

A New Framework for Matching Forensic Composite Sketches With Digital Images

Chethana H. T., Vidyavardhaka College of Engineering, *Mysuru, India*

Trisiladevi C. Nagavi, S. J. College of Engineering, JSS Science and Technology University, Mysore, India

ABSTRACT

Face sketch recognition is considered as a sub-problem of face recognition. Matching composite sketches with its corresponding digital image is one of the challenging tasks. A new convolution neural network (CNN) framework for matching composite sketches with digital images is proposed in this work. The framework consists of a base CNN model that uses swish activation function in the hidden layers. Both composite sketches and digital images are trained separately in the network by providing matching pairs and mismatching pairs. The final output resulted from the network's final layer is compared with the threshold value, and then the pair is assigned to the same or different class. The proposed framework is evaluated on two datasets, and it exhibits an accuracy of 78.26% with extended-PRIP (E-PRIP) and 69.57% with composite sketches with age variations (CSA) respectively. Experimental analysis shows the improved results compared to state-of-the-art composite sketch matching systems.

KEYWORDS

Convolution Neural Network (CNN), Dropout, E-PRIP Dataset, Exponential Linear Unit, Face Sketch Recognition, Leaky Rectified Linear Unit, Sigmoid, Swish Activation Function

INTRODUCTION

Face recognition is a problem that is being solved from decades by now. It is one of the most challenging tasks and plays a very significant role in forensics. Another important subtask of face recognition problem is identifying a suspect by a sketch. In crime scenarios, the suspect's photo is not available in most cases. Only the eye-witness description can be relied on for the identification of the victim or the criminal. A sketch of the suspect's face can be drawn by a forensic artist (Wang and Tang.,2008) or even generated by facial software from the details provided by the eye-witness (Han et al., 2012). After the sketch is generated, the next important step is to identify who that person is in the sketch. This is performed by matching the sketch to the photos of mugshots/suspects in the criminal database. The automation of this process is useful in identifying the suspects in less time (Cheragi and Lee., 2019).

Numerous challenges can occur while matching a sketch and a photo. While drawing a sketch various complementary information such as hair color, skin color, and ethnicity may not be noticeable. Another critical challenge is the phenomenological gap between a sketch and a photo, such as illumination, facial expressions, color background, brightness, etc. In addition to this, only a single sketch is available unlike face recognition. Therefore, there is a chance that can lead to the misidentification of the suspect (Peng et al., 2018). Information from other sources like a description

DOI: 10.4018/IJDCF.20210901.oa1

This article, published as an Open Access article on July 2nd, 2021 in the gold Open Access journal, theInternational Journal of Digital Crime and Forensics (converted to gold Open Access January 1st, 2021), is 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.

from the multiple eye-witnesses and surveillance camera footage can be used to improve the performance of suspect identification (Best-Rowden et al., 2014).

Forensic sketches can be categorized into three types such as viewed, forensic and composite sketches. The viewed sketch is drawn by looking at a photo, and no description is provided for drawing these types of sketches. Forensic sketches are drawn by forensic artists hearing the description provided by the eye-witnesses. These sketches are used for investigation since 19th century. Composite sketches are generated using software tools such as evoFIT, FACES etc. Here different facial parts are selected to draw the composite sketch. Almost 80% of law enforcement agencies use software oriented facial suspects to identify the suspects (Klum et al., 2013). Comparing the three types of sketches, composite sketches require less time, effort and experience. So composite sketches are used in the proposed framework to identify the corresponding digital images.

In order to provide identification based on composite sketch, many artificial neural network (ANN) models are proposed. CNN (Zhang, H. et al., 2019) are also a category of ANN model which are represented in a fully connected manner and thus avoids over fitting of the data. So a new CNN framework is proposed in this work to identify matching pairs and mismatching pairs of a composite sketch and a digital image. A CNN model uses Swish as the activation function in the hidden layers. The model is tested with multiple activation functions and the Swish activation function proved to provide the best performance among all other activation functions. The input provided to the system is a pair of composite sketch and a digital image. The model trains both the images and an absolute difference is taken as final value between composite sketch F_s and digital image F_d . The final output matching score obtained from the network is compared with the threshold value and decision is made to identify whether the input pair belongs to the same class or different class.

The key contributions of this work are stated below:

- A new framework is proposed to perform matching of composite sketches and digital images.
- To classify whether an input pair of composite sketch and a digital image belongs to the same or different class, a new CNN framework is proposed.
- A detailed analysis of the proposed framework is provided.
- The result achieved from the proposed framework outperforms many state-of-the-art face sketch recognition systems.

The organization of the paper is as follows. A brief discussion of previously developed face sketch recognition systems in the form of literature survey is provided in section 2. The dataset description is provided in section 3. The methods used for developing the proposed face sketch recognition system are discussed in section 4. Section 5 provides detailed analysis and results obtained from the proposed framework. The section 6 presents the conclusion of the paper.

LITERATURE SURVEY

Many systems have been developed to perform face sketch recognition. Some of the traditional methods use features such as Eigen and Local Binary Patterns, Histogram oriented gradient (HOG) and scale-invariant feature transform (SIFT) feature representations. A windows based real-time application was proposed to perform face recognition by Haji and Varol (2016). The method used feature representations extracted from algorithms such as Eigen and Local Binary Patterns which eliminates the light exposure impact on the facial images. Further, an algorithm was proposed by Mittal et al. (2017) which performs the combinations of various feature representations by considering the facial attributes such as gender, ethnicity and color of the skin to develop a good classifier. The feature combination includes the fusion of DAISY and HOG features.

To perform a composite sketch to photo matching, a component-based framework was proposed by Han et al. (2012). An active shape model (ASM) was used to detect facial landmarks. Multiscale Local Binary Pattern (MLBP) was used to extract the features from the images. These features are then matched and fused for obtaining matching results. Liu et al. (2016) developed a composite sketch based face recognition method by performing feature extraction on two images, fusing these two feature representations and then combining the facial parts with different weights. A novel technique for matching composite sketch with the images captured by unmanned aerial vehicle under uncontrolled environment was proposed by Fernandes et al., (2015). Here image quality assessment is performed to detect whether image captured from unmanned aerial vehicle has varying conditions. Suitable feature extraction method and classification approach is applied to perform mapping of composite sketch to digital images.

A deep learning based architecture was proposed by Galea et al., (2017) in which transfer learning is applied to both sketches and photos. Proposed framework reduces the error rate by 80.7% for viewed sketches and reduces the retrieval rank by 32.5% for forensic sketches. A robust framework called local feature based discriminant analysis was proposed by Klare et al., (2010). The proposed framework outperforms the other existing approaches in case of matching viewed sketches. An evolutionary genetic algorithm was proposed by Bansode et al., (2016) which generates face sketch from face description.

A composite sketch recognition system was proposed by Shivaleela Patil and Shibhangi (2020) stating composite sketches provide better results than forensic sketches in face recognition. The methods used are Weber Local Descriptor (WLD), MLBP, Tchebichef moments and ANN. In recent trends, there is a lot of interest is developed in deep learning methods proposed by Gadekallu T. R et al.(2020) for performing face sketch recognition. A Sketch Photo Network (SP-Net) was proposed by Cheragi and Lee (2019). Here network uses VGG-Face net as the base model, which is followed by two branches S-Net and P-Net. The network provided good results by comparing the matching and mismatching pairs of sketches and photos.

A coupled deep neural network architecture was proposed by Iranmanesh et al. (2018), which considers facial attributes such as eye, hair, skin color and ethnicity. A joint loss function was also introduced which performs the identification of facial attributes and also verifies the subspace of a sketch and a photo. The main limitation of using state-of-the-art methods for sketch- photo recognition is the unavailability of a huge dataset. To tackle this limitation, transfer learning can be applied by using pre-trained models. This method was adopted by Galea and Farrugia (2017) to perform face sketch recognition. To avoid the over-fitting of the model, 3-Dimensional morphable software was used for synthesizing new images artificially.

A deep convolution network was used by Kazemi et al., (2018) to perform sketch photo recognition. The model was used to learn deep shared features (R M et al.,2020) between a sketch and a photo. An attribute-centered loss function was proposed in their work for training the network. The network matched the facial attributes of the sketch and the photo. After comparing the previously developed face sketch recognition systems, it is found that the deep learning model's (Venkatraman S., et al., 2019) performance is better than the traditional models. Hence in proposed system, a new CNN framework is proposed to perform face sketch recognition. A brief summary on different methods applied for forensic face recognition is illustrated in the Table 1.

PROPOSED METHODOLOGY

A face sketch recognition system is developed in the proposed work. It consists of three stages: pre-processing, classification and output. The input provided to the proposed framework is the pair of composite sketch and its corresponding digital images. The steps involved in the pre-processing of these images are mentioned in the image pre-processing section. The process of collecting composite sketches is a very challenging and tedious task. Hence the input images are limited. Therefore, to get

Table 1. A Summary on different methods for forensic face recognition

| Author | Methodology | Advantages |
|-------------------------------|--|---|
| Haji et al.(2016) | Eigen with Local Binary Patterns | Eliminates the light exposure on facial images |
| Mittal et al.(2017) | Combination of DAISY and HOG features | Facial attributes such as gender, ethnicity and color of the skin were considered |
| Han et al.(2012) | Combination of active shape model with multi scale local binary pattern | Here features were fused to get the matching results |
| Shivaleela Patil et al.(2020) | Weber Local Descriptor (WLD), MLBP, Tchebichef moments and Artificial Neural Network (ANN) | Provides better results for composite sketches rather than forensic sketches |
| Iranmanesh et al.(2018) | Joint Loss function | Facial attributes such as eye, hair, skin color and ethnicity were considered |
| Kazemi et al.(2018) | Attribute centered loss function | The model used here will learn the deep features between sketch and photo |

a better performance of the system, more input images are necessary. To tackle this issue, the images are heavily augmented in the training phase. Therefore, the network learns the input images well.

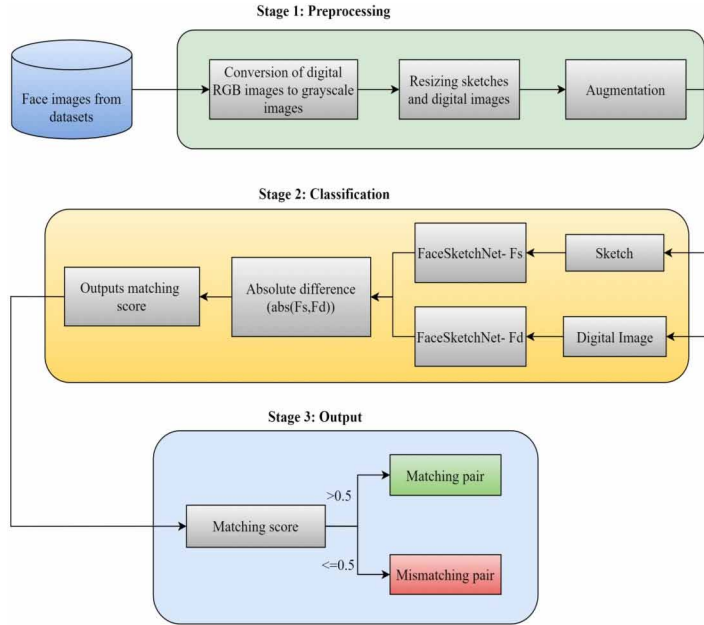
The most important section of the architecture is the identification of the matching face for the sketch provided. Therefore, in the classification section, the composite sketches and digital images are trained separately in proposed framework. The absolute difference is computed between the composite sketch F_s and digital image F_d which is treated as outputs matching score. The matching score obtained in the classification stage is compared with the threshold value in the output stage. The images are trained in terms of a matching pair and mismatching pair. The threshold value decides whether the given input images belong to matching pair and mismatching pair. The schematic illustration of the proposed framework is shown in Figure 1.

Dataset

In the proposed face sketch recognition framework, two datasets are selected. The first dataset used is E-PRIP dataset (Mittal et al., 2014). It consists of 123 composite sketches and its respective digital images. The digital images used in the E-PRIP dataset are taken from AR dataset (Wang and Tang, 2008). The original PRIP dataset was developed by (Han et al., 2017) and the E-PRIP dataset was developed by (Mittal et al., 2014). The dataset is created by four different sets created by multiple users. Using FACES software, an American artist created one set. Two sets were created by an Asian artist using FACES and Identi-kit software (Identi-kit, 2020). Another set is created by an Indian artist using FACES software that is the contribution in the extended dataset (Mittal et al., 2014).

The second dataset used for the evaluation of the proposed work is Composite Sketch with Age Variation dataset collected from IIIT, Delhi (Chugh et al., 2017). The dataset consists of multiple age-separated digital images for a sketch image. There are a total of 3529 sketches and digital images available in the dataset collected from 150 individuals. The dataset is collected from multiple sources, which includes 52 from FG-NET Aging dataset (Lanitis, 2008), 82 from IIIT-D Aging Dataset (Yadav et al., 2014) and the rest subjects are taken from the Internet, and their corresponding composite sketches are generated from FACES software (Faces, 2003). The dataset consists of three groups of different age variations ranging from 1 to 65 years namely younger age, same age and older age. In proposed framework, only the same age group individual digital images and sketches are considered.

Figure 1. Block diagram of proposed framework



The number of digital images, along with their corresponding composite sketches in the dataset is 98(belonging to same age variation category). The illustration of sample images from both the datasets are shown in Figure 2.

Image Pre-Processing

The images provided in the datasets may vary in size, colour format etc. This may affect the performance of the recognition system (Bhattacharya et al., 2020). Therefore, it is necessary to bring all these images in to common uniform format. The digital images in the dataset are provided in RGB format. However, the composite sketches are in grayscale format. Hence, all digital images are first converted into grayscale images. Both digital images and sketches are resized into a dimension of 32 x 32. The illustration of the images in the datasets and the pre-processed images is shown in Figure 3.

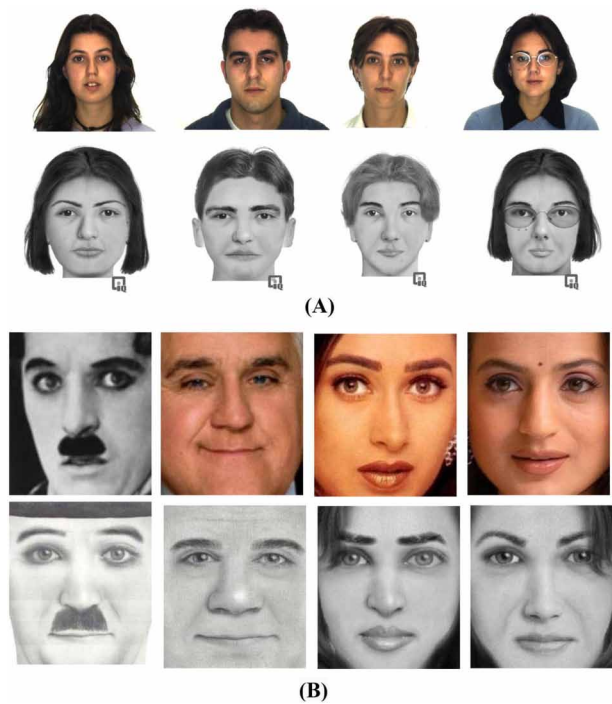
To train a deep model (Alazab et al., 2020) the dataset must be large. But, the dataset available for composite sketch matching is very less. Therefore, to generate more number of images, augmentation must be performed. In the proposed framework, the sketch images are augmented using multiple operations such as horizontal flipping, vertical flipping, image transformation by scaling the images by -10% or +10%, and rotating the images by +20 degrees or -20 degrees. The augmentation of the images is performed only for the training and validation phase.

Classification

The main aim of the classification network proposed in this work is to estimate whether a given sketch matches the same individual or not. The network was evaluated on multiple CNN layer configurations and also fine tuned using various parameters. The results obtained from proposed framework are described in this section. The architecture diagram of the proposed model is illustrated in Figure 4.

The proposed model is a six layer CNN. Two sets of image pairs are provided to the proposed model, namely, a matching pair and a mismatching pair. The matching pair is a composite sketch with its corresponding digital image and the mismatching pair is the composite sketch, which has a

Figure 2. Illustration of composite sketches and digital images from: (a) E-PRIP dataset (b) CSA dataset



digital image of a different individual. Composite sketches and digital images are trained separately by proposed model. The network learns feature representations that discriminate the composite sketches from the digital images. Feature visualization is the process of visualizing each layers in neural networks and see how they work. The proposed model is a six layer CNN. CNN are kind of deep neural networks which operates on two dimensional image data which preserves the spatial relationship based on the learning performed by the model. Filters and activation maps are used to perform feature visualization. Filters are nothing but the weights which can be used to detect the type of features that the proposed model is using and activation maps are used to detect the features generated from the given input image.

Figure 3. Illustration of sample images before and after pre-processing step (before augmentation)

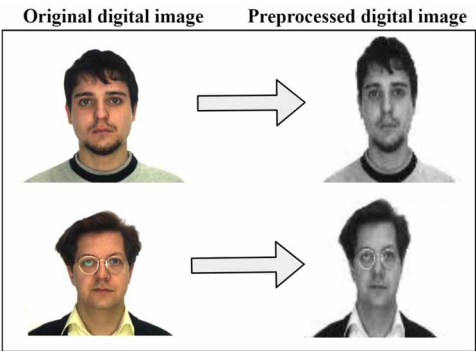
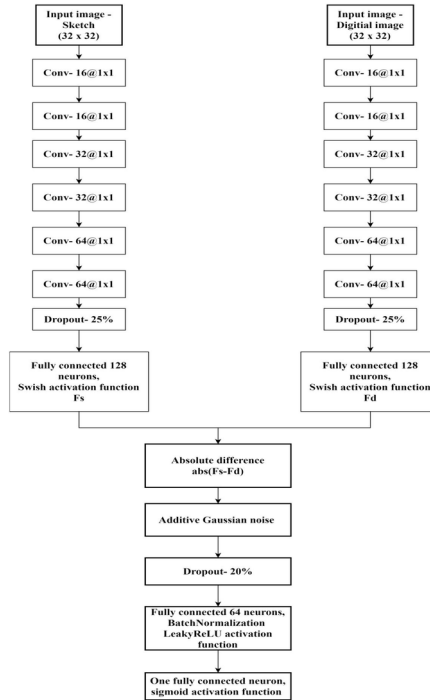


Figure 4. Architecture diagram of proposed framework for matching composite sketches and digital images



CNN model is a good model for feature visualization because it possess a uniform structure with a combination of convolution, pooling, flatten and dense layers.

Steps to Perform Feature Visualization in the Proposed Model

- Review the filters in the proposed model by specifying the layer details including layer name and shape of filters in the layer.
- Layers in proposed model are well defined and organized in to blocks as shown in the Figure 5.
- Given input image is provided with three channels red, green and blue. Each filter can be visualized as a plot with three images.
- Retrieve the filters from each layer. The weight values should be normalized in the range 0-1 which makes the process of visualization easy. Plotting of six filters for each of the layers from proposed model is shown in the Figure 6.

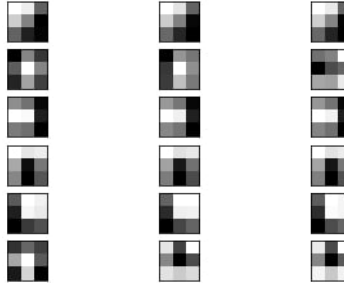
The absolute difference between the matching and mismatching pairs is adopted in the network to find the variant visual structures in the sketch and the digital image. Based on the final output resulting in the network's final layer compared with the threshold value, the pair is assigned to the same or different class.

The proposed model consists of convolution blocks, max pooling layer and hidden convolution layer. A new activation function named “Swish” (Ramachandran et al., 2017) is used in the hidden convolution layer. The reason for choosing the Swish activation function is stated in the upcoming section. The formula of the Swish activation function is represented in Equation 1.

Figure 5. Layers in proposed model organized in to blocks

| Layer (type) | Output Shape | Param # | Connected to |
|----------------------------------|-------------------|---------|-------------------------------------|
| sketch (InputLayer) | (None, 1, 32, 32) | 0 | |
| digital_image (InputLayer) | (None, 1, 32, 32) | 0 | |
| model_1 (Model) | (None, 512) | 1470432 | sketch[0][0] digital_image[0][0] |
| merge_1 (Merge) | (None, 512) | 0 | model_1[1][0] model_1[2][0] |
| gaussiannoise_1 (GaussianNoise) | (None, 512) | 0 | merge_1[0][0] |
| dropout_7 (Dropout) | (None, 512) | 0 | gaussiannoise_1[0][0] |
| dense_3 (Dense) | (None, 256) | 131328 | dropout_7[0][0] |
| dense_4 (Dense) | (None, 64) | 16448 | dense_3[0][0] |
| dense_5 (Dense) | (None, 32) | 2080 | dense_4[0][0] |
| batchnormalization_2 (BatchNorma | (None, 32) | 128 | dense_5[0][0] |
| leakyrelu_1 (LeakyReLU) | (None, 32) | 0 | batchnormalization_2[0][0] |
| dense_6 (Dense) | (None, 1) | 33 | leakyrelu_1[0][0] |
| activation_3 (Activation) | (None, 1) | 0 | dense_6[0][0] |
| Total params: 1,620,440 | | | |
| Trainable params: 1,619,361 | | | |
| Non-trainable params: 1,008 | | | |

Figure 6. Plot of six filters from the proposed model



$$f(x) = x * \text{sigmoid}(x) = x * \frac{1}{(1 + e^{-x})} \quad (1)$$

The dimension of the both the input images are 32 x 32. The network architecture consists of six convolutional layers with varying filter sizes. The size of the filters at first two layers is 16, next two layers are 32 and last two layers are 64. The image size (m) after convolution is represented in Equation 2 where n*n denotes image size and f*f denotes filter size.

$$m = ((n - f) + 1) * ((n - f) + 1) \quad (2)$$

The convolution operation can be stated as the dot product of the weights in the filter and the input image. The feature representations from the input images are learnt in the convolution layers of the network.

The absolute difference of the output obtained from the training of composite sketch (Fs) and digital image (Fd) is taken. This can be represented in Equation 3.

$$\text{Absolutedifference} = \text{abs}(F_s - F_d) \quad (3)$$

This output is again provided to the Gaussian noise layer in order to make the model more robust. Gaussian noise layer is a regularization layer which takes only real values and also adds additional noise to the input. This property of adding noise to the input makes the network more robust by not over fitting the input. Dropout layer is also a regularization layer. As the name suggests, the dropout

layer randomly drops certain nodes and trains only the remaining nodes. This property of dropout layer, does not learn the input completely and avoids the over fitting of the model.

In the proposed model we have considered dropout as 20% to 25% i.e., out of 100 neurons it will drops 20 to 25 neurons which are redundant and passes the data of remaining 75 neurons to the next layer. In the final layer, the sigmoid activation function is used to generate a score. If the score generated is greater than the threshold value 0.5, the input pair is assigned as the same class, else if the score generated is less than or equal to the threshold value 0.5, the input pair is assigned as a different class.

RESULTS AND DISCUSSION

In the proposed framework, the face sketch recognition system is designed. As mentioned earlier, two composite sketch datasets are used for evaluating the proposed framework. They are E-PRIP and CSA datasets. The system is trained by providing input pairs of sketch and a digital image. The datasets are split into training, validation and testing. The split of the datasets for E-PRIP and CSA are stated in the Table 2 and 3. It can be observed the number of pairs for training the system is less for a deep model. Therefore, both training and validation data is heavily augmented to learn different variations of sketches and digital images which is explained in image pre -processing section.

The use of activation function in the hidden layers of the activation function plays an important role in the network's performance. Therefore, the activation function must be chosen carefully. In this work, three different activation functions are tested for obtaining the best performance. The activation functions considered for the hidden layers in the network are Exponential Linear Unit (ELU) (Clevert et al., 2015), Leaky Rectified Linear Unit (LeakyReLU) (Mittal et al., 2015), and Swish activation function [(Ramachandran et al., 2017).

ELU activation function is one of the activation functions that tend to converge faster which produces better and accurate results. Compared to the ReLU activation function, ELU results in a smooth curve. It accepts the negative inputs as well, unlike ReLU. Therefore, ELU is considered in the hidden layers of the network. LeakyReLU activation function is a variant of ReLU activation function. In this, a small negative slope is added to overcome the “dying ReLU” issue. It also proved to provide better results in this work. However, the use of Swish activation functions in proposed model proved to provide the best results for both the datasets. Swish activation function results in a smooth function; that is, there is no abrupt change in the direction. Therefore, this activation function

Table 2. Dataset split of E-PRIP dataset

| Dataset split | #Pairs |
|---------------|--------|
| Training | 75 |
| Validation | 25 |
| Testing | 23 |

Table 3. Dataset split of CSA dataset

| Dataset split | #Pairs |
|---------------|--------|
| Training | 65 |
| Validation | 10 |
| Testing | 23 |

is chosen for the final network in the proposed framework. The performance of different activation functions on the two datasets is shown in Figure 7.

The performance of the two datasets obtained is analyzed through various metrics. Confusion matrix is one of the metric which describes the performance of classification model based on test data for which true values are known. It is also used to calculate True Positive (TP), True Negative (TN), False Positive (FP) and False Negative(FN). The confusion matrices obtained for E-PRIP and CSA datasets are shown in the Figure 8 and Figure 9.

It can be observed from the matrices that the proposed framework provides a Rank-1 accuracy of 78.26% on E-PRIP dataset and 69.57% on CSA dataset. From the confusion matrix, various performance metrics can be calculated to evaluate the system. Those performance metrics considered for evaluating the proposed framework are accuracy, precision, Recall (True Positive Rate-TPR), F1-score and False Positive Rate(FPR). The results of different performance metrics obtained for both the datasets is tabulated in Table 4.

Figure 7. Evaluation of proposed framework with different activation functions in the hidden layers of the network

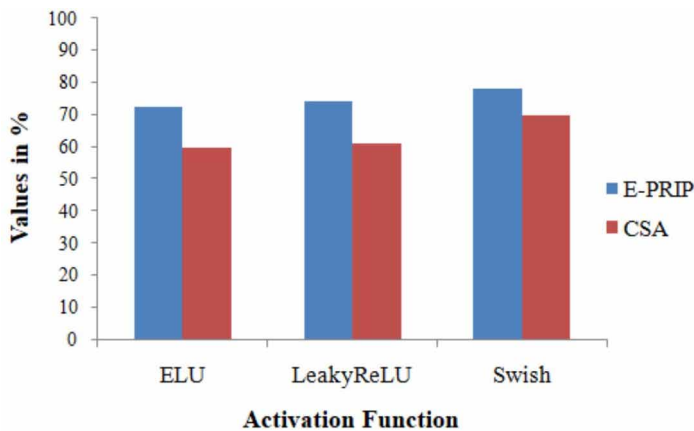


Figure 8. Confusion matrix obtained on testing data for E-PRIP dataset

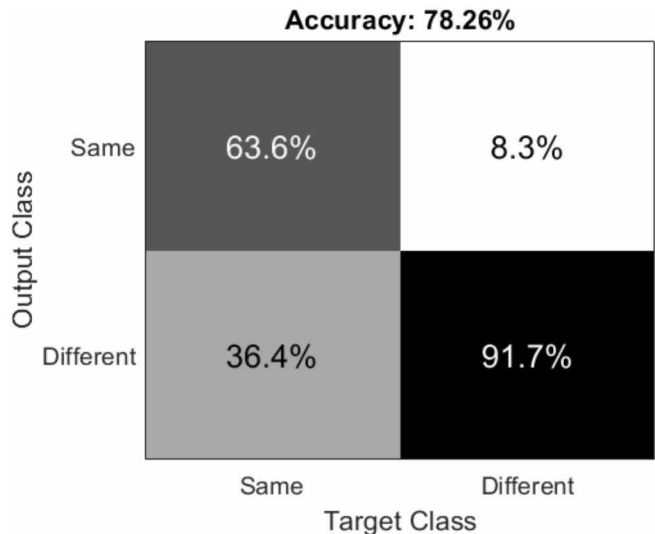
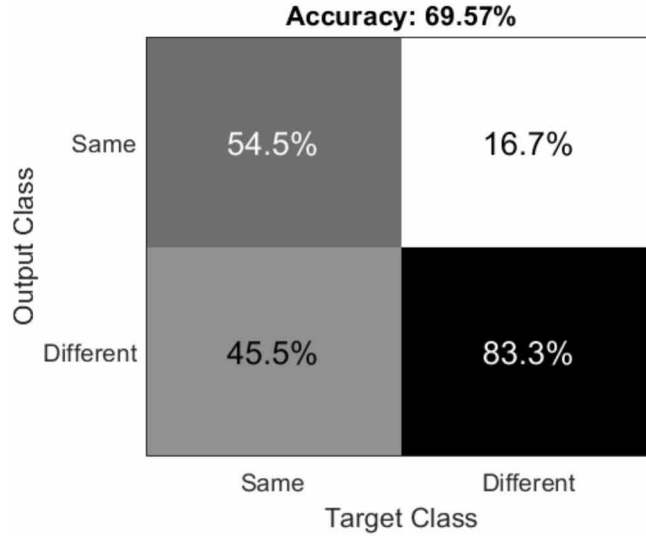


Figure 9. Confusion matrix obtained on testing data for CSA dataset



Accuracy can be defined as closeness of measurement to a specific value which is represented in Equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precision can be defined as the measure of exactness which is represented in Equation 5.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall can be defined as percentage of certain class that is correctly being identified which is represented in Equation 6.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Table 4. Performance evaluation of the proposed framework on E-PRIP and CSA datasets

| Performance Measures | E-PRIP | CSA |
|----------------------|--------|--------|
| Accuracy | 78.26% | 69.57% |
| Precision | 63.64% | 54.55% |
| Recall (TPR) | 87.50% | 75% |
| F1-score | 73.68% | 63.16% |
| FPR | 0.2667 | 0.3333 |

F1- score is a score which balances between precision and recall which is represented in Equation 7.

$$F1 - score = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (7)$$

False Positive rate (FPR) is the probability of falsely rejecting the hypothesis which is represented in Equation 8.

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

The performance measures on E-PRIP and CSA dataset is shown in the Figure 10. In this plot, values in X- axis denotes different performance measures like accuracy, precision, recall(TPR), F1-score, FPR and values in Y-axis denotes the results for both the datasets.

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the performance of classification model. In this plot, a value in X-axis indicates False Positive Rate (FPR) and values in Y-axis indicates True Positive Rate (TPR). The ROC curve is plotted based on TPR and FPR obtained from the confusion matrix for both the datasets is shown in the Figure 11 and Figure 12.

The performance of the system is compared to other face sketch recognition systems. The results achieved in the proposed framework outperform the previous recognition systems. The comparison of the proposed framework with other face sketch recognition systems for both the datasets is indicated in Table 5 and Table 6.

The performance analysis of existing approaches and proposed framework on both the datasets is illustrated in Figure 13 and Figure 14. In both the plots, values in X-axis indicates the names of different methods and values in Y-axis indicates the recognition accuracy.

From the comparative study it is observed that proposed framework gives the Rank 1- accuracy of 78.6% on E-PRIP dataset and 69.57% on CSA dataset and it outperforms the other existing approaches. An observation can be made that the proposed framework resulted in good performance compared to many previously developed face sketch recognition systems and outperform many state-of-the-art face sketch recognition systems (Cheragi et al.,2019, Mittal et al.,2015, Iranmanesh et al.,2018, Mittal et al.,2017, peng et al.,2019).

The proposed model consists of a new framework which includes training of the composite and digital images with two individual networks and by the difference obtained from the two networks, the pair of the digital image and composite image is assigned to a matching or non-matching class. The use of this type of framework is not performed in any of the previous works and also the use of Swish activation function in the network is a first for composite sketch matching. In the previous works, networks such as coupled deep learning model, deep belief networks, combination of saliency features and texture features etc. are used. The accuracies obtained from the previous works are comparatively less from the accuracy obtained for the proposed model.

In the proposed model, the digital and the composite images are fed in to two different CNNs. This mechanism allows the learning of information of the two sets of images individually. The use of Swish activation function provided better results as compared to ELU and LeakyRelu activation functions. Swish function is an unbounded value activation function, i.e., there is no bound on the maximum value generated from the hidden layers. Swish activation function proved to outperform other activation function with varied batch sizes. In addition to this, the proposed model is robust as it outperformed even when additional noise is incorporated in the images. Considering all the above reasons, the proposed model is better than other previously proposed systems for composite sketch matching.

Figure 10. List of performance measures on E-PRIP and CSA dataset

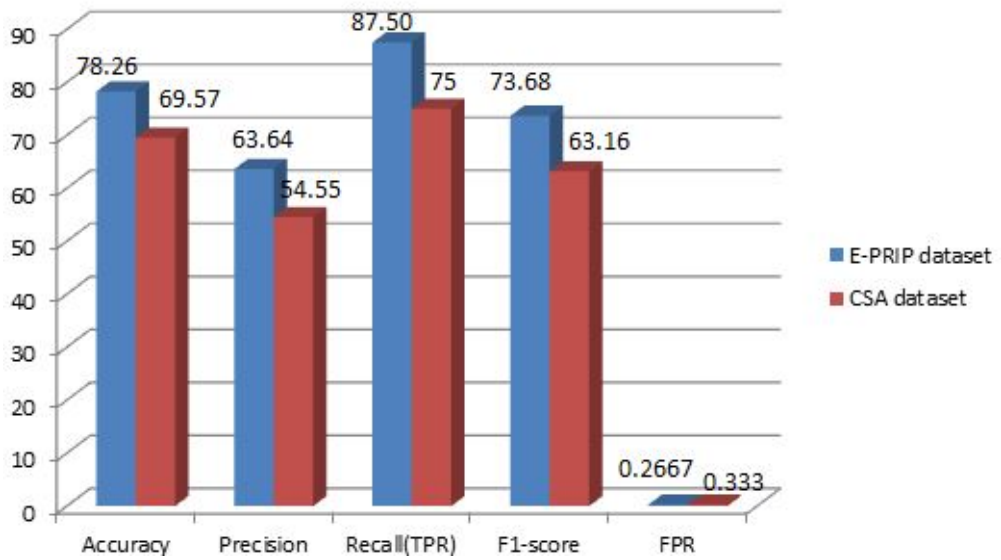


Figure 11. ROC curve for E-PRIP dataset

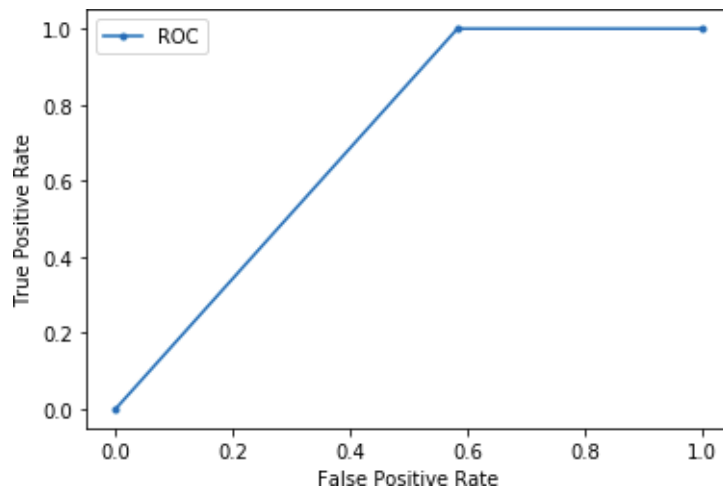


Figure 12. ROC curve for CSA dataset

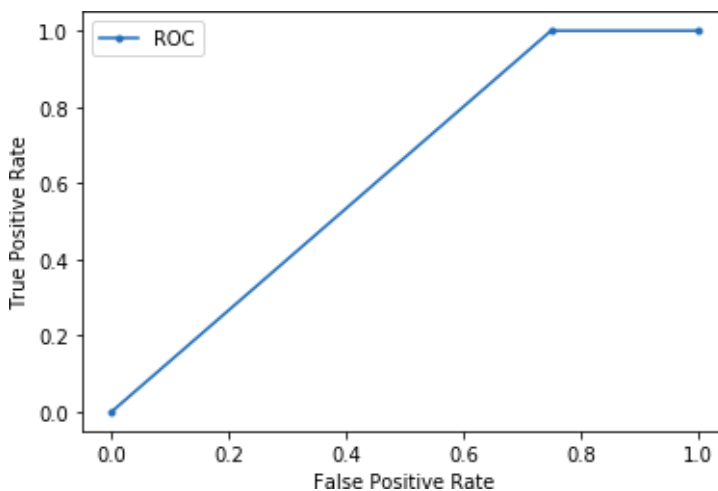


Table 5. Comparison of previous works with the proposed work for face sketch recognition on E-PRIP dataset

| Authors | Methods used | Rank-1 Accuracy |
|-----------------------------|---|-----------------|
| Cheragi. (2019) | SP-Net | 28.1% |
| Mittal et al. (2015) | Deep Belief Network | 58% |
| Iranmanesh et al. (2018) | Coupled deep neural network | 76.4% |
| Mittal et al. (2017) Visual | Visual saliency and combination of texture features | 58.4% |
| Peng et al. (2019) | Sparse Graphical representation based discriminant analysis | 70% |
| Proposed framework | Swish Activation function with 6 layer CNN | 78.6% |

Table 6. Comparison of previous works with the proposed work for face sketch recognition on CSA dataset

| Authors | Methods used | Rank-1 Accuracy |
|-------------------------|--|-----------------|
| Paritosh M et al.(2015) | Single Shot Detector (SSD) with Gentle Boost KO classifier | 58.6% |
| Roy et al. (2019) | RBPLQ with neural networks | 60.72% |
| Proposed framework | Swish Activation function with 6 layer CNN | 69.57% |

Figure 13. Performance analysis of existing approaches and proposed framework on E-PRIP dataset

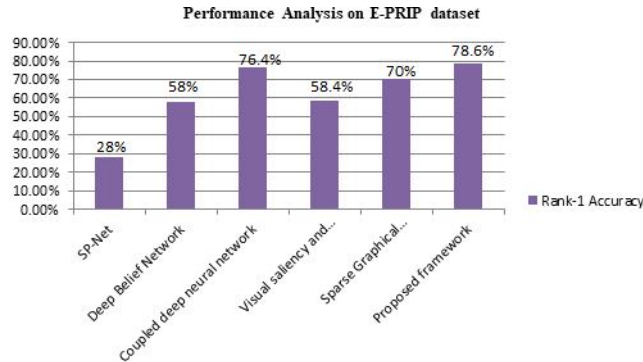
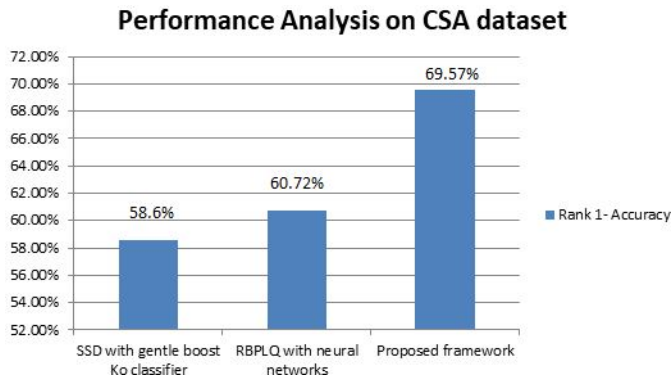


Figure 14. Performance analysis of existing approaches and proposed framework on CSA dataset



CONCLUSION

A new framework for performing face sketch recognition is proposed in this work. A CNN model using swish activation function in the hidden layers of the network is used for training the sketch and digital image. The network is separately trained for matching and mismatching pairs. An absolute difference between the output generated from the sketch and the digital image is adapted to find the variant visual structures. Gaussian noise is added to the network to make the network more robust. The proposed framework extracts discriminative facial features from the sketch and photo. The absolute difference is used in the network to learn the similarity or dissimilarity in the matching and mismatching pairs of sketch and a digital image. The network's final layer provides a score that is compared with a threshold value to assign it to the same class or different class for the matching and mismatching pairs. To prove the reliability of the proposed framework, the framework is tested on two datasets. The model resulted in state-of-the-art accuracy for E-PRIP and CSA datasets (Cheragi et al., 2019, Mittal et al., 2015, Iranmanesh et al., 2018, Mittal et al., 2017, peng et al., 2019). The framework resulted in promising results by outperforming most of the state-of-the-art face sketch recognition systems which can be adopted in real-time application as well. The proposed model's performance can be improved by experimenting with multiple light-weight CNN models and also by increasing the size of the dataset. Combination of different neural network architecture can be performed to improve the system's performance. However, the datasets can be combined to increase the number of images thereby improving the performance of the proposed model. The limitation of this work is

that the performance of the proposed model can still be improved by testing on larger datasets. The deep networks' performance increases with more data. Therefore, if more number of images is given to the network, the performance of the proposed model can be improved. In the future, the advanced neural network models can be used for training the sketches and digital images with larger datasets.

ACKNOWLEDGMENT

Authors would like to thank the researchers Chugh T, Singh M, Nagpal S, Singh R, Vatsa M, Noore A, Majumdar A, 2017 and Martinez A M, Benavente R, 1998 for providing their datasets (composite sketch with age variation and AR face database) to carry out the experimentation.

REFERENCES

- Alazab, M., Huda, S., Abawajy, J., Islam, R., Yearwood, J., Venkatraman, S., & Broadhurst, R. (2014). A hybrid wrapper-filter approach for malware detection. *Journal of Networks*, 9(11), 2878-2891.
- Alazab, M., Suleman, K., & Somayaji, S. (2020). A Multidirectional LSTM Model for Predicting the Stability of a Smart Grid. *IEEE Access: Practical Innovations, Open Solutions*, 8, 85454–85463. doi:10.1109/ACCESS.2020.2991067
- Bansode, N. K., & Sinha, P. K. (2016). Face Sketch generation using evolutionary computing. *International Journal of Soft Computing*, 7(4), 01–10. Advance online publication. doi:10.5121/ijsc.2016.7401
- Best-Rowden, L., Han, H., Otto, C., Klare, B. F., & Jain, A. K. (2014). Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, 9(12), 2144–2157. doi:10.1109/TIFS.2014.2359577
- Bhattacharya, S., Kaluri, R., Singh, S., Alazab, M., & Tariq, U. (2020). A Novel PCA-Firefly based XGBoost classification model for Intrusion Detection in Networks using GPU. *Electronics (Basel)*, 9(2), 219. doi:10.3390/electronics9020219
- BiometrixI. Q. (2003). http://www.iqbiometrix.com/products_faces_40.html
- Cheraghi, H., & Lee, H. J. (2019). SP-Net: A Novel Framework to Identify Composite Sketch. *IEEE Access: Practical Innovations, Open Solutions*, 7, 131749–131757. doi:10.1109/ACCESS.2019.2921382
- Chugh, T., Singh, M., Nagpal, S., Singh, R., & Vatsa, M. (2017). Transfer learning based evolutionary algorithm for composite face sketch recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. doi:10.1109/CVPRW.2017.90
- Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). *Fast and accurate deep network learning by exponential linear units (elus)*. arXiv preprint arXiv:1511.07289.
- Fernandes, S. L., & Bala, G. J. (2015). Developing a Novel Technique to Match Composite Sketches with Images captured by Unmanned Aerial Vehicle. *Proceedings of International Conference on Information Security and Privacy (ICISP2015)*.
- Gadekallu, T. R., Rajput, D. S., Reddy, M. P. K., Lakshmana, K., Bhattacharya, S., Singh, S., & Alazab, M. (2020). A novel PCA – whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *Journal of Real Time Image Processing, Real-Time Image Processing*, 1-14.
- Galea, C., & Farrugia, R. A. (2017). Forensic face photo-sketch recognition using a deep learning-based architecture'. *IEEE Signal Processing Letters*, 24(11), 1586–1590. doi:10.1109/LSP.2017.2749266
- Han, H., Klare, B. F., Bonnen, K., & Jain, A. K. (2012). Matching composite sketches to face photos: A component-based approach. *IEEE Transactions on Information Forensics and Security*, 8(1), 191–204. doi:10.1109/TIFS.2012.2228856
- Haji, S., & Varol, A. (2016). Real time face recognition system (RTFRS). *Proceedings of 4th International Symposium on Digital Forensic and Security (ISDFS)*.
- Identi-Kit. (2020). <https://www.identikit.net/>
- Iranmanesh, S. M., Kazemi, H., Soleymani, S., Dabouei, A., & Nasrabadi, N. M. (2018). Deep sketch-photo face recognition assisted by facial attributes. *Proceedings of IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*.
- Klare, B., Li, Z., & Jain, A. K. (2011). Matching Forensic sketches to mug shot photos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3), 639–645. doi:10.1109/TPAMI.2010.180 PMID:20921585
- Klum, S., Han, H., Jain, A. K., & Klare, B. (2013). Sketch based face recognition: Forensic vs. composite sketches. *Proceedings of IEEE International Conference on Biometrics (ICB)*.
- Kazemi, H., Soleymani, S., Dabouei, A., Iranmanesh, M., & Nasrabadi, N. M. (2018). Attribute centered loss for soft-biometrics guided face sketch-photo recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. doi:10.1109/CVPRW.2018.00091

- Lanitis, A. (2008). Comparative evaluation of automatic age-progression methodologies. *EURASIP Journal on Applied Signal Processing*, 2008(1), 239480. Advance online publication. doi:10.1155/2008/239480
- Liu, D., Peng, C., Wang, N. N., Li, J., & Gao, X. (2016). Composite face sketch recognition based on components. *Proceedings of 8th International Conference on Wireless Communications and Signal Processing (WCSP)*.
- Martinez, A. M., & Benavente, R. (1998). *The AR Face Database*. CVC Technical Report, #24.
- Mittal, P., Jain, A., Goswami, G., Vatsa, M., & Singh, R. (2017). Composite sketch recognition using saliency and attribute feedback. *Information Fusion, Elsevier*, 33, 86–99. doi:10.1016/j.inffus.2016.04.003
- Mittal, P., Jain, A., Goswami, G., Singh, R., & Vatsa, M. (2014). Recognizing composite sketches with digital face images via SSD dictionary. *Proceedings of IEEE International Joint Conference on Biometrics*. doi:10.1109/BTAS.2014.6996265
- Mittal, P., Vatsa, M., & Singh, R. (2015). Composite sketch recognition via deep network- a transfer learning approach. *Proceedings of International Conference on Biometrics (ICB)*. doi:10.1109/ICB.2015.7139092
- Nagpal, S., Singh, M., Singh, R., Noore, A., & Majumder, A. (2017). Face Sketch Matching via Coupled Deep Transform Learning. *Proceedings of International Conference on Computer Vision (ICCV)*. doi:10.1109/ICCV.2017.579
- Paritosh, M., Vatsa, M., & Singh, R. (2015). Composite Sketch Recognition via Deep Network-A Transfer Learning Approach. *International Conference on Biometrics*.
- Peng, C., Gao, X., Wang, N., & Li, J. (2018). Face recognition from multiple stylistic sketches: Scenarios, datasets, and evaluation. *Pattern Recognition*, 84, 262–272. doi:10.1016/j.patcog.2018.07.014
- Peng, C., Gao, X., Wang, N., & Li, J. (2019). Sparse graphical representation based discriminant analysis for heterogeneous face recognition. *Signal Processing*, 156, 46–61. doi:10.1016/j.sigpro.2018.10.015
- Ramachandran, P., Zoph, B., & Le, Q. V. (2017). *Searching for activation functions*. arXiv preprint arXiv:1710.05941.
- RM, S. P., Maddikunta, P. K. R., Parimala, M., Koppu, S., Reddy, T., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. *Computer Communications*.
- Roy H., & Bhattacharjee D. (2019). Heterogeneous Face Matching Using Robust Binary Pattern of Local Quotient: RBPLQ. *Advances in Intelligent Systems and Computing*, 883.
- Shivaleela Patil, D., & Shibhangi, D. C. (2020). Composite Sketch Based Face Recognition Using ANN Classification. *International Journal of Scientific and Technology Research*, 9, 42–50.
- Venkatraman, S., Alazab, M., & Vinayakumar, R. (2019). A hybrid deep learning image-based analysis for effective malware detection. *Journal of Information Security and Applications*, 47, 377–389. doi:10.1016/j.jisa.2019.06.006
- Wang, X., & Tang, X. (2008). Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11), 1955–1967. doi:10.1109/TPAMI.2008.222 PMID:19762924
- Yadav, D., Singh, R., Vatsa, M., & Noore, A. (2014). Recognizing age-separated face images: Humans and machines. *PLoS One*, 9(12), e112234. doi:10.1371/journal.pone.0112234 PMID:25474200
- Zhang, H., Jolfaei, A., & Alazab, M. (2019). A face emotion recognition method using convolutional neural network and image edge computing. *IEEE Access: Practical Innovations, Open Solutions*, 7, 159081–159089. doi:10.1109/ACCESS.2019.2949741

Chethana H. T. is working as assistant professor in the Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru. She is currently a PhD scholar in the Department of Computer Science and Engineering at the JSS Science & Technological University, Mysuru. She has received her B E degree in Computer Science and Engineering from Coorg Institute of Technology, Ponnampet during 2011 and M.Tech degree in Software Engineering from Visvesvaraya Technological University during 2015. She has total 5 years of teaching experience. Her research interests include Image Processing, Machine Learning, Deep Learning and Digital Forensics. She has published more than 10 research papers in various international journals and conferences. She is a life member of Indian Society for Technical Education.

Trisiladevi C. Nagavi is working as Assistant Professor in the Department of CS &E, S. J. College of Engineering, JSS Science & Technology University Mysuru was conferred with a Ph.D. in Computer Engineering by VTU Belagavi for her research work on Retrieval of Music through Melody. She completed her M.Tech in Software Engineering and secured "Second Rank" from VTU Belagavi. She obtained her B.E. in CS&E from Karnataka University Dharwad. She has an expertise in the area of Audio, Music, Speech and Image Signal Processing, Information Retrieval, Machine Learning and Digital Forensics. Her research outcome resulted in an AndroidApp to play favourite tunes. She is an Active member of IEEE India Special Interest Group on Communications Disability and worked on Real Time Text to Speech Conversion and Translation System in Collaboration with All India Institute of Speech and Hearing (AIISH) Mysore and IEEE Standards Association. Reviewer for peer reviewed Journals and conferences.