Understanding deep convolutional neural networks: A biometric based case study[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], Global Customer Service[b,]*

[a] *1600 John F Kennedy Boulevard, Philadelphia*
[b] *360 Park Avenue South, New York*

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

Keywords: `elsarticle.cls`, LaTeX, Elsevier, template

2010 MSC: 00-01, 99-00

Contents

1 Introduction **2**

2 Multi class oculus classification **4**

2.1 Related Work: 5

2.2 Contribution: 6

2.3 Model Selection and Justification 7

2.3.1 Proposed Network Design(CLResNet) : 8

2.3.2 Proposed Network Design(GHCLNet): 9

2.3.3 Systematized ResNet Pruning based on Performance . . . 12

2.4 Database and Testing Protocol 14

2.4.1 IIT-K 14

[☆] Fully documented templates are available in the elsarticle package on CTAN.

[*] Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)

[1] Since 1880.

2.4.2 ND 14
2.4.3 IIIT-D 14
2.5 Experimental Results and Discussion 15
2.5.1 Intra sensor validation 15
2.5.2 Inter sensor validation 18
2.5.3 Multi sensor validation 20
2.6 Combined-sensor validation testing generalization ability 20
2.7 Comparative Analysis 21
2.8 Layer Specific Feature Analysis of GHCLNet 24
2.9 Conclusion 25

3 Fingerprint Sensor Classification 25
3.1 Related Work 26
3.2 Proposed Architecture 28
3.2.1 Shallow Network(VGG-19 variant) 28
3.2.2 Deep Network(ResNet50 variant) 29
3.3 Experimental Analysis 32
3.4 Single Slap Fingerprint Experimentation: 33
3.4.1 Intra-sensor Classification 33
3.4.2 Multi-sensor Classification 34
3.4.3 Comparative Analysis with the state of art techniques : . 34
3.5 Four Slap Fingerprint Experimentation : 35
3.6 Robustness or Generalization Analysis 37
3.7 Layer Specific Feature Analysis 38
3.8 Conclusion 39

1. Introduction

In today's world no field in image processing and computer vision is untouched without deep-learning. Indeed it is right worthy to declare the last five years as the golden era in the history of deep-learning. The main reason behind its success is it's impressive emprically exceptional performance, particularly

in the field of supervised learning. It allows data-processing models that are composed of heterogeneous cascade units of non-linear transformations to learn discriminative feature representation of data belonging to different classes with several levels of abstraction. This is possible today only due to the presence of large gigantic datasets and massive computational power at our disposal. Many researchers have contributed to enrich this field with novel ideas and concepts like Deep-Convolutional Neural Networks(DCNN), Generative Adversial networks(GAN), Recurrent Neural Networks(RCNN) and many more. Particularly the work done by Hiltons in backpropagation and Lecuns in convolutional neural network is praise worthy. In spite of its spectacular success almost in every field we are lagging far behind in realizing the scientific theory underlying it. Mathematical postulate that will summarize the peculiar properties of deep neural networks is still unaccounted. Our understanding of DCNN's is still far from complete.

Universal approximation theorem posulates that a wide variety of computational functions can be represented quite well by a simple neural network given a right melange of hyper-parameters[1]. But some of the questions related to deep neural networks are still unanswered like (i) Is it really true that depth in conjunction with pre-training provides a good prior for model weights and thus helps in generalization? (ii)Why do the deep-networks despite being trained on massive amount of data and highly over-parameterized, still predicts remarkablly[2]? Some of the researchers in the past tried to answer this question by relating the complex architecture of visual cortex in human brain with the massive layers of deep-neural networks. They suggested that like in case of human-brain thousands of neurons are dying and many of neurons are born each day, but still we have an outstanding recognition capability. In a similar manner deep architectures exploit the appropriate structure of compositional functions, in fact they perform a combinatorial swindle by replacing the exponentiation operation with multiplication operation to achieve outrageous results. But still there is a enormous difference in thinking ability of human brain and the processing ability of machines this is mainly because of the different ways

of encapsulating the patterns, Fig.1 captures this dichotomy between humans and computers.

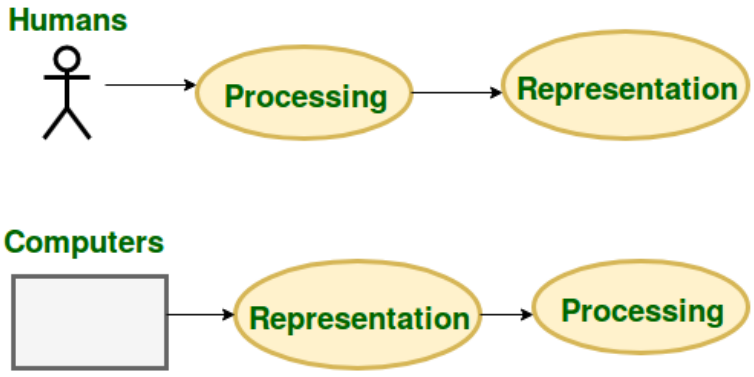

Figure 1: Dicotomy between human thinking and computer processing ability

Problem Statement In our work we are trying to find answers to some of the above mentioned questions by investigating the dichotomy between the deep-neural networks and the shallow neural networks under various scenarios for this we have considered following biometric based case studies:

- Multi class oculus classification into no-lens, soft-lens and cosmetic-lens
- Fingerprint sensor classification
- Fingerprint spoof detection

2. Multi class oculus classification

Iris serves as one of the best biometric modality owing to its remorselessness to circumvent. But still it can be easily dodged using contact lenses and spurious eyeballs. Complex textual iris patterns are the decisive features that help in distinguishing different persons. It has been proved in the literature that even identical twins have different iris patterns. It is clearly evident by the work done by [3] that the performance of any iris based recognition system debase

considerably in the presence of contact lenses particularly the cosmetic contact lenses. Hence, it is crucial to detect the ubiety of contact-lenses before going for true iris recognition. Cosmetic lens changes the highly unique textual patterns of iris and thus can be easily detected as clearly evident from Fig.2 , however differentiating between soft lens and no lens is still a challenging task.

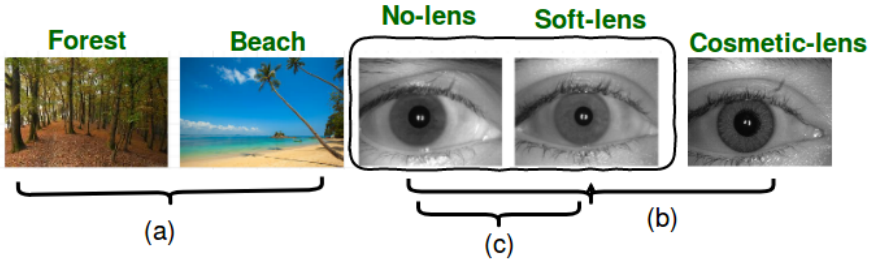

Figure 2: Level of Classification (a) Coarse level (b) Fine level (c) Finer-level (Soft Vs No contact lens)

2.1. Related Work:

In the past several work has been done in the field of iris based biometric authentication, but still the work done by Daugman[4] to distinguish between genuine iris and fabricated iris on the basis of frequency spectrum analysis is considered as one of the best contribution in the history of iris based biometric systems. Another notable work has been done by Lee et al to distinguish between a real and fabricated iris images on the basis of the purkinje image formation. Ring et al analysed the iris bitcodes to detect the regions of local distortion within the iris mainly due to the presence of contact lenses. Doyel et al ensembled 14 different classifiers jointly for three class oculus detection into no-lens, cosmetic-lens and soft-lens. In another interesting work Daksha et al proposed a variant of Local Binary Patterns for studying the effects of cosmetic lenses on the iris recognition. Lovish et al. proposed a method based on Local Phase Quantization(LPQ) and Binary Gabor Patterns(BGP) for investigating the effects of textured patterns in case of cosmetic lenses. Recently, Ragvendra

et al. proposed ContlensNet which is an architecture based on Deep- Convolutional Neural Network for lens detection. Table 1 summarizes the related work done uptill now in the field of lens detection.

Author	Proposed Approach
Daugman [4]	Frequency Spectrum Analysis
Lee [5]	Purkinje Image Formation
Zang [6]	Weighted Local Binary Patterns
Daksha [3]	Variants of Local Binary Patterns
Lovish [5]	Local phase quantization for detecting cosmetic lens
Ragvendra[7]	Deep Neural Network based architecture (ContlensNet)

Table 1: Summarizing the work done in the field of lens detection

2.2. Contribution:

In this paper, we have proposed two network architectures (i) CLResNet(Contact-lens detection network based on ResNet) (ii) GHCLNet(Generalized Hierarchically tuned Contact Lens detection Network); for three class ocular classification namely no-lens, soft lens and cosmetic lens. The main contribution of this paper is four fold, that is summarized as follows :

- A novel architecture based on Deep Convolutional Neural Network (CLResNet) is proposed and tested over two set of experimentation. Firstly on raw iris images and secondly on binarized iris images.
- Hierarchical Deep Convolutional Network (GHCLNet) for three class ocular classification namely no-lens, soft lens and cosmetic lens has been proposed. The prodigious strength of this network lies in the fact that it works on full holistic contact lens features without any pre-processing and segmentation prerequisite.
- Generalized deep convolutional neural network based architecture has been proposed.

- To ensure the generalization ability of the proposed network, multi-sensor and combined sensor validation has been performed over benchmark datasets and compared with the state-of-the-art methods.

2.3. Model Selection and Justification

It is clearly evident from Fig.2 that classifying a forest image from a beach image is a coarse-level classification but classifying a no-lens from soft-lens is a finer level classification as it requires finest level of granular feature extraction. As the complexity of the classification problem increases one requires much more complex functions to be learned at finest granularity level appropriately. We have done extensive experimentation in order to decide the suitable network for our oculus three class classification problem. Mainly we have considered two sets of experimentation. In the first set of experimentation we fed the binarized iris images as input to one of our convolutional neural network architecture. Multi-Scale Line Tracking(MSLT) algorithm have been used for converting raw iris images into binarized images. Fig.3 shows the output of MSLT algorithm.

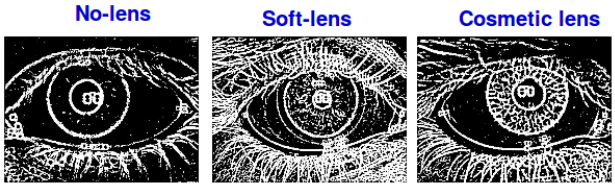

Figure 3: Example of binarized iris images with no lens, soft lens and cosmetic lens

In the second set of experimentation we fed the raw iris images without any pre-processing to the convolutional neural network architecture. The main reason for doing these two variants of experimentation was to ensure the high feature extraction ability of convolutional neural networks without any need of pre-processing which is one of the main prodigious strength of our proposed architecture. Initially we start with a simple 5 Conv layer VGG network architecture. Here, VGG network with 5 Conv layers can be thought of as a shallow

network. After conducting rigorous experimentation on our test data, we conclude that for better feature learning, and particularly to distinguish between no-lens and soft-lens a much deeper architecture is required as it is a very fine grain classification.

During model designing phase we have considered two options (i) either to use an augmented version of a VGG network (ii) or to use a deep model variant like ResNet. For our problem we have chosen network architecture based on ResNet50 model as it is a residual network model uses identity connections between the layers so as to address vanishing gradient problem as well as uniform deep network training. We started with a single network model named as CLResNet based on ResNet50 and trained that network over our dataset and have conducted several experimentation. Our proposed CLResNet architecture has been showed in Fig.4. We have observed that due to very fine grain classification problem and highly similar features between no-lens and soft-lens network performance was poor. This inspired us to propose a hierarchical model that should learn highly generalizable features and uses the full holistic contact lens image. Our proposed hierarchical network is named as Generalized Hierarchically tuned Contact Lens detection Network(GHCLNet)and its shown in Fig.5

2.3.1. Proposed Network Design(CLResNet) :

The first layer of CLResNet is an input layer which takes an input image of size $224 * 224$. Given the iris image and its corresponding label, the first convolutional layer $CONV1$, filters the input image of size $224 * 224$ using 64 kernels and pools it to an output of size $112 * 112$. The filters of $CONV1$ layer basically detect the low level features like edges and corners. The output of the $CONV1$ is connected to max-pooling layer. After that $Branch-2$, $Branch-3$ and $Branch-4$ blocks of ResNet50 architecture has been used in our proposed architecture. At the end a fully connected layer with 2048 neuron has been added. Finally $softmax$ function has been used to generate the probability distribution by minimizing the categorical cross-entropy loss-function. All the

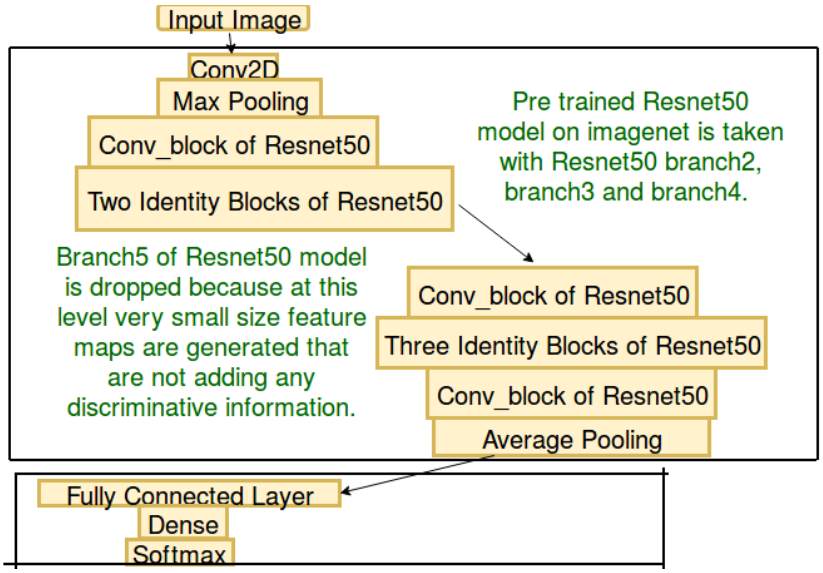

Figure 4: Proposed Diagram of CLResNet architecture based on Deep- Convolutional Neural Network

parameters that are used in this work are calculated empirically over a small validation set and are shown in table 2.

2.3.2. Proposed Network Design(GHCLNet):

The proposed hierarchical GHCLNet is composed of two parts or models. The first part of the hierarchical network has been exclusively trained for classifying iris images into "textured" and "non-textured" and is named as TexNet. Second part has been exclusively trained for classifying iris images into "lens" and "no-lens" and is named as LensNet. Both the parts has been using the pre-trained ResNet-50 models on ImageNet images with the first part re-trained on textured-lens images (comprises of cosmetic-lens) and non-textured-lens images (comprises of no-lens and soft-lens). The second part has been re-trained only on soft-lens images and no-lens images. The internal architecture of both TexNet and LensNet is shown in Fig.6.

Parameter	Value
Optimizer	Adam optimizer
Learning Rate	0.0001
$beta_1$	0.8
$beta_2$	0.888
Mini batch size	64
Epoches	50

Table 2: Summarizing the CLResNet & GHCLNet Architecture Parameters

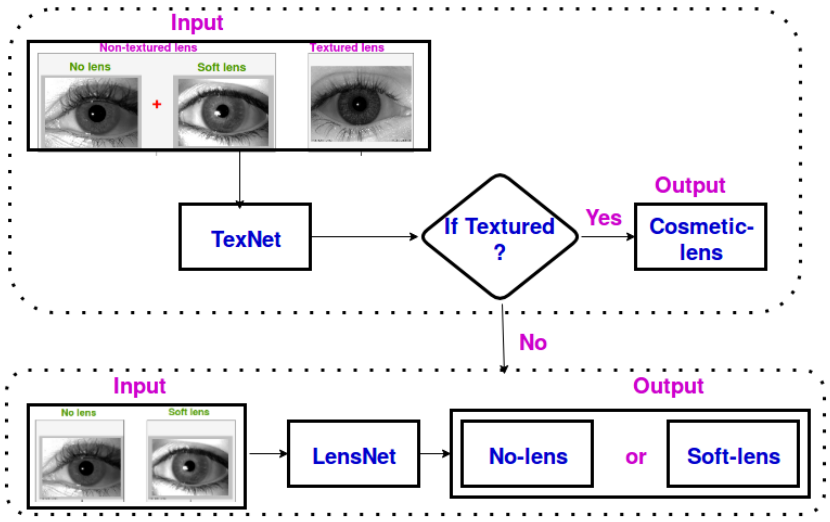

Figure 5: Generalized Hierarchically Tuned Contact Lens Detection Network (GHCLNet) architecture

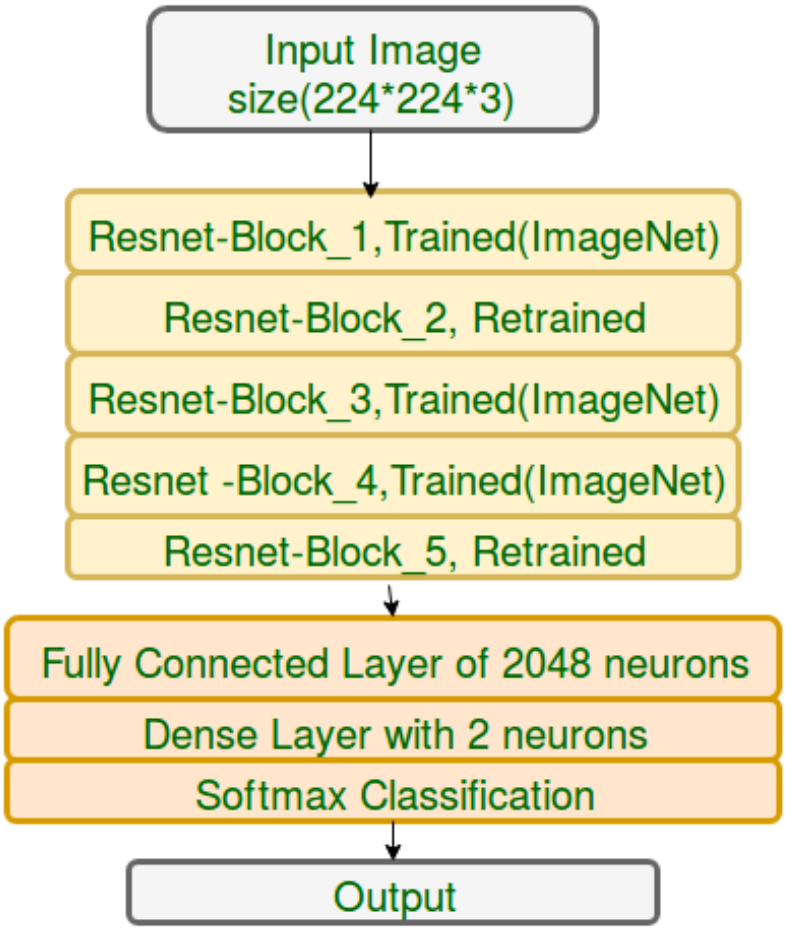

Figure 6: Internal Architecture of TexNet and LensNet

2.3.3. Systematized ResNet Pruning based on Performance

The ResNet-50 model is a recently proposed deep convolutional neural network model consisting of five blocks as explained below:

- Block-1 is the initial branch which gets the input RGB image of size $224 * 224$. The input image is convolved with 64 kernels to give a feature map of $112 * 112$, which is then passed to the max-pooling layer to reduce its size to $55 * 55$.

- Block-2 comprises of three sub-blocks : $block-2a$, $block-2b$ and $block-3c$. The output feature map of block-2 is of size $55 * 55$.

- Block-3 comprises of four sub-blocks : $block - 3a$, $block - 3b$, $block - 3c$ and $block - 3d$. The final output of block-3 is a $28 * 28$ feature map.

- Block-4 consists of six sub-blocks namely : $block-4a$, $block-4b$, $block-4c$, $block - 4d$, $block - 4e$ and $block - 4f$. In the end the output feature map is of the size $14 * 14$.

- Block-5 is the last block which consists of three sub-blocks : $block - 5a$, $block - 5b$ and $block - 5c$. The output of block-5 is a feature map of size $7 * 7$.

ResNet-50 is a very deep network with almost 50 weight layers pre-trained on the ImageNet database that contains 1000 classes. ResNet is quiet famous for its notorious training. In our oculus three class classification problem we have only three classes in hand but very limited number of training images per class. We have to be very careful while showing images as it can easily be over-trained. Hence we decided to systematically prune our model such a deep model in order to adjust the network capacity so that it can learn well the complicated features.

[a] Redundant block removal : During experimentation over our NIR contact lens iris images, we have found that similar performance has been

achieved while re-training the $3rd$ and $5th$ blocks of ResNet-50 instead of retraining all the blocks. Re-training of any of the $1st$, $2nd$ and $4th$ blocks resulted in a drastic drop in performance for our specific contact-lens data. This is mainly due to the fact that ResNet is a very deep network and require huge amount of training data for training it from scratch. Due to the limited amount of training data at our disposal we are incapable of re-training it from scratch. Hence we have inferred that $3rd$ and $5th$ blocks are redundant for our set of images and one can drop any of them. Also $1st$, $2nd$ and $4th$ blocks have learned very useful and generalized features that can be reused even for our contact lens classification. Later we have visualized the features discussed in Section 2.8, and shown in Fig. 7, so as to reinforce our inferences.

After extensive experimentation it was found that for the four databases (discussed in Section 2.4) viz. IITK, Cogent, Vista and ND2, maximum performance along with minimal training time has been achieved when only the $3rd$ and $5th$ blocks of the ResNet-50 models have been trained. Since the database of ND1 is very small in number as compared to other databases, the training was only restricted to the sub-block $5(c)$ in order to avoid over-fitting. We have explicitly chosen a small mini-batch size in order to avoid complexities involved in parallelizing the task for large mini-batch size. Also it has helped *Adam* optimizer to carefully avoid local minima.

While testing the output of the TexNet(first part) of the hierarchical network is first checked. If it classifies the image as textured, the image is assigned the label of cosmetic-lens. However, if it classifies the image as non-textured, the output of the LensNet(second part) of the network is considered. If it classifies it as lens, the image is put in the soft-lens category and if it classifies it as no-lens, the image is put in the no-lens category.

[b] Network Implementation Details : The proposed generalized hierarchically tuned contact-lens detection network has been implemented using python and keras[8] library using tensorflow[9] as its backend. All the implementation has been done on Intel(R) Xenon(R) CPU E5-2630 V4 at 2.20 GHz, with 32 GB RAM and NVIDIA 1080 Ti gpu with 8 GB RAM.

2.4. Database and Testing Protocol

In this section, we present the details about the database and the testing protocols used. We have tested our proposed network on IIT-Kanpur contact lens iris database (IIT-K) and on two publicly available iris databases namely, Notre Dame cosmetic contact lens 2013 database (ND) and IIIT-Delhi contact lens iris database (IIIT-D). The detail description about these databases are presented in the following sub section and in Table 3.

2.4.1. IIT-K

This database consists of a total of $12,828$ iris images corresponding to 50 subjects captured using Vista Imaging FA2 sensor. Since iris images are available for both the left and right eye there are 100 unique instances of iris present in this database. All the soft lenses used in this database are manufactured by Johnson & Johnson [10] and Bausch & Lomb [11] and all the cosmetic lenses are manufactured by CIBA Vision [12], Flamymboyout, Oxycolor and FreshLook. This database has been provided with an evaluation protocol that comprises of 80 subjects for training and remaining 20 subjects for testing.

2.4.2. ND

This database is divided into two databases namely : ND-I and ND-II. The ND-I dataset comprises of 600 training set images and 300 testing set images, all captured using IrisGuard AD-100 sensor [13]. The ND-II dataset comprises of $3,000$ training set images and $1,200$ testing set images all captured using LG-4000 [14] sensor. CIBA Vision [12], Johnson & Johnson [10] and Cooper Vision [15] are the three main manufacturers of cosmetic contact lenses in this database. In this work for evaluating the proposed framework we have follow the evaluation protocol as recommended in [16].

2.4.3. IIIT-D

This database consists of a total of $6,570$ iris images corresponding to 101 subjects captured using two sensors namely : Cogent dual iris sensor (CIS 202)

and Vista FA2E single iris sensor. All the soft lenses used in this database are manufactured either by CIBA Vision [12] or Bausch & Lomb [11]. This database is provided with an evaluation protocol that comprises of 50 subjects for training and remaining 51 subjects for testing as suggested in [16],[7].

Database	Sensor Model	Total	Train	Test
IITK [5]	Vista Imaging FA2	12,823	10,258	2,565
IIITD [3]	Cogent Dual Vista FA2E	6,570	3,285	3,285
ND-I [17]	IrisGuard-AD100 [13]	900	600	300
ND-II [17]	LG4000 [14]	4,200	3,000	1,200

Table 3: Databases used in the proposed GHCLNet architecture

2.5. Experimental Results and Discussion

We have trained and tested our both proposed networks CLResNet and GHCLNet on various experiments. For CLResNet architecture we have performed Intra sensor , Inter sensor and Multi sensor validation on both raw iris images and on binarized iris images. For GHCLNet architecture we have performed four different experiments namely (i) Intra sensor validation (ii)Inter sensor validation (iii) Multi sensor validation and (iv) Combined sensor validation. The experimental results are evaluated using the correct classification rate(CCR%), its higher value indicates better performance.

2.5.1. Intra sensor validation

In this type of validation taining and testing has been done on data seized from single sensor. The proposed network results has been showed in Table4 and compared with other state-of-the art techniques like Statistically Independent Filters, Deep Image Representation and ContlensNet. The following key observations were being made after the analysis.

- Cosmetic lenses are highly textual in nature and thus easily detected by all three experimentation in consideration in comparision to no-lens and soft-lens as clearly indicated by the Table4 with high CCR(%) in case of cosmetic-cosmetic lenses.

- The results obtained on IIITD cogent database with the proposed GHCLNet network shows an hike of around 6% in CCR% with a total CCR% of 93.71% as compared to the current state-of-the-art algorithm of ContlensNet [7].

- The results on IIITD vista database shows remarkable performance on GHCLNet network with a total CCR% of 95.49% . The results obtained show a total hike of 8% in CCR% as compared with second best ContlensNet model [7].

- The results obtained on ND-I database are slightly less than the available state-of the art results. The main reason behind this can be attributed to the lesser number of training images available in this database. As our network is quite deep it requires large number of training images. Over such small dataset more training epochs may lead to better results but we have avoided to over train the network.

- The results obtained on ND-II database are comparable to the available state of the art technique.

- The results obtained on IITK databases are exceptional with a total CCR% of 99.67% . The accuracy of all the classes have been found to be above 99% in CCR% . This can be attributed to the fact that large number of training samples made the network learn better discriminative intrinsic information corresponding to each class.

- It is clearly evident from Table 4 that performance of GHCLNet outperforms the performance of the CLResNet on both binarized as well as on raw iris images.

Database	Type	[18]	[19]	[7]	CLResNet(Binarized)	CLResNet(Raw)	GHCLNet
IIITD-Cogent	N-N	64.16	35.50	68.68	71.30	78	89.86
	S-S	66.45	98.21	93.62	67.90	81.20	91.26
	C-C	100	73.00	100	90.11	97.70	100
	Avg.	76.87	69.05	86.73	76.43	86.63	93.71
IIITD-Vista	N-N	68.89	60.80	74.50	70.39	88.28	94.6
	S-S	75.63	98.30	87.50	53.66	83.40	91.88
	C-C	100	55.88	100	99.27	99.28	100
	Avg.	81.50	72.08	87.33	74.44	90.32	95.49
ND-I	N-N	76.50	84.50	93.25	69.07	88	91.67
	S-S	84.50	73.75	97.50	92.52	39	87.50
	C-C	100	99.75	100	100	97	100
	Avg.	87.00	86.00	96.91	87.46	74.67	93.05
ND-II	N-N	79.50	73.00	88.00	78.7	81.19	95.24
	S-S	62.00	65.00	97.00	82.68	66.84	89.74
	C-C	100	97.00	100	97.5	99.5	99.75
	Avg.	80.50	78.33	95.00	86.61	82.75	94.91
IITK	N-N	-	-	-	95.49	82.02	99.78
	S-S	-	-	-	86.34	92.06	99.24
	C-C	-	-	-	99.13	98.23	100
	Avg.	-	-	-	93.65	90.77	99.67

Table 4: Intra-Sensor comparative performance analysis in terms of CCR(%) on CLResNet,GHCLNet architecture(where N-N is No lens-No lens,S-S is Soft lens-Soft lens, C-C is Cosmetic lens-Cosmetic lens, SIF is Statistically Independent Features[18] and DIR is Deep Image Representation[19]), Green colour represents significant rise in CCR(%), Red colour indicates no rise in CCR(%)

2.5.2. Inter sensor validation

In this case of validation pair-wise comparision of sensors belonging to the same database is done. Here the network is trained on one sensor and testing is done on the another sensor of the same database. Table 5 shows the qualitative performance of the proposed network on the four resultant cases.

- When training data has been drawn from IIITD Vista sensor and tested over IIITD Cogent sensor, an accuracy of 82.61% in CCR% is obtained on the GHCLNet network. One can see that such inter training results in lesser performance but that is obvious due to variation in training and testing data.

- When training of the proposed network has been done on data from IIITD Cogent sensor and testing is done on data from IIITD Vista sensor an accuracy of 92.01% in CCR % has been noted.

- In both of the above cases our performance drops as compared to [7], but for the remaining cases GHCLNet network performed better than it.

- Case when the training data has been chosen from ND-II sensor and testing data is chosen from ND-I sensor an accuracy of 91.51% in CCR% has been obtained. The results show an increase in accuracy of 3.5% in CCR % as compared to the previous state-of-the-art algorithm of ContlensNet [7].

- Finally, when the training data has been chosen from ND-I sensor and testing data from ND-II sensor, the best performance of 90.58% in CCR% is obtained from the proposed network. The result shows a very nominal hike in accuracy of 0.13 % in CCR % as compared to ContlensNet [7].

Comparative Generalization Ability Analysis and Justifications : The remaining two strategies namely multi-sensor and the combined dataset are the ones that really can demonstrate the generalizable performance of any system. We have used a deep hierarchical network just to get better and much more generalizable network that can be trained once and used overall the databases.

Comment	Type	[18]	[19]	[7]	CLResNet(Binarized)	CLResNet(Raw)	GHCLNet
Train on VISTA Test on COGENT	N-N	57.67	48.67	87.75	57.68	72.12	96.74
	S-S	66.06	42.25	87.75	26.31	27.56	65.73
	C-C	100	38.15	78.91	71.87	88.27	85.36
	Avg.	74.57	43.08	84.80	52.72	63.4	82.61
Train on COGENT Test on VISTA	N-N	66.91	06.00	96.19	64.08	60.11	93.40
	S-S	56.96	45.47	88.23	41.23	57.12	83.37
	C-C	97.09	89.61	100	74.09	90.42	99.25
	Avg.	73.65	45.51	94.80	60.01	68.98	92.01
Train on ND-II Test on ND-I	N-N	72.66	75.00	68.50	76.63	87	81.25
	S-S	54.00	65.00	98.00	41.12	55	93.27
	C-C	100	94.00	97.50	92.13	98	100
	Avg.	75.33	78.00	88.00	68.65	80	91.51
Train on ND-I Test on ND-II	N-N	57.64	80.00	81.33	33.83	62.38	91.9
	S-S	73.64	49.00	90.03	80.31	55.79	81.84
	C-C	94.85	97.00	100	31.03	76	98.00
	Avg.	75.37	75.33	90.45	50.07	64.83	90.58

Table 5: Inter-Sensor comparative performance analysis in terms of CCR(%) on GHCLNet architecture(where N-N is No lens-No lens,S-S is Soft lens-Soft lens, C-C is Cosmetic lens-Cosmetic lens, SIF is Statistically Independent Features[18] and DIR is Deep Image Representation[19]), Green colour represents significant rise in CCR(%), Red colour indicates no rise in CCR(%)

We have observed that the proposed network performed way better than [7] hence exhibiting the better generalization ability.

2.5.3. Multi sensor validation

In this kind of validation testing data from same databases(captured through different sensors) are combined together resulting in two databases namely: IIITD-Combined & ND-Combined. To ensure modality training and testing data has been combined separately. Table6 indicates the CCR% of the proposed networks along with the available state of the art techniques.

- Training and testing of ND combined data(both ND-I and ND-II) on the proposed network show an excellent accuracy of 95.57% in CCR% . The results show a hike of 3% from the previous state-of-the-art network of ContlensNet [7].
- Training and testing of IIITD combined data(both Vista and Cogent) show an accuracy of 94.82% in CCR% obtained from the proposed network. The results obtained show an improvement of 0.17% in CCR% as compared to the previous state-of the art ContlensNet [7].

2.6. Combined-sensor validation testing generalization ability

In this testing strategy *(train only once)*, training data of all the databases are combined to form a single large train database while testing is performed on individual test databases. The combined database constitutes of: 1) IITK, 2) IIITD (Cogent and Vista) and 3) ND (ND-I and ND-II). The result obtained by experiments has ben recorded in Table 7. The key observations are as follows : testing the network on IITK , ND-I, ND-II, IIITD Vista and IIITD Cogent achieved very good accuracy of 99.14%, 92.87%, 94.93%, 95.69% and 95.43% respectively.

Here, we have not performed any comparative analysis because this kind of validation has not been done earlier. The main aim of performing combined sensor validation is to show that our GHCLNet is trained in such a manner that

Database	Type	[19]	[7]	CLResNet(Binarized)	CLResNet(Raw)	GHCLNet
ND-Combined	N-N	77.40	95.40	76.50	72.52	91.67
	S-S	71.40	82.40	68.77	83.62	95.04
	C-C	99.60	100	98.92	98.26	100
	Aggregate	82.80	81.17	85.35	92.60	95.57
IIITD- Combined	N-N	47.55	96.56	68.81	61.37	91.87
	S-S	97.99	88.90	59.25	85.62	92.85
	C-C	61.07	98.50	97.51	98.21	99.73
	Aggregate	69.28	94.65	75.59	81.92	94.82

Table 6: Multi-Sensor qualitative performance in CCR(%) on GHCLNet architecture(where N-N is No lens-No lens,S-S is Soft lens-Soft lens , C-C is Cosmetic lens-Cosmetic lens and DIR is Deep Image Representation), Green colour represents significant rise in CCR(%), Red colour indicates no rise in CCR(%)

it encapsulates different variations quite well and hence depicts high generalizability.

Database	Classes	IITK	ND-I	ND-II	VISTA	COGENT	Avg
Combined dataset from all databases	N-N	99.67	84.38	94.52	94.8	95.19	93.71
	S-S	97.86	94.23	90.26	92.28	91.43	93.21
	C-C	99.88	100	100	100	99.67	99.91
	Aggregate	99.14	92.87	94.93	95.69	95.43	95.61

Table 7: Combined-sensor qualitative performance in CCR(%) on GHCLNet architecture(where N-N is No lens-No lens,S-S is Soft lens-Soft lens , C-C is Cosmetic lens-Cosmetic lens)

2.7. Comparative Analysis

Accuracy Vs Generalization : There is always a trade-off between accuracy and generalization ability of any learning model. Here in our work we mainly focused on designing a network that is highly generalizable. It can be

inferred from Table 4, Table 5, Table 6, Table 7 that our proposed architecture(GHCLNet) is performing far better in terms of CCR% as compared to the algorithms of pre-deep learning era[18], [19]. To the best of our knowledge ContlensNet[7] a recent research paper, is the only architecture based on deep convolutional neural network for contact lens detection. In few experiments, we are marginally lagging behind ContlensNet[7] as shown in Table 4, Table 5. One can easily improve this by modifying the hyper parameters of our network and tune them perfectly so as to achieve state-of-the-performance, but this could lead to over-fitting, that is why we have avoided that. Through out this work, our major aim has been to achieve higher accuracy results for multi-sensor validation as well as combined-sensor validation ensuring better generalizable network that has been trained only once but performers substantially well over all the datasets. Any one later can use this network for their dataset after just fine tuning it over their data that do not require much significant time and hardware resources.

Without any Pre-processing : In the recent state of the art technique ContlensNet[7] OSIRIS V4.1, a publicly available segmentation tool is used for iris segmentation and normalization. Due to various enviornmental factors like occlusion, illumination and many more the performance of this tool is limited. The available state of the art technique [7] takes training input as a segmented and normalized patch of iris region of size $32 * 32$. The main advantage of taking this iris patch is that it is not affected by occlusion due to eyelashes.

Pros and Cons of the proposed GHCLNet : The main advantage of our proposed GHCLNet is that it is not using any kind of pre-processing and segmentation and still giving comparable results and in many cases even better. Our network is trained in such a way that it is able to handle illumination, occlusion and other external environmental factors in a quite remarkable manner. We are marginally lagging behind ContlensNet at few places as discussed previously. The main reason for this is the poor quality of raw iris images as it can be seen in fig.8. It is clearly visible from fig.8, that some of the iris images are illuminated to a large extent which distorts their textual patterns and some

of the images are highly occluded. It is very difficult even for a human being to distinguish between no-lens, soft-lens and cosmetic-lens in such kind of images. In our proposed architecture GHCLNet as we are using the entire input raw image without segmentation these kinds of factors effect our network performance. But as our network is quite deep, when we are combining the data of all data-sets in consideration we are getting an exceptional high performance as depicted in Table 6, this indicates a high generalization ability of our network.

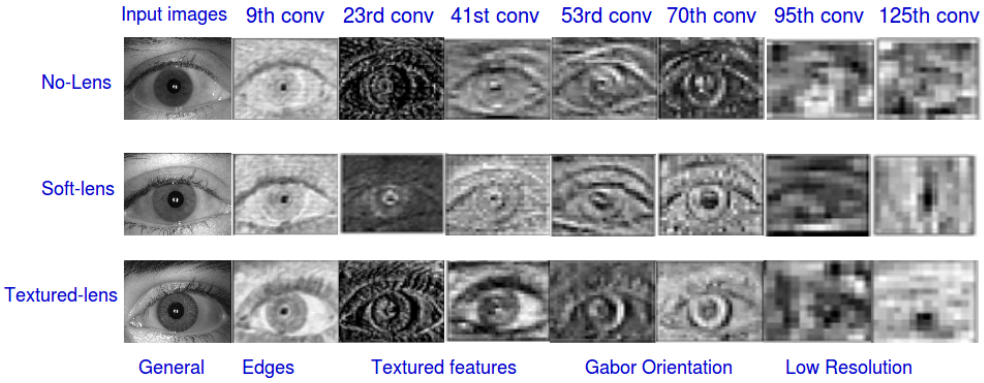

Figure 7: Features learned by different convolutional layers corresponding to [a] no lens [b] soft lens [c] cosmetic lens

Result Summarization : We have finally summarized our network performance for different testing protocols as follows : In all kinds of validation the results on GHCLNet outperforms the results on CLResNet architecture this shows that indeed deep architecture is required for multi-class oculus classification which is a very fine grain classification problem. It is the complexity of the problem in hand which decides the dictotomy between the shallow and deep networks.

- **Intra-Sensor Validation** GHCLNet performance is quite high in case of IIITD-Cogent and IIITD-Vista but it is less in case of ND-I and ND-II mainly because of less amount of training data available in these datasets and since our network is deep it requires large amount of data for predict-

ing good results.

- **Intra-Sensor Validation** It can be observed that GHCLNet performance is quite high from SIF[18] and DIR [19]and marginally lagging behind ContlensNet[7], that too in few cases mainly because of the poor quality, occluded input images.

- **Multi -Sensor Validation & Combined Sensor Validation** Table 6 and Table 7 depicts the highly generalized ability of GHCLNet which outperforms all the available state-of-the art techniques.

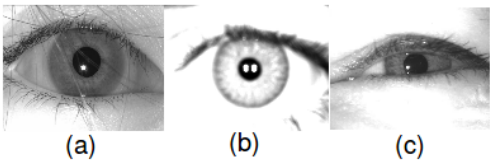

Figure 8: Categorized poor quality input dataset images (a) Occlusion (b) Illumination (c) Sleepy eye

2.8. Layer Specific Feature Analysis of GHCLNet

Fig.7 shows layer specific feature analysis of no-lens, soft-lens and cosmetic lens images. It is clearly evident from Fig.7, that initial convolutional layers learn general specific features. The main reason behind this is that initial convolutional layers look directly at the raw pixels which makes them more interpretative, while as we go deeper features corresponding to no-lens, soft-lens and cosmetic-lens are learned. Features learned by initial layers like $Conv-2$ layer are very basic features, but as we move deeper in the network more specific learning is been performed like in the $Conv-9$ layer which detects edges and lines. $Conv-23$ layer is playing a major role in differentiating no-lens image from cosmetic-lens image. In this layer textual features are learned. Interestingly, we have observed that our network automatically learns state-of-the-art

gabor filter like features at different orientations. The lower layers of the network learn high level aggregated discriminative features, as shown in Fig. 4, like $Conv-95$ and $Conv-125$ layers. In the lower layers of the network, the resolution and features become mostly an encoding of few discriminative intrinsic information.

2.9. Conclusion

For increasing the accuracy and reliability of iris based biometric systems it is paramount to detect the presence of contact lenses, especially cosmetic lens as they muddle the iris texture to a large extent. In this paper, we proposed two architectures (i) CLResNet and (ii) GHCLNet(Generalized Hierarchically tuned Contact Lens detection Network) in order to find the best network for our multi-class oculus classification. Extensive experimentation has been carried out with three publicly available databases using four testing strategies: intra-sensor validation, inter-sensor validation, multi-sensor validation, and combined-sensor validation to find the best among the two proposed architectures. The consistent CCR(%) improvements in case of GHCLNet over CLResNet makes him the winner. One important point to note is that deep neural networks like GHCLNet with multiple layers learns features at different levels of abstraction and thus they are highly generalizable if trained carefully in comparision to shallow networks . To the best of our knowledge combined sensor testing that we performed is not done by anyone so far. The main strength of the proposed GHCLNet architecture lies in the fact that it is not using any kind of pre-processing and iris segmentation, and still giving remarkable results. This saves lots of computational time and can thus be integrated very easily as the first step in any iris recognition system to increase its performance.

3. Fingerprint Sensor Classification

In heterogeneous sensor environment, it is crucial to identify the source sensor by which the acquired image is captured. This is essentially required to

handle sensor inter-operability issues and further in identifying various attacks on biometric systems, where biometric templates can be modified or mis-used. Another interesting application of sensor identification is in establishing the sequence of commands for law enforcement to identify spurious activities in online systems. An image can be altered or fabricated during the acquisition phase, transmission or during storage. In order to understand whether the image has been fabricated or not it is necessary to know the source that generates the image.

Fingerprint sensors can be classified into various categories *e.g.* (i) basis of imaging technology they are classified as optical, capacitive and thermal; (ii) basis of user interaction they are classified as press, sweep and non-contacted ones. Fig.9 shows fingerprint images captured from different types of sensors. It is evident from Fig.9, that image quality largely depends upon underlying sensor employed.

Figure 9: Example of fingerprint images taken from different sensors (a) Futonic, (b) Lumidigm, (c) SecuGen

3.1. Related Work

The existing sensor identification techniques can be grouped into three main categories based on : (i) hand crafted features, (ii) sensor pattern noise and (iii) colour filter. Bayrem et al. [20] proposed a method for sensor identification based on measuring the interpolation artifacts occurred in image using color filter arrays. Lukas et al. [21] proposed a technique in which sensor is identified by measuring the pixel non uniformity (PNU) noise of each image using wavelet

Author	Significant Contribution
Ross & Jain [26]	Optical versus Solid State
Bartlow [22]	Photo Response Non Uniformity
Modi [27]	False non match rate, minutia count
Jia [28]	Cross Database(Fingerpass)
Lungini [29]	Optical fingerprint sensor inter-operability
Agarwal [23]	Combining handcrafted features
Debiasi [30]	Multiple PRNU enhancements

Table 8: Summarized sensor classification literature review

based de-noising. Further Barlow et al. [22] used a variant of PNU technique known as photo response non-uniformity ($PRNU$) for fingerprint sensor identification. Agarwal et al. [23] used handcrafted features that includes features based on entropy, texture, image quality and statistics for sensor recognition. Recently, Sudipta et al. [24] identified sensors from NIR iris images. In their work they have reported that enhanced Sensor Pattern Noise Scheme (SPN), works better for detecting image sensor than maximum likelihood and phase based SPN methods. Uhl and Holler [25] have also used $PRNU$, to identify NIR iris sensor from their images. Table 8, summarizes the related work done in fingerprint sensor classification.

Contribution: In this paper we have proposed a Convolutional Neural Network based architecture for fingerprint sensor classification. The main contribution of this paper is three fold, that is summarized in the following section.

1. An architecture based on Deep Convolutional Neural Network is proposed that is capable of detecting input fingerprint sensor by systematically pruning and training two different types of convolutional neural networks VGG and ResNet50 namely.

2. In-depth feature analysis is done to understand the real-insight of features learned by different layers.
3. A highly generalized deep convolutional neural network based architecture has been proposed.

3.2. Proposed Architecture

In an image classification problem the task is to predict a label of the input image among the set of the pre-defined labels. Traditional methods of classification uses hand-crafted features like HOG and SIFT, but these methods encode low-level characteristics and thus not able to distinguish well in case of fine grain classification problems. In present world deep learning based methods are the state-of the art methods that are capable of encoding higher level characteristics.

Extensive experimentation has been done in order to decide the suitable network for our fine grain finger-print sensor classification problem. We have considered two networks (a) A shallow network (VGG) (b) A deep network (ResNet50) in order to understand the type of features, classification accuracy and generalization ability trade-off between shallow and deeper networks. As our finger-print sensor classification problem is not a trivial one, we are required to extract features at granular level. Another important point of consideration here is that our fingerprint image size is small and using a network having large kernel size will not be too useful. Considering all the above points in mind we conducted two set of experimentation.

3.2.1. Shallow Network(VGG-19 variant)

In the first set of experimentation we have used a variant of VGG-19 a popular deep-convolutional neural network model as shown in Fig.10. The foremost advantage of using this network is its small kernel filter size of $3 * 3$ which tries to learn high-level features at granular levels. VGG-19 network is divided into 5-blocks with 19 weight layers. It takes an input image of size $224 * 224$. Due to the limited amount of training data at our disposal we are incapable of re-

training it from scratch. After doing experimentation we have found that the block-5 is not adding any discriminative information for our fingerprint sensor classification problem so we systematically prune it and add a dropout layer also to avoid overfitting . While fine tuning this network we have used *adam* optimizer with a mini-batch size of 64, initial learning rate has been set as 0.001 for 15 epoches. All the parameters that are used in this work are calculated empirically over a small validation set.

3.2.2. Deep Network(ResNet50 variant)

In the second set of experimentation we have considered variant of ResNet50 architecture. ResNet50 network was developed by Microsoft research, this network is very deep consisting of 50 weight layers, it is approximately 2.6 times deeper than the VGG-19 model. We have deliberately chosen this network in order to find whether the knowledge of predecessor input to every layer gives it an upper hand in learning discriminative features essential for fingerprint sensor classification or not. Detail description about ResNet50 model is already explained in section 2.3.3. We have named this network FPSensorNet, its detailed diagram is shown in Fig. 11.

[a] Redundant block removal : After doing extensive experimentation we have found that $Block-2$ and $Block-4$ of ResNet50 are extremely important for learning discriminative information, which is necessary for differentiating between images acquired from different fingerprint sensors. We also have observed that $Branch-5$ and $Branch-3$ have "similar" contribution in final classification for our specific fingerprint sensor classification problem. Hence one can drop anyone layer, but dropping both have caused drastic performance deterioration. Hence we have dropped $Branch-5$ because generated feature map size was only $7*7$. Fig. 11 shows the proposed network architecture, which has been designed in a manner so that it can classify commonly used fingerprint sensors. In this model, we have dropped $Branch-5$, since in our case this branch was not learning much discriminative information. By doing so, we have decreased the computation time while retaining the performance.

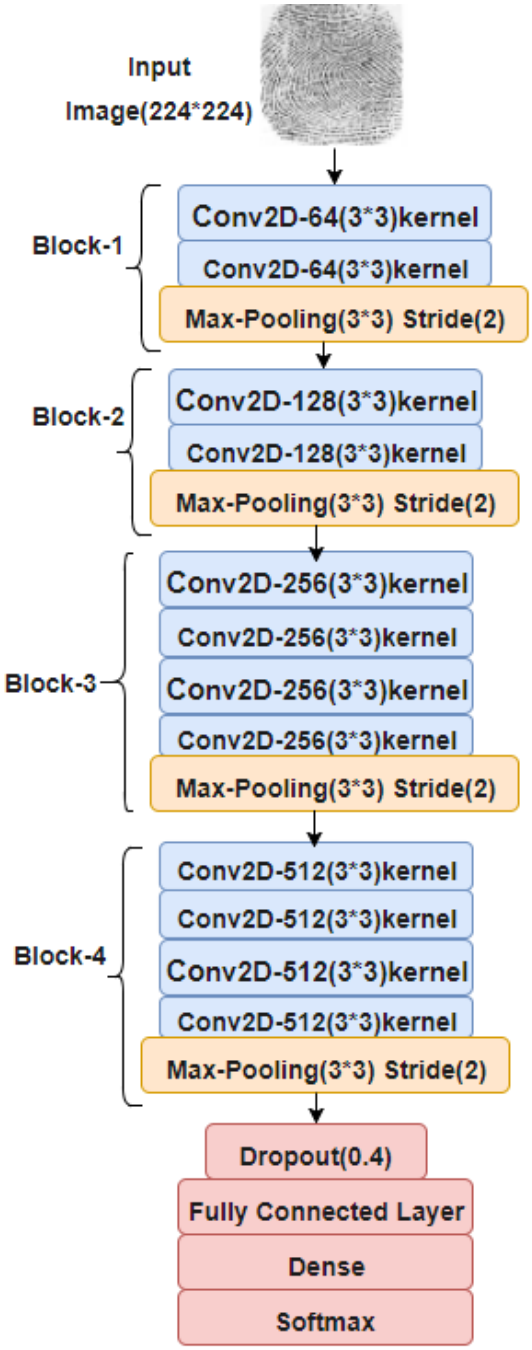

Figure 10: VGG-19 based shallow network for fingerprint sensor classification

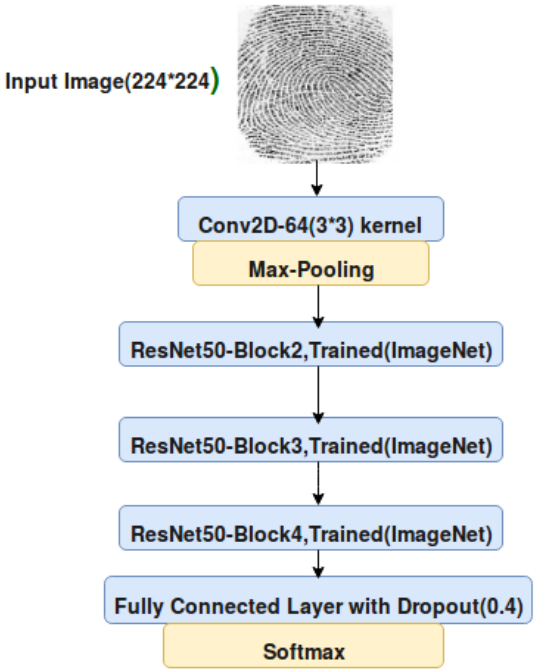

Figure 11: ResNet50 based deep network(FPSensorNet) for fingerprint sensor classification

[b] Major Observation: While implementing the two set of models described above we have observed that although the ResNet is much deeper model than the VGG but the model size of ResNet is much smaller than the VGG due to the usage of global average pooling layers in comparison to the fully connected layers.

[c] Network Implementation Details: All implementation is done in python using keras[8] library and tensorflow[9] as a backend. All the implementation has been done on Intel(R) Xenon(R) CPU E5-2630 V4 at 2.20 GHz, with 32 GB RAM and NVIDIA 1080 Ti gpu with 8 GB RAM.

3.3. Experimental Analysis

In this section, we provide the details about the database and the testing protocol used in this work. The proposed two architectures described in the previous section has been tested upon four publicly available benchmark single fingerprint databases *viz.*, (i) FVC2002 [31], (ii) FVC2004[32] and (iii) FVC2006 [33] (iv) IITD-MOLF [34] and on IITK database which is not publicly available and the largest dataset comprising of more than $40,000$ images so far . Few sample images of FVC datasets are shown in Fig.12. We have used single fingerprint images as well as four slap fingerprint images of IITK dataset for testing our proposed network.

The IITK dataset of single fingerprints consists of $41,129$ images collected using three different types of sensors *viz.*, (i) Futronic (FS88H), (ii) Lumidigm V310 (V31X) and (iii) SecuGen Hamster I. All of them are of same 500 DIP, but the light source and the image size generated is different. The IITK dataset of four slap fingerprints consists of $26,215$ images collected using three different types of sensors *viz.*, (i) Crossmatch, (ii) Sagem and (iii) Morpho. All FVC dataset consists of images collected from four different types of sensors. The detail description about the dataset considered has been given in Table 9. For our work we have considered only the set-A of FVC datasets. More information about these databases can be obtained from the reference papers [33], [31], [32]. We have trained the proposed model by taking only 10% of the single finger

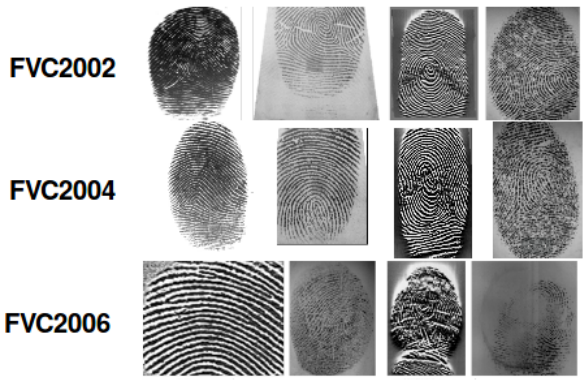

Figure 12: Sample images from FVC dataset.

print images as training set and rest 90% as testing set, thus adopting a very difficult protocol.

3.4. Single Slap Fingerprint Experimentation:

In this section we present the classification results of the proposed two different architectures (deep, shallow) on single fingerprint images. The results are computed for two different testing strategies namely (a) Intra sensor classification (b) Multi sensor classification [7]. The Correct Classification Rate (CCR%) is computed for performance evaluation, higher the value better is the result.

3.4.1. Intra-sensor Classification

In this type of classification training and testing is performed on images acquired from same type of sensor model. Table 10, indicates the computed performance of two different proposed architectures on various datasets. The main phenomenon that we observed is that we are getting almost same type of CCR(%) inspite of the fact of using two different kind of varied architectures. This can be attributed to the fact that fingerprint classification problem is a fine grain problem which does not require to learn features at the finest granular level and thus even shallow network is giving remarkable performance.

3.4.2. Multi-sensor Classification

In multi-sensor classification, data is fused together from various fingerprint sensors. We did this purposefully in order to check the generalizability of our proposed architectures. We have merged data from all FVC datasets *i.e.* FVC 2002[31], FVC 2004[32] and FVC 2006[33]. The combined dataset consists of $13,063$ images resulting from 12 different sensors and trained our network with 12 output neurons. We have trained our proposed model on 1306 images and tested on remaining $11,757$ images (*i.e.* 10% training and 90% testing). Table 11 indicates the computed performance in terms of CCR(%). We have observed that the obtained results are quite remarkable which depicts the high generalizability of our proposed architectures. In this case also shallow networks are learning discriminative features quite well.

3.4.3. Comparative Analysis with the state of art techniques :

In[23], fingerprint sensor classification has been done *via.* combination of handcrafted features this is the latest work done in this domain. They have combined four databases namely FVC 2002, FVC 2006, IIIT-D MOLF and CASIA Cross Sensor Fingerprint dataset. In total they have $29,320$ images out of which they have used 3000 images for training and remaining for testing. They have achieved an accuracy of about 96.52%.

In our experimentation we have also considered training on 10% dataset and testing on remaining 90% dataset. We have also considered FVC 2004 dataset which is quite challenging, and four- slap finger-print dataset of IIT-K . In all the cases the achieved aggregated accuracy is more than 98% and above. We are not comparing our results with [23] because of the unavailability of CASIA Cross Sensor fingerprint dataset .To the best of our knowledge this is the first attempt in which Deep Convolutional Neural Network is used for identifying the sensor of the underlying fingerprint image.

3.5. Four Slap Fingerprint Experimentation :

We have used one state of the art localization network *viz.*, Faster R-CNN [35] that is based on region proposals for extracting the ROI from four slap fingerprints. The ground truth for finger bounding boxes have been generated using method suggested in [36] which was around 92% accurate. Fig. 14 shows the ROI extracted on Four Slap fingerprint dataset. Extensive experimentation has been done in order to select the appropriate threshold and other network parameters for extracting the correct ROI. We have trained Faster R-CNN from scratch, for four slap fingerprint. Fig. 13, shows the graph that can summarize the performance of the the trained localization network where x axis is threshold applied over IOU (Intersection over Union) computed between predicted and actual bounding box, while y axis represents the accuracy at some threshold.

Four Slab Sensor Detection : For detecting the sensor of four slap fingerprints we have used the same proposed architecture (as discussed previously). We fed a four slap fingerprint image (sized 1672×1572) to our trained model as well as four single fingerprint images (extracted using the Faster-RCNN). Finally $softmax$ function has been used to generate the probability distribution. Each single fingerprint have probability distribution corresponding to its sen-

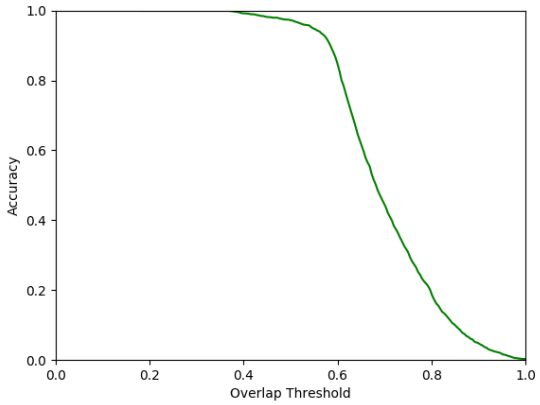

Figure 13: Accuracy Graph

sors. In order to predict the corresponding sensor for four slap fingerprint the aggregated sum of all probabilities has been calculated and maximum among them is assigned as its label class.

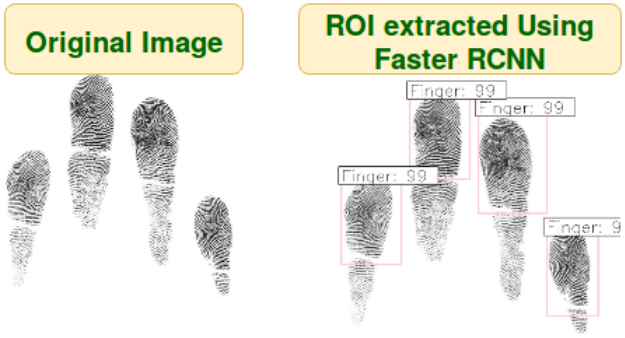

Figure 14: Four Slap Fingerprint (a) Original image, (b) Corresponding ROI extracted using FPSegNet.

Four Slap Fingerprint Results : We have used the same deep network for predicting the sensor of the four slap fingerprint images. In total we have $26,215$ images of four slap fingerprints, out of these, when we have used $15,729$ images as whole for training the network, obtained CCR(%) are as follows Crossmatch-Crossmatch 99.98, Sagem-Sagem 96.82, Morpho-Morpho 80.9. In the second set of experiment, we have used only 4080 images for training the network. In this case we have only shown the four single fingerprint images (extracted using retrained- Faster-RCNN), corresponding to the single four slap finger print for training the network. In this case the CCR(%) are as follows, Crossmatch-Crossmatch 100, Sagem-Sagem 94.76, Morpho-Morpho 65.2. In the first set of experiments we are getting low CCR(%) for Morpho due to big resizing, one have to do before feeding it to our network (from $1672 * 1572$ to $224 * 224$), hence a lot of discriminative information got lost during size reduction. In order to address this we have consider the second set of experimentation but due to the restricted computational resources at the dispose we have to train our network only over 1360 images per sensor.

3.6. Robustness or Generalization Analysis

In order to perform indepth analysis of the proposed shallow and deep network architecture described in section 3.2.1 and 3.2.2 respectively; we have introduced some artifacts in the original image in the form of rotation, random-noise, occlusion and translation. All these artifacts has been done on a small validation set comprising of 500 images of each sensor corresponding to IITK single fingerprint dataset, detailed description about all these artifacts is described in the following section. Fig.**??** and Fig.**??** shows the artifacts on fingerprint images captured through Futonics and Lumidigim sensors respectively.

[a] **Rotation** In order to check the robustness of the proposed architectures the fingerprint images of the validation set have been rotated with different angles like 2, 4, 6, 8 and 10 both in clockwise and anticlockwise direction. The computed CCR% on both shallow and deep network is shown in Table 12.

[b] **Random Noise** In real life scenarios it is very difficult to get a perfect condition. Most of the time we end up with noise affecting our ideal conditions. In such cases it is essential that our architecture is robust for noise upto certain range. An input image of size $224 * 224$ comprises of 50176 pixels. We introduce noise in percentage form like 0.01 percent, 0.05 percent and 0.1 percent. All these percentage is calculated over the pixels of the input image for example in case of 0.01 percent noise out of the total 50176 pixels 5 pixels are actually random noise. The computed CCR% on both shallow and deep network is shown in Table 12.

[c] **Occlusion** In order to test whether our proposed architecture is invariant to occlusion or not. We consider a patch of size $90 * 90$ at the center of the input image of size $224 * 224$ and occluded 1, 5 and 10 percent of this central patch with a black region of size $4 * 4$ for example 1 percent of $90 * 90$ patch means there will be 81 pixels which are occluded giving rise to 5 occlusion blocks of size $4 * 4$ each. The computed CCR% on both shallow and deep network is shown in Table 12.

[d] **Translation** While taking fingerprints it often happens that it is not positioned at the center of the sensor mostly it is tilted when taken in the un-

controlled enviornment. Hence it is foremost important to check the robustness of the proposed architectures when the input image is translated. In order to study this artifact we have translated the input image along x-axis as well as along y-axis. We have consider four set of experiments under this as (i) 1 to 4 (ii)5 to 8 (iii)9 to 12 (iv)14 to 16. Here 1 to 4 means any random number is generated between 1 to 4 for x-axis and for y-axis, this random number is different for both the axis and according to this generated random numbers pixels of the input image is translated along each axis; say if for x-axis random number 2 is generated and for y-axis random number 3 is generated than input image pixels will get tanslated 2 units along x-direction and 3 units along y-direction . Same stategy is followed for other set of experiments with random number ranges.The computed CCR% on both shallow and deep network is shown in Table 12.

3.7. Layer Specific Feature Analysis

As it is clearly evident from Table 13 that on increasing the amount of occlusion on our input images the performance of our proposed network is not degrading. This compels us to think some interesting questions: like what exactly our network is learning? Is there something wrong in its learning? Is there any kind of prestidigitation in our network? In order to answer these questions we try to visualize the feature maps learned by different layers of our proposed shallow network. For visualizing the feature-maps we took the SecuGen sensor of IIIT-K single fingerprint images. Part (a) of Fig.15 shows the features learned by our proposed shallow network. It is clearly evident from the Fig.15 , that initial convolutional layers are learning general specific features, while as we go deeper and deeper more sensor specific features, localized and sparse features are learned. It can be observed in Part (a) of Fig.15 that layer 3 and layer 6 are trying to understand the textual patterns of the image, but as we move deeper in the network like in layer 9 and layer 14 more emphasis is given to learn the shape and the background of the image rather than its textual patterns. By observing this we realize that our network is smart enough instead of learning finer granular level discriminative information to distinguish between

different sensors it learned the background and shape of the image instead of its textual features.

This could be the main reason for the exceptionally high performance of the network in case of occluded images. In order to prove it we did another set of experimentation in this we crop an input image from the middle in the size of 90∗ 90 patch and paste it over a white background. By doing this we have forcefully made the background of all the images same, in-order to force our network to learn textual features instead of background. It is clearly evident from Fig.15 that now our network is learning textual patterns instead of background. It is also proved by the results that we obtained after adding occlusion on this set of images. Table 15 shows the results after adding occlusion on this set of images. It is clearly visible as the amount of occlusion increases in our image the performance of the network decreases. By visualization the features it has been observed that at the lower layers of the network, the resolution and features becomes mostly an encoding of few discriminative intrinsic information.

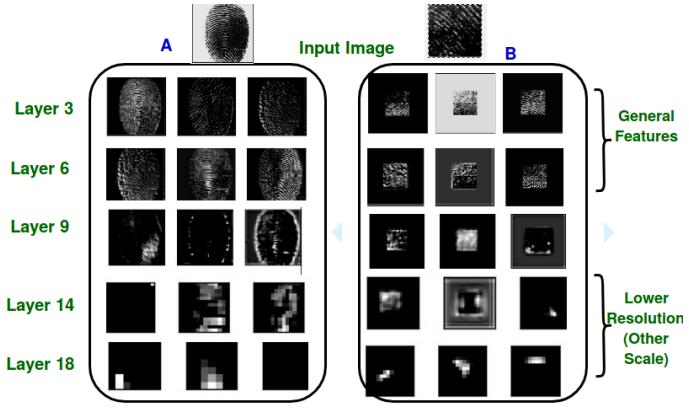

Figure 15: Comparative Feature Analysis (a)Input image with original background (b) Input image with forced background

3.8. Conclusion

The main contribution of this paper is in-depth analysis of the feature-map generated by different convolutional neural networks and proposing an archi-

tecture for identifying different fingerprint sensors, which is indeed important to handle sensor-inter-operability issues. We have observed that convolutional-neural networks are smart enough to learn different details under different scenarios. We have also seen that the choice of the network for any classification problem depends upon the complexity of the underlying problem. Extensive experiments have been performed on FVC-2002, FVC-2004 , FVC-2006, IIITD-MOLF and over IIT-K single slab fingerprints and multi-slab fingerprints. The proposed architecture yields correct classification rate above 98% with a very rigorous training and testing protocol compressing of only 10% of available images for training and rest 90% for testing.[34]

References

[1] H. Mhaskar, Q. Liao, T. A. Poggio, When and why are deep networks better than shallow ones?, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA., 2017, pp. 2343–2349.

[2] H. Mhaskar, T. A. Poggio, Deep vs. shallow networks : An approximation theory perspective, CoRR abs/1608.03287.

[3] D. Yadav, N. Kohli, J. S. D. Jr., R. Singh, M. Vatsa, K. W. Bowyer, Unraveling the effect of textured contact lenses on iris recognition, IEEE Trans. Information Forensics and Security 9 (2014) 851–862. doi:10.1109/TIFS.2014.2313025.

[4] J. Daugman, Demodulation by complex-valued wavelets for stochastic pattern recognition, Int. J. Wavelets, Multiresolution Inf. Process. 1 (2003) 1–17. doi:10.1016/0003-4916(63)90068-X.

[5] Lovish, A. Nigam, B. Kumar, P. Gupta, Robust contact lens detection using local phase quantization and binary gabor pattern, in: 16th International Conference, Computer Analysis of Images and Patterns CAIP, 2015, pp. 702–714. doi:10.1007/978-3-319-23192-1_59.

[6] E. C. Lee, K. R. Park, J. Kim, Fake iris detection by using purkinje image, in: International Conference on Biometrics,IAPR, 2006, pp. 397–403.

[7] R. Raghavendra, K. B. Raja, C. Busch, Contlensnet: Robust iris contact lens detection using deep convolutional neural networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision,WACV , Santa Rosa, CA, USA, 2017, pp. 1160–1167.

[8] F. Chollet, et al., Keras, `https://github.com/fchollet/keras` (2015).

[9] A. Mart, et al., Tensorflow: Large-scale machine learning on heterogeneous systems, `https://www.tensorflow.org/`.

[10] Johnson& johnson, skillman, nj, usa. (2013, apr.). acuvue2 colours, available: `http://www.acuvue.com/products-acuvue-2-colours`.

[11] Bausch & lomb, rochester, ny, usa. (2014, jan.). bausch & lomb lenses, available: `http://www.bausch.com`.

[12] Cibavision, duluth, ga, usa (2013, apr.) freshlook colorblends, available:`http://www.freshlookcontacts.com`.

[13] Irisguard, washington, dc, usa. (2013, apr.). ad100 camera, available: `http://www.Irisguard.com/uploads/AD100ProductSheet.pdf`.

[14] Lg, riyadh, saudi arabia. (2011, oct.). lg 4000 camera, available: `http://www.lgIris.com`.

[15] C. vision. (2013, apr.). expressions colors, available: `http://coopervision.com/contact-lenses/expressions-color-contacts`.

[16] J. Doyle, K. Bowyer, P. Flynn, Variation in accuracy of textured contact lens detection based on sensor and lens pattern, in: 6th IEEE International Conference on Biometrics, Technol., Appl., Syst., 2013, pp. 1–7. `doi:10.1109/BTAS.2013.6712745`.

[17] J. Daugman, How iris recognition works, in: Proceedings of the 2002 International Conference on Image Processing, ICIP, 2002, pp. 33–36.

[18] R. Raghavendra, K. B. Raja, C. Busch, Ensemble of statistically independent filters for robust contact lens detection in iris images, in: Proceedings of Indian Conference on Computer Vision Graphics and Image Processing, ICVGIP 14, 2014, pp. 241–247.

[19] P. Silva, E. Luz, R. Baeta, H. Pedrini, A. X. Falcao, D. Menotti, An approach to iris contact lens detection based on deep image representations, in: Graphics, Patterns and Images (SIBGRAPI), 2015, pp. 157–164.

[20] S. Bayram, H. Sencar, N. Memon, I. Avcibas, Source camera iden- tification based on cfa interpolation, in: IEEE International Conference on Image Processing, Vol. 3, 2005, pp. 69–72.

[21] J. Lukas, J. Fridrich, M. Goljan, Digital camera identification from sensor pattern noise, IEEE Transactions on Information Forensics and Security (2006) 205–214.

[22] N. Bartlow, N. D. Kalka, B. Cukic, A. Ross, Identifying sensors from fingerprint images, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops , Miami, FL, 2009, pp. 78–84.

[23] A. Agarwal, R. Singh, M. Vatsa, Fingerprint sensor classification via mélange of handcrafted features, in: 23rd International Conference on Pattern Recognition, ICPR, 2016, pp. 3001–3006.

[24] S. Banerjee, A. Ross, From image to sensor: Comparative evaluation of multiple prnu estimation schemes for identifying sensors from nir iris images, in: 5th International Workshop on Biometrics and Forensics (IWBF), (Coventry, UK), 2017.

[25] A. Uhl, Y. Hller, Iris-sensor authentication using camera prnu fingerprints, in: 5th IAPR International Conference on Biometrics (ICB), 2012, pp. 230–237.

[26] A. Ross, A. K. Jain, Biometric sensor interoperability: A case study in fingerprints, in: Biometric Authentication, ECCV International Workshop,BioAW, Prague, Czech Republic, 2004, pp. 134–145.

[27] S. K. Modi, S. J. Elliott, H. Kim., Statistical analysis of fingerprint sensor interoperability performance, In IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems.

[28] X. Jia, X. Yang, Y. Zang, N. Zhang, J. Tian, A cross-device matching fingerprint database from multi-type sensors, in: 21st International Conference on Pattern Recognition, 2012, pp. 3001–3004.

[29] L. Lugini, E. Marasco, B. Cukic, I. Gashi, Interoperability in fingerprint recognition: A large-scale empirical study, in: 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop, 2013.

[30] L. Debiasi, A. Uhl, Blind biometric source sensor recognition using advanced prnu fingerprints, in: 23rd European Signal Processing Conference (EUSIPCO),2015, 2015, pp. 779–783.

[31] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, A. K. Jain, Fvc2002: Second fingerprint verification competition, in: 16th International Conference on Pattern Recognition,2002, Vol. 3, 2002, pp. 811–814.

[32] D.Maio, D.Maltoni, R.Cappelli, J. Wayman, A.K.Jain, Combining multiple matchers for fingerprint verification: A case study in FVC2004, in: 13th International Conference, Image Analysis and Processing ICIAP, 2005, pp. 1035–1042.

[33] R. Cappelli, M. Ferrara, A. Franco, D. Maltoni, Fingerprint verification competition 2006, in: Biometric Technology Today,2007, Vol. 15, 2007, pp. 7–9.

[34] A. Sankaran, M. Vatsa, R. Singh, Multisensor optical and latent fingerprint database, in: IEEE Access, 2015, pp. 653 – 665.

[35] S. Ren, K. He, R. B. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. (2017) 1137–1149.

[36] N. Singh, A. Nigam, P. Gupta, P. Gupta, Four slap fingerprint segmentation, in: 8th International Conference, Intelligent Computing Theories and Applications ICIC, Huangshan, China, 2012, pp. 664–671.

Database	Sensor Model	Images
FVC 2002 [31] (*SingleFinger*)	Optical Sensor "TouchView" Optical Sensor "FX2000" Capacitive Sensor "100SC" Synthetic fingerprint generation	3,200
FVC 2004 [32] (*SingleFinger*)	Optical Sensor "V300" Optical Sensor "U4000" Thermal Sweeping Sensor Synthetic fingerprint generation	3,143
FVC 2006 [33] (*SingleFinger*)	Electric Field Sensor Optical Sensor Thermal sweeping Sensor SFinGe V3.0	6,720
IIITD-MOLF [34] (*SingleFinger*)	Lumidigm Venus IP65 Shell Secugen Hamster-IV CrossMatch L-Scan Patrol	16,400
IITK-S Dataset (*SingleFinger*)	Futronic FS88H Lumidigm V310(V31X) SecuGen Hamster I	41,129
IITK-F Dataset **(Four Slap Finger)**	Crossmatch Sagem Morpho	26,215

Table 9: Detail Database Specifications

Database	Sensor Type	Shallow	Deep
FVC 2002 [31]	DB1	100	100
	DB2	99.17	96.79
	DB3	100	99.72
	DB4	99.72	99.45
	Aggregate	**99.72**	**98.99**
FVC 2004[32]	DB1	100	100
	DB2	100	100
	DB3	99.85	93.78
	DB4	100	99.17
	Aggregate	**99.96**	**98.23**
FVC 2006[33]	DB1	100	99.87
	DB2	100	99.28
	DB3	100	99.80
	DB4	99.93	100
	Aggregate	**99.98**	**99.73**
IIITD-MOLF	DB1	100	100
	DB2	100	100
	DB3	99.92	100
	Aggregate	**99.97**	**100**
IITK S Dataset	Futronic	99.45	99.34
	Lumidigm	100	100
	SecuGen	100	100
	Aggregate	**99.82**	**99.78**

Table 10: Intra-Sensor qualitative performance in CCR(%) on proposed architectures (where IITK S dataset is single fingerprint dataset and DB1 means the images from the first sensor of the corresponding dataset and same nomenclature is used for other datasets)

Database	Sensor Type	Shallow	Deep
FVC Combined [31]	2DB1	98.75	99.29
	2DB2	97.63	98.37
	2DB3	100	99.57
	2DB4	99.58	99.72
	4DB1	99.31	99.16
	4DB2	100	100
	4DB3	100	99.40
	4DB4	100	97.47
	6DB1	100	100
	6DB2	100	100
	6DB3	99.86	98.56
	6DB4	99.80	100
	Aggregate	**99.58**	**99.29**

Table 11: Multi-sensor qualitative performance in CCR(%) on proposed architectures (here in xDBy: x stands for FVC dataset number and y stands for sensor type of the corresponding dataset(eg. x=2 & y=1 means first sensor of FVC-2002 dataset))

Rotation	2	4	6	8	15
Futonic	99	99	99	99	99
Lumidigm	100	100	100	100	100
SecuGen	94.5	84.5	80	76	54

Table 12: Rotation qualitative performance in CCR(%) on validation set for shallow network architecture (here 2, 4, 6, 8 and 15 represents the angle of rotation in clockwise as well as in anticlockwise direction)

Occlusion	0.1%	1%	5%	10%
Futonic	99	99	98	98
Lumidigm	100	100	100	99
SecuGen	99	100	100	100

Table 13: Occlusion qualitative performance in CCR(%) on validation set for shallow network architecture (here 0.1%,1%,5% and 10% are the percentage of pixels occluded in a region of $90 * 90$ patch)

Noise	0.01%	0.05%	0.1%	1%	5%
Futonic	99	99	99	98	93
Lumidigm	100	94	50	0	0
SecuGen	99	98	98	99	98

Table 14: Random noise qualitative performance in CCR(%) on validation set for shallow network architecture (here 0.01%, 0.05%, 0.1%, 1% and 5% are the percentage of the random noise inserted in the input image of size $224 * 224$)

Occlusion	0.1%	1%	5%	10%
Futonic	93	37	10	0.8
Lumidigm	87	2	0	0
SecuGen	99	100	100	98

Table 15: Occlusion qualitative performance in CCR(%) on experimental set for shallow network architecture (here 0.1%,1%,5% and 10% are the percentage of pixels occluded in a region of $90 * 90$ patch)