DeFraudNet:End2End Fingerprint Spoof Detection using Patch Level Attention

B.V.S Anusha
IIT Bombay
bvs.anusha26@gmail.com

Sayan Banerjee
IIT Bombay
sayan91.ban@gmail.com

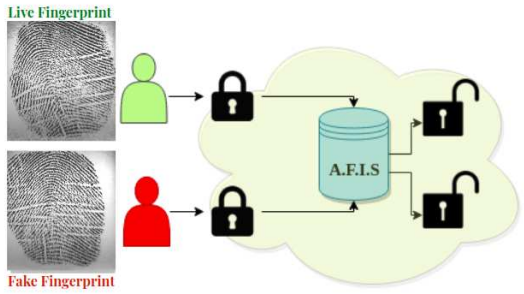
Subhasis Chaudhuri
IIT Bombay
sc@ee.iitb.ac.in

Abstract

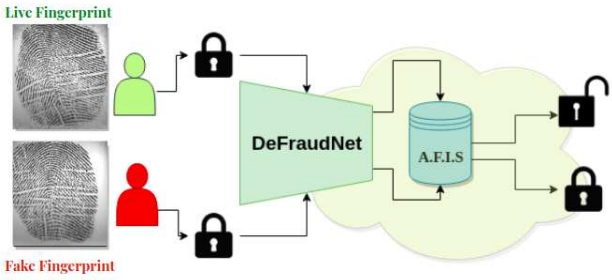
*In recent years, fingerprint recognition systems have made remarkable advancements in the field of biometric security as it plays an important role in personal, national and global security. In spite of all these notable advancements, the fingerprint recognition technology is still susceptible to spoof attacks which can significantly jeopardize the user security. The cross sensor and cross material spoof detection still pose a challenge with a myriad of spoof materials emerging every day, compromising sensor interoperability and robustness. This paper proposes a novel method for fingerprint spoof detection using both global and local fingerprint feature descriptors. These descriptors are extracted using DenseNet which significantly improves cross-sensor, cross-material and cross-dataset performance. A novel patch attention network is used for finding the most discriminative patches and also for network fusion. We evaluate our method on four publicly available datasets: LivDet 2011, 2013, 2015 and 2017. A set of comprehensive experiments are carried out to evaluate cross-sensor, cross-material and cross-dataset performance over these datasets. The proposed approach achieves an average accuracy of **99.52%, 99.16%** and **99.72%** on LivDet 2017, 2015 and 2011 respectively outperforming the current state-of-the-art results by **3%** and **4%** for LivDet 2015 and 2011 respectively.*

1. Introduction

Over the past few years, with an exponential increase in the usage of IoT devices, biometrics have played a key role in maintaining user confidentiality for various operations. According to a report [34, 35], the fingerprint is the most widely used biometric identity over any other existing biometrics. It's extensive global usage makes it vulnerable to several security threats like, identity theft, account hacking, unauthorized access and many more. One of the many threats that can severely compromise fingerprint security is fingerprint spoofing. It is a technique in which fake fingerprint impressions are created to fool the fingerprint sensor



(a)



(b)

Figure 1: (a) Any Automatic Fingerprint Identification System (A.F.I.S) without spoof detection system is susceptible to identity theft as the Fake and Live fingerprint have similar properties. (b) The presence of Spoof Detection Systems like DeFraudNet, prevents such threats by filtering out the Fake fingerprints and allow only the Live fingerprints into the A.F.I.S.

to make unauthorized access into the fingerprint system.

To overcome this problem, over the years various spoof detection methods have been developed. But as the fingerprint sensing technology advances, so does the spoofing technology, which increases the degree of difficulty for organizations to protect their biometric systems from being compromised. The spoof fingerprints can be fabricated using various materials like latex, ecoflex, clay, wood glue, gum etc. Visually, no clear distinction can be observed be-

tween the spoofed and live fingerprint on the sensor imagery as shown in Figure 1. Therefore, it is necessary to extract textural, anatomical or physiological features for spoof detection.

Traditional fingerprint spoof detection methods [2, 25, 3, 13] extract handcrafted texture features and use them to classify fingerprints into live/spoof classes. These methods require high resolution images and exhaustive feature tuning. As a result, they become sensitive to the computed features and input noise. To address this challenge, Menotti *et al.* developed a CNN based network called '*spoofnet*' [22] and trained the network on LivDet 2013 dataset. This supervised network learns robust high level features which improves the performance by a significant margin. Following the same path, Nouguira *et al.* [26] and Pala *et al.* [29] applied standard CNN networks (VGGNet, AlexNet) pretrained on large ImageNet dataset and fine tuned on LiveDet dataset. Use of a pre-trained network with transfer learning further enhances the performance. Inspired by these works, a lot of CNN based fingerprint spoof detection methods have been proposed [26, 15, 39, 31, 7]. However, even though these methods provide affirmative results on fingerprint spoof detection on a single sensor data, their cross material or cross sensor performance is very poor.

To improve the performance of spoof detection across cross-material and cross-sensor, we propose a novel CNN based end-to-end model which takes global images and the corresponding overlapping local patches to classify the fingerprints as live or fake. It should be noted that T. Chugh *et al.* in their paper [8] also used fingerprint patches but they are computed using significant minutiae points. This method obtains significantly improved results but has a few drawbacks. Firstly, minutiae point extraction requires high resolution input fingerprint images ($>$ 500dpi) and secondly, due to this preprocessing step, the model does not operate in end-to-end fashion. Taking these problems into consideration, the proposed method extracts suitable patches implicitly using a novel attention mechanism, referred as patch attention, along with channel attention and spatial attention modules. The role of the channel and spatial attention modules are to highlight useful information from each patch for live/spoof detection which together helps the patch attention module to identify appropriate patches from a pool of patches. The salient part is that, all of these attentions are learned using different neural networks with back propagation of the gradient of the main loss function. Hence, the proposed patch extraction method can be integrated with any neural network based system. DenseNet has been used as the base network for feature computation of the input image and patches. It is lighter in terms of memory usage than the existing state-of-the-art networks even with the presence of attention modules. One of the most challenging part for any patch based method is to develop an efficient fusion strategy to integrate information obtained across individual patches. Instead of using the standard approaches which predominantly use strategies like majority voting the proposed method uses patch attention network. It first determines importance of each patch to the final objective of live/spoof classification and then aggregates this information based upon their individual importance scores. By this process, the network itself learns to perform better decision fusion which also indirectly helps to learn better features for identifying spoof fingerprints.

Extensive experiments and ablation studies have been performed on various challenging fingerprint spoof detection datasets: LivDet 2011,2013,2015,2017, which demonstrate significant improvement over state-of-the-art methods. The main contributions of this paper are:

- With the use of local patch features and global contextual image features, the proposed method obtains significantly better performance than the existing state-of-the-art fingerprint spoof detection methods across cross material and cross datasets.
- The proposed method also exploits handcrafted features (i.e. LBP and Gabor filters) which are integrated along with the input image. The combination of handcrafted features and deep high level semantic features show a significant improvement over cross material and cross datasets.
- The proposed novel patch attention network learns highly discriminative patches with additional channel and spatial attention modules using gradients of the live/spoof classification error. Therefore unlike the existing methods, the proposed method does not require any separate intermediate step for patch discrimination. It learns to identify useful patches by itself and the complete network can be operated in an end-to-end manner.
- The proposed model employs a novel feature fusion strategy using patch attention. It learns to aggregate information across patches which in turn makes it less susceptible to input noise and error in initial patch computation.
- The proposed network reduces the memory usage by at least fifty percentage as compared to the state-of-the-art networks. Therefore, it can be easily embedded into tiny low powered, low memory IoT devices (for example: mobile phones).

2. Related Work

In this section, we briefly review the existing work on fingerprint spoof detection, image classification using

DenseNet and use of attention network for binary classification.

Fingerprint spoof detection has been extensively studied and experimented over the centuries as it poses a huge threat to security. The spoof detection methods can be categorised into hardware based methods and software based methods. Hardware based methods involve external fingerprint sensing devices by adding sensors to detect living traits like blood pressure, blood sugar, skin distortion or odor [33, 6]. The software based methods involve extracting various handcrafted features from the sensor image of fingerprint and then classify them as live/fake. These handcrafted features can be broadly classified as outer anatomical features like ridge strength, pore location and their distribution, continuity and clarity [21] etc. or physiological features like perspiration patterns [20] and texture based features or statistical features like Weber Local Descriptors (WLD) [14], rotation invariant Local Phase Quantization (LPQ) [11] features or Binarized Statistical Image Features (BSIF) [10] or Local Binary Patterns [27] etc. or a combination of these features. We use textural features (i.e LBP and Gabor filters) in our approach. LBP [27] provides good textural variation for liveliness detection. In 2015 LiveDet competition, Noguira *et al.* [26] obtained a state-of-the-art accuracies by using LBP and transfer learning for binary classification. In this method, the fingerprint data was preprocessed by extracting LBP features and then classified using pre-trained standard networks VGG, AlexNet. Pala *et al.* [29] and Menotti *et al.* [22] also use similar methods for spoof results and obtain significantly better results than spoof detection using only handcrafted features. However, all of these methods perform poorly on cross-sensor and cross material tests. To overcome this, T.Chugh *et al.* [8] developed a robust spoof detection method using local minutiae based patches and trained them using MobileNet-v1 [17]. They obtained state-of-the-art results for intra-sensor, inter-sensor, cross-material and cross sensor over three datasets (LivDet 2011,2013,2015). But their method was not end to end as it involved two stage training process.

DenseNet [18] is being extensively used for various applications like image classification [43, 42], segmentation [41], image super-resolution [37] etc. This can be attributed to its memory efficiency, computational efficiency and feature re-usability properties. This network alleviates vanishing gradient problem and strengthens feature propagation. It's feature reusability ensures memory and computational efficiency. George *et al.* in their paper [9] used DenseNet to counterfeit presentation attack on human faces. Huang *et al.* [19] employed DenseNet along with LSTM for audio spoof attack detection. But with the best of our knowledge, DenseNet has not been used for fingerprint spoof detection.

Channel and Spatial Attention networks highlights the salient features from visual data [36, 5, 38]. Attention has been used for face anti-spoofing [4] but as far as we know, attention has never been exploited for fingerprint spoof detection. In this paper we use attention in a similar way as mentioned in [39].

Patch Attention Network is a novel attention module proposed in this paper, which learns the most discriminative patches from a set of input patches. The patch-based training always poses a problem of fusing the patch-level predictions considering the fact that not all patches are equally informative. Le Hou *et al.* [16] overcame this challenge by using a novel Expectation-Maximization (EM) based method that automatically locates discriminative patches robustly by utilizing the inter-spatial relationship of patches.

3. Proposed Approach

We propose an end-to-end network which combines both the global fingerprint image features and local patch based image features to obtain the final binary classification results. The whole procedure can be summarized in 6-steps. The first step involves preprocessing where the gray scale images of LiveDet datasets are converted into 3-channel images by adding LBP and Gabor feature channels. The second step is patch extraction followed by simultaneous training of two DenseNets for whole image and patch feature extractions. After the feature map extraction, channel, spatial and patch attention are performed on the patches. This is followed by the final step of patch feature fusion with the whole image and classification. An overview of the proposed model is shown in Figure 2.

3.1. Preprocessing

As stated in the above section, we use preprocessing to convert the gray scale fingerprint sensor images to 3-channel images. As the fingerprints sensor images visually do not have any discriminative features, these additions facilitate the network to obtain more robust and better classification results on intra-sensor, inter-sensor, cross-material and cross-sensor datasets. We use two preprocessing methods, the first one is Local Binary Pattern and second is Gabor filters.

Local Binary Patterns(LBP) are texture descriptors made popular by the work of Ojala *et al.* [27] in 1994. An LBP operator divides an image into cells of n$\times$n and a label is assigned to each cell after thresholding the neighboring pixel with the center pixel. The end result is an 2^n-bit code representing all the 2^n possible combinations. As the comparison of the central pixel is made with the neighborhood pixels, LBP is an illumination invariant descriptor. We use an extension of this, called uniform or rotation-invariant LBP descriptor [28]. This is used to reduce the length of feature vector as it contains only two bit transitions; 0 or 1, which in turn reduces the memory requirement. The rotation-

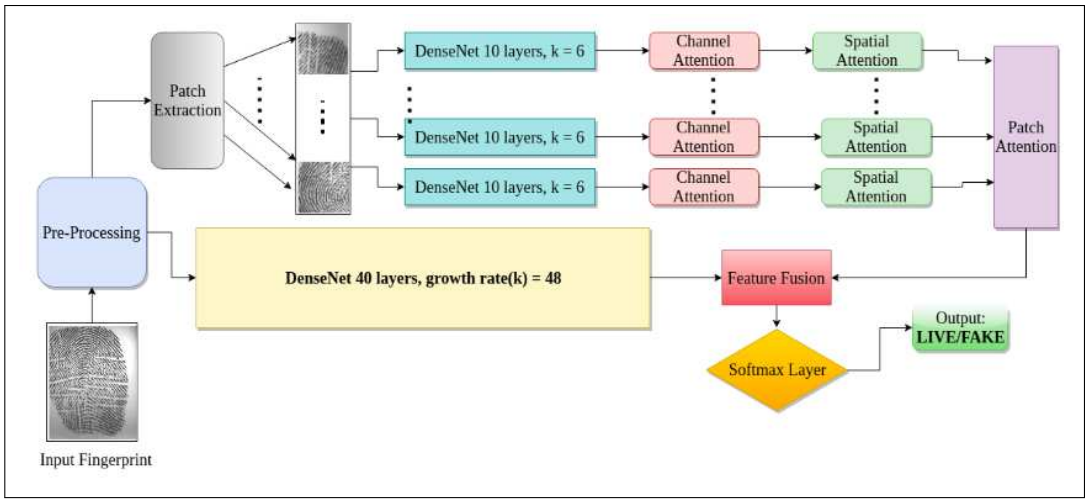


Figure 2: Complete overview of the proposed model. It should be noted that in this figure the dotted lines signify continuation of the blocks.

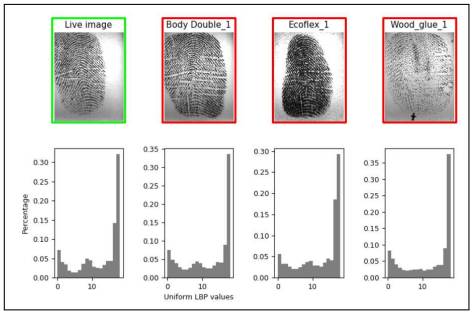


Figure 3: LBP histograms for Live fingerprint and its corresponding fake counterparts for user 001_1_25 on Greenbit sensor from LiveDet 2017 dataset.

invariant LBP histograms of live and it's corresponding fake fingerprints are shown in Figure 3.

We create the third channel using the method of Gabor filtering.

Gabor filter is extensively used for various image processing applications like, edge detection, feature extraction, texture analysis etc. These filters can be viewed as special classes of band pass filters that have been known to possess optimal localization properties in both spatial and frequency domains. This makes them suitable for texture classification problems. We use a Gabor filter with kernel size of 51 and theta of 11.55 degrees to obtain the edges and texture intricacies of the fingerprint images. The output of Gabor filtered fingerprint is shown in the Figure 4.

Data Augmentation is also added on the dataset to increase classifier robustness. We add standard augmentation methods like image resizing, random affine transforms and random crop on the data. These augmentation methods considerably reduce the memory requirements of the network without compromising its performance.

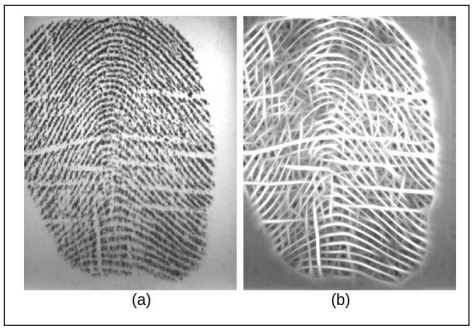


Figure 4: (a) Original image , (b) Gabor filtered output for kernel size 51 and theta 11.55 degrees.

3.2. Implemented Network Details

Base Network: Any of the standard network architectures like VGG [31], ResNets [15], Google Inception [32], MobileNet [7], DenseNets [18] etc. can be used to extract feature maps for both the whole image and the patches. We chose DenseNet as our backbone network because of its several compelling advantages like: (i) DenseNet has successfully obtained state-of=the-art results with high memory efficiency and less computation. (ii) It also has very few parameters for example; It has 0.8 million parameter for 100-layers with a growth rate k = 12 as compared to other networks like VGG which has 138 million parameter and ResNet-50 which has 25 million parameters. (iii)

Due to the presence of densely connected layers, it alleviates the vanishing gradient problem and also strengthens feature propagation. (iv) It is easier to export this network on to hardware devices like mobile phones, FPGAs etc. due to it's small number of parameters.

Our model consists of two standard DenseNets which are trained from scratch. We chose DenseNet with 40 layers and growth rate(k) 48 as network 1. This is obtained after several trails with different network parameters like DenseNet-121, DenseNet-169, DenseNet-190 etc.as it was optimal in terms to memory efficiency and accuracy. The second DenseNet consists of 10 layers and has a growth rate of $k = 6$. The results of Network-1 and Network-2 when trained on LivDet 2015 dataset separately are summarized in the Table 1.

Training Dataset	Testing Dataset	Network-1 ACE	Network-2 ACE
CrossMatch 2015	CrossMatch 2015	**0.21**	1.70
CrossMatch 2015	Hi Scan 2015	3.67	**2.13**
CrossMatch 2015	GreenBit 2015	4.12	**3.27**
CrossMatch 2015	Digital Persona 2015	6.87	**2.67**
GreenBit 2015	GreenBit 2015	**0.88**	2.75
GreenBit 2015	Hi Scan 2015	**4.23**	**3.48**
GreenBit 2015	Digital Persona 2015	**5.75**	**2.87**
GreenBit 2015	CrossMatch 2015	**2.64**	3.72
Digital Persona 2015	Digital Persona 2015	1.37	**0.67**
Digital Persona 2015	Hi Scan 2015	5.81	**2.43**
Digital Persona 2015	GreenBit 2015	**3.52**	4.72
Digital Persona 2015	CrossMatch 2015	**2.66**	5.43

Table 1: Comparison of intra sensor and cross sensor ACE between Network-1 and Network-2 for LivDet 2015 dataset

Attention Module: We propose a novel patch attention model along with channel and spatial attention networks which identifies the most discriminative patches amongst the n given patches and allocates corresponding weights to each patch. As we can see from the overview of the complete proposed model in Figure 2, once feature maps are obtained from each patch after training it with the second DenseNet (10 layer, growth rate = 6), they are passed through a channel attention network.

Channel Attention Network [39] is used to obtain the channel attention map by exploiting the inter channel relationship. Using channel attention network we get **what** is the most important feature of an image. For obtaining this we compress the feature map in spatial dimensions ($H \times W$) using average pooling and max pooling techniques. Let the feature vector at the end of each patch be defined as f, and let the image dimensions be $(C \times H \times W)$. The feature vector obtained after average pooling and max pooling can be given as f_{avg} and f_{max} respectively. These feature maps are forwarded to a shared layer in the channel attention network which is a multi-layer perceptron with one hidden layer. The hidden layer reduces the channel parameters by a defined reduction ratio r. The operation of the channel attention model is mathematically summarized in equation (1).

$$F^c = \sigma(MLP(f_{avg}) + MLP(f_{max})) \tag{1}$$

where σ denotes the sigmoid function and F^c denotes channel attention output. This channel attention is followed by a spatial attention module with the output of channel attention network is given as an input to it.

Spatial Attention Network [39] highlights **where** the important information exists in the the patch. This network extracts the most informative regions in the feature map using inter-spatial relationship. Like channel attention, to obtain spatial attention, the channel information is restored and compressed by average pooling (F^c_{avg}) and max pooling (F^c_{max}) across the channel dimension. These generate two 2-D maps of dimensions $(1 \times H \times W)$. The two 2-D feature maps are concatenated and convoluted with a standard convolution layer to obtain a 2-D spatial attention map. The operation of the spatial attention model is mathematically summarized in equation (2).

$$F^s = \sigma(L([F^c_{avg}; F^c_{max}])) \tag{2}$$

where, L is a filter, σ denotes the sigmoid function and F^s denotes spatial attention output. The output of spatial attention model is the input for the patch attention network.

Patch Attention Network is a novel attention module proposed in this paper, which highlights the most discriminative patches among the n patches in the network. The input for the patch attention model is a concatenated output map of all the maps obtained after channel attention and spatial attention modules. Therefore, the dimension of the final input to the patch attention model is $n \times C \times H \times W$ where, n is the number of patches. Similar to the earlier attention network, the features of both the channel attention and the spatial attention are aggregated using two pooling operations. This ensures that the vital information obtained from the channel and spatial attention modules is preserved. We first apply average pooling across spatial dimension which gives us a map M_s of dimension $n \times C$. After average pooling across spatial map, average pooling across channel map is performed which gives an output map M_c of dimension $n \times 1$. It is then passed through a multi-layer perceptron with one hidden layer. The hidden layer reduces the patch parameters by a given reduction ratio r. The operation of the patch attention network is mathematically shown in equation(3) and equation(4).

$$F^p = \sigma(MLP(M_c(M_s(F^s)))) \tag{3}$$

In terms of weights this can be given as:

$$F^p_i = \sigma(W_i(M_c)) \tag{4}$$

Where, σ denotes sigmoid, and W_i is the weight of i^{th} patch such that $W_i \in R$

Dataset Year	Training Sensor	Spoof Materials used for training	Testing Sensor	Spoof Materials used for testing	ACE = ((FPR + FNR) / 2)*100 %	
					Current S.O.T.A	Proposed Network
LiveDet 2017	DigitalPersona U.are.U 5160	Wood Glue, Ecoflex and Body Double	DigitalPersona U.are.U 5160	Latex,Gelatine, Liquid Ecoflex	–	**0.71**
			GreenBit DactyScan84c	Latex,Gelatine, Liquid Ecoflex	–	**2.19**
			Orcanthus Certis2 Imag	Latex,Gelatine, Liquid Ecoflex	–	**3.21**
	GreenBit DactyScan84c	Wood Glue, Ecoflex and Body Double	GreenBit DactyScan84c	Latex,Gelatine, Liquid Ecoflex	–	**0.68**
			Orcanthus Certis2 Imag	Latex,Gelatine, Liquid Ecoflex	–	**2.73**
			DigitalPersona U.are.U 5160	Latex,Gelatine, Liquid Ecoflex	–	**5.32**
	Orcanthus Certis2 Imag	Wood Glue, Ecoflex and Body Double	Orcanthus Certis2 Imag	Latex,Gelatine, Liquid Ecoflex	–	**0.03**
			GreenBit DactyScan84c	Latex,Gelatine, Liquid Ecoflex	–	**6.77**
			DigitalPersona U.are.U 5160	Latex,Gelatine, Liquid Ecoflex	–	**7.46**
Average					NA	**0.48**
LiveDet 2015	Cross-Match L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	Cross-Match L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	1.525[8]	**0.23**
			Cross-Match L SCAN GUARDIAN	Gelatin, OOMOO	2.475[8]	**0.37**
			DigitalPersona U.are.U 5160	Ecoflex, Gelatin, Latex, Wood Glue	–	**1.12**
			GreenBit DactyScan84c	Ecoflex, Gelatin, Latex, Wood Glue	–	**1.97**
			HiScan-PRO	Ecoflex, Gelatin, Latex, Wood Glue	–	**1.78**
			Cross-Match L SCAN GUARDIAN - 2013	Ecoflex, Gelatin, Latex, Wood Glue	–	**1.25**
	HiScan-PRO	Ecoflex, Gelatin, Latex, Wood Glue	HiScan-PRO	Ecoflex, Gelatin, Latex, Wood Glue	–	**0.63**
			HiScan-PRO	Liquid−Ecoflex,RTV	–	**2.57**
			DigitalPersona U.are.U 5160	Ecoflex, Gelatin, Latex, Wood Glue	–	**1.87**
			GreenBit DactyScan84c	Ecoflex, Gelatin, Latex, Wood Glue	–	**2.05**
			Cross-Match L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**2.31**
	GreenBit DactyScan84c	Ecoflex, Gelatin, Latex, Wood Glue	GreenBit DactyScan84c	Ecoflex, Gelatin, Latex, Wood Glue	3.9[8]	**1.81**
			GreenBit DactyScan84c	Liquid−Ecoflex,RTV	5.65[8]	**2.82**
			DigitalPersona U.are.U 5160	Ecoflex, Gelatin, Latex, Wood Glue	–	**2.51**
			HiScan-PRO	Ecoflex, Gelatin, Latex, Wood Glue	–	**2.09**
			Cross-Match L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**2.35**
	DigitalPersona U.are.U 5160	Ecoflex, Gelatin, Latex, Wood Glue	DigitalPersona U.are.U 5160	Ecoflex, Gelatin, Latex, Wood Glue	7.85[8]	**1.72**
			DigitalPersona U.are.U 5160	Liquid−Ecoflex,RTV	7.05[8]	**2.63**
			GreenBit DactyScan84c	Ecoflex, Gelatin, Latex, Wood Glue	–	**4.67**
			HiScan-PRO	Ecoflex, Gelatin, Latex, Wood Glue	–	**2.94**
			Cross-Match L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**3.95**
Average					0.97*[8]	**0.84***
LiveDet 2013	Biometrika FX2000	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	Biometrika FX2000	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	**0.20[8]**	0.24
			Biometrika FX2000 -2011	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	31.16[8]	**11.90**
			Italdata ET10	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	**1.5[29]**	6.46
			Crossmatch L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**2.32**
	Italdata ET10	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	Italdata ET10	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	**0.30[8]**	0.32
			Crossmatch L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**1.35**
			Biometrika FX2000	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	2.30[26]	**1.75**
	Crossmatch L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	Crossmatch L SCAN GUARDIAN	Body Double,Ecoflex,Playdoh	–	**0.34**
			Crossmatch L SCAN GUARDIAN-2015	Body Double,Ecoflex,Playdoh	–	**1.58**
			Italdata ET10	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	–	**2.47**
			Biometrika FX2000	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	–	**2.65**
Average					**0.25***	**0.28***
LiveDet 2011	Digital 4000B	Gelatin, Latex, Playdoh, Silicone, WoodGlue	Digital 4000B	Gelatin, Latex, Playdoh, Silicone, WoodGlue	**1.61[8]**	2.43
			Biometrika FX2000	Ecoflex, Silgum, Gelatin, WoodGlue, Latex	–	**6.21**
			Italdata ET10	Ecoflex, Gelatin, Latex, Silgum, WoodGlue	–	**5.17**
			Sagem MSO300	Gelatin, Latex, Playdoh, Silicone, Wood Glue	–	**11.89**
	Biometrika FX2000	Ecoflex, Silgum, Gelatin, WoodGlue, Latex	Biometrika FX2000	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	1.24[8]	**0.19**
			Biometrika FX2000	Ecoflex, Silgum, Gelatin, WoodGlue, Latex	7.60[8]	**2.18**
			Italdata ET10	Ecoflex, Gelatin, Latex, Silgum, WoodGlue	25.35[8]	**2.13**
			Sagem MSO300	Gelatin, Latex, Playdoh, Silicone, Wood Glue	–	**2.43**
			Digital 4000B	Gelatin, Latex, Playdoh, Silicone, WoodGlue	–	**3.48**
	Sagem MSO300	Gelatin, Latex, Playdoh, Silicone, Wood Glue	Sagem MSO300	Gelatin, Latex, Playdoh, Silicone, Wood Glue	1.23[29]	**0.96**
			Italdata ET10	Ecoflex, Gelatin, Latex, Silgum, WoodGlue	–	**2.46**
			Biometrika FX2000	Ecoflex, Silgum, Gelatin, WoodGlue, Latex	–	**6.78**
			Digital 4000B	Gelatin, Latex, Playdoh, Silicone, WoodGlue	–	**11.35**
	Italdata ET10	Ecoflex, Gelatin, Latex, Silgum, WoodGlue	Italdata ET10	Ecoflex, Gelatin, Latex, Silgum, WoodGlue	2.45[8]	**1.06**
			Italdata ET10 - 2013	Ecoflex, ,Gelatin, Modasil, Latex, Wood Glue	6.70[8]	**2.54**
			Biometrika FX2000	Ecoflex, Silgum, Gelatin, WoodGlue, Latex	25.21[8]	**9.90**
			Digital 4000B	Gelatin, Latex, Playdoh, Silicone, WoodGlue	–	**13.47**
			Sagem MSO300	Gelatin, Latex, Playdoh, Silicone, Wood Glue	–	**5.93**
Average					1.63*	**1.16***

Table 2: The overall performance comparison between different sensors across different datasets. * The average ACE takes only the intra sensor errors into consideration and for LivDet 2013, only Italdata and Biometrika are considered for comparison purpose only.

4. Experimental Results

4.1. Implementation Details

We train our model from scratch and implement it with a batch size of $n + 1$. Each fingerprint is of size 224×224 from which n patches of size 56×56 are extracted and trained on Network-2 which is a 10-layer DenseNet with a growth rate of $k = 6$. The whole fingerprint image is trained on Network-1 which is a 40 layer DenseNet with a growth rate of $k = 48$. Standard cross-entropy loss is used for training both the networks. To optimize the loss, we use the SGD optimizer with nesterov momentum [30] with a learning rate initialized to 0.006 and a weight decay of 1e-4. The network is trained for 500 epochs with a momentum initialized at 0.9. Our model is implemented on Pytorch platform and has 2.74M parameters. It is implemented on GeForce RTX 2080 Ti GPU. All our experiments on different datasets follow the same setting as above.

4.2. Evaluation Metric

The proposed approach is evaluated using the performance evaluation metrics used for LiveDet[1] The following metrics are evaluated:
$F_{errlive}$: Percentage of misclassified live fingerprints
$F_{errfake}$: Percentage of misclassified fake fingerprints
The Average Classification Error (ACE) is given as:

$$ACE = \frac{F_{errlive} + F_{errfake}}{2} \tag{5}$$

4.3. Datasets

The proposed approach is trained and tested on four datasets provided by the Liveness Detection Competition (LivDet) in the years of 2011 [40], 2013 [12] 2015 [23] and 2017 [24]

LivDet 2011 comprises 16,000 images obtained using four different sensors: Biometrika FX2000, Digital 4000B, Italdata ET10, and Sagem MSO300, each having 2000 images of fake and real fingerprints.

LivDet 2013 comprises 16,000 images acquired from four different sensors: Biometrika FX2000, Crossmatch L SCAN GUARDIAN, Italdata ET10, and Swipe, each having approximately 2,000 images of fake and real fingerprints.

LivDet 2015 comprises of 19,431 fingerprint images acquired from four different sensors: CrossMatch L SCAN GUARDIAN, Digital Persona U are U 5160, HiScan-PRO, GreenBit DactyScan26 each having approximately 1000 fingerprints for training and 1000 for testing. 8983 fingerprints both Live and Fake are used for training and 10,448 are used for testing purposes.

LivDet 2017 comprises of 17,820 fingerprint images acquired from three different sensors:Digital Persona U are U 5160, Orcathus Certis2 Image, GreenBit DactyScan2, each having approximately 2200 fingerprints for training and 3740 for testing. For each of the above mentioned fingerprint dataset, we evaluate the same-material ACE, cross-material ACE, cross-sensor ACE and cross-dataset ACE. The complete evaluation is summarized in Table 2.

4.4. Analysis

An exhaustive experimental analysis is provided for the four datasets LivDet 2011, 2013, 2015 and 2017 which is summarized in Table 2. For each sensor in the dataset, same-material, cross-material, cross-sensor and cross-dataset error metrics are obtained.

Intra-Sensor error metric is obtained when the network is trained and tested on the same dataset with same spoof materials(same-material) or different spoof materials(cross-material). We can see from Table 2. that for almost all datasets there is a considerable reduction in error when tested on both same material and different material. Our proposed method obtained better intra-sensor accuracies for almost all datasets. There is a striking **85%** decrease in the average classification error when trained and tested on same spoof materials and also different spoof materials for CrossMatch 2015 dataset when compared with the previous state-of-the-art approaches. This can be attributed to the presence of both global features and local features learned by our network. If we compare the same material average classification error with the results given by Network-1 and Network-2 in Table 1 we can see that with only preprocessing and DenseNet, Network-1 outperforms the state-of-the-art. Cross Material error is considerably reduced due to the addition of Network-2. The 10-layer DenseNet in Network-2 which is trained on patches, learns the local features which along with attention network improves the cross-material accuracy. The Digital Persona fingerprints are of smaller resolution as compared the other sensors but, this did not pose a problem with our method as we include both local patch features and the global features. The average classification error is significantly lower with our method as compared to the previous state-of-the-art methods. The plots for intra-sensor ACE is shown in Figure 5.

Cross-Sensor error metric is obtained when the training images belong to one sensor and the testing images belong to a different sensor of the same dataset. We can infer from the summary table that our approach provides exemplary results in case of cross-sensor metric also. If we compare the cross-sensor results of LivDet 2015 dataset. The cross-sensor ACE for CrossMatch sensor when tested with GreenBit, Digital Persona and Hi Scan data is 1.97, 1.12, 1.78 respectively which is relatively lower compared to that of previous state-of-the-art approaches. This is due the innate property of our network to learn common characteristics among the datasets which help in the subsequent clas-

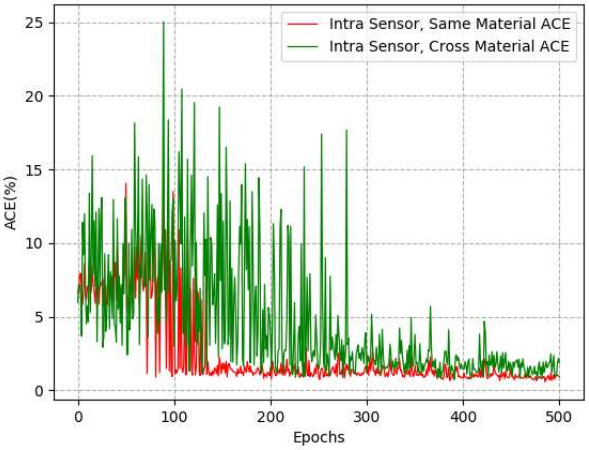


Figure 5: Intra-Sensor average classification error for network trained on CrossMatch L SCAN GUARDIAN 2015 sensor and tested on data having same testing spoof materials and different testing spoof materials.

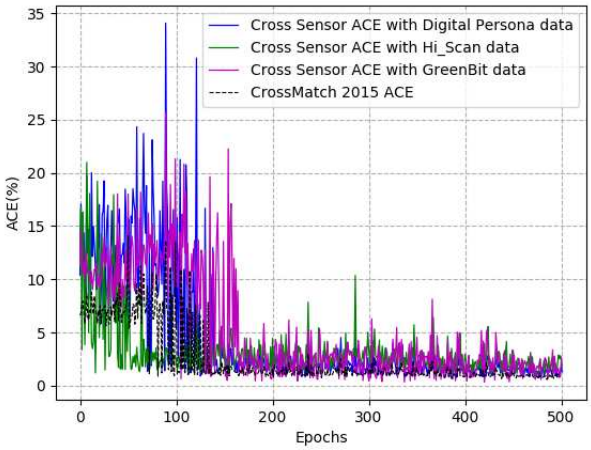


Figure 6: Cross-Sensor average classification error for network trained on CrossMatch L SCAN GUARDIAN 2015 sensor and tested with the data of Hi Scan, GreenBit and Digital Persona sensors.

sification. The cross-sensor metrics outperforms the state of the art for all datasets except for LivDet 2013. Figure 6. shows the average classification error plot for our network trained on CrossMatch sensor and tested on GreenBit, Hi Scan and Digital Persona.

Cross-Dataset error metric is obtained when the training images belong to the same sensor of one dataset and testing images belong to same sensor of different datasets. For example, network is trained on Biometrika sensor from LivDet 2011 and tested on Biometrika sensor data from LivDet 2013. As our network adeptly learns the common characteristics to classify a fingerprint as live or fake, we obtain considerably good cross dataset results. If we compare the cross dataset result of Italdata 2011 when tested with Italdata 2013, we can see that there is a **60%** reduction in the average classification error. Even though the cross sensor and cross dataset classification errors are significantly lower with our approach, there is still a considerable amount of scope of improvement in this area. Figure 7. shows the cross dataset ACE for CrossMatch 2015.

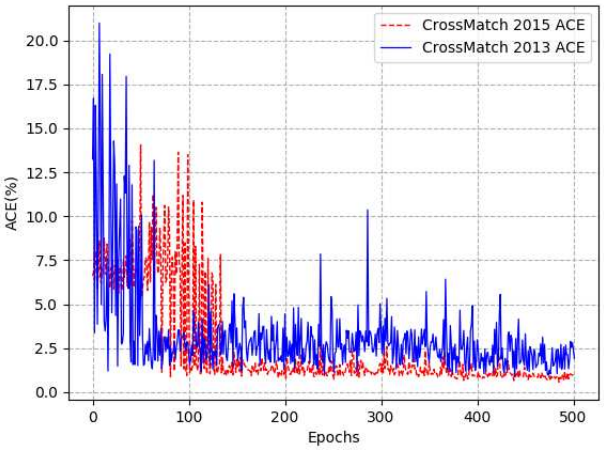


Figure 7: Cross-Dataset average classification error for network trained on CrossMatch 2015 dataset and tested on CrossMatch 2013 dataset.

5. Conclusion

In this paper, we propose a novel end-to-end fingerprint spoof detection network. The proposed network automatically extracts informative patches using a novel patch based attention mechanism. Use of DenseNet as the base network optimizes memory requirement. Furthermore, use of global fingerprint image along with the fingerprint patches helps in improving network robustness and generalizes it's performance across cross sensor, cross material and cross dataset. The effectiveness of the proposed network is validated through extensive set of experiments carried over various datasets, where a significant improvement from the current state-of-the-art is achieved. In future, we plan to integrate this proposed network with fingerprint authentication and recognition networks to obtain a complete fingerprint recognition system with inbuilt spoof detector.

6. Acknowledgement

This research work is supported and funded by the Department of Science and Technology(DST) under Indo-Korean bilateral scheme.

References

[1] Review of the fingerprint liveness detection (livdet) competition series. *Image Vision Comput.*, 58(C):110–128, Feb. 2017.

[2] A. Abhyankar and S. Schuckers. Fingerprint liveness detection using local ridge frequencies and multiresolution texture analysis techniques. In *2006 international conference on image processing*, pages 321–324. IEEE, 2006.

[3] Z. Akhtar, C. Micheloni, and G. L. Foresti. Correlation based fingerprint liveness detection. In *2015 International Conference on Biometrics (ICB)*, pages 305–310. IEEE, 2015.

[4] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson, and S. Z. Li. Attention-based two-stream convolutional networks for face spoofing detection. *IEEE Transactions on Information Forensics and Security*, 15:578–593, 2019.

[5] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5659–5667, 2017.

[6] H. Choi, R. Kang, K. Choi, and J. Kim. Aliveness detection of fingerprints using multiple static features. In *Proc. of World Academy of Science, Engineering and Technology*, volume 22, 2007.

[7] F. Chollet. Xception: Deep learning with depthwise separable convolutions, 2016. cite arxiv:1610.02357.

[8] T. Chugh, K. Cao, and A. K. Jain. Fingerprint spoof buster. *ArXiv*, abs/1712.04489, 2017.

[9] A. George and S. Marcel. Deep pixel-wise binary supervision for face presentation attack detection. In *International Conference on Biometrics*, number CONF, 2019.

[10] L. Ghiani, A. Hadid, G. Marcialis, and F. Roli. Fingerprint liveness detection using binarized statistical image features. pages 1–6, 09 2013.

[11] L. Ghiani, G. L. Marcialis, and F. Roli. Fingerprint liveness detection by local phase quantization. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 537–540, Nov 2012.

[12] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli, and S. Schuckcrs. Livdet 2013 fingerprint liveness detection competition 2013. In *2013 International Conference on Biometrics (ICB)*, pages 1–6, June 2013.

[13] C. Gottschlich, E. Marasco, A. Y. Yang, and B. Cukic. Fingerprint liveness detection based on histograms of invariant gradients. In *IEEE international joint conference on biometrics*, pages 1–7. IEEE, 2014.

[14] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. Fingerprint liveness detection based on weber local image descriptor. 09 2013.

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[16] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2424–2433, June 2016.

[17] 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*, abs/1704.04861, 2017.

[18] G. Huang, Z. Liu, and K. Q. Weinberger. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.

[19] L. Huang and C.-M. Pun. Audio replay spoof attack detection using segment-based hybrid feature and densenet-lstm network. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2567–2571. IEEE, 2019.

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

[21] G. L. Marcialis, F. Roli, and A. Tidu. Analysis of fingerprint pores for vitality detection. In *2010 20th International Conference on Pattern Recognition*, pages 1289–1292, Aug 2010.

[22] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):864–879, 2015.

[23] V. Mura, L. Ghiani, G. L. Marcialis, F. Roli, D. A. Yambay, and S. A. Schuckers. Livdet 2015 fingerprint liveness detection competition 2015. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sep. 2015.

[24] 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. *CoRR*, abs/1803.05210, 2018.

[25] S. B. Nikam and S. Agarwal. Fingerprint anti-spoofing using ridgelet transform. In *2008 IEEE second international conference on biometrics: theory, applications and systems*, pages 1–6. IEEE, 2008.

[26] R. Nogueira, R. Lotufo, and R. Machado. Fingerprint liveness detection using convolutional neural networks. *IEEE Transactions on Information Forensics and Security*, 11:1–1, 06 2016.

[27] T. Ojala, M. Pietikäinen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. *Proceedings of 12th International Conference on Pattern Recognition*, 1:582–585 vol.1, 1994.

[28] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, July 2002.

[29] F. Pala and B. Bhanu. *Deep Triplet Embedding Representations for Liveness Detection*, pages 287–307. Springer International Publishing, Cham, 2017.

[30] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Netw.*, 12(1):145–151, Jan. 1999.

[31] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

[33] B. Tan and S. C. Schuckers. New approach for liveness detection in fingerprint scanners based on valley noise analysis. *Journal of Electronic Imaging*, 17(1):011009, 2008.

[34] D. Thakkar. Top five biometrics:face, fingerprint, iris, palm and voice. `https://www.bayometric.com/biometrics-face-finger-iris-palm-voice/`.

[35] D. Thakkar. What makes fingerprint the most popular biometric modality. `https://www.bayometric.com/fingerprint-popular-biometric-modality/`.

[36] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017.

[37] L. Wang, L. Qiu, W. Sui, and C. Pan. Reconstructed densenets for image super-resolution. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3558–3562. IEEE, 2018.

[38] W. Wang, S. Zhao, J. Shen, S. C. Hoi, and A. Borji. Salient object detection with pyramid attention and salient edges. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1448–1457, 2019.

[39] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon. Cbam: Convolutional block attention module. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[40] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. Livdet 2011 fingerprint liveness detection competition 2011. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 208–215, March 2012.

[41] Y. Yuan, W. Qin, X. Guo, M. Buyyounouski, S. Hancock, B. Han, and L. Xing. Prostate segmentation with encoder-decoder densely connected convolutional network (ed-densenet). In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 434–437. IEEE, 2019.

[42] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding. Multiple feature reweight densenet for image classification. *IEEE Access*, 7:9872–9880, 2019.

[43] K. Zhang, Y. Guo, X. Wang, J. Yuan, Z. Ma, and Z. Zhao. Channel-wise and feature-points reweights densenet for image classification. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 410–414. IEEE, 2019.