Chapter 6

EXPERIMENTS AND EVALUATION

Objective
To discuss the various experiments and verify the results with face recognizer and prior work.

In this chapter we have discussed various experiments of predicted older images which are found by our proposed method, and the results are evaluated and compared with the prior works.

6.1 Evaluation Techniques

We have verified predicted faces with ground truth images using two methods (a) based on human feedback and (b) based on a face recognition tool. Both the techniques are described in following sub-sections.

Ans	Category	Score (S_i)
A	Very High	[80, 100]
B	High	[70, 79]
C	Average	[60, 69]
D	Low	[50, 59]
E	Very low	[0, 49]

Table 6.1: Response Scoring Table of Human Feedback System

6.1.1 Human Feedback Score for Comparing two Face Images

The human feedback score has been computed based on a question given to a group of persons $(X_1, X_2, \ldots X_N)$. Two face images P and Q are shown to each person X_i, who is asked a question **"How much face image P is similar to face image Q?"**. The response from the person X_i is to be recorded in the form of a Scoring Table having 5 options A to E as shown in Table 6.1.

For every responses from A to E, an associated response score S_i is assigned from the table, where S_i is the i^{th} feedback score of the person X_i. Then similarity score $SimilarityScore = \frac{1}{N}\sum_{i=1}^{N} S_i$ is calculated for each experiments.

6.1.2 Face Recognition Tool using OpenCV and Python

R. Raja [51] has described a method to implement a face recognition program using OpenCV and Python which is based on Local Binary Patterns Histograms (LBPH) [11, 61, 64, 85] face recognizer technique. The face recognition process has been divided into following three steps:

1. **Prepare Training Data:** All the available images of FG-NET dataset are used for training purpose. For an image of one person one FaceID is assigned, again a serial number is assigned preceding with _ sign for multiple images of a same person. For example 1P_1, 1P_2, 1P_3 are the FaceIDs of person No. 1.

2. **Train Face Recognizer:** The LBPH recognizer procedure is trained with the data prepared in step 1. By this process, the LBPH recognizer generates a histogram for each new image fed into the system.

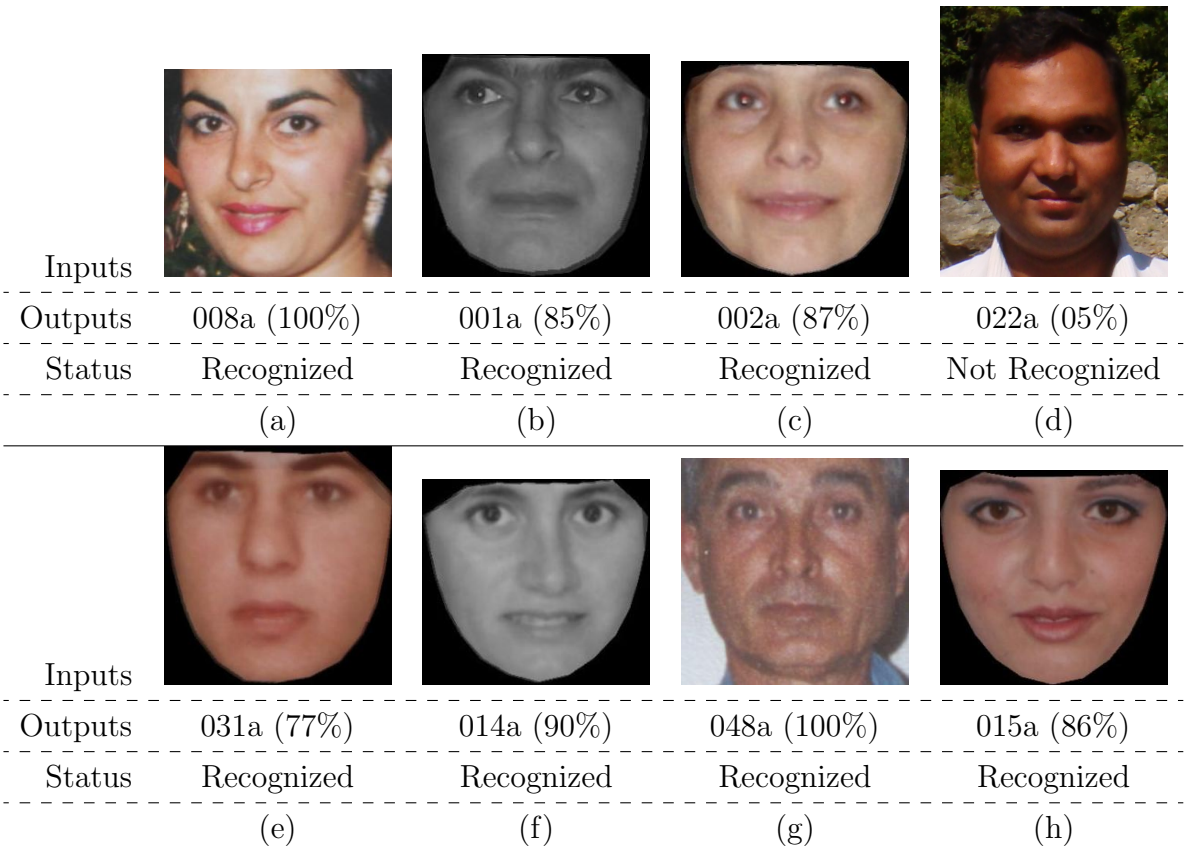

Figure 6.1: Experiments for face recognition [17, 51]. First row images are the input images to be recognized by the system, Second row is the outputs, which are the recognized image IDs with highest confidence score. Third row is the validated result whether the faces are recognized or not.

3. **Prediction/ Recognition:** For recognition of any given input image, the face recognizer tests whether it recognize them correctly or not. It computes the histogram for any new input image and compares that histogram with the histograms it already has. Finally, it finds the best match with the highest confidence score, and returns the person label (FaceID) associated with that best matching confidence score.

6.1.2.1 Experiment on the Face Recognition Tool

We have performed an experiment where 8 different test images are recognized by using the face recognition tool. In Figure 6.1, the test images are shown along with the confidence score shown beside the imageID. It can be noticed that when test image itself is present in the dataset confidence score is 100%. If the test image is not present in the dataset then a confidence score lower than 100% is returned depending upon the similarity to any image available in the dataset.

6.2 Evaluation of the Predicted Future Appearance of a Child's Face Image

The FG-NET dataset contains face images of 82 different persons, 6-18 images per person. We have done 150 different experiments covering each person at least once and at most twice. A child's face image is taken as input and his/her aged face image is predicted as output at some target age groups like 11-15, 16-20, 21-25 *etc.* Some sample results are shown in Figure 6.2.In order to evaluate the performance of our method we need to compare the output image to the ground truth image available in the dataset. The comparison of output images with ground truth images are done by human feedback system as well as automated face recognition system as discussed in previous section. The observed similarity scores of predicted and ground-truth images are shown graphically in Figure 6.3. The outcome of the experiment is summarised below:

- **Manual system:** 90% of times recognition level is higher than 70%.
- **Automatic system:** 89% of times recognition level is higher than 70%.

Here similarity score threshold of 70% is set for recognition of a better predicted face.

6.2.1 Performance Comparison of with Previous Works

Xiangbo Shu et al. has reported a facial aging method named BDL-PAP in one of the recent works [73]. They have used 8 input images of FG-NET dataset to compare their method to another existing method BDL-AP [73]. We also used the same 8 images as input to our method and compare the results with BDL-PAP and BDL-AP as reported by the authors in [73]. The comparative results have been shown in Figure 6.4 and 6.5.

Here we have used face recognizer tool to get the confidence similarity scores of output results of [BDL-AP, BDL-PAP, our proposed mehod] with ground truth images, the respective scores of Figure 6.5 shows that the outputs of our method are comparatively better than the recent BDL-AP and BDL-PAP methods.

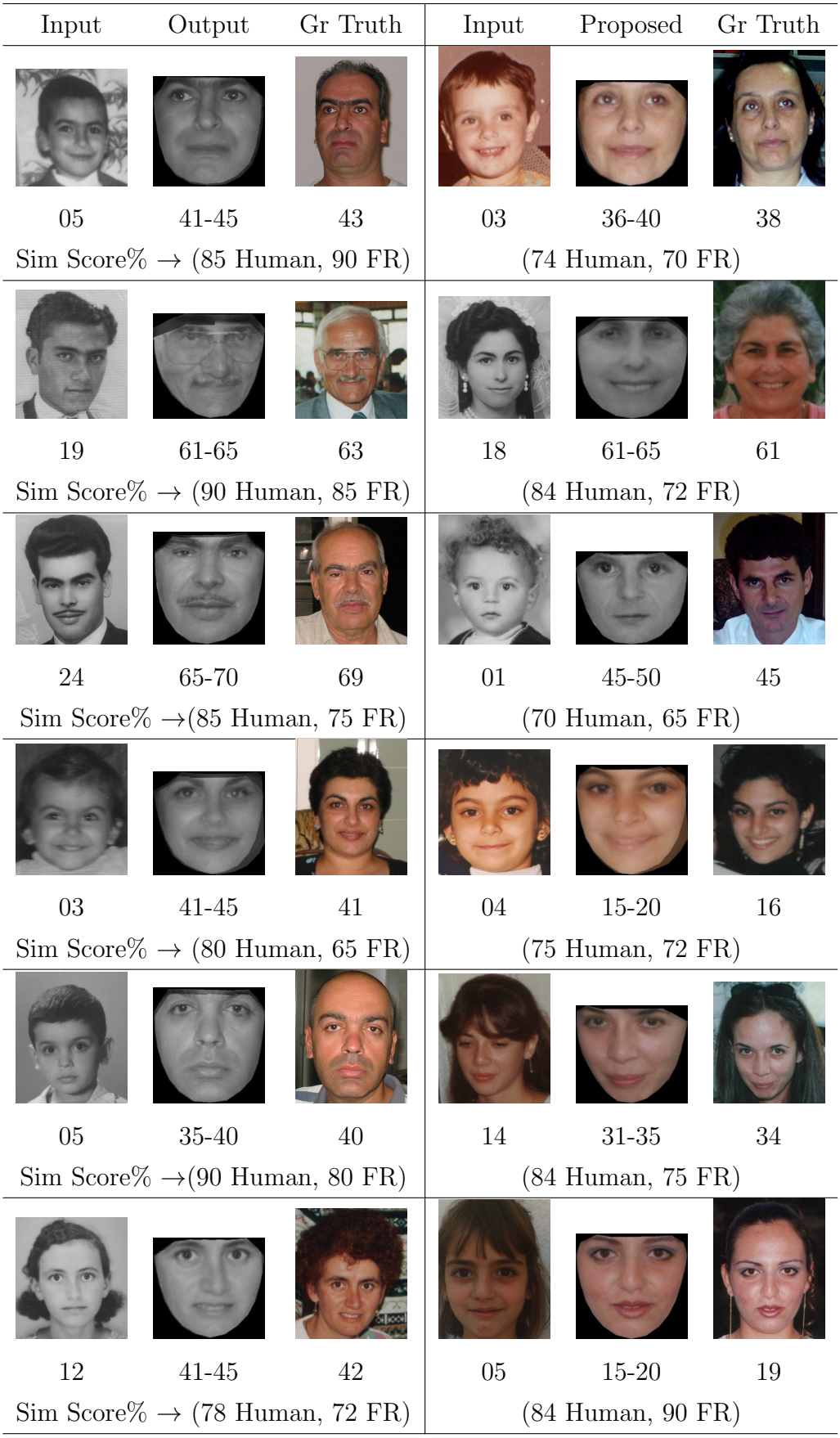

Figure 6.2: Comparison with ground truth images. Firs row- Input, output and ground truth images. Second row- corresponding ages of the images. Third row- two similarity scores for each experiment, first score is of human feed back based and second score is of face recognizer tool based.

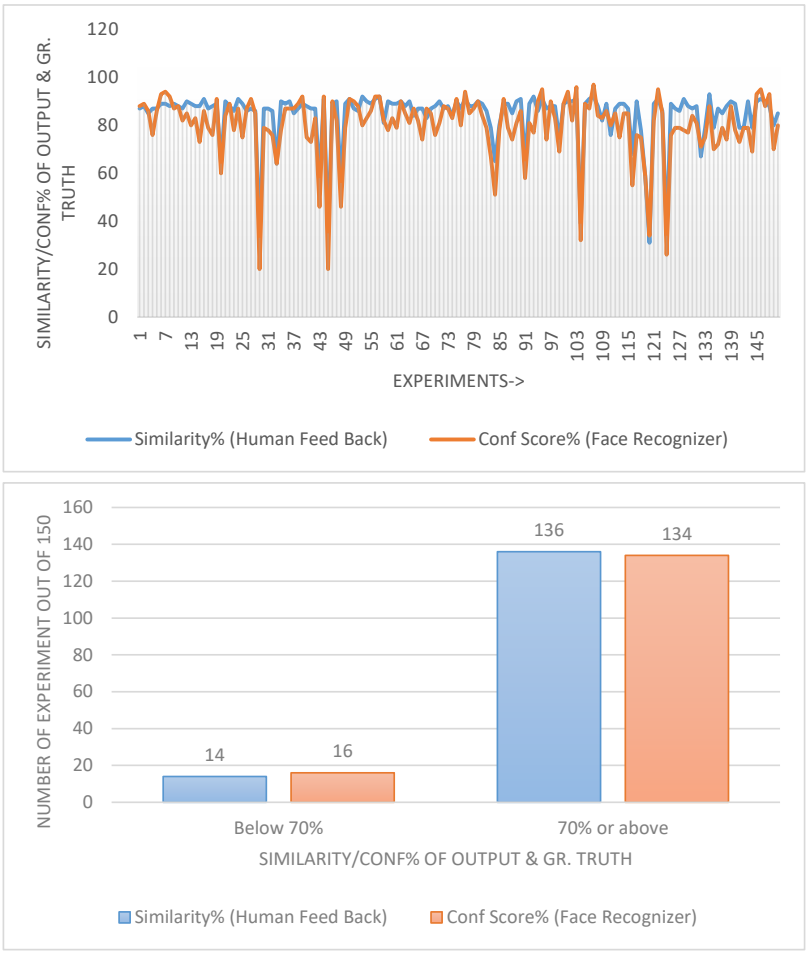

Figure 6.3: Result analysis of predicted images. Comparison of Human feedback based and Face recognition based result analysis of 150 experiments.

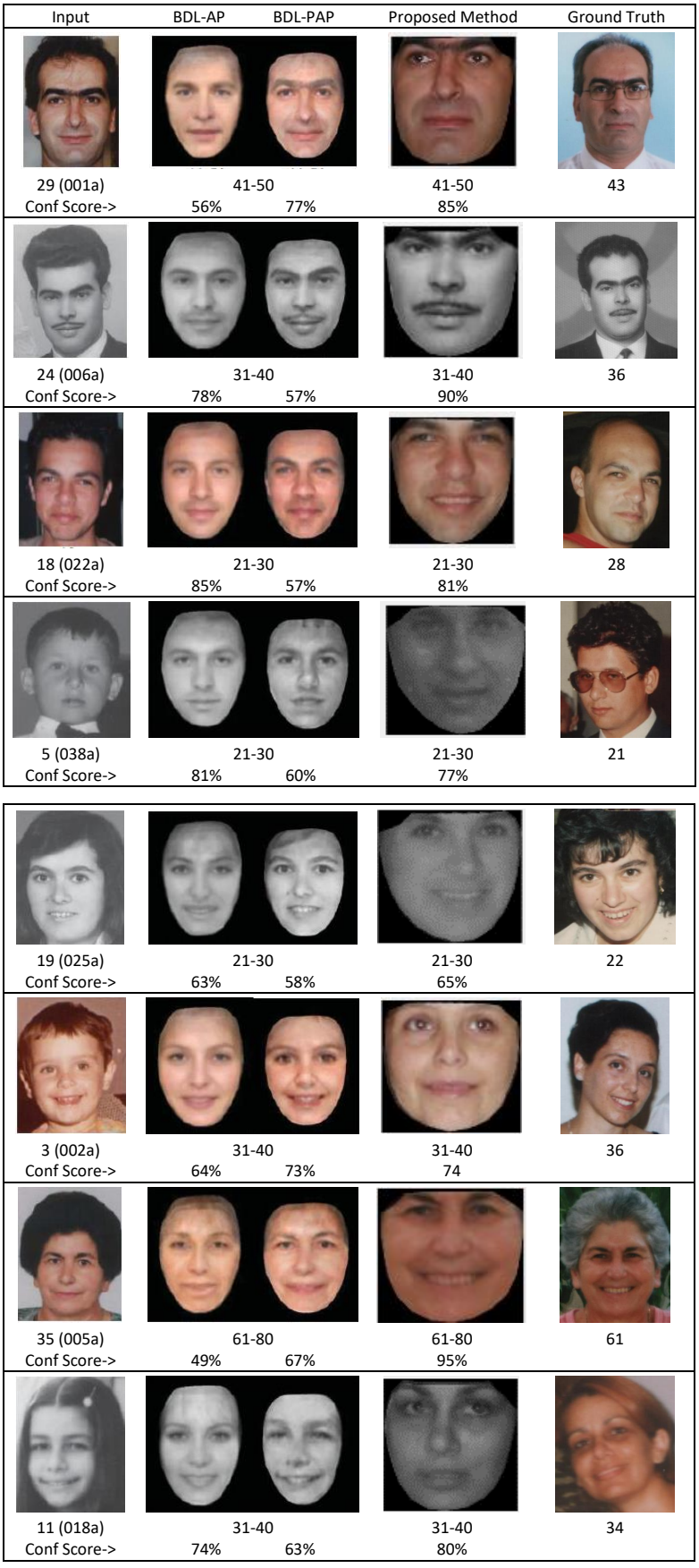

Figure 6.4: The comparisons of BDL-AP, BDL-PAP (very recent work by Xiangbo Shu et al. [73]) and proposed method on FG-NET. First row- Images of Input (first column), outputs of existing and proposed (second, third column for existing and fourth column for proposed) and ground truth (last column). Second row- ages of respective input and outputs. Third row- matching score of output and ground truth images.

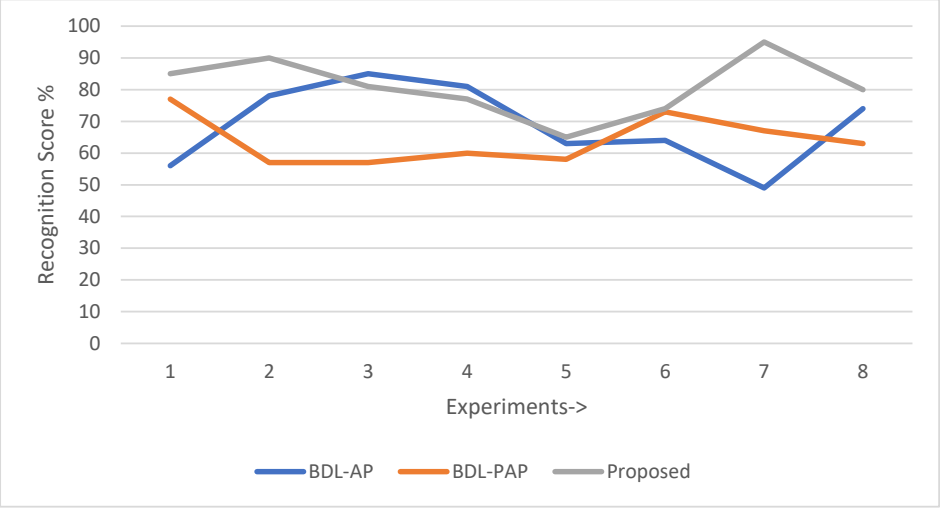

Figure 6.5: The comparisons of BDL-AP, BDL-PAP and proposed method on FG-NET (results shown in Figure 6.4).