Super-Low Resolution Face Recognition using Integrated Efficient Sub-Pixel Convolutional Neural Network (ESPCN) and Convolutional Neural Network (CNN)

Mohammed Ahmed Talab
Department of Engineering of Computer Technology
Al-maarif University College
Ramadi, Iraq
mmss_ah@yahoo.com

Suryanti Awang
Soft Computing & Intelligence Systems (SPINT)
Faculty of Computer Systems & Software Engineering
Universiti Malaysia Pahang
Kuantan, Malaysia
suryanti@ump.edu.my

Saif Al-din M. Najim
College of Computer Science and Information Technology
University of Anbar
Ramadi, Iraq
sayf73@gmail.com

***Abstract*—Several deep image-based models which depend on deep learning have shown great success in the recorded computational and reconstruction efficiencies, especially for single high-resolution images. In the past, the use of super-resolution was commonly characterized by interference, and hence, the need for a model with higher performance. This study proposed a method for low to super-resolution face recognition, called efficient sub-pixel convolution neural network. This is a convolutional neural network which is usually employed at the time of image pre-processing to increase the chances of recognizing images with low resolution. The proposed Efficient Sub-Pixel Convolutional Neural Network is used for the conversion of low-resolution images into a high-resolution format for onward recognition. This conversion is based on the features extracted from the image. Using several evaluation tools, the proposed Efficient Sub-Pixel Convolutional Neural Network recorded a higher performance in terms of image resolution when compared to the performance of the benchmarked traditional methods. The evaluations were carried out on a Yale face database and ORL dataset faces. For Yale and ORL datasets, the obtained accuracy of the proposed method was 95.3% and 93.5%, respectively, which were higher than those of the other related methods.**

Keywords—Super-Resolution (SR), Face Recognition, Low Resolution (LR), Deep Learning

I. INTRODUCTION

One of the tasks with great interest in digital image processing is the generation of higher resolution (HR) images from their low resolution (LR) variants. This process, also known as super-resolution (SR), is greatly important due to its direct usefulness in several applications like HDTV, satellite imaging [1], face recognition [2], medical imaging [3, 4], and surveillance [2]. A range of methods, such as multi-image SR assumed that multiple images only exist as LR instances with varying perspectives and specified prior affine transformations [5, 6].

Several methods present multiple images in a low-resolution version with myriads of instances using the same scene with different approaches. These are classified as multi-image super-resolution methods. They explicitly explore redundancy using a constraint with added information in an attempt to invert the downsampling process. Be that as it may, the quality of an outcome is mainly affected by the accuracy of the method adopted. However, a computationally complex image registration and fusion stages are required for these methods. The single image super-resolution (SISR) methods are the suitable alternative as they can implicitly learn the inherent redundancy in most data sourced from both high and single low-resolution images. These depict the temporal correlations for videos and local spatial correlations for images. As such, there is a need for pre-information for the solution space in reconstruction. In recent times, the use of artif cial intelligence (AI) has assumed a vantage position with great technological relevance. An important aspect of this is the use of AI in biometric authentication for face recognition [7, 8].

Face recognition is important because it has several advantages compared to the other authentication or biometric methods like iris and speaker recognition [9, 10]. Studies on biometric authentication have recently focused on face recognition. Biometrics refers to the instant identification of the attributes of an individual based on the physiological features. Several modifications and improvements have been witnessed in the field of identification technology in recent times compared to other traditional identification methods such as surveillance, security systems, and credit card authentication. A modified feature selection technique is used in the linear discriminant analysis (LDA) for face and signature at the feature extraction stage [11, 12]. Certain recognition frameworks can use sample domains such as Discrete Wavelet Transform (DWT) to significantly alter the attributes of an original image [13, 14]. When using most super-resolution methods, redundancy is often assumed to be predominant in high-frequency data and that it is possible to reconstruct such redundancy accurately from low-resolution low-frequency components. Super-resolution is therefore characterized by inference problem which makes it dependent on our model in consideration.

The objective of this study is to evaluate the performance of a combined Gaussian filter-based (ESPCN) model to generate super-resolution images from their low-resolution

versions. In this investigation, related research on the merits of super-resolution was succinctly presented.

To guide the reader of this article, the remaining sections of this study are organized thus: the related works in SR were presented in the second section, while the study methodology was presented in the third section. In the fourth section, the results of the evaluation and their discussions were presented, while the conclusions derived from the study were presented in Section 5.

II. Related Work

In the real practice of face recognition, there are several changes in the attributes of captured face images due to the intensified degradation in their face recognition performances. In such cases, the objects are generally far from the

surveillance camera and this results in the capturing of images with smaller faces than the original size. Such small-sized images often yield low image recognition performance. This situation is referred to as low-resolution face recognition (LRFR). Being that the LRFR algorithms often presents with

low efficiency, coupled with their limited availability for facial feature recognition, video processing is usually done using convolution neural network when striving to achieve images with better quality from LR images [15]. The method presented in this study had a better efficiency compared to the efficiency of LRFR due to its HR characteristics. The LR images decrease the image descriptor features, and thus, only CNN and Remove Noise were applied to the original LR image prior to the application of Super-resolution. The application of deep network cascade (DNC) in LR images has been reported. The added network layer was also reported to record better recognition and visual quality performance [16, 17]. The LRFR model is currently based on DL which generally divides problems of face recognition into clusters of interest, and it is still not known how to classify the two clusters of interest.

It is, therefore, imperative that the problem of LR face detection is a complicated one, and the performance criteria in the building of new models are extremely high. Majority of the problems of SR are encountered in the form of LR, blurred vision, as well as noisy and down-scaled HR data variants. The problems are characterized by the loss of relevant data information during data subsampling, as well as during non-invertible low-pass f ltering processes. Non-linear mapping is used in the super-resolution method for the deduction of the inherent features that support the higher recognition efficiency of Nearest Neighbor (NN) classifiers for single LR face image recognition [18]. In the proposed framework, there are two aspects of Deep CNN for the nonlinear transformation of LR and HR face images into a common space [19]. The advancements in this field have led to improvements in the DL structure, and more models have been developed which focused on the optimization of the training and process methods. Although the accuracy of LRFR is improving, the runtime is proportionally reducing, thus, supporting its suitability for practical applications. Surveillance cameras usually supply face images for forensic studies, and these images are poor in quality and with LR. This often results in a significant reduction in the accuracy of face recognition. High-resolution images without distortions can be obtained from a simple up-sampling algorithm. To obtain good quality and HR images, there is a need to deploy a reconstructed super-resolution framework. Meanwhile, most SR frameworks require to estimate image motion, and in most cases, this is not achievable due to the required dimension for a proper feature extraction process. Hence, a major task in HR-LR face recognition is the development of a technology for image reconstruction which can directly generate HR images from their LR variants. The summary of the previous works related to this study is presented in Table 1.

Author	Technique	Advantages	Dataset
[16]	Deep network cascade (DNC)	Enhance high-frequency texture details of the partitioned patches in the input image.	
[18]	RBF model used to construct the nonlinear mapping relationship between the coherent features and CCA was applied to the classical PCA features	Coherent subspaces between the holistic features of HR and LR face images,	FERET ORL
[19]	Two branches of CNNs to map the high and low-resolution face images into a common space with nonlinear transformations	significant improvement and robustness against variations in expression, illumination, and age.	FERET
[20]	Locality Constrained Regression LLRLCR and Locality-constrained matrix	Better discriminative ability by obtaining full underlying structural information of gallery and probe data.	AR, Yale B
[21]	2D LDA.Classification using Nearest neighbor (NN)	Loss of the spatial information problem.	ORL
[22]	Neighborhood information and local geometric structure	It's very useful for discriminative analysis.	Extend Yale-B, CMU PIE
[23]	5Convolution layer and 3 FCL	Better results with R-CNN even without using annotations of face landmarks.	PASCAL

TABLE I. Summary of related work

III. METHODOLOGY

There are two phases of the ESPCN + CNN in this study, they are the super-resolution and recognition phases involved in face-recognition (Fig. 1). During the super-resolution phase, the proposed ESPCN was deployed to transform the LR image into an HR image, and during the recognition phase, CNN was used. The subsequent subsections provided a better explanation of these phases.

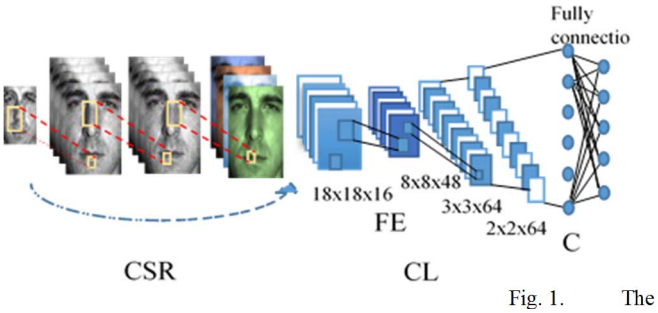

Fig. 1. The schematic of the suggested method (FE = feature extraction, CL = convolutional layer, C= classification, CSR = convolutional super resolution).

A. Super-Resolution Phase

Generally, single image super-resolution (SISR) is used to decompose an original HR image I^HR into its LR variant I^LRimage prior to the generation of the super-resolution image I^SR. For our evaluations with the ESPCN, the following setting was used: l = 3, (f1; n1) = (5; 64), (f2; n2) = (3; 32), and f3 = 3. These parameters were selected based on the inspiration from SRCNN's 3-layer 9-5-5 model and Eq. 1 and 2. The 17r _ 17 r pixel sub-images were extracted during the training phase from the training ground truth images I^HR, where r represents an upscaling factor. This downscaling process is notorious for producing I^{LR} from I^{HR}, and it's deterministic. At first, a Gaussian f lter was used to convolve the I^{HR} in a way that simulated a cameras' spread function. Then, the image was minimized by a factor r (called upscaling ratio). Both I^{HR} and I^{LR}can generally have C color channels; hence, are presented as tensor values of H ×W ×C for I^{HR}, and rH ×rW ×C for I^{LR} in size. A description of the f rst L−1 layers in a network comprised of L layers can be made as shown in (1) and (2).

$$f^1(I^{LR}; W_1, b_1) = \emptyset(w_1 * I^{LR} + b_1) \quad (1)$$

$$f^l(I^{LR}; W_{1:l}, b_{1:l}) = \emptyset(w_1 * I^{l-1}(I^{LR}) + b_1); \quad (2)$$

where

$W_l, b_l, l \in (1, L-1)$ = network biases and weights that can be learned;

W_l = 2-D convolution tensor of $n_{l-1} X n_l X\ k_l X k_l size$, where

n_l = number of features contained in layer l,

n_0= C, and k_l = size of the f lter at layer l.

The network bias b_l is a vector of nl length.

$\emptyset$ = fixed non-linearity function found in each element.

The LR feature maps are converted to an HR imageI^{SR} by the last layer f^Las illustrated in Fig. 2.

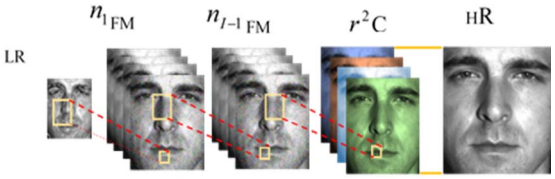

Fig. 2. Step in the conversion of LR images HR images using 2 convolution layers for feature map extraction

B. Convolution Neural Network (CNN)

The CNN is made up of three (3) processes, which include feature extraction (FE), feature mapping (FM), and sub-sampling (S) as shown in Fig. 3. The process of FM is performed in each network layer as it is the backbone of network layers. The other two processes (FE and S) are performed in the S layer but the convolution process is done in the region between S and the convolution layers. In this architecture, an 'L' layer CNN was initially applied directly to the low-resolution image, and thereafter, to a sub-pixel convolution layer which later helps in upscaling the low-resolution version maps to produce ISRCNN (an NN model which is made up of complex neuronal planes). In the CNN, the connections between two proximal layers have unsaturation attributes, and neurons contained in the same layer can have equal weights.

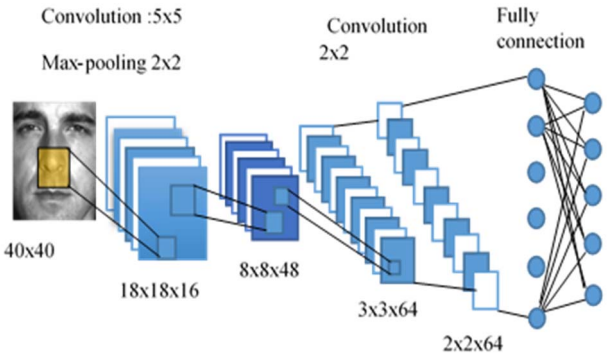

Fig. 3. The structure of the CNN model

IV. EXPERIMENTAL AND RESULTS

The method suggested in this study was evaluated based on its performance on YALE Face Database (YFD) and ORL dataset. During the evaluation, the deployed YALE database was composed of 16,128 face images of 28 different persons under 9 poses and 64 lightning conditions (Fig. 4) and ORL dataset consist of 280 face images of 20 different persons; each class has 14 persons as shown in (Fig. 5). First, the data was partitioned into two sets, one set for training and the other for testing. Furthermore, the training dataset was partitioned into two percentages (66% and 70%) in order to determine the appropriate training data percentage that will offer the best performance. The evaluations were carried out using the ESPCN + CNN and benchmarked in terms of performance against other methods (DCN, Twi Branches of CNN, and LLRLCR). The results in Table 2 showed that the proposed framework had a better performance accuracy compared to the benchmarked methods due to the ability of the proposed method to smoothen and convert the LR image to an HR image prior to recognition.

$$SD = \sqrt{\frac{\Sigma|x-\mu|}{N}} \quad (3)$$

where Σ = sum of, μ = dataset mean value, x = a value in the data set, and N = number of data points in the population.

Regarding the average value, the following mathematical formula was used for the arithmetic mean:

$$A = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4)$$

A = average mean
n = number of terms averaged
x_i = value of each item in the averaged numbers

Fig. 4. Face images contained in the YFD

TABLE II. COMPARISON OF THE ACCURACY PERFORMANCE UNDER DIFFERENT % OF TRAINING DATASET (YALE DATASET)

Method	Percentage of the training dataset					
	66%	***67%***	***68%***	***69%***	***70%***	***80%***
DCN [16]	93.00	93.50	94.50	94.00	92.10	94.50
Two branches of CNN [19]	91.50	91.80	90.90	92.90	93.10	93.10
LLRLCR [20]	90.60	91.90	92.90	92.30	91.90	92.90
ESPCN + CNN	95.20	95.30	94.50	94.20	94.60	95.30

Our models were evaluated using PSNR as a performance metric. The PSNR of SRCNN and Chen's models used as a benchmark was calculated using Matlab code. The models and their PSNR values were presented under each figure. The table showed that the evaluated method in each of the training dataset percentages attained the highest accuracy compared to the other methods. The highest accuracy achieved by the proposed method was 95.30% when 67% and 80% of the dataset were used as the training dataset. Also, the proposed method showed the most consistent performance in this experiment based on the values of the standard deviation in Table 3.

TABLE III. MEAN AND STANDARD DEVIATIONS OF ALL METHODS (YALE DATASET)

Method	Mean	Standard Deviations
DCN [16]	92.20	0.88
Two branches of CNN [19]	93.60	0.86
LLRLCR [20]	92.10	0.71
ESPCN + CNN	94.90	0.22

Table 4 showed the obtained accuracy performance when ORL dataset was used in the experiment. The dataset was partitioned in a similar manner to the previous experiment.

Fig. 5. Face images contained in the ORL dataset

TABLE IV. COMPARISON OF THE ACCURACY PERFORMANCE UNDER DIFFERENT % OF TRAINING DATASET (ORL DATASET)

Method	Percentage of the training dataset					
	66%	***67%***	***68%***	***69%***	***70%***	***80%***
DCN [16]	89.8	88.4	87.3	89.2	87.5	89.8
Two branches of CNN [19]	91.6	90.4	91.6	92.7	90.4	92.7
LLRLCR [20]	89.5	90.5	88.5	90.5	91.4	91.4
ESPCN + CNN	92.5	91.6	93.5	91.5	92.1	93.5

TABLE V. MEAN AND STANDARD DEVIATIONS OF ALL METHODS (ORL DATASET)

Method	Mean	Standard Deviations
DCN [16]	88.44	1.07
Two branches of CNN [19]	91.34	0.97
LLRLCR [20]	90.08	1.11
ESPCN + CNN	92.24	0.81

From Table 4, the proposed method achieved the highest accuracy performance (between 91.5% to 93.5%) compared to the other methods which achieved an accuracy range of 88.4% to 92.7%. However, the accuracy of the proposed method was lower than the previous experiments. This is possibly due to the various inferences such as angles, and expressions in the face images in ORL dataset. Also, Table 5 showed the proposed method to have achieved the most consistent performance compared to the other methods as evidenced by its low standard deviation of 0.81.

V. CONCLUSION

This study concisely investigated and addressed the problem of low-resolution images and the model techniques to mitigate its effects in face recognition. The DL technique was therefore proposed for direct image super-resolution and accurate face recognition. The suggested method for super-resolution was based on the use of sub-CNN coupled with a Gaussian filter for image smoothing prior to recognition. This method is a combination of these two methods (ESPCN and CNN). The performance of the ESPCN + CNN was thereafter evaluated and compared to other traditional methods. The evaluation results indicated that the developed model achieved better accuracy when compared to the benchmark methods.

ACKNOWLEDGMENT

This research was supported by Universiti Malaysia Pahang Research grant PGRS180306 and RDU190315 and Al-maarif University College.

REFERENCES

[1] M.W. Thornton, P.M. Atkinson, and D. Holland, "Sub˜ pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super˜ resolution pixel˜ swapping". International Journal of Remote Sensing, 2006. 27(3): p. 473-491.

[2] B.K. Gunturk, "Eigenface-domain super-resolution for face recognition. IEEE transactions on image processing". 2003. 12(5): p. 597-606.

[3] S. Peled and Y. Yeshurun, "Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging". Magnetic resonance in medicine, 2001. 45(1): p. 29-35.

[4] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, K. and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch." In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 9-16). Springer, Berlin, Heidelberg, 2013.

[5] S. Borman and R.L. Stevenson. "Super-resolution from image sequences-a review. in Circuits and Systems", 1998. Proceedings. 1998 Midwest Symposium on. 1998. IEEE.

[6] S. Farsiu, M. D. Robinson, M. Elad and P. Milanfar. "Fast and robust multiframe super resolution". IEEE transactions on image processing, 13(10), 1327-1344. 2004.

[7] S. Awang, J. Sulaiman, N. Noor, K. Mohd and L. Bayuaji, L. "Comparison of accuracy performance based on normalization techniques for the features fusion of face and online signature". Advanced Science Letters, 23(11), 2017, 11233-11236.

[8] M. Furqan, A. Embong, S. Awang, S. W. Purnami and S. Sembiring. "Smooth support vector machine for face recognition using principal component analysis". Proceeding 2nd International Conference On Green Technology and Engineering (ICGTE), 2009, 193-198.

[9] S. Awang, R. Yusuf and R. Arfa. "Multimodal biometrics system: A feature level fusion of physical and behavioral biometric to improve person's identity recognition" ICIC Express Letters, 6(2), 2012, 543-548

[10] Bowyer, K.W., K.P. Hollingsworth, and P.J. Flynn, "A survey of iris biometrics research: 2008–2010, in Handbook of iris recognition". 2016, Springer. p. 23-61.

[11] S. Awang, R. Yusof, M. F. Zamzuri, M. F. and R. Arfa, "Feature level fusion of face and signature using a modified feature selection technique". in Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on. 2013. IEEE.

[12] M. A. Talab, S.N.H.S. Abdullah, and M.H.A. Razalan. "Edge direction matrixes-based local binary patterns descriptor for invariant pattern recognition". in Soft Computing and Pattern Recognition (SoCPaR), 2013 International Conference of. 2013. IEEE.

[13] S. Abe and T. Inoue, "Fuzzy support vector machines for multiclass problems", in ESANN. 2002. p. 113-118.

[14] T. G. Kumar. and V.P. Vijayan."A multi-agent optimal path planning approach to robotics environment". in Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on. 2007. IEEE.

[15] L. Shaoxin, S. Shiguang, and K. Meina, "Application of cross-attitude face recognition based on multi-view unified subspace". Journal of Police Technology, 2014: p. 8-11.

[16] Z. Cui, H. Chang, S. Shan, B. Zhong and X. Chen. "Deep network cascade for image super-resolution". in European Conference on Computer Vision. 2014. Springer.

[17] M. A. Talab, H. Tao, and A.A.M. Al-Saffar, "Review on Deep Learning-Based Face Analysis". American Scientific Publishers, 2017. 24(10): p. 7630-7635.

[18] H. Huang and H. He, "Super-resolution method for face recognition using nonlinear mappings on coherent features". IEEE Transactions on Neural Networks, 2011. 22(1): p. 121-130.

[19] E. Zangeneh, M. Rahmati and Y. Mohsenzadeh, "Low Resolution Face Recognition Using a Two-Branch Deep Convolutional Neural Network Architecture". arXiv preprint arXiv:1706.06247, 2017.

[20] G. Gao, Z. Hu, P. Huang, M. Yang, Q. Zhou, S. Wu and D. Yue. "Robust low-resolution face recognition via low-rank representation and locality-constrained regression". Computers & Electrical Engineering, 2018.

[21] D. Zhao, Z. Chen, C. Liu and Y. Peng. "Two-dimensional Linear discriminant analysis for low-resolution face recognition". in Chinese Automation Congress (CAC), 2017. 2017. IEEE.

[22] J. Jiang, R. Hu, Z. Wang and Z. Cai. "CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images". Signal Processing, 2016. 124: p. 162-172.

[23] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation". IEEE Transactions on Multimedia, 2015. 17(11): p. 2049-2058.

[24] W.W. Zou and P.C. Yuen, "Very low resolution face recognition problem". IEEE Transactions on Image Processing, 2012. 21(1): p. 327-340.