Proceedings of the 12th International Conference of Control Systems, and Robotics (CDSR 2025)
July 13, 2025 - July 15, 2025 | Imperial College London Conference Center, London, United Kingdom
Paper No. 103
DOI: 10.11159/cdsr25.103

Enhancing Low-Resolution Face Recognition with Deep Learning Techniques

Chia-Sheng Horng[1], Pin-Siang Huang[2], J. Supardi[3], and S. Horng[4]
[1]Department of Chemical Engineering, National Taiwan University of Science and Technology, Taipei Taiwan
43, Section 4, Kee-Lung Road, Taipei 106, Taiwan
b10706149@mail.ntust.edu.tw; d11115006@mail.ntust.edu.tw; julian@unsri.ac.id; horngalan@gmail.com
[2] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei Taiwan
43, Section 4, Kee-Lung Road, Taipei 106, Taiwan
[3]Department of Informatics Engineering, Sriwijaya University, Indonesia
[4]Department of Computer Science and Information Engineering, Asia University, Taichung Taiwan

Abstract **-** Human-robot collaboration in factory automation improves efficiency and safety, especially in chemical plants. A framework proposed in this paper combining super-resolution, facial recognition, and image segmentation monitors personnel, prevents hazardous area entry, and sends real-time alerts. The proposed system enhances low-resolution face images using two neural networks for feature extraction and reconstruction, optimized with momentum, Adam, and RMSprop for fast convergence. Evaluated on open-set datasets, it outperforms existing methods in both PSNR and SSIM, respectively. Facial recognition is then implemented via FaceNet with the reconstructed high-resolution images and then combined with image segmentation to prevent human entry into hazardous areas, and send real-time alerts, improving automation and reducing costs while offering scalability to other industries.

Keywords**:** Chemical plants, Super-resolution, Face hallucination, Image reconstruction, Low-resolution face images,

1. Introduction

Usually, in chemical plants, where numerous pieces of equipment and experimental chemicals are installed, safety is a paramount and critical issue, especially when personnel accidentally enter hazardous areas. Despite the use of traditional monitoring and recording devices, common human factors, such as forgetting to wear safety badges or failing to use monitoring devices correctly, may not be addressed in real time. To enhance risk control and safety in the plant, employing facial recognition technology to track personnel will make chemical plants more automated and globally advanced.

Using facial recognition to monitor hazardous areas in chemical plants can be carried out through the following steps: First, the hazardous areas should be classified. Hazardous areas within the plant must be meticulously classified and assessed by evaluating their accessibility to determine if unauthorized personnel can easily enter, classifying areas based on their level of danger (e.g., control rooms, safety zones, and gas storage areas), analyzing the speed of potential changes in hazardous incidents to identify areas requiring real-time monitoring, and examining detailed maps and personnel flow during busy periods to define entry and exit regulations as well as monitoring requirements. Through the above analysis, the plant's hazardous areas can be categorized into different safety levels, laying the foundation for subsequent operations. Then adopt segmentation technology to precisely divide internal areas of the plant. By analyzing maps and surveillance images, the plant can be segmented into safe areas and hazardous areas. Then, integrate facial recognition technology to input plant personnel data into the monitoring system, enabling real-time tracking of whether personnel have entered hazardous areas. Notifications can alert workers to avoid danger, and warnings can be activated. By combining facial recognition and segmentation technologies, human-factor-related safety risks can be significantly reduced, and the safety and efficiency of chemical plants during busy operations can be greatly improved.

The height of factories is typically quite high, causing objects to lose clarity as the distance increases. To enhance image clarity, high-resolution cameras can be used, but they are costly. To reduce costs, standard cameras can be utilized in

combination with super-resolution techniques to improve image quality. This study adopts the latter approach to minimize factory construction costs.

We can use super-resolution techniques to reconstruct high-resolution images from the corresponding low-resolution images. Super-resolution technique is one of the major research topics in image processing as it can be widely used in medical imaging, satellite imaging, face recognition, high-definition television (HDTV) and so on.

Face hallucination is categorized into three types: interpolation-based [1], reconstruction-based [2], and learning-based methods [3]. Interpolation-based methods are simple but produce poor-quality results. Reconstruction-based methods offer better quality but are time-consuming. Learning-based methods deliver the best results but require a lengthy training phase and a testing phase.

Convolutional Neural Network (CNN) is popularly used in the learning-based method. For image processing, we can use CNN to process the super-resolution image. The Super Resolution Convolution Neural Networks (SRCNN) method was proposed by Dong, et. al. [4], then successively followed by many other researchers. The Very Deep Super Resolution (VDSR) method [5] and the Deeply-Recursive Convolutional Network (DRCN) [6] were proposed by Kim et. al. Yamanaka et. al. proposed Deep CNN with Residual Net, Skip Connection and Network in Network (DCSCN) [7]. To further improve PSNR/SSIM of the facial hallucination obtained from small size images, some researchers had proposed their methods using other various convolution neural networks including iterative back-projection [8], and recurrent network and reinforcement learning [9]. The image quality obtained by these methods was better than that achieved by CNNs previously.

Building on the successes of deeper CNN layers by Kim et al. [5] [6] and skip connections by Yamanaka et al. [7], these methods improve super-resolution image quality. Color information also aids in generating better results, and deep CNNs effectively model complex details. This study proposes a CNN architecture for face hallucination, consisting of two parts: a feature extraction network and a reconstruction network, both with deeper layers and new loss functions for training.

As for exactly identifying the internal areas of the plant, deep learning architecture is also used. Y. Wang et al. [32] reviewed both traditional and modern object segmentation approaches, comparing their strengths, weaknesses, and utilities. For deep learning, X. Lyv et al. [33] introduced DeepMerge that integrates deep learning and region adjacency graphs for effective segmentation of large very high spatial-resolution remote sensing images.

2. Related Work

2.1. Super-Resolution Image

The higher the resolution of an image, the denser the pixels in the image. Visually, high-resolution images look more detailed than low-resolution images. Fig. 1. shows the visualization of a face image with different resolutions.

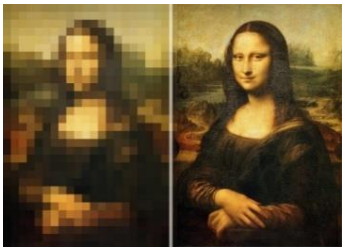


Fig. 1. The visualization of a face image with different resolutions (https://sapc.co.uk/high-resolution-images/)

Let χ be a high-resolution image and ψ the corresponding low-resolution image, where S and θ are the sampling operator and low pass filter. The relation is then expressed in Eq. (1).

$$\psi = S\,\theta\,\chi \tag{1}$$

Until now, there are a lot of methods which have been proposed to produce super-resolution images. Especially for deep learning, SRCNN [4] is the first method using CNN for super-resolution. This method is often used as a benchmark in developing methods to produce super-resolution using deep learning.

The next method named as Efficient Sub-Pixel Convolutional Neural Network (ESPCN) was proposed in [10]. It expands the output feature channels to save extra points for increasing resolution and then reset these points to get high resolution output via certain mapping criteria. Following up, the Fast Super Resolution Neural Networks (FSRCNN) was proposed in [11]. This method is the first to use normal deconvolution to reconstruct high-resolution images from a corresponding low-resolution feature folder. Furthermore, Kim et. al. proposed two methods, Very Deep Super Resolution (VDSR) [5] and Deeply-recursive Convolutional Network [6]. The former uses a very deep convolution network, which consists of 20 CNN layers, while the latter also uses very deep and recursive processing networks. The next method is the Deep Recursive Residual Network (DRRN) which was proposed in [12]. It rearranges unit residues in the recursive topology to form recursive blocks. The next method is the Laplacian Super-Resolution Network (LapSRN) which was proposed in [13]. In this method, to produce the high-resolution (HR) image output, the Laplacian pyramid structure is used and LapSRN is composed by two major branches; one for feature extraction and the other is for image reconstruction. Furthermore, in [14], they proposed two methods using the generative adversarial network (GAN) and named as SRGAN and Super Resolution Residual Network (SRResNet). Both PSNR and SSIM obtained from SRResNet are better than those of SRGAN. The next method named as Memory Network (MemNet) was proposed in [15]. Here, the normal convolution on basic DenseNet blocks has been replaced by a recursive residual unit. In [16], they proposed a method named as Residual Dense Network (RDN). This method has been developed using RDN blocks. The basic convolution unit in the block has densely connected like DenseNet, and at the end of the RDN block, the bottleneck layer is used to follow residual learning throughout the block. In [17], Lim and Lee proposed a method named as Enhanced Deep Residual Networks (EDSR). This method makes improvement to the previous method from three aspects, namely: (1) eliminating BN; (2) increasing the number of output parameters on a large scale for each layer; and (3) initializing parameters with a pre-trained x2 network when training x3 and x4 models. In [18], Haris et. al. proposed the Deep Back Projection Network (DBPN). This method uses the deep CNN architecture to perform iterative back-projection simulations and further improve performance with dense connections. Note that this method can reach the x8 scale. A fast and accurate image super resolution method with shallow network was proposed by Yamanaka et. al. in [7]. This method uses a relatively shallow network, which only has eleven CNN layers, seven layers for feature extraction and four layers for reconstruction. The PSNR and SSIM of the methods listed above are shown in Table 1.

Tabel 1. PSNR and SSIM values for deep learning-based methods applied to a super-resolution image with a scale factor of 4.

Models	**PSNR/SSIM**	**Train data**	**Layers**
SRCNN [4]	*30.49/0.8628*	*ImageNet Subset*	*3*
ESPCN [10]	*30.90/-*	*ImageNet Subset*	*3*
FSRCNN [11]	*30.71/0.8657*	*G100+Yang91*	*5*
VDSR [5]	*31.35/0.8838*	*G200+Yang91*	*20*
DRCN [6]	*31.53/0.8838*	*Yang91*	*20*
DRRN [12]	*31.68/0.8888*	*G200+Yang91*	*54*
LapSRN [13]	*31.54/0.8855*	*G200+Yang91*	*24*
SRResNet [14]	*32.05/0.9019*	*ImageNet Subset*	*16*
SRGAN [14]	*30.64/0.8701*	*ImageNet Subset*	*37*
MemNet [15]	*31.74/0.8893*	*G200+Yang91*	*80*
RDN [16]	*32.61/0.9003*	*DIV2K*	*149*
EDSR [17]	*32.62/0.8984*	*DIV2K*	*69*
MDSR[17]	*32.60/0.8982*	*DIV2K*	*166*
DBPN [18]	*32.47/0.898*	*DIV2K+Flickr+ImageNetb*	*46*
DCSCN [7]	*33.05/0.9126*	*Berkeley Segment Dataset*	*11*

2.2. Face Hallucination

Researchers propose alternatives to SR methods as they often fail on very low-resolution face images. Some of them are listed in the following. Cao et al. [19] used deep reinforcement learning. Zhang et al. [20] used Super-Identity CNN, and Grm et al. [21] adopted GANs.

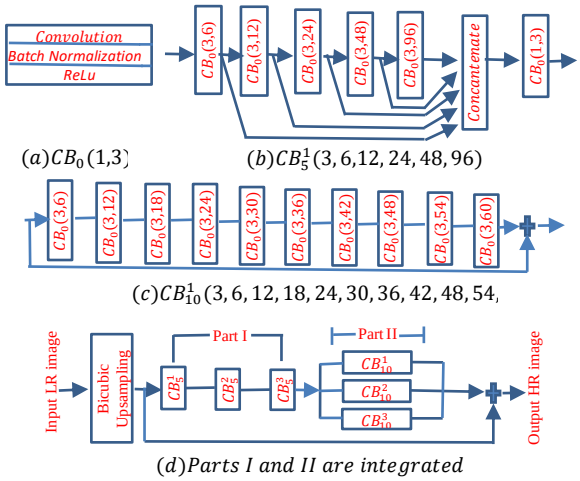
3. Face Hallucination Using Deep Learning

3.1. The Architecture

As mentioned previously, the proposed method in this paper consists of two main components: the feature extraction network and the reconstruction network. The former uses convolution layers to capture the most relevant image features, while the latter employs deep layers to generate the face hallucination.

First, some basic convolution blocks are defined, each consisting of a convolutional layer, a batch normalization layer, and a ReLU activation function. These blocks can vary based on the kernel size and the number of channels used. $CB_0(r, c)$ is a control block whose kernel size is $r \times r$and channel number is c. For example, $CB_0(1,3)$ means $r = 1$ and $c = 3$. CB_5^i $(r, 6i, 12i, 24i, 48i, 96i)$ represents a cascade of five CB_0 blocks, each with a kernel size of $r \times r$ and the channel numbers are $6i$, $12i$, $24i$, $48i$, and $96i$. CB_{10}^i $(r, 6i, 12i, 18i, 24i, 30i, 36i, 42i, 48i, 54i, 60i)$ represents a cascade of five CB_0 blocks, each with a kernel size of $r \times r$ and the channel numbers are $6i$, $12i$, $18i$, $24i$, $30i$, $36i$, $42i$, $48i$, $54i$, and $60i$.

The detailed description of each part is shown in Figure 2; *PartI (Feature Extraction Network)*: The feature extraction network is composed of three cascaded blocks: CB_5^1, CB_5^2 and CB_5^3 and *Part II (Reconstruction Network)*: The reconstruction network is composed of three cascaded blocks: CB_{10}^1, CB_{10}^2 and CB_{10}^3.



(a) $CB_0(1,3)$ (b) $CB_5^1(3, 6, 12, 24, 48, 96)$

(c) $CB_{10}^1(3, 6, 12, 18, 24, 30, 36, 42, 48, 54,$

(d) *Parts I and II are integrated*

Fig. 2. System Architecture

3.2. Training

We used the database from Yang et.al. [22] and the database from Berkeley Segmentation [23] for training and both databases contain 96 images and 200 images, respectively. Our training process is divided by a scale of factors 2, 3, 4, 8, and 16. Each scale represents the desired level of image resolution enhancement. For example, a scale factor of 2 is used for doubling the input image resolution (x2), a scale factor of 3 for tripling it (x3), and so on.

In general, the training is similar to other CNN training processes by minimizing the loss function $\mathcal{L}$. Here, we use the loss function defined in Eq (2).

$$\mathcal{L} = (\sum_{i=1}^{n} \|a(x) - g\|)^2 \tag{2}$$

where $a(x)$ is the output of the network, g is the ground truth image, and n is the number of data.

We used the RAdam optimizer for training; the RAdam optimizer was configured with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and eps = 1e-8, and weight decay was set to 1e-4. Training ran for 16,000 epochs, batch size 16, with the learning rate starting at 1e-3 and reducing to 5e-4. The learning rate was reduced to 90% if the loss did not change for 20 epochs, stopping when the final rate was reached.

4. Experimental Results

4.1. Data Sets

In this research, we evaluate the proposed method using six publicly available databases. These databases encompass a wide range of variations, including pose, expression, lighting, background, and illumination. A summary of the six databases is provided below:

- PubFig [24] is a face database with 58,797 images of 200 people. The images vary in size, pose, and background.
- BioID [25] has 1,521 grayscale images (each of 384x286 pixels) with varied backgrounds, face sizes, and lighting.
- LFW database [26] has 13,233 images (each of 512x512) pixels with varied backgrounds and poses.
- CelebA [27] has 10,177 identities and 202,599 face images.
- SCface [28] has 4160 static images of 130 subjects in visible and infrared spectrum.
- The extended Yale Face Database B [29] has 16128 images of 28 human subjects, 9 poses and 64 illumination conditions.

4.2. Experimental Setting

To evaluate the proposed method, we conduct experiments using two distinct settings. The first is limited to the face area, adhering to the protocol established by [8]. The second is expanded to include both the face and neck, following the protocol outlined by [9]. These protocols are designed to ensure a fair comparison between the proposed method and previously existing approaches.

For the first experiment, we use the LFW database to evaluate the proposed method, following the protocol in [8]. The original 250x250 images are cropped to 128x128 for the face area. Downsampling is then applied with scales of 2, 3, 4, and 8, resulting in low-resolution images of 64x64, 42x42, 32x32, and 16x16 pixels, respectively. Each downsampled image is reconstructed to the ground truth size of 128x128 pixels for face hallucination.

4.3. Comparison with Existing Methods in the First Experimental Setting

We compared our method to state-of-the-art face hallucination methods with a scale factor of 4, using released codes and consistent settings. We used the bicubic as a base benchmark and compared it to some other existing methods i. e. PRDRN [3], SRCNN [4], VDSR [5], DCSCN [7], SRCNN-IBP [8], SRGAN [14], DRL [19], ScSR [22]. The comparison results in terms of PSNR and SSIM are shown in Table 2, while the visual comparisons are shown in Fig. 3.

Figure 3 and Table 2 show that our method outperforms previous methods in PSNR, SSIM, and visual clarity of reconstructed face images.

GT Ours NN Bicubic RCNN SRCNN-IBP

Fig. 3. The visual comparisons between the proposed method and some existing methods by setting scale factor to 4.

Table 2. Comparison of the proposed method with other approaches using downsampling at a scale factor of 4.

Method	PSNR	SSIM
Bicubic	29.75	0.831
DCSCN [7]	30.38	0.814
DRL [19]	32.93	0.910
PRDRN [3]	31.26	0.9049
ScSR [22]	29.77	0.841
SRCNN [4]	30.84	0.856
SRCNN-IBP [8]	30.90	0.859
SRGAN [14]	30.94	0.8936
VDSR [5]	31.14	0.9033
Our Method	**33.16**	**0.9582**

4.4. Comparison with Existing Methods in the Second Experimental Setting

Following the protocol outlined by Shi et al. [9], six public databases: CelebA, BioID, LFW, PubFIG, SCFace, and Extended Yale B were used for the second experiment. Here, the size of image cropping obtained is 160x120 pixels for all

databases except LFW database which is 128x128 pixels. Furthermore, we use downsampling factors 4, 8, and 16 to make LR images. Thus, the images input to system for each scale are 40x30, 20x15, and 10x8 except images from LFW database whose input sizes are 32x32, 16x16, and 8x8, respectively. To obtain high-resolution (HR) images, the LR inputs were upscaled using magnification factors of x4, x8, and x16, respectively.

The performance of our results was compared with several existing methods, including Bicubic, SRCNN [4], VDSR [5], BiCNN [30], GLN [31], SRGAN [14], Attention-FH [9], based on PSNR and SSIM metrics.

The PSNR and SSIM results are presented in two separate tables. As shown in Tables 3, the proposed method outperforms the previous methods in PSNR, producing reconstructed images closer to the ground truth. Meanwhile, Table 4 shows the SSIM of the reconstructed image using the proposed method. Compared to the SSIM values obtained by previous methods, some of our results show slight improvements, while overall, all cases demonstrate better performance than the previous methods.

Table 3. Comparison of PSNR values with other methods.

Database	**Scale**	**Method**							
		Bicubic	**SRCNN**	**VDSR**	**BiCNN**	**GLN**	**SRGAN**	**Attention-FH**	**Our**
CelebA	x4	25.76	26.01	28.38	24.93	29.92	28.17	30.56	32.401
	x8	21.84	22.64	24.46	21.32	25.48	24.37	26.14	27.461
	x16	18.78	19.80	20.07	19.01	21.20	21.99	22.63	22.845
BioID	x4	24.59	25.71	29.38	23.16	26.71	28.16	33.38	33.851
	x8	20.24	21.71	23.95	20.14	22.03	23.23	27.81	27.852
	x16	17.15	18.45	19.41	17.11	19.85	21.59	23.48	23.577
LFW	x4	26.79	28.94	32.11	26.60	30.34	31.41	32.93	33.169
	x8	21.92	23.92	24.12	22.62	24.51	25.49	27.81	28.215
	x16	19.95	21.34	22.4	20.82	22.44	23.01	23.13	23.535
PubFIG	x4	24.76	25.21	28.05	24.35	26.12	27.44	28.87	28.895
	x8	20.75	21.31	21.94	20.69	21.33	23.45	23.59	23.612
	x16	17.89	18.48	19.28	18.13	18.73	20.25	20.31	20.968
SCFace	x4	26.15	26.54	31.59	26.04	29.84	30.10	34.01	34.121
	x8	20.83	21.68	24.12	20.14	24.10	24.72	26.04	26.313
	x16	17.15	18.45	20.32	17.11	18.28	20.63	22.19	22.316
Yale B	x4								33.021
	x8	24.67	25.69	26.57		25.98	27.32	28.89	29.011
	x16								22.618

Table 4. Comparison of SSIM values with other methods.

Database	Scale	Method							
		Bicubic	SRCNN	VDSR	BiCNN	GLN	SRGAN	Attention-FH	Our
CelebA	x4	0.8565	0.8634	0.8890	0.9148	0.8956	0.8956	0.9211	0.93038
	x8	0.7514	0.7762	0.8077	0.8290	0.8290	0.8066	0.8411	0.84381
	x16	0.6511	0.7872	0.6980	0.7331	0.7331	0.7552	0.7628	0.76928
BioID	x4	0.8538	0.8767	0.9066	0.8926	0.8926	0.9012	0.9571	0.95803
	x8	0.7405	0.7800	0.7977	0.7803	0.7803	0.8162	0.9072	0.90768
	x16	0.6487	0.7055	0.7104	0.7379	0.7379	0.7849	0.8336	0.83621
LFW	x4	0.8947	0.9069	0.9300	0.8982	0.9151	0.9384	0.9571	0.95823
	x8	0.7824	0.8314	0.8391	0.7903	0.8405	0.8278	0.9072	0.90821
	x16	0.6771	0.7454	0.7595	0.7243	0.7788	0.7722	0.7824	0.78387
PubFIG	x4	0.8454	0.8581	0.8942	0.8748	0.8748	0.8892	0.9099	0.91001
	x8	0.7337	0.7604	0.7587	0.7669	0.7669	0.8081	0.8121	0.82208
	x16	0.6360	0.7872	0.7207	0.7151	0.7151	0.7297	0.7346	0.73572
SCFace	x4	0.8813	0.8906	0.9321	0.9185	0.9185	0.9205	0.9513	0.95162
	x8	0.7518	0.7893	0.8195	0.8318	0.8318	0.8437	0.8675	0.86889
	x16	0.6487	0.7055	0.7236	0.6955	0.6955	0.7586	0.7977	0.79883
Ext Yale B	x4								0.9233
	x8								0.8721
	x16								0.8152

4.5. Testing to Face Recognition

Developed by Google, FaceNet utilizes a deep convolutional neural network to map facial images into a compact Euclidean space, enabling highly accurate facial recognition and outperforming many contemporary models. In our experiment, we first retrained FaceNet using a high-resolution dataset to enhance its performance. Considering that people at a distance have lower resolution, we restricted the face ROI to 8×8 pixels during testing. Despite this limitation, the recognition rate degraded by up to 7% compared to high-resolution, full-size face images. Nevertheless, the performance remains practical and viable for real-world applications.

5. Image Segmentation

Numerous image segmentation techniques have been introduced so far [32], [33]. The goal of this paper is to utilize existing image segmentation techniques to divide the interior of a factory into regions, then classify these regions based on safety levels. Subsequently, the method proposed in this study will be used to accurately identify the specific locations of personnel and send timely warning messages as needed. Using the segmentation technique, the internal structure of a plant is segmented, and a warning system is established by integrating the previously stated face recognition method. A plant with greater safety is designed and illustrated in Fig. 4.

(a) (b)

Fig. 4. A plant with greater safety system (a) a plant is segmented (b) workers are identified.

6. Conclusion

In chemical factories, safety is critical, especially in preventing unauthorized access to hazardous areas. This study enhances monitoring using super-resolution for clearer images, facial recognition for identifying personnel, and image segmentation to track locations and send real-time alerts. The approach improves automation, safety, and scalability to other factories.

References

[1] J. Choi, M. Kim, and S. Member, "Super-interpolation with edge-orientation-based mapping kernels for low complex 2 × upscaling," *IEEE Tran. Image Process.*, vol. 25, no. 1, 2016, pp. 469-483.

[2] Y. Zhang, Q. Fan, F. Bao, Y. Liu, and C. Zhang, "Single-image super-resolution based on rational fractal interpolation," *IEEE Trans. Image Process.*, vol. 27, no. 8, 2018, pp. 3782-3797.

[3] T. A. O. Lu, Y. Zhang, and K. A. I. Liu, "Parallel region-based deep residual networks for face hallucination," *IEEE Access*, vol. 7, 2019, pp. 81266-81278.

[4] C. Dong, C. C. Loy, and K. He, "Image super-resolution using deep convolutional networks," 2015. pp. 1-14. Retrieved from https://arxiv.org/pdf/1501.00092.pdf.

[5] J. Kim, J.K. Lee., & K.M. Lee., "Accurate image super-resolution using very deep convolutional networks," 2016. Retrieve from https://arxiv.org/abs/1511.04587.

[6] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," 2016. Retrieve from: https://arxiv.org/abs/1511.04491.

[7] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super-resolution by deep CNN with skip connection and network in-network," pp. 1-9. 2017. https://arxiv.org/ftp/arxiv/papers/1707/1707.05425.pdf

[8] D. Huang and H. Liu, "Face hallucination using convolutional neural network with iterative back projection," *Proc. 11th Chinese Conference*, CCBR 2016, vol. 9967, pp. 167-175, 2016.

[9] Y. Shi, G. Li, Q. Cao, K. Wang, and L. Lin, "Face hallucination by attentive sequence optimization with reinforcement

learning," *IEEE Trans. Pattern Anal. Mach. Intell. (early access)*, May 2019. pp. 1-15.
[10] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," 2016. Retrieve from: https://arxiv.org/abs/1609.05158
[11] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network." 2016. Retrieve from: https://arxiv.org/abs/1608.00367
[12] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," *IEEE Conference on Computer Vision and Pattern Recognition Image*, pp. 2790-2798, 2017.
[13] W. Lai, N. Ahuja, M. Yang, and V. Tech, "Deep laplacian pyramid networks for fast and accurate super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell*, Vol.41, No. 11, 2019, pp. 2599-2613.
[14] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681-4690. 2017.
[15] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet : A persistent memory network for image restoration," *IEEE International Conference on Computer Vision (ICCV)*, pp. 4539-4547, 2017.
[16] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution." *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
[17] B. Lim and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.
[18] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1664-1673, 2017.
[19] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li, "Attention-aware face hallucination via deep reinforcement learning," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, pp. 1656-1664, 2017.
[20] K. Zhang, Z. Zhang, C. Cheng, H. Winston, Y. Qiao, and T. Zhang, "Super-identity convolutional neural network for face hallucination," *ECCV 2018*, pp. 1-16, 2018.
[21] K. Grm, S. Member, W. J. Scheirer, and S. Member, "Super-resolution and identity priors," *IEEE Trans. Image Process.*, vol. 29, 2020, pp. 2150-2165.
[22] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, 2010, pp. 2861-2873.
[23] M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, 2011, pp. 898-916.
[24] PubFIG: N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 365-372.
[25] O. Jesorsky, K. J. Kirchberg, and R. Frischholz, "Robust face detection using the hausdorff distance," in AVBPA, 2001, pp. 90-95.
[26] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in ICCV, 2007.
[27] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in ICCV, 2015, pp. 3730-3738.
[28] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," Multimedia tools and applications, vol. 51, no. 3, pp. 863-879, 2011.
[29] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on PAMI*, vol. 23, no. 6, pp. 643-660, 2001.
[30] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild,"AAAI, 2015, pp. 3871-3877.
[31] O. Tuzel, Y. Taguchi, and J. R. Hershey, "Global-local face upsampling network," arXiv:1603.07235, 2016.
[32] Y. Wang, U. Ahsan, H. Li, M. Hagen, "A comprehensive review of modern object segmentation approaches," Foundations and Trends in Computer Graphics and Vision: Vol. 13: No. 2-3, 2023, pp 111-283.
[33] X. Lyu, C. Persello, X. Huang and D. Ming, "DeepMerge: Deep-learning-based region-merging for image segmentation," 2023. DOI: 10.48550/arXiv.2305.19787