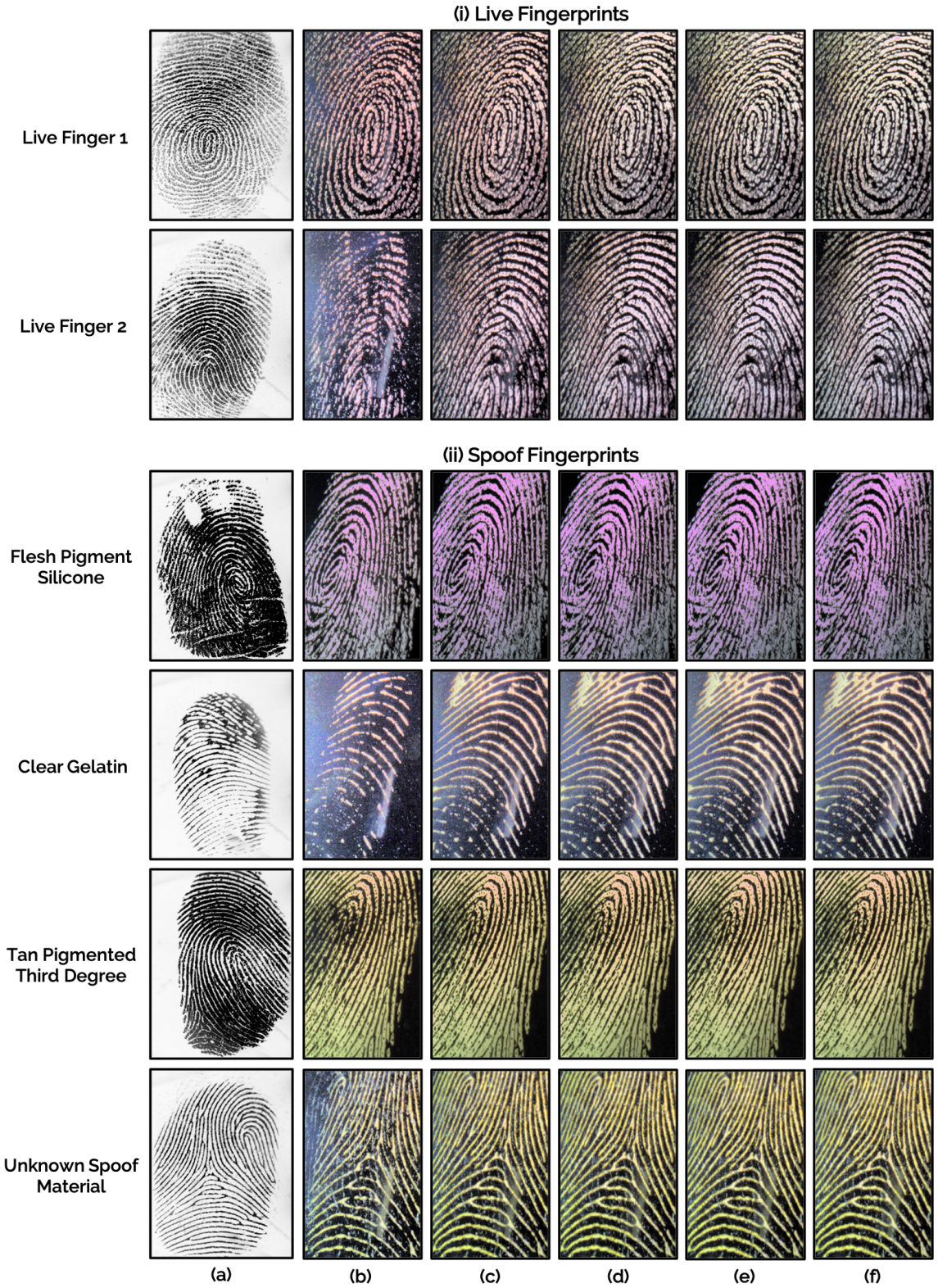


Figure 4. Examples of (i) live and (ii) spoof fingerprint images. (a) Grayscale 1000 ppi image, and (c)-(g) the first five (colored) frames captured by SilkID SLK20R Fast Frame Rate reader. Live frames exhibit the phenomenon of blanching of the skin, *i.e.* displacement of blood when a live finger is pressed on the glass platen changing the finger color from red/pink to pale white.

### 2.4. Network Architecture

Several Convolutional Neural Network (CNN) architectures, such as VGG [28], Inception-v3 [29], MobileNet-v1 [12] etc., have been shown to achieve state-of-the-art performance for many vision-based tasks, including fingerprint spoof detection [22, 5]. Unlike traditional approaches where spatial filters are hand-crafted, CNNs can automatically learn salient features from the given set of training images. However, as CNNs are feed-forward networks, they are not well-suited to capture the temporal dynamics involved in a sequence of images. On the other hand, a Recurrent Neural Network (RNN) architecture with feedback connections can process a sequence of data in order to learn the temporal features.

With the goal of learning discriminative and generalizable spatio-temporal features for fingerprint spoof detection, we utilize a joint CNN-RNN architecture that can extract deep spatial features from each frame, and learn the temporal relationship across the sequence. One of the most popular RNN architectures is Long Short-Term Memory [11] that can learn long range dependencies from the input sequences. The proposed network utilizes a time-distributed MobileNet-v1 CNN followed by a Bi-directional LSTM layer[17] and a 2-unit softmax layer for the binary classification problem *i.e.* live vs. spoof. See Figure 2.

MobileNet-v1 is a low-latency network with only $4.24$M trainable parameters compared to other networks, such as Inception-v3 (23.2M parameters) and VGG (138M parameters), which achieve comparable performance in large-scale vision tasks [26]. In low resource requirements such as smartphones and embedded devices, MobileNet-v1 is well-suited for real-time spoof detection. Most importantly, it has been shown to achieve state-of-the-art performance for fingerprint spoof detection [6] on publicly available datasets [9]. It takes an input color image of size $224 \times 224 \times 3$, and outputs a 1024-dimensional feature vector (bottleneck layer). We resize the local patches from $192 \times 192$ to $224 \times 224$ as required by the MobileNet-v1 input. For the purposes of processing a sequence of images, we utilize a Keras' TimeDistributed wrapper to utilize the MobileNet-v1 architecture as a feature extractor with shared parameters across different frames in the sequence.

### 2.5. Implementation Details

The network is designed in the Keras framework[18] and trained from scratch on a Nvidia GTX 1080Ti GPU. We utilize the MobileNet-v1 architecture without its last layer wrapped in a Time-Distributed layer. The Bi-directional LSTM layer contains 256 units and has a dropout rate of $0.25$. We utilize Adam [14] optimizer with a learning rate of $0.001$ and a binary cross entropy loss function. The network is trained end-to-end for $80$ epochs with early-stopping ($patience$ = 20) and a batch size of $4$ sequences.

[17]Experiments with uni-directional LSTM layer achieved lower or similar performance compared to when using bi-directional layer.

[18]`https://keras.io/`

## 3. Experimental Results

### 3.1. Database

In this study, we utilize $26,650$ live frames from $685$ subjects ($1,333$ unique fingers), and $32,930$ spoof frames of $7$ materials ($14$ variants) collected on SilkID SLK20R fingerprint reader. This database is constructed by combining fingerprint images collected from two sources. First, as part of the IARPA ODIN program [23], a large-scale Government Controlled Test (GCT-3) was conducted at Johns Hopkins University Applied Physics Laboratory (JHUAPL) facility in Nov. 2019, where a total of $685$ subjects with diverse demographics (in terms of age, profession, gender, and race) were recruited to present their real (live) as well as spoof biometric data (fingerprint, face, and iris). The spoof fingerprints were fabricated using 5 different spoof materials (11 variants) and a variety of fabrication techniques, including use of dental and 3D printed molds. For a balanced live and spoof data distribution, we utilize only right thumb and right index fingerprint images for the live data. Second, we collected spoof data in a lab setting[19] using dental molds casted with three different materials, namely Ecoflex (with flesh tone pigment), Crayola model magic (red and white colors), and dragon skin (with flesh tone pigment). The details of the combined database are summarized in Table 2.

### 3.2. Evaluation Metrics

The performance of the proposed approach is evaluated by measuring, (i) *True Detection Rate (TDR), i.e.,* percentage of spoofs correctly classified as spoofs for a fixed threshold, and (ii) *False Detection Rate (FDR), i.e.,* percentage of lives incorrectly classified as spoofs for a fixed threshold.

Based on the guidelines of the IARPA ODIN Program [23], we utilize the evaluation metric of TDR (%) @ FDR of $0.2\%$, which represents[20] the percentage of correctly classified spoofs, *i.e.*, unable to breach the biometric system security, when the reject rate of legitimate users (lives) $\leq$ $0.2\%$, i.e., no more than 2 out of $1,000$ users. Additionally, we also report TDR (%) @ FDR = $1.0\%$.

[19]This database will be made accessible to the interested researchers after signing a license agreement.

[20]This metric is equivalent to $[100-APCER]$ @ $[BPCER] \leq 0.2\%$, as per *ISO/IEC 30107-3* standard for the evaluation of spoof detection approaches [13], where $APCER$ (Attack Presentation Classification Error Rate) is the percentage of misclassified presentation attacks (spoofs) for a fixed threshold, and $BPCER$ (Bonafide Presentation Classification Error Rate) is the percentage of misclassified bonafide (live) presentations for a fixed threshold.

Table 3. Performance comparison (**TDR (%) @ FDR = 0.2% and 1.0%**) between the proposed approach and two state-of-the-art methods [6, 35] for **known-material scenario**, where the spoof materials used in testing are also known during training.

| Study | Approach | Architecture | TDR ($\pm$ s.d.) (%) @ FDR = 0.2% | TDR ($\pm$ s.d.) (%) @ FDR = 1.0% |
|---|---|---|---|---|
| **Baseline** | Static (Whole Image) | CNN (MobileNet-v1) | $96.90 \pm 0.78$ | $97.64 \pm 0.55$ |
| **Zhang et al. [35]** | Static (Center of Gravity Patches) | CNN (Slim-ResCNN) | $98.05 \pm 0.38$ | $98.44 \pm 0.30$ |
| **Chugh et al. [6]** | Static (Minutiae Patches) | CNN (MobileNet-v1) | $99.11 \pm 0.24$ | $99.15 \pm 0.24$ |
| **Proposed** | Dynamic (Whole Frames) | CNN-LSTM (MobileNet-v1) | $98.94 \pm 0.44$ | $99.04 \pm 0.43$ |
| | Dynamic (Center of Gravity Patches) | CNN-LSTM (Slim-ResCNN) | $99.04 \pm 0.26$ | $99.30 \pm 0.28$ |
| | Dynamic (Minutiae Patches) | CNN-LSTM (MobileNet-v1) | $\mathbf{99.25 \pm 0.22}$ | $\mathbf{99.45 \pm 0.16}$ |

Table 4. Performance comparison (**TDR (%) @ FDR = 0.2% and 1.0%**) between the proposed approach and two state-of-the-art methods [6, 35] for three **cross-material scenarios**, where the spoof materials used in testing are unknown during training.

| | Baseline Static Approaches (CNN) | | | Proposed Dynamic Approaches (CNN-LSTM) | | |
|---|---|---|---|---|---|---|
| **Unknown Material** | **Whole Image (Grayscale)** | **Slim-ResCNN [35]** | **Fingerprint Spoof Buster [6]** | **Sequence of Whole Images** | **Sequence of CoG Patches** | **Sequence of Minutiae-based Patches** |
| | **TDR @ FDR = 0.2%** | | | | | |
| **Third Degree** | 43.83 | 75.32 | 79.20 | 80.44 | 83.22 | 84.50 |
| **Gelatin** | 50.74 | 76.84 | 76.52 | 73.88 | 83.10 | 82.81 |
| **Ecoflex** | 77.37 | 87.39 | 89.23 | 87.55 | 90.94 | 91.28 |
| **Mean $\pm$ s.d.** | $57.31 \pm 17.71$ | $79.85 \pm 6.57$ | $81.65 \pm 6.70$ | $80.62 \pm 6.84$ | $85.75 \pm 4.49$ | $\mathbf{86.20 \pm 4.48}$ |
| | **TDR @ FDR = 1.0%** | | | | | |
| **Third Degree** | 60.25 | 86.15 | 89.11 | 88.10 | 94.22 | 96.20 |
| **Gelatin** | 66.40 | 90.10 | 89.00 | 89.50 | 96.38 | 96.08 |
| **Ecoflex** | 85.31 | 93.27 | 94.90 | 93.27 | 98.00 | 98.20 |
| **Mean $\pm$ s.d.** | $70.65 \pm 13.06$ | $89.84 \pm 3.57$ | $91.00 \pm 3.37$ | $90.29 \pm 2.67$ | $96.20 \pm 1.90$ | $\mathbf{96.83 \pm 1.19}$ |

## 3.3. Results

To demonstrate the robustness of our proposed approach, we evaluate it under two different settings, namely *Known-Material* and *Cross-Material* scenarios.

### 3.3.1 Known-Material Scenario

In this scenario, the same set of spoof materials are included in the train and test sets. To evaluate this, we utilize five-fold cross-validation, splitting the live and spoof datasets in 80/20 splits for training and testing, with no overlapping subjects. In each of the five folds, there are $21,320$ live and $26,400$ spoof frames in training and rest in testing. Table 3 presents the results achieved by the proposed approach on known-materials compared to a state-of-the-art approach [6] that utilizes minutiae-based local patches from static grayscale images. The proposed approach improves the spoof detection performance from TDR of $99.11\%$ ($99.15\%$) to $99.25\%$ ($99.45\%$) @ FDR = $0.2\%$ ($1.0\%$).

### 3.3.2 Cross-Material Scenario

In this scenario, spoof materials used in the test set were not included in the training set. This scenario is simulated by adopting a leave-one-out protocol, where one material (including all its variants) is removed from training, and is then used for evaluating the trained model. It is a more challenging and practical setting as it evaluates the cross-material generalizability of a spoof detector against materials that are unknown during training. For instance, in one of the cross-material experiments, we exclude Third Degree silicone spoof material, including its variants (pigmented, tan, beige, and medium) from training, and use them for testing. The live data is randomly divided in a $80/20$ split, with no subject overlap, for training and testing, respectively.

Table 4 presents the performance achieved by the proposed approach, on three cross-material experiments, compared to two state-of-the-art[21] methods [6, 35]. We observe that utilizing sequence of whole images significantly improves the performance achieved by static whole images (from TDR = $57.31\%$ ($70.65\%$) to TDR = $80.62\%$ ($90.29\%$) @ FDR = $0.2\%$ ($1.0\%$)). However, it is slightly lower than the performance achieved by the static patch-based approaches, *i.e.*, TDR = $81.65\%$ ($91.00$) @ FDR = $0.2\%$ ($1.0\%$). This could be due to the drawbacks of utilizing whole images compared to local patches [6] for training a deep neural network, namely, (i) directly resizing whole images with large blank areas around friction ridge area, from $630 \times 390$ to $224 \times 224$, results in the friction ridges occupying less than $20\%$ of the original image size, (ii) resizing a rectangular image to a square image leads to different amounts of information retained in the two spatial dimensions, and (iii) downsizing an image typically leads

[21] The algorithm by Zhang et al. [35], Slim-ResCNN, was the winner of the LivDet 2017 competition [21].

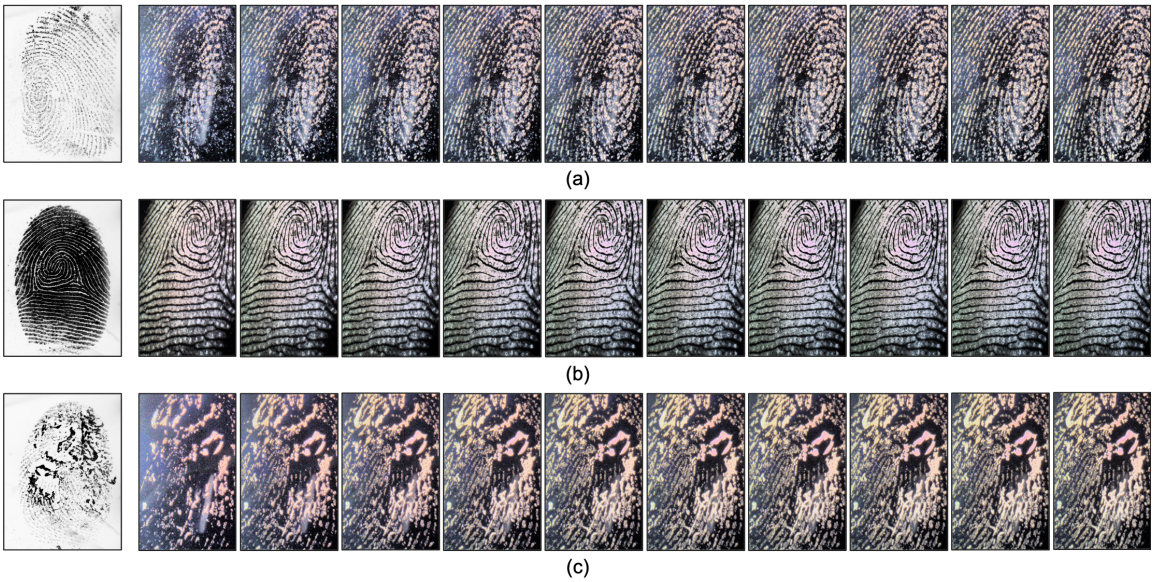


Figure 5. Examples of challenging cases in a cross-material scenario where Third Degree silicone spoofs are left-out from training. (a) A Third Degree silicone spoof and (b) a live fingerprint, are incorrectly classified as live and spoof, respectively, by the static (grayscale) state-of-the-art approach, and is correctly classified by the proposed dynamic approach utilizing sequence of frames. (c) A live fingerprint, with visible damages to the friction ridges, is incorrectly classified by both state-of-the-art static and the proposed dynamic approaches.

to significant loss of discriminatory information. However, these drawbacks are addressed by using a sequence of local patches in the proposed approach, which achieves a superior cross-material spoof detection performance of TDR = $86.20\%$ ($96.83\%$) @ FDR = $0.2\%$ ($1.0\%$). Figure 5 presents three challenging cases where the Third Degree silicone spoof is left out from training, and is used in testing.

### 3.4. Spoof Detection Times

The proposed network architecture takes around $4-6$ hours to converge when trained with sequences of whole frames, and $24-30$ hours with sequences of minutiae-based local patches, using a Nvidia GTX 1080Ti GPU. An average of 11 (13) sequences of minutiae-based local patches are extracted from the live (spoof) frames. After the sequence of fingerprint frames are captured, the average spoof detection time, including preprocessing, minutiae-detection, patch extraction, and sequence generation and inference, on a Nvidia GTX 1080 Ti GPU is $58$ms for full frame-based sequences, and $393$ms for minutiae-based patch sequences[22].

## 4. Conclusions

A robust and generalizable spoof detector is pivotal to assure the security and privacy of fingerprint recognition systems against unknown spoof attacks. In this study, we utilized a sequence of local patches centered at detected minutiae from ten color frames captured at $8$ fps as the finger is presented on the sensor. We posit that the dynamics involved in the presentation of a finger, such as skin blanching, distortion, and perspiration, provide discriminating cues to distinguish live from spoofs. We utilize a jointly learned CNN-LSTM model to learn the spatio-temporal dynamics across different frames in the sequence. The proposed approach improves the spoof detection performance from TDR of $99.11\%$ ($99.15\%$) to $99.25\%$ ($99.45\%$) @ FDR = $0.2\%$ ($1.0\%$) in known-material scenarios, and from TDR of $81.65\%$ ($91.00\%$) to $86.20\%$ ($96.83\%$) @ FDR = $0.2\%$ ($1.0\%$) in cross-material scenarios. In future, we will explore firmware changes to (i) incorporate a quality check on captured frames during data collection to reduce the required sequence length, and (ii) capture a smaller sensing area than $630 \times 390$ for a faster frame rate.

## 5. Acknowledgement

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017 - 17020200004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

[22]It takes 1.25 seconds to capture a sequence of 10 frames using SilkID SLK20R reader. This time could be further reduced by capturing a smaller sensing area than $630 \times 390$ and/or utilizing a shorter sequence.

## References


[1] A. Antonelli, R. Cappelli, D. Maio, and D. Maltoni. Fake finger detection by skin distortion analysis. *IEEE Transactions on Information Forensics and Security*, 1(3):360–373, 2006.

[2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[3] K. Cao and A. K. Jain. Hacking mobile phones using 2D Printed Fingerprints. MSU Tech. report, MSU-CSE-16-2 `https://www.youtube.com/watch?v=fZJI_BrMZXU`, 2016.

[4] K. Cao, D.-L. Nguyen, C. Tymoszek, and A. K. Jain. End-to-end latent fingerprint search. *IEEE Transactions on Information Forensics and Security*, 15:880–894, 2019.

[5] T. Chugh, K. Cao, and A. K. Jain. Fingerprint Spoof Detection using Minutiae-based Local Patches. In *IEEE International Joint Conference on Biometrics (IJCB)*, 2017.

[6] T. Chugh, K. Cao, and A. K. Jain. Fingerprint Spoof Buster: Use of Minutiae-centered Patches. *IEEE Transactions on Information Forensics and Security*, 13(9):2190–2202, 2018.

[7] T. Chugh and A. K. Jain. OCT Fingerprints: Resilience to Presentation Attacks. *arXiv preprint arXiv:1908.00102*, 2019.

[8] J. J. Engelsma, S. S. Arora, A. K. Jain, and N. G. Paulter. Universal 3D wearable Fingerprint Targets: Advancing Fingerprint Reader Evaluations. *IEEE Transactions on Information Forensics and Security*, 13(6):1564–1578, 2018.

[9] L. Ghiani, D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli, and S. A. Schuckers. Review of the Fingerprint Liveness Detection (LivDet) competition series: 2009 to 2015. *Image and Vision Computing*, 58:110–128, 2017.

[10] L. J. González-Soler, M. Gomez-Barrero, L. Chang, A. Pérez-Suárez, and C. Busch. Fingerprint Presentation Attack Detection Based on Local Features Encoding for Unknown Attacks. *arXiv preprint arXiv:1908.10163*, 2019.

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

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*, 2017.

[13] International Standards Organization. ISO/IEC 30107-1:2016, Information Technology—Biometric Presentation Attack Detection—Part 1: Framework. https://www.iso.org/standard/53227.html, 2016.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] J. Kolberg, M. Gomez-Barrero, and C. Busch. Multi-algorithm benchmark for fingerprint presentation attack detection with laser speckle contrast imaging. In *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2019.

[16] X. Li, B. Gunturk, and L. Zhang. Image demosaicing: A systematic survey. In *Visual Communications and Image Processing*, volume 6822. International Society for Optics and Photonics, 2008.

[17] E. Marasco and C. Sansone. Combining perspiration-and morphology-based static features for fingerprint liveness detection. *Pattern Recognition Letters*, 33(9):1148–1156, 2012.

[18] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, editors. *"Handbook of Biometric Anti-Spoofing: Presentation Attack Detection"*. Springer, 2 edition, 2019.

[19] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. In *Proc. SPIE*, volume 4677, pages 275–289, 2012.

[20] Y. Moolla, L. Darlow, A. Sharma, A. Singh, and J. Van Der Merwe. Optical coherence tomography for fingerprint presentation attack detection. In *Handbook of Biometric Anti-Spoofing*. Springer, 2019.

[21] V. Mura, G. Orrù, R. Casula, A. Sibiriu, G. Loi, P. Tuveri, L. Ghiani, and G. L. Marcialis. LivDet 2017 Fingerprint Liveness Detection Competition 2017. In *Proc. IAPR International Conference on Biometrics (ICB)*, pages 297–302, 2018.

[22] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. Fingerprint Liveness Detection Using Convolutional Neural Networks. *IEEE Transactions on Information Forensics and Security*, 11(6):1206–1213, 2016.

[23] Office of the Direction of National Intelligence (ODNI), IARPA. IARPA-BAA-16-04 (Thor). `https://www.iarpa.gov/index.php/research-programs/odin/odin-baa`, 2016.

[24] S. T. Parthasaradhi, R. Derakhshani, L. A. Hornak, and S. A. Schuckers. Time-series Detection of Perspiration as a Liveness Test in Fingerprint Devices. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(3):335–343, 2005.

[25] R. Plesh, K. Bahmani, G. Jang, D. Yambay, K. Brownlee, T. Swyka, P. Biometrics, P. Johnson, A. Ross, and S. Schuckers. Fingerprint Presentation Attack Detection utilizing Time-Series, Color Fingerprint Captures. In *IEEE International Conference on Biometrics (ICB)*, 2019.

[26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *Proc. International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[27] S. Schuckers and P. Johnson. Fingerprint Pore Analysis for Liveness Detection, Nov. 14 2017. US Patent 9,818,020.

[28] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proc. IEEE CVPR*, pages 2818–2826, 2016.

[30] P. Thévenaz, T. Blu, and M. Unser. Image Interpolation and Resampling. *Handbook of Medical Imaging, Processing and Analysis*, 1(1):393–420, 2000.

[31] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia. Biometric Presentation Attack Detection: Beyond the Visible Spectrum. *IEEE Transactions on Information Forensics and Security*, 2019.

[32] Z. Xia, C. Yuan, R. Lv, X. Sun, N. N. Xiong, and Y.-Q. Shi. A Novel Weber Local Binary Descriptor for Fingerprint Liveness Detection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[33] D. Yambay, L. Ghiani, G. L. Marcialis, F. Roli, and S. Schuckers. Review of Fingerprint Presentation Attack Detection Competitions. In S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, editors, *Handbook of Biometric Anti-Spoofing*. Springer, 2019.

[34] W.-Y. Yau, H.-T. Tran, E.-K. Teoh, and J.-G. Wang. Fake finger detection by finger color change analysis. In *International Conference on Biometrics*, pages 888–896. Springer, 2007.

[35] Y. Zhang, D. Shi, X. Zhan, D. Cao, K. Zhu, and Z. Li. Slim-ResCNN: A Deep Residual Convolutional Neural Network for Fingerprint Liveness Detection. *IEEE Access*, 7:91476–91487, 2019.