

# Crime Detection and Criminal Recognition to Intervene in Interpersonal Violence Using Deep Convolutional Neural Network With Transfer Learning

Mohammad Reduanul Haque, Daffodil International University, Bangladesh

Rubaiya Hafiz, Daffodil International University, Bangladesh


Alauddin Al Azad, Daffodil International University, Bangladesh

Yeasir Adnan, Daffodil International University, Bangladesh

Sharmin Akter Mishu, Daffodil International University, Bangladesh

Amina Khatun, Jahangirnagar University, Bangladesh

Mohammad Shorif Uddin, Jahangirnagar University, Bangladesh

 <https://orcid.org/0000-0002-7184-2809>

## ABSTRACT

Interpersonal violence, such as physical and sexual abuse, eve-teasing, bullying, and taking hostages, is a growing concern in our society. The criminals who directly or indirectly committed the crime often do not go into the trial for the lack of proper evidence as it is very tough to collect photographic proof of the incident. A subject's corneal reflection has the potentiality to reveal the bystander images. Motivated with this clue, a novel approach is proposed in the current paper that uses a convolutional neural network (CNN) along with transfer learning in identifying crime as well as recognizing the criminals from the corneal reflected image of the victim called the Purkinje image. This study found that off-the-shelf CNN can be fine-tuned to extract discriminative features from very low resolution and noisy images. The procedure is validated using the developed datasets comprising six different subjects taken at diverse situations. They confirmed that it has the ability to recognize criminals from corneal reflection images with an accuracy of 95.41%.

## KEYWORDS

Criminal Identification, Deep Convolutional Neural Network, Interpersonal Violence, Purkinje Image

## 1. INTRODUCTION

Interpersonal violence, usually referred to as the intentional use of physical force or power, threatening, psychological torture, or deprivation by someone against another individual (Mayo Clinic, 2017). It is a growing concern in our society as children and adolescents are more likely to be targeted, but anyone can be a victim of this brutality including women, children, and disabled persons. It might

DOI: 10.4018/IJACI.20211001.oa1

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

not be easy to identify these types of violence because it is very hard and sometimes impossible to collect the evidence during the offense. Also, these types of crimes often start subtly and get worse over time. For these reasons, the victims may suffer lifelong consequences. For example, kids who get bullied are more likely to suffer from clinical anxiety, depression, and feelings of social isolation. They usually tend to avoid school, and eventually, these continuous chronic strain makes them physically sick (Nansel, Craig, Overpeck, Saluja & Ruan, 2004). Moreover, the chances of being suffered from agoraphobia, panic disorder, and anxiety are 3 to 5 times higher for adults due to these reasons (Copeland, Wolke, Angold & Costello, 2013).

Social awareness programs as well as necessary laws are enforced but criminals are often getting unpunished for the lack of proper evidence. In all the above-mentioned incidents, photographs play a very crucial role in criminal investigations. If photographs (face images) can be taken during the incident, then it gives us a visual knowledge of the offender as well as the surrounding environment that is desirable to identify the crime as well as criminals. Recently, Jenkins and Kerr (2013) performed a primitive investigation through the conventional method on the detection of bystanders from the corneal reflections of a subject known as Purkinje images (Sigut & Sidha, 2011). However, a detailed study of practical perception is not performed yet. Due to technological advancements capturing these images is possible through an IoT-based tiny camera and can be processed using deep learning. Motivated by this research gap we are trying to investigate on criminal identification from the corneal images with advanced technologies. The current paper focuses on the very crucial work on detection and recognition of faces of the criminals from the Purkinje images as these are of low resolution and poor quality.

Moreover, there are lots of variations in image appearance, such as pose variation (front, non-front), occlusion, image orientation, illuminating condition, and facial expressions. Nowadays, computer vision has numerous applications in every sector of our life, such as load forecasting for smart grid (Mukherjee, Mukherjee, Dey, De & Panigrahi, 2020), prediction of possible health issues (Mukherjee & Mukherjee, 2019), intelligent sensing (Mukherjee, Panja & Dey, 2020), etc.

Recently, deep learning through deep convolutional networks (CNNs) show promising results in object detection and classification (Chen, Li & Li, 2020). The deep CNN approach has been gained significant improvement in face detection (Garcia & Delakis, 2002; Osadchy, Cun & Miller, 2007), facial point detection (Sun, Wang & Tang 2013), human attribute inference (Zhang, Paluri, Ranzato, Darrell & Bourdev, 2014), plant phenotyping (Jiang & Li, 2020), organism detection (Huang et al., 2019) and in many other applications. These models achieve state-of-the-art results on several data sets and can enhance the success rate of criminal identification as compared to conventional as well as other shallow machine learning methods. However, currently, no work has done so far using deep CNNs. So, a method is proposed here that takes face images from the corneal reflections and after pre-processing trains using the off-the-self CNN along with transfer learning.

The contributions of this work are summarized as follows:

1. Criminal identification and recognition through benchmark CNN methods have been implemented to reduce interpersonal violence.
2. Transfer learning is applied to retrain CNN models to handle the problem with fewer efforts and improved accuracy.
3. Developed a dataset of 6000 unconstrained images of six different subjects (1000 of each category).
4. Robustness of the technique is tested using noise, occlusion, rotation, and low-resolution images.

The rest of the paper is assembled as follows: a literature review is presented in Section II. Section III consists of methodology along with the description of the dataset. Experimental results and discussions are described in Section IV. Eventually, it is concluded in Section V.

## 2. LITERATURE REVIEW

As photographs are very crucial in criminal investigations and thus cameras are seized routinely during investigations. Sometimes it is desirable to identify the photographer or other individuals who were present at the scene or surrounding but not directly present in the photographed images. Criminal identification and recognition are very important particularly in those cases where criminal photographs the victim.

In this perspective, Nishino & Nayar (2006) et. al. has proposed a comprehensive framework that is capable to extract different types of information embedded within a single catadioptric image of the eyes (the image that can be built by combining the cornea and a camera viewing at the eyes). They computed an environment map from the reflections of the cornea as well as the iris. This map could be able to recover the surrounding world of the person which reveals what the person is looking at. As a result, their proposed model could be able to determine the location and circumstances of the person at the time of image capturing. Environment map obtained from the corneal reflection enables several applications such as scene panorama reconstruction (Nakazawa, Nitschke, & Nishida, 2016), biometrics (Bowyer, Hollingsworth & Flynn, 2008), creation of a 3D model (Nishino & Nayar, 2006), super-resolution imaging (Nitschke & Nakazawa, 2012), illumination normalization in face recognition (Nakazawa & Nitschke, 2012), point of gaze estimation (Kar & Corcoran, 2017; Nitschke, Nakazawa & Nishida, 2013; Ogawa, Nakazawa & Nishida, 2018), etc. A study was conducted by Wan and his co-authors to remove reflection components from a mixture image to recover reflection scenes from an image (Wan, Shi, Li, Duan & Kot, 2020). This will be beneficial especially for surveillance or criminal investigations.

A huge number of psychological and image-based researches has been done to identify humans from the low-resolution photographs as well as videos if the identifier is previously familiar with that particular face (Burton, Wilson, Cowan, & Bruce, 1999; Jenkins & Kerr, 2013; Marciniak, Dabrowski, Chmielewska & Weychan, 2012). Yip and Sinha (2001) have proposed a model that could identify dim photographs of familiar faces with a low resolution of  $7 \times 10$  pixels. For extracting information hidden on an image, Jenkins and Kerr (2013) used corneal reflection images. They experimented with constraint environment (high-resolution) images. From the reflections of the subject eyes, they have segmented out the images of bystanders. They imposed manual experiments and got 71% accuracy if the participants were unfamiliar with the bystander's face. But in the case of familiarity, they got an average accuracy of 84%.

CNN or ConvNet, one of the most extensively used deep learning architecture, have achieved state-of-the-art accuracy for any type of face detection, recognition, and classification tasks (Albu, 2009; Khalajzadeh, Mansouri & Teshnehlal, 2014; Jiang & Learned-Miller, 2017; Khan, Sohail, Zahoor & Qureshi, 2019). Xinhua and Qian (2015) have proposed a deep neural network-based face detection method to improve the biometric recognition technique. They have experimented with their proposed model using more than 13,000 face images and achieved a classification accuracy of more than 97%. Kamencay (Kamencay, Benčo, Miždoš & Radil, 2017) has proposed a CNN-based face recognition method that can be able to identify faces with a recognition accuracy of 98.3%. Kwolek (2005) proposed a combined architecture of Gabor filter and CNN and obtained a detection accuracy of 87.5% on a sample data-set containing 1000 face and 10000 non-face images. A study found that image degradation dropped the classification accuracy of CNN models significantly (Pei, Huang, Zou, Zhang & Wang, 2019). Some studies aim at identifying faces from tiny (Chang, Lu, Liu, Zhou & Qiao, 2020), distorted (Dodge & Karam, 2016), or noisy (Wu, He, Sun & Tan, 2018) images. However, there are limited works done on criminal detection from corneal reflection images. Hence, the present paper tries to fill this gap using a deep CNN-based automated system along with transfer learning. A summary of these works is given in Table I.

**Table 1. Summary of related face recognition works. Only the first one (Jenkis and Kerr, 2013) used corneal reflection images**

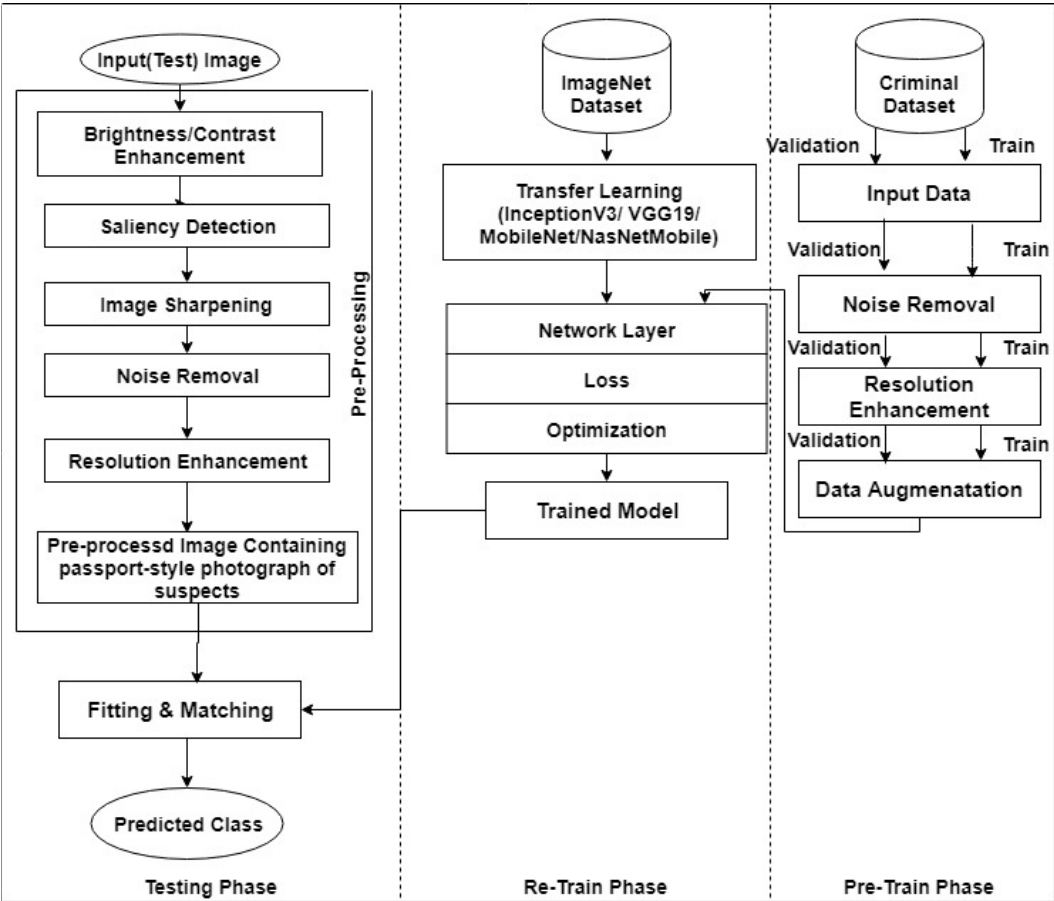
References	Methodology	Corneal Reflection Image Used (Yes/No)	No of Images/ Dataset Used	Accuracy
(Jenkis and Kerr, 2013)	Manual experiments	Yes	16	84% (For familiar faces) and 71% (for unfamiliar faces)
(Xinhua and Qian, 2015)	Deep Neural Network	No	13,000	97%
(Jiang & Learned-Miller, 2017)	Faster Region-based Convolutional Neural Networks (R-CNN)	No	WIDER, FDDB, and IJB-A dataset	Fast-RCNN gives state-of-the-art result
(Kamencay, Benčo, Miždoš & Radil, 2017)	Principal Component Analysis (PCA)/ Local Binary Patterns Histograms (LBPH)/ K-Nearest Neighbour (KNN) / CNN	No	400 from ORL face database	CNN gives the best accuracy of 98.3%
(Kwolek, 2005)	Gabor Filter and CNN	No	1000 face and 10000 non-face images	87.5%
(Dodge & Karam, 2016)	Deep Neural Network	No	ImageNet Dataset	VGG-16 gives the best result
(Chang, Lu, Liu, Zhou & Qiao, 2020)	YOLOv3-C (Enhanced YOLOv3)	No	12,877 images from Wider Face data set	YOLOv3-C gives the best result with a mean average precision (mAP) of 51.9%
(Wu, He, Sun & Tan, 2018)	Light CNN (Light CNN-4 Light CNN-9 Light CNN-29)	No	LFW, YouTube Face (YTF), YouTube Celebrities (YTC), Celebrity1000, MegaFace, IJB-A, IJB-B, CACD-VS, Multi-PIE, and the CASIA NIR-VIS 2.0 databases	Light CNN-29 outperform all state-of-the-art method.
(Wan, Shi, Li, Duan & Kot, 2020)	Train the network with a shift-invariant loss	No	Training: 2500 Evaluated: 30	It gives promising results compared to other methods.

### 3. METHODOLOGY

#### 3.1. The Proposed Scheme

Our work aims to recognize criminals from the Purkinje images. To do this, at first, we have prepared the training and testing datasets. We fed our training dataset directly into the augmentation process to increase the data size. Figure 1 shows the step-wise schematic diagram of our proposed methodology.

Figure 1. Schematic diagram of the System



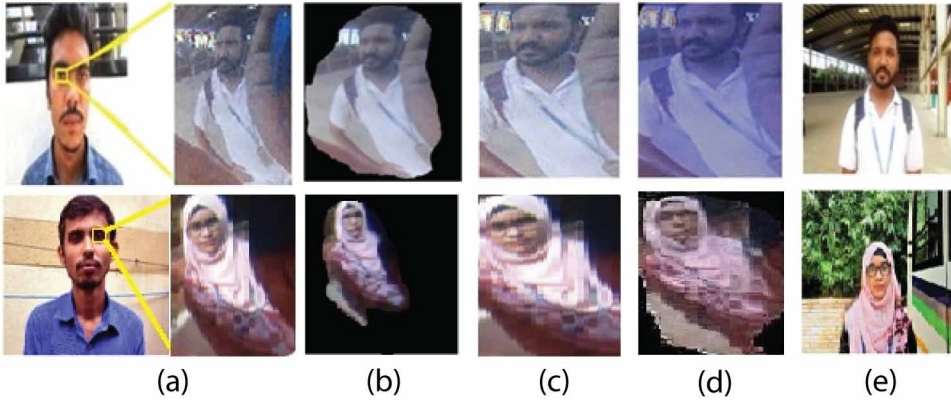
### 3.2. Dataset Preparation

For the experiment, six volunteers (5 male, 1 female, and age between 23 to 26 years) are selected. They gave their consent to use and publish their photographs only for research purposes. We have collected a total of 6000 unconstrained images of six different subjects (1000 of each). 80% of these images are used for training and the remaining 20% for the validation dataset to pursue 5-fold cross-validation. For constructing the test dataset, a total of 3000 images (500 images for each subject) are captured in such a way that every image contains a criminal on the subject's corneal region. Some sample images along with the sequential flow diagram of the preparation of the testing image are shown in Figure 2 and Figure 3, respectively.

Figure 2. Sample images from the test dataset obtained from the victim's eyes



Figure 3. Flow diagram of the preparation of testing images. (a) Original scene image of the criminal that is obtained from the cornea region of the victim after zooming and cropping, (b) Image after employing saliency detection, (c) Resultant passport-style photograph of the criminal after removing backgrounds, (d) Pre-processed image after performing Gaussian noise reduction and discrete wavelength transform-based resolution enhancement, (e) Final image for the input to the deep CNN.



Only the training data is augmented by data augmentation techniques for improving the accuracy and reduce the overfitting problem. For each category cropping, shifts, random rotations, and flips are applied while performing augmentation. After augmentation, the size of the training dataset becomes 48,000.

The dataset developed in this work is available at <https://github.com/reduan/Criminal-Identification-Dataset.git>

### 3.3. Data Pre-Processing

After preparing the datasets, all test images are pre-processed as they include a huge amount of noise including iris texture contamination, shadows of eyelash and eyelid, non-linear distortion from the reflection of a curved surface, etc.

The graph-based visual saliency (GBVS) method helps us to highlight the “remarkable” position where the image is informative (Harel, Koch & Perona, 2007). Regions that contain 60% saliency are cropped out and the rest of the regions are considered as the background and these are removed. After that, a Gaussian high pass filter is used to enhance the object present in the scene. If  $D(a, b)$  denotes the distance of the point  $(a, b)$ ,  $\sigma$  is the cutoff frequency, then the filter is defined by Eq. (1) (Asadi, Jamzad & Sajedi, 2008):

$$P(a, b) = 1 - e^{-D^2(a, b)/2\sigma^2} \quad (1)$$

Eq. (2) is imposed to generate high pass filtered image  $X'$  from input image  $X$  where  $F\{X\}$  is the Fourier Transform of arguments:

$$X'(a, b) = F\{X\}P(a, b) \quad (2)$$

A fuzzy filter is shown in Eq. (3) with kernel size 3 x 3 pixels is imposed for reducing Gaussian noise while keeping other features intact (Rahman, Haque, Rozario & Uddin, 2014):

**Table 2. A brief architecture of 4 well-known deep CNNs for transfer learning**

	InceptionV3	VGG19	MobileNet	NasNetMobile
<b>Input Size</b>	227x227	224x224	224x224	224x224
<b>Conv. Layers</b>	21	19	28	-
<b>Filter Size</b>	1,3,5,7	3	1,3	1,3,5
<b>Stride</b>	1,2	1,2	2	1,2
<b>Parameter</b>	23M	155M	5855942	6929562
<b>FC Layer</b>	1	1	4	4
<b>Size</b>	92MB	549MB	17MB	-
<b>Depth</b>	159	26	88	-

$$Fp = \begin{cases} 1, & f_p = 0 \text{ or } 255 \\ \exp\left(-\frac{(f_p - f_{max})^2}{2 \times 8 \times \sigma}\right), & \text{Otherwise} \end{cases} \quad (3)$$

where  $f_{max}$  and  $\sigma$  denote the maximum intensity value among the 8 neighboring pixels and the standard deviation of all intensity values respectively. Since color images are used here, this filter is applied separately for each of the R, G, and B components.

A discrete wavelet transform based resolution enhancement is applied to enhance the resolution of the images (Khair & Shelkikar, 2013). Finally, the resultant pre-processed images contain only the passport-style photographs of the criminal are fed to the training and testing processes.

### 3.3. Deep CNN With Transfer Learning

Augmented training and pre-processed test images are sent to retrain the models using transfer learning. Deep learning models need huge labeled data to train for achieving high recognition accuracy. However, it is very expensive to collect huge amounts of labeled data. Due to this reason, it is a suitable alternative to train CNNs with scarce datasets through transfer learning, initiated by Thrun (1996). He investigated the role of previously learned features for applications with scarce training data. The dataset is retrained by 4 well-known deep CNNs: VGG19, InceptionV3, MobileNet, and NasNetMobile. A brief architecture of these 4 deep CNNs is given in Table 2.

### 3.4. Performance Metric

The overall performance of any system is usually measured using the confusion matrix through accuracy. This is calculated as the ratio of the sum of correct clusters to the total clusters as per the binary classifier and the equation is:

$$\begin{aligned} \text{Accuracy}(\%) &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \\ &= \frac{\text{Correctly classified number of images}}{\text{Total number of images}} \times 100 \end{aligned} \quad (4)$$

where:

True Positive (TP) – A criminal is identified as a criminal

True Negative (TN) – A good person is identified as a good person

False Positive (FP) – A good person is identified as a criminal

True Positive (FN) – A criminal is identified as a good person

## 4. EXPERIMENTAL RESULT AND DISCUSSION

### 4.1. Experimental Result

The system is tested on varying conditions such as noise, illumination, occlusion pose, etc. to check the performance and robustness of different CNNs.

### 4.2. Discussion

Table 3 shows the comparative performance of InceptionV3, VGG19, MobileNet, and NasNetMobile models for rotated (both clockwise and anticlockwise) images. MobileNet shows the highest performance. Table 4 demonstrates how CNN models perform for criminal identification and recognition under uneven illumination settings along with pose variations (full profile, and 3/4 profile). This table shows that the recognition rate decreases by almost 10% when illumination condition changes for the 3/4 profile and full profile poses. Table 6 describes the relative performance of InceptionV3, NASNetMobile, VGG19, and MobileNet methods for images with Salt and Pepper noise. This result validated that VGG19 and MobileNet perform better (on the basis of noise density) than InceptionV3 and NASNetMobile. Table 7 depicts a similar result for Gaussian noise where MobileNet shows the best performance. The experimental result confirms two things: (i) MobileNet shows better accuracy in comparison with other methods in terms of diverse illumination, occlusion, noise, and pose, (ii) robustness of the system is ensured. Some visual results are illustrated in Figure 4 and Figure 7. The confusion matrix for MobileNet is shown in Table 5 as it gives the best performance.

Figure 5 shows some wrong recognition results. The reason for getting confused results is mainly due to the difficulties of acquiring high-resolution test images. The images contain noise including iris texture contamination, shadows of eyelash and eyelid, non-linear distortion from the reflection

Table 3. Maximum rotation limit tolerable by different CNNs for subject recognition

Maximum Rotation Limit			
InceptionV3	NasNetMobile	VGG19	MobileNet
3.5°	6°	8°	8°

Table 4. Recognition rate (%) of different CNNs under varying illumination and occlusion

Illumination	Varying	
Occlusion Pose	3/4 profile	Full Profile
InceptionV3	78.77%	86.93%
VGG19	86%	94%
MobileNet	88%	95.41%
NasNetMobile	55%	68%



**Table 5. Confusion matrix of recognition by MobileNet model using a full profile of the criminal**

		Predicted Class					
Actual Class		Sub1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6
	Sub1	<b>97.2</b>	0	0	0	0	2.98
	Sub2	0	<b>97.98</b>	0	0	0	2.02
	Sub3	2.04	4.08	<b>93.88</b>	0	0	0
	Sub4	1.25	2.75	0	<b>93.45</b>	2.55	0
	Sub5	0	0	0	4.3	<b>95.7</b>	0
	Sub6	4.7	1	0	0	0	<b>94.3</b>
	<b>Accuracy 95.41%</b>						

**Table 6. Recognition under noisy situations using the criminal dataset (with pre-processing and for Salt and Pepper Noise with density, d)**

InceptionV3	NASNetMobile	VGG19	MobileNet
$d < 0.2$	$d < 0.1$	$d < 0.4$	$d < 0.4$

**Table 7. Recognition under noisy situations using the criminal dataset (with pre-processing and for Gaussian Noise with mean = 0 and variance,  $\nu$ )**

InceptionV3	NASNetMobile	VGG19	MobileNet
$\nu < 0.18$	$\nu < 0.12$	$\nu < 0.45$	$\nu < 0.5$

**Figure 4. Images of an object under different noise levels: (a) original image; (b) image captured from cornea (c) corrupted image by Gaussian noise with density,  $d = 0.5$  (MobileNet and VGG19 recognized this face correctly, but InceptionV3 and NASNetMobile failed.); (d) corrupted image by Salt & Pepper noise with mean = 0 and variance,  $\nu = 0.4$  (MobileNet and InceptionV3 recognized this face correctly, but VGG19 and NASNetMobile failed).**



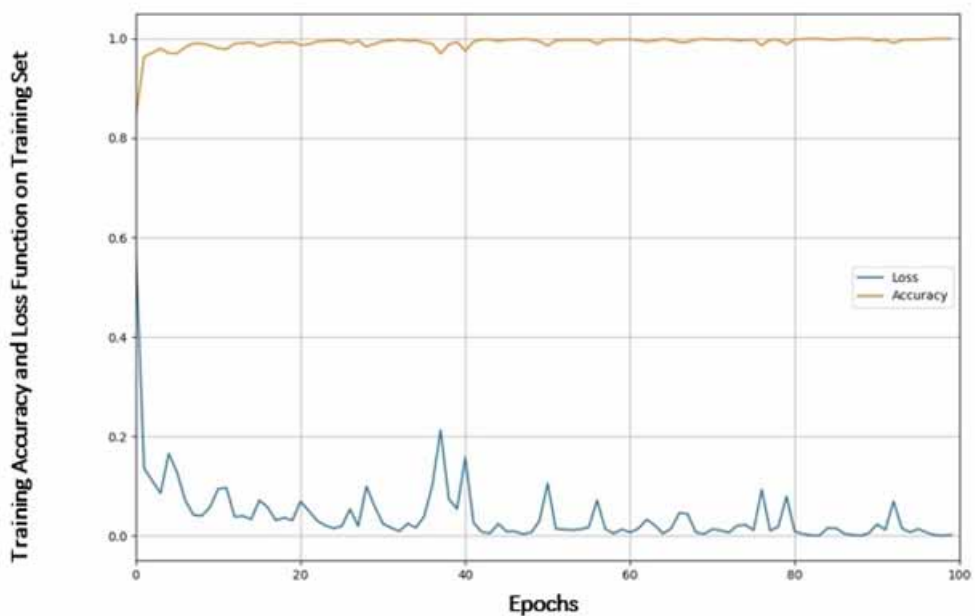
of a curved surface, and inter-class similarities. The accuracy and loss graphs for MobileNet are shown in Figure 6.

The overall classification accuracy of our proposed criminal identification system for different CNN models after employing the transfer learning is shown in Table 8. The accuracy of the test dataset is calculated using with and without pre-processing. It is found that after pre-processing the accuracy is improved a lot.

Figure 5. Some error-prone images from the test dataset. Subjects (a), (b), and (c) are classified correctly by the MobileNet but failed to recognize the subjects (d) and (e) due to noise and partial occlusion.



Figure 6. Accuracy and loss function vs. Training epoch graph for MobileNet.



Through our extensive experiment, it is found that MobileNet gives a better result than other CNNs. First of all, The proposed methodology is used non-contact advanced technology. Therefore, this method is better than conventional methods. Moreover, MobileNet has fewer parameters than others. As a result, MobileNet easily avoided overfitting problems but it is not assured for other CNNs. Also, its computation time is much smaller than the other off-the-shelf CNNs.

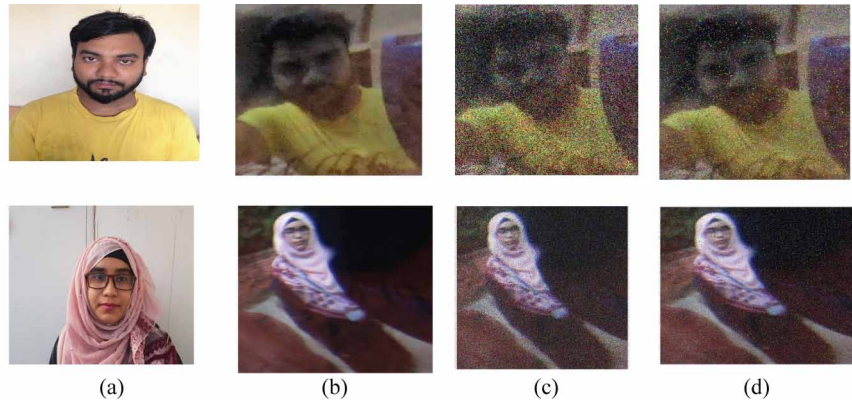
## 5. CONCLUSION

As interpersonal violence is one of the major threats to our society, attention is given to intervening through image processing techniques. This paper utilized transfer learning techniques to identify crime and recognize criminals. According to our experimental analysis, it is evident that the pre-trained CNN models, which are originally trained for other tasks of object recognition, can be effectively fine-tuned to identify criminals through transfer learning. We have done experiments with

**Table 8.** Recognition accuracy using our criminal dataset (with and without pre-processing) obtained after employing five different CNNs along with transfer learning. Here, the accuracy of the traditional method of Jenkis & Kerr (2013) is also shown for comparison, as this is the only method that uses corneal reflection images.

	Traditional Method (Jenkis and Kerr, 2013)	Percentage of Recognition by Neural Network-Based Methods			
		InceptionV3	NASNetMobile	VGG19	MobileNet
Without Pre-processing	-	74.44%	65.77%	80.77%	<b>84.27%</b>
With Pre-processing	84% (for familiar faces) & 71% (for unfamiliar faces)	86.93%	70.22%	93.61%	<b>95.41%</b>

**Figure 7.** Error-prone images of two subjects from the test dataset under different noise levels: (a) original image; (b) image captured from cornea (c) corrupted image by Gaussian noise with density,  $d = 0.7$ ; (d) corrupted image by Salt & Pepper noise with mean = 0 and variance,  $v = 0.7$ . (none of InceptionV3, MobileNet, VGG19 and NASNetMobile recognized this face)



different CNN models, such as VGG19, InceptionV3, MobileNet, and NasNetMobile. Among the 4 pre-trained models, MobileNet has attained the highest accuracy (95.41%) than other techniques. Nevertheless, there are still some challenges for applying deep learning to identify criminals using the corneal reflection images: (i) stressing on more robustness of the scheme (especially from the distant image capturing) to enhance its identification accuracy, (ii) integration of images from both eyes to enhance the quality of retrieved images as well as the identification accuracy, (iii) investigation on the practicability of an IoT-based tiny camera system to acquire and detect interpersonal violence by analyzing human emotions.

### 5.1. Informed Consent

Consent has been obtained from the six volunteers to use and publish their photographs only for research purposes.

### 5.2. Funding

No funding was received.

### 5.3. Conflict of Interest

The authors declare that this research work has no conflict of interest to anyone.

## REFERENCES

- Albu, R. D. (2009). Human face recognition using convolutional neural networks. *Journal of Electrical and Electronics Engineering (Oradea)*, 2(2), 110–113.
- Asadi, N., Jamzad, M., & Sajedi, H. (2008, August). Improvements of image-steganalysis using boosted combinatorial classifiers and Gaussian high pass filtering. In *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (pp. 1508-1511). IEEE.
- Bowyer, K. W., Hollingsworth, K., & Flynn, P. J. (2008). Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2), 281–307.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10(3), 243–248.
- Chang, J. Y., Lu, Y. F., Liu, Y. J., Zhou, B., & Qiao, H. (2020). *Long-distance tiny face detection based on enhanced YOLOv3 for unmanned system*. arXiv preprint arXiv:2010.04421
- Chen, W., Li, Y., & Li, C. (2020). A Visual Detection Method for Foreign Objects in Power Lines Based on Mask R-CNN. *International Journal of Ambient Computing and Intelligence*, 11(1), 34–47.
- Copeland, W. E., Wolke, D., Angold, A., & Costello, E. J. (2013). Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry*, 70(4), 419–426. PMID:23426798
- Dodge, S., & Karam, L. (2016, June). Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)* (pp. 1-6). IEEE.
- Garcia, C., & Delakis, M. (2002, August). A neural architecture for fast and robust face detection. In *Object recognition supported by user interaction for service robots* (Vol. 2, pp. 44–47). IEEE.
- Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545-552). Academic Press.
- Huang, H., Zhou, H., Yang, X., Zhang, L., Qi, L., & Zang, A. Y. (2019). Faster R-CNN for marine organisms detection and recognition using data augmentation. *Neurocomputing*, 337, 372–384.
- Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 650-657). IEEE.
- Jenkins, R., & Kerr, C. (2013). Identifiable images of bystanders extracted from corneal reflections. *PLoS One*, 8(12), e83325. PMID:24386177
- Jiang, Y., & Li, C. (2020). Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics*, 2020, 4152816.
- Kamencay, P., Benčo, M., Miždoš, T., & Radil, R. (2017). *A new method for face recognition using convolutional neural network*. Academic Press.
- Kar, A., & Corcoran, P. (2017). A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access: Practical Innovations, Open Solutions*, 5, 16495–16519.
- Khalajzadeh, H., Mansouri, M., & Teshnehlab, M. (2014). Face recognition using convolutional neural network and simple logistic classifier. In *Soft Computing in Industrial Applications* (pp. 197–207). Springer.
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 1–62.
- Khaire, M. G., & Shelkikar, R. P. (2013). Resolution Enhancement of images with interpolation and DWT–SWT Wavelet domain components. *International Journal of Application or Innovation in Engineering and Management*, 2.
- Kwolek, B. (2005, September). Face detection using convolutional neural networks and Gabor filters. In *International Conference on Artificial Neural Networks* (pp. 551-556). Springer.

Marciniak, T., Dabrowski, A., Chmielewska, A., & Weychan, R. (2012, May). Face recognition from low resolution images. In *International Conference on Multimedia Communications, Services and Security* (pp. 220-229). Springer.

Mayo Clinic. (2017). *Healthy Lifestyle Adult health*. Retrieved from <https://www.mayoclinic.org/healthy-lifestyle/adult-health/in-depth/domestic-violence/art-20048397>

Mukherjee, A., Mukherjee, P., Dey, N., De, D., & Panigrahi, B. K. (2020). Lightweight sustainable intelligent load forecasting platform for smart grid applications. *Sustainable Computing: Informatics and Systems*, 25, 100356.

Mukherjee, A., Panja, A. K., & Dey, N. (2020). *A Beginner's Guide to Data Agglomeration and Intelligent Sensing*. Academic Press.

Mukherjee, P., & Mukherjee, A. (2019). Advanced processing techniques and secure architecture for sensor networks in ubiquitous healthcare systems. In *Sensors for Health Monitoring* (pp. 3–29). Academic Press.

Nakazawa, A., Nitschke, C., & Nishida, T. (2016). Registration of eye reflection and scene images using an aspherical eye model. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 33(11), 2264–2276. PMID:27857447

Nakazawa, A., & Nitschke, C. (2012, October). Point of gaze estimation through corneal surface reflection in an active illumination environment. In *European Conference on Computer Vision* (pp. 159-172). Springer.

Nansel, T. R., Craig, W., Overpeck, M. D., Saluja, G., & Ruan, W. J. (2004). Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Archives of Pediatrics & Adolescent Medicine*, 158(8), 730–736. PMID:15289243

Nishino, K., & Nayar, S. K. (2006). Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1), 23–40.

Nitschke, C., & Nakazawa, A. (2012). Super-Resolution from Corneal Images. In *BMVC* (pp. 1-12). Academic Press.

Nitschke, C., Nakazawa, A., & Nishida, T. (2013, November). I see what you see: point of gaze estimation from corneal images. In *2013 2nd IAPR Asian Conference on Pattern Recognition* (pp. 298-304). IEEE.

Osadchy, M., Cun, Y. L., & Miller, M. L. (2007). Synergistic face detection and pose estimation with energy-based models. *Journal of Machine Learning Research*, 8(May), 1197–1215.

Ogawa, T., Nakazawa, A., & Nishida, T. (2018). Point of gaze estimation using corneal surface reflection and omnidirectional camera image. *IEICE Transactions on Information and Systems*, 101(5), 1278–1287.

Pei, Y., Huang, Y., Zou, Q., Zhang, X., & Wang, S. (2019). Effects of image degradation and degradation removal to CNN-based image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PMID:31689183

Rahman, T., Haque, M. R., Rozario, L. J., & Uddin, M. S. (2014, December). Gaussian noise reduction in digital images using a modified fuzzy filter. In *2014 17th International Conference on Computer and Information Technology (ICCIT)* (pp. 217-222). IEEE.

Sigut, J., & Sidha, S. A. (2010). Iris center corneal reflection method for gaze tracking using visible light. *IEEE Transactions on Biomedical Engineering*, 58(2), 411–419. PMID:20952326

Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3476-3483). IEEE.

Thrun, S. (1996). Is learning the n-th thing any easier than learning the first? In *Advances in neural information processing systems* (pp. 640-646). Academic Press.

Wan, R., Shi, B., Li, H., Duan, L. Y., & Kot, A. C. (2020). Reflection Scene Separation From a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2398-2406). Academic Press.

Wu, X., He, R., Sun, Z., & Tan, T. (2018). A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11), 2884–2896.

Xinhua, L., & Qian, Y. (2015). Face recognition based on deep neural network. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(10), 29–38.

Yip, A., & Sinha, P. (2001). *Role of color in face recognition*. Academic Press.

Zhang, N., Paluri, M., Ranzato, M. A., Darrell, T., & Bourdev, L. (2014). Panda: Pose aligned networks for deep attribute modeling. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637-1644). IEEE.

Mohammad Reduanul Haque received his Master of Science and Bachelor of Science degrees in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka in 2012 and 2011, respectively. He is currently served as a Senior Lecturer in the Department of Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh. His research interests includes Computer Vision, Deep Learning and Image Processing.

Rubaiya Hafiz received her B.Sc. and M. Sc. in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh. She is currently working as a Senior Lecturer at Daffodil International University, Dhaka, Bangladesh. Her research interests include Artificial Intelligence, Machine Learning, Image Processing and HCI.

Mohammad Shorif Uddin completed his Doctor of Engineering (Ph.D.) at Kyoto Institute of Technology in 2002, Japan, Master of Technology Education at Shiga University, Japan in 1999, Bachelor of Electrical and Electronic Engineering at Bangladesh University of Engineering and Technology (BUET) in 1991 and also Master of Business Administration (MBA) from Jahangirnagar University in 2013. He began his teaching career as a Lecturer in 1991 at Chittagong University of Engineering and Technology (CUET). In 1992, he joined the Computer Science and Engineering Department of Jahangirnagar University and at present, he is a Professor of this department. In addition, he is an Adviser of the School of Science and Engineering, ULAB, and Teacher-in-Charge of ICT Cell of Jahangirnagar University. He served as the Chairman of the Computer Science and Engineering Department of Jahangirnagar University from June 2014 to June 2017. He undertook postdoctoral researches at Bioinformatics Institute, Singapore, Toyota Technological Institute, Japan and Kyoto Institute of Technology, Japan, Chiba University, Japan, Bonn University, Germany, Institute of Automation, Chinese Academy of Sciences, China. His research is motivated by applications in the fields of artificial intelligence, imaging informatics, and computer vision. He holds two patents for his scientific inventions and has published extensively in international journals and conference proceedings. He had delivered a remarkable number of keynotes and invited talks and also acted as a General or TPC Chair or Co-Chair of many international conferences. He received the Best Paper award in the International Conference on Informatics, Electronics & Vision (ICIEV2013), Dhaka, Bangladesh, and Best Presenter Award from the International Conference on Computer Vision and Graphics (ICCVG 2004), Warsaw, Poland. He was the Coach of Jahangirnagar University ACM ICPC World Finals Teams in 2015 and 2017 and supervised a good number of doctoral and Master theses. He is the Editor-in-Chief of ULAB Journal of Science and Engineering and an Associate Editor of IEEE Access.