Impact of applying super resolution to low resolution face images on the performance of deep neural networks

Rajeshwar Moghekar1 and Sachin Ahuja2
1 Associate Professor, HITAM, India
2 Professor, Chitkara University, Punjab, India.

E-mail: rajeshwarm.cse@hitam.org

Abstract: Face images captured in unconstrained environment differ in various aspects such as expression, illumination, resolution, occlusion, pose etc. which makes face recognition task difficult. The face images captured by the camera from a distance will have low resolution and lack many finer details that makes face recognition a challenging task. Super resolution (SR) is a process of generating high resolution (HR) images from one or more images. In this work, we apply super resolution to low resolution (LR) images of faces to find the impact on the deep models performance. To achieve this, we create dataset with face images captured in unconstrained environment. Later we designed a CNN model with eight layers and trained on the dataset created. Our deep model with low memory requirement and less parameters achieves an accuracy of 99.75% on test dataset and outperforms fine-tuned VGGFace by a small margin. The performance of our deep neural network and fine-tuned VGGFace was observed on low resolution images pre and post-super resolution. The deep neural network-based model available in OpenCV, SRGAN super resolution model and INTER_CUBIC interpolation are used to generate HR images. The HR images generated by OpenCV, SRGAN are better than INTER_CUBIC interpolation. The results show that HR images generated by applying SR to low resolution face images improve the image quality in terms of Mean squared error (MSE), Structural similarity index measure (SSIM) and Peak to signal noise ratio (PSNR). However, the results indicate that improvement in the image quality does not significantly improve performance of deep model.

Key words: Super resolution, low resolution, Face recognition, fine tuning, deep neural network, transfer learning

1. Introduction
In the field of computer vision, Face recognition has grabbed the attention of researchers as it finds many applications in law enforcement, biometrics, law enforcement and so forth [1]. Face recognition in unconstrained environment is still a challenging task as the face images vary in terms of pose, illumination, resolution, occlusion etc.[2-4]. Most early approaches in face recognition extract the features like SIFT [5], LBP [6], PCA [7], HOG [8] from the face images and train the classifier to recognize the different classes. But the approach to perform face recognition has changed after AlexNet a deep neural network won the ImageNet competition by a large margin [9]. These deep convolutional networks have multiple layers which extract the features and learn representations that have different levels of abstraction. The face images captured from surveillance camera has variations in terms of pose, resolution, expression etc. Generally, camera is placed to cover wide angle, so the face images captured at longer distance will have low resolution and lacks many finer details.
Figure 1 shows the face image captured by surveillance camera with wide viewing angle has lot of image details missing. Face recognition with low resolution images is an active research area as demand for surveillance camera-based applications increased. Generally, to solve face recognition of low-resolution face images they construct high resolution faces by applying super resolution. Threshold resolution or minimal resolution above which the face recognition performance is steady, but if resolution decreases below then that the performance degrades [10]. The quality of an image can be assessed in two different ways i.e. subjective or objective. Subjective methods depend on the perception of the human whereas objective methods are related to computational models. Objective image quality measurement models are classified into three types based on the information available about the reference image.[11]

Full-reference methods - They compare test image with ground truth or reference image.

Reduced reference methods - Test image quality is assessed by comparing the features extracted from it with that of the reference image.

No-reference methods - The quality of test image is assessed without any reference image.

Image quality techniques such as Mean Square Error (MSE), Peak signal to noise ratio (PSNR) and Structured similarity index method (SSIM) are generally used to assess the quality of the image.[12] MSE (Mean Squared Error) - MSE also referred as Mean squared deviation (MSD) gives the average of squared difference between the ground truth and the estimated value. The higher the MSE value indicates that the image quality is degraded.

PSNR (Peak signal to noise ratio) – It is defined as ratio between the maximum power of the signal to the power of the noise that impacts the quality. It is expressed in decibel (DB) form for comparison between two images. The higher the PSNR value indicates the image quality is good.

SSIM (Structured similarity index method) - It measures the similarity between the reconstructed image and the actual image. The SSIM value lies between 0 to 1, with 1 indicating high similarity (same as the original) and 0 low similarities.

The process of obtaining single or multiple HR images from one or more low resolution (LR) images is referred as super resolution (SR) [13]. To study the pre- and post-of applying of super resolution to the low-resolution face images impact the performance of deep neural network, the paper discusses about the dataset created with variations in pose, illumination, resolution, occlusion etc. in section 3. In section 4 the performance of deep neural network designed with eight convolutional layers followed by fully connected neural network and VGGFace fine-tuned model on the dataset created is observed. We study the impact of low-resolution images, pre and post of applying super resolution on the deep neural network designed and fine-tuned VGGFace in section 5.

2. Related work

Broadly super resolution methods are catagorized into (i) Multi-image super resolution (ii) Example-Based super resolution [14]. Many applications like aerial and satellite imaging, processing of medical images, ultra sound imaging [15], improving facial and text images, enhancing video and compressed images, iris recognition make use of super resolution [16]. To obtain high resolution images they
increase the pixels in each unit area of the image to get finer details [17]. There is difference between SR, Interpolation and Image restoration. Interpolation does not store high frequency details in contrast to SR which does and it is generally applied to single image [18]. In image restoration the quality of the image is improved but the size of the input and output are same whereas in SR the output image size increases. SR algorithms can be classified based on the factors like how many LR images used, domain and reconstruction method. Based on the domain they are classified into frequency or spatial domain. In the same way they are classified into single image or multiple images based on the number of LR images used. The assessments of SR algorithms are carried out using both the methods, subjective and objective. SR algorithms are also applied to facial images which have the advantage due to standard and established dataset as compared to applications like processing the medical images, satellite imaging etc. They found that CAS-PEAL database provides the ground truth of resolutions which is expected for SR [19]. To improve the performance Face recognition with very low resolution they designed a data constraint for super resolution [20].

3. Dataset
Dataset with face images of 80 subjects is created in unconstrained environment with half of the subjects images downloaded from internet and the other half subjects captured from the camera placed at a fixed location in the college campus. Deep neural network available in OpenCV is used for face detection. Figure 2 show face images of a particular subject captured by the camera with variations in resolutions. We split the dataset in the ratio of 70 : 15 : 15 (training set : validation set : testing set).

Data augmentation was applied to increase the face images of subject which were less in number by performing simple transformations like flip, scale and noise as shown in Figure 3.

4. Deep neural network model
A deep model with eight layers of convolution was designed, which is followed by a fully connected layer. The model accepts input image of size 224x224 pixels and uses activation function Rectified linear unit. The Figure 4 shows architecture of deep neural network.
Figure 4. Architecture of deep neural network with eight convolutional layers.

VGGFace deep neural network trained on huge dataset consisting of 2.6M face images with 2622 different subjects [21]. Instead of training VGGFace from scratch, we fine tune the model by freezing the top layers which learn the generic features. Figure 5 shows the fine-tuned VGGFace where the top layers weights are freezed.

Table 1 shows the deep neural network performs marginally better than the fine-tuned VGGFace on the dataset created with the advantage of having less parameter with low memory space requirement. Our deep neural network achieved 99.94% training accuracy and 99.75% testing accuracy, whereas fine-tuned VGGFace achieved 99.82% training accuracy and 99.68% testing accuracy on our dataset. Our model has around 1/8th of the VGGFace parameters and requires less memory requirement.

Table 1. Results obtained our dataset

| Model           | Training accuracy | Testing accuracy | Number of parameters | Model size in MB |
|-----------------|-------------------|------------------|----------------------|------------------|
| Fine-tuned VGGFace | 99.82             | 99.68            | 40,487,824           | 293              |
| Our Model       | 99.94             | 99.75            | 5,671,824            | 68               |
5. Impact of applying super resolution to low resolution images on the deep model’s performance
To study the impact of super resolution on the deep model trained on our dataset, deep neural network-based model available in OpenCV and Super resolution model based on SRGAN [22] are used to generate HR face images from the low-resolution face images. To evaluate the effectiveness of the model the images were reduced to one fourth of the actual using interpolation INTER_CUBIC method available in OpenCV. Figure 6 (a) is sample face image from our dataset with actual dimension 300x392 and Figure 6 (b) with reduced dimension of 75x98. Figure 6 (c) another sample face image from our dataset with actual dimension 156x204 and Figure 6 (d) with reduced dimension of 39x51.

(a) 300 X 392  (b) 75 X 98  (c) 156 X 204  (d) 39 X 51

Figure 6. Shows the actual image and the corresponding image after resizing with INTER_CUBIC.

The reduced image or images with low resolution is converted back to high resolution by up scaling with a factor of four using INTER_CUBIC, deep neural network model available in OpenCV and SRGAN based super resolution. Figure 7 shows the a sample image of one of the subject from our dataset, with the image on the left image being the actual image and the others enhanced version using INTER_CUBIC interpolation, the deep neural network in OpenCV and SRGAN based super resolution. Figure 8 shows a sample image of one of the subjects from our dataset, with the image on the left being the actual image and the others enhanced version using INTER_CUBIC interpolation, the deep neural network in OpenCV and SRGAN based super resolution.

Figure 7. High resolution of the images generated with INTER_CUBIC, deep neural network in OpenCV and SRGAN.
Figure 8. High resolution of the images generated with INTER_CUBIC, deep neural network in OpenCV and SRGAN.

Figure 9. (a) Original Vs Original (b) Original Vs Original
MSE: 0.00, SSIM: 1.00
MSE: 0.00, SSIM: 1.00

Figure 10. Original vs INTER_CUBIC (b) Original vs INTER_CUBIC
MSE: 177.44, SSIM: 0.89
MSE: 271.73, SSIM: 0.78

Figure 11. a) Original vs OpenCV based SR (b) Original vs OpenCV based SR
MSE: 141.51, SSIM: 0.92
MSE: 205.91, SSIM: 0.82
Figure 9 (a) and (b) shows that when the original image is compared with itself the MSE value is 0 indicating there is no deviation between the two images and SSIM is 1 which indicates both the images are similar to each other. Figure 10 (a) and 10 (b) show the MSE and SSIM values, when the original image is compared with the enhanced image using INTER_CUBIC. Figure 11 (a) and 11 (b) show the MSE and SSIM values when the original image is compared with the HR image generated using deep neural network in OpenCV. Figure 12 (a) and 12 (b) show the MSE and SSIM values when the original image is compared with the HR image generated using SRGAN based SR.

Figure 12. (a) Original vs SRGAN based SR       (b) Original vs SRGAN based SR

(a) Original vs SRGAN based SR       (b) Original vs SRGAN based SR

Figure 13. MSE values of the high-resolution Image generated using INTER_CUBIC, deep neural network in Opencv and SRGAN based SR.
Figure 14. SSIM values of the high-resolution Image generated using INTER_CUBIC, deep neural network in OpenCV and SRGAN based SR.

Figure 13 (a) and 13(b) show that of DNN in OpenCV performs better as the MSE value is low compared to INTER_CUBIC and SRGAN based SR. Figure 14 (a) and 14(b) show that DNN in OpenCV performs better as the SSIM value is high compared to INTER_CUBIC and SRGAN based SR.

Figure 15. PSNR values of the enhanced images using INTER_CUBIC, SRGAN based SR and deep neural network in OpenCV.

Figure 16. PSNR values of the enhanced images using INTER_CUBIC, SRGAN based SR and deep neural network in OpenCV.
Table 2. PSNR values of the HR images generated using INTER_CUBIC, SRGAN based SR and deep neural network in OpenCV

| Method          | PSNR of HR image of image in 6(d) | PSNR of HR image of image in 6(b) |
|-----------------|-----------------------------------|-----------------------------------|
| INTER_CUBIC     | 22.68                             | 22                                |
| SRGAN based SR  | 22.73                             | 22.26                             |
| OpenCV based DNN| 23.18                             | 22.83                             |

Figure 15 and 16 show the PSNR values of the HR images generated using INTER_CUBIC, SRGAN based SR and deep neural network in OpenCV. Table 2 show that the PSNR value of the HR images generated by DNN in OpenCV outperforms INTER_CUBIC and SRGAN based SR.

The Figure 17 (a) shows the low-resolution face image, which is not recognized by our Model but after generating the HR image using SRGAN based SR as shown in figure 17 (b) the model was able to correctly classify. By using the same image shown in figure 17 (a) when HR image is generated using DNN in Opencv as shown in figure 17 (c) it was not correctly classified.

Figure 18 (a) shows the low-resolution face image the which is not recognized by our Model but after generating the HR image using SRGAN based SR and DNN available in OpenCV as shown in figure 18 (b) and 18 (c) they were correctly classified. The figure 19 (a) shows the low-resolution face image the which is recognized by fine-tuned VGGFace but after generating the HR image using SRGAN based SR and DNN available in OpenCV as shown in figure 19 (b) and 19 (c) they were mis classified by it.

Figure 17. (a) LR image (b) HR image (SRGAN) (c) HR image (DNN in OpenCV)

Figure 18. (a) low resolution image. (b) HR image generated using SRGAN based SR. (c) HR
image generated using DNN available in OpenCV.

(a)                                    (b)                                              (c)

Figure 19. (a) low resolution image. (b) HR image generated using SRGAN based SR. (c) HR image generated using DNN available in OpenCV.

(a)                                    (b)                                              (c)

Figure 20. (a) LR images (b) HR images (SRGAN) (c) HR images (DNN OpenCV)

Figure 20 (a) shows low resolution images that are not recognized correctly by fine-tuned VGGFace, but with HR images generated using SRGAN based SR 20 (b) and DNN in OpenCV 20 (c) are correctly classified by it. A few low-resolution images have been collected from Indian movies face database IMFDB database [23], where the face images were manually cropped from the movie videos. The LR image from IMFDB in figure 21 (a) is correctly classified by our Model but when it is enhanced using SRGAN 21 (b) and DNN in OpenCV 21(c) our Model misclassified them.

(a)                                    (b)                                              (c)

Figure 21. (a) LR image (b) HR Image (SRGAN) (c) HR image (DNN Opencv)
Figure 22 (a) shows the LR images from IMDBF which are not recognized by our Model, but the HR images generated using SRGAN based SR 22(b) and the DNN in OpenCV 22 (c) were correctly classified by our Model.

Table 3. Impact of pre and post of applying super resolution to low resolution images

| Model            | Low resolution image | HR image INTER_CUBIC | HR image SRGAN | HR image DNN in Opencv |
|------------------|----------------------|----------------------|----------------|------------------------|
| Our model        | 168/200              | 165/200              | 173/200        | 175/200                |
| Fine-tuned VGGFace | 173/200             | 173/200              | 178/200        | 180/200                |

As shown in the Table 3 when 200 low resolution images (few images from IMDB and few from our data set) were given as it is our model could classify 168 correctly and Fine-tuned VGGFace classified 174 correctly. After converting to HR using INTER_CUBIC the performance slightly reduced. There were cases where low resolution image which was correctly classified by the model was later misclassified when we applied super resolution.

6. Conclusion

A deep neural network with eight CNN layers has been developed and trained on the dataset created with face images captured in unconstrained environment with variations in pose, resolution, occlusion etc. Our CNN model with less numbers of parameters performs marginally better than fine-tuned VGGFace on our dataset. As Low-resolution face images captured lack many details which pose challenge for face recognition, so they were converted to high resolution images using super resolution and pre & post analysis has been done. We generated HR images using SRGAN and Deep neural network available in Opencv. Generating HR images improves the image quality in terms of MSE, SSIM and PSNR but the performance on recognition rate cannot be assured. As in a few cases LR images which are not correctly recognized by the Deep learning models after converting them to HR images using the SR algorithms, they were correctly classified, at the same time we observed a few low-resolution images which are correctly classified by the Deep models were later mis classified by the Deep models, when those low-resolution images were converted to HR using the SR algorithms used. We found only marginal improvement in the performance after enhancing the low-resolution models using super resolution model. However, we have to explore other super resolution methods in the future to come to conclusion.

References
[1] Deng, J., Guo, J., Zhang, D., Deng, Y., Lu, X. and Shi, S., 2019. Lightweight face recognition challenge. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 0-0).
[2] He, M., Zhang, J., Shan, S., Kan, M. and Chen, X., 2020. Deformable face net for pose invariant face recognition. Pattern Recognition, 100, p.107113.
[3] Guo, G. and Zhang, N., 2019. A survey on deep learning based face recognition. *Computer Vision and Image Understanding*, 189, p.102805.

[4] Lahasan, B., Lutfi, S.L. and San-Segundo, R., 2019. A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. *Artificial Intelligence Review*, 52(2), pp.949-979.

[5] Geng, C. and Jiang, X., 2009, November. Face recognition using sift features. In *2009 16th IEEE international conference on image processing (ICIP)* (pp. 3313-3316). IEEE.

[6] Ahonen, T., Hadid, A. and Pietikäinen, M., 2004, May. Face recognition with local binary patterns. In *European conference on computer vision* (pp. 469-481). Springer, Berlin, Heidelberg.

[7] Turk, M. and Pentland, A., 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), pp.71-86.

[8] Déniz, O., Bueno, G., Salido, J. and De la Torre, F., 2011. Face recognition using histograms of oriented gradients. *Pattern recognition letters*, 32(12), pp.1598-1603.

[9] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[10] Wang, Z., Miao, Z., Wu, Q.J., Wan, Y. and Tang, Z., 2014. Low-resolution face recognition: a review. *The Visual Computer*, 30(4), pp.359-386.

[11] Lahoulou, A., Larabi, M.C., Beghdadi, A., Viennet, E. and Bouridane, A., 2016. Knowledge-based taxonomic scheme for full-reference objective image quality measurement models. *Journal of Imaging Science and Technology*, 60(6), pp.60406-1.

[12] Hore, A. and Ziou, D., 2010, August. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition* (pp. 2366-2369). IEEE.

[13] Miravet, C. and Rodri, F.B., 2007. A two-step neural-network based algorithm for fast image super-resolution. *Image and Vision Computing*, 25(9), pp.1449-1473.

[14] Glashner, D., Bagon, S. and Irani, M., 2009, September. Super-resolution from a single image. In *2009 IEEE 12th international conference on computer vision* (pp. 349-356). IEEE.

[15] Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H. and Zhang, L., 2016. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128, pp.389-408.

[16] Cui, J., Wang, Y., Huang, J., Tan, T. and Sun, Z., 2004, August. An iris image synthesis method based on PCA and super-resolution. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 2, pp. 471-474)*. IEEE.

[17] Su, K., Tian, Q., Xue, Q., Sebe, N. and Ma, J., 2005, July. Neighborhood issue in single-frame image super-resolution. In *2005 IEEE international conference on multimedia and expo* (pp. 4-pp). IEEE.

[18] Gotot, T. and Okutomi, M., 2004, June. Direct super-resolution and registration using raw CFA images. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. (Vol. 2, pp. II-II)*. IEEE.

[19] Nasrollahi, K. and Moeslund, T.B., 2014. Super-resolution: a comprehensive survey. *Machine vision and applications*, 25(6), pp.1423-1468.

[20] Zou, W.W. and Yuen, P.C., 2011. Very low resolution face recognition problem. IEEE Transactions on image processing, 21(1), pp.327-340.

[21] Parkhi, O.M., Vedaldi, A. and Zisserman, A., 2015. Deep face recognition.

[22] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).