Postprint: Research on Low-Resolution Face Recognition Algorithms for Security Surveillance Scenarios

Authors: Lu Feng, Zhou Lin, Cai Xiaohui

Date: 2020-09-28T00:00:00+00:00

Abstract

To address the problem of low face recognition accuracy caused by poor-quality face images and loss of detail information in security monitoring scenarios, we propose a low-resolution face recognition algorithm based on super-resolution reconstruction. This algorithm comprises two sub-networks: a super-resolution reconstruction sub-network and a face recognition sub-network, which respectively implement super-resolution reconstruction of low-resolution face images and extraction of face features. The algorithm first achieves wide activation by increasing the number of feature maps before the activation function in the super-resolution reconstruction sub-network, ensuring effective transmission of information flow to reconstruct high-resolution face images containing richer detailed information. Then, during training, it combines image content loss and identity loss to preserve more identity information while reconstructing images, thereby enabling the extracted face features to possess stronger discriminability. Experimental results demonstrate that the algorithm improves the accuracy of low-resolution face recognition and outperforms traditional algorithms on the surveillance face dataset QMUL-SurFace.

Full Text

Preamble

Vol. 38 No. 3
Application Research of Computers
Accepted Paper

Research on Low-Resolution Face Recognition Algorithm for Security Surveillance Scenarios

Lu Feng[1], Zhou Lin[2]†, Cai Xiaohui[2]
(1. Unit 78090, Chinese People' s Liberation Army, Chengdu 610054, China;
2. State Key Laboratory of Integrated Services Networks, Xidian University, Xi' an 710071, China)

Abstract: To address the problem of low face recognition accuracy caused by poor image quality and loss of detailed information in face images obtained from security surveillance scenarios, this paper proposes a low-resolution face recognition algorithm based on super-resolution reconstruction. The algorithm comprises two sub-networks: super-resolution reconstruction and face recognition, which respectively perform super-resolution reconstruction of low-resolution face images and extraction of facial features. The algorithm first increases the number of feature maps before the activation function in the super-resolution reconstruction sub-network to achieve wide activation, ensuring effective transmission of information flow and reconstructing high-resolution face images containing more detailed information. During training, it combines image content loss and identity loss to retain more identity information while reconstructing images, making the extracted facial features more discriminative. Experimental results demonstrate that the algorithm improves the accuracy of low-resolution face recognition and outperforms traditional algorithms on the surveillance face dataset QMUL-SurFace.

Keywords: security surveillance; super-resolution reconstruction; wide activation; identity loss

0 Introduction

With the rapid development of society, security technology has become an indispensable component in building smart cities. As one of the most representative biometric recognition technologies, face recognition has been widely deployed in security applications due to its simplicity and efficiency. Existing face recognition algorithms have achieved high accuracy in scenarios where subjects actively cooperate. However, unlike constrained face recognition, security surveillance face recognition focuses more on performance in real-world unconstrained scenarios [1]. Face images captured by surveillance videos are predominantly low-resolution, and directly applying existing face recognition algorithms leads to a significant drop in accuracy, rendering them impractical. Therefore, how to perform fast and accurate identity authentication for low-resolution faces in surveillance videos is a major challenge currently facing the face recognition field.

Four main approaches address the low-resolution face recognition problem: (1) downsampling high-resolution face images in the database, (2) mapping both

low-resolution query images and high-resolution gallery images to a unified feature space before recognition [2], (3) extracting resolution-robust features unaffected by resolution variations, and (4) performing super-resolution reconstruction on low-resolution face images. Among these, downsampling causes loss of useful information, unified feature space mapping introduces noise, and resolution-robust feature extraction performs poorly when resolution differences are substantial. Consequently, using super-resolution algorithms to reconstruct low-resolution face images for recognition has become the mainstream solution.

Face image super-resolution algorithms, known as "face hallucination," were first proposed by Baker [3], who used Bayesian formulas to estimate spatial gradient distributions from Gaussian and Laplacian pyramids as prior information for frontal faces to obtain high-resolution images. Wang et al. [4] reconstructed low-resolution input images by weighting training set face images using Principal Component Analysis (PCA). Tappen and Liu [5] warped high-resolution face images from the training set using SIFT flow and then employed a Bayesian framework for reconstruction. Song et al. [6] proposed a method that generates facial components via CNN and synthesizes fine-grained facial structures through component enhancement. FSRNet [7] introduced a face super-resolution generative adversarial network to produce more realistic face images and incorporated adversarial loss for end-to-end training. This algorithm first constructs a coarse super-resolution network to recover rough high-resolution images, then extracts image features and computes facial landmark heatmaps and parsing maps as face prior information, which are finally fed into a fine super-resolution decoder to restore high-resolution images.

Low-resolution face recognition algorithms based on super-resolution reconstruction fall into two categories: direct and indirect methods. Direct methods connect the super-resolution reconstruction network with the recognition network for joint training, updating the super-resolution network parameters in a direction beneficial to face recognition [8]. Indirect methods train the super-resolution and face recognition networks separately. While simple and effective, indirect methods may reconstruct realistic faces that are not necessarily recognition-friendly and could introduce additional noise.

To address these issues, this paper adopts a direct approach and proposes a low-resolution face recognition network structure based on identity preservation and super-resolution reconstruction. Identity loss is introduced during network training to guide the super-resolution reconstruction network toward preserving identity information. Experimental results demonstrate that the proposed method significantly improves low-resolution face recognition accuracy.

1 Algorithm Framework

The flowchart of the low-resolution face recognition system for security surveillance scenarios is shown in Figure 1 [Figure 1: see original paper]. In this figure, CenterFace [9] is employed for face detection in low-resolution images, and the

content within the dashed box represents the core of the entire system—the primary focus of this paper' s research.

Figure 1 Flow chart of low-resolution face recognition system in security surveillance scene

The proposed algorithm mainly consists of two sub-networks: the Super-Resolution Network (SRNet) and the Face Recognition Network (FRNet). SRNet reconstructs high-resolution face images from low-resolution inputs, while FRNet extracts facial features from the reconstructed images.

1.1 Wide-Activation SRNet

The core idea of deep convolutional neural network-based image super-resolution reconstruction is to find an optimal mapping function between low-resolution and high-resolution images through convolution and activation functions. For single-image super-resolution tasks, it is essential to learn as much information as possible from the input low-resolution image and convolution-generated feature maps while ensuring that this information propagates to deeper network layers during forward propagation—that is, guaranteeing information flow transmission.

In super-resolution networks, ReLU activation functions can impede information flow transmission. To mitigate this effect, this paper applies the wide activation concept to super-resolution algorithms. The most straightforward approach to wide activation is increasing the channel count of all feature maps, which improves performance but introduces substantial parameters. Building upon the network structure in [10], this paper constructs a super-resolution reconstruction network using wide activation by increasing the number of feature maps before activation functions, enabling more information to propagate to deeper layers. To avoid introducing excessive parameters, the base number of feature maps is reduced while increasing pre-activation feature maps, thereby controlling parameter growth while achieving wide activation.

The SRNet architecture is illustrated in Figure 2 [Figure 2: see original paper]. The input low-resolution image first passes through two network branches, undergoes identical upsampling operations, and the results are directly summed to produce the reconstructed high-resolution image. To reduce network parameters, redundant convolutional layers from [10] have been removed.

Figure 2 Network structure of SRNet

Common upsampling methods include bilinear interpolation and transposed convolution. However, transposed convolution introduces excessive artificial factors in super-resolution reconstruction. Therefore, this paper employs sub-pixel convolution for feature map upsampling, where the interpolation function is implicitly included in the convolutional layer and can be learned automatically. Since batch normalization (BN) restricts network flexibility by normalizing features,

this paper' s SRNet removes BN layers when using residual networks for image super-resolution reconstruction and adopts weight normalization (WN) to reduce computational cost.

Assuming the output is $y = \mathbf{w} \cdot \mathbf{x} + b$, where $\mathbf{w}$ is a k-dimensional weight vector, b is a scalar bias, and $\mathbf{x}$ is a k-dimensional input feature vector. Weight normalization reparameterizes weights by decoupling the length and direction of the weight vector, constraining weights within a specific range. Literature [11] demonstrates that this decoupling accelerates neural network convergence, enabling training with larger learning rates. The process is shown in Equation (2):

$$\mathbf{w} = \frac{g}{\|\mathbf{v}\|}\mathbf{v}$$

where $\mathbf{v}$ is a k-dimensional weight vector, g is a scalar, and $\|\mathbf{v}\|$ is the Euclidean norm of $\mathbf{v}$.

1.2 Depthwise-Separable-Convolution-Based FRNet

After low-resolution face images are reconstructed into high-resolution images via super-resolution, they must pass through a face recognition network to extract features for subsequent identity verification. This paper' s FRNet adopts the lightweight face recognition network MobileFaceNet proposed in [12] to achieve fast and accurate face recognition.

Most visual recognition networks employ global average pooling (GAP), which assigns equal weights to all neurons in feature maps. However, the receptive fields corresponding to central and peripheral regions differ, with central regions having greater impact on output. GAP treats all units equally, degrading network performance. To address this, MobileFaceNet uses global depthwise convolution (GDConv) instead of GAP, enabling different weights for different feature map units.

GDConv is implemented through depthwise separable convolution, which decomposes standard convolution into two steps: depthwise convolution and pointwise convolution. The standard convolution process is shown in Figure 3 [Figure 3: see original paper].

Figure 3 Operation process of ordinary convolution

Assuming the input is a 64$\times$64$pixelthree-channelcolorimagethatpassesthroughaconvolutionallayerwith4fil$ kernels, the output consists of 4 feature maps with the same dimensions as the input. The parameter count for standard convolution is $4 \times 3 \times 3 \times 3 = 108$. Under the same conditions, the depthwise separable convolution process is shown in Figure 4 [Figure 4: see original paper].

Figure 4 Operation process of deep separable convolution

The input image first undergoes convolution in the 2D plane, producing 3 feature maps (matching the input depth). These 3 feature maps then undergo a second convolution operation—depthwise convolution—where weighted combinations are performed along the depth dimension to obtain the final result. The parameter count for depthwise separable convolution is $3 \times 3 \times 3 + 1 \times 1 \times 3 \times 4 = 39$, approximately one-third that of standard convolution. Therefore, depthwise separable convolution significantly reduces network parameters and accelerates feature extraction.

1.3 Low-Resolution Face Recognition Algorithm with Joint Losses

Current deep learning-based single-image super-resolution reconstruction algorithms primarily focus on subjective visual quality, using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as evaluation metrics. When applied to low-resolution face recognition, these algorithms may sacrifice identity information while preserving image details, limiting face recognition performance.

To improve identity preservation during super-resolution reconstruction, this paper proposes a low-resolution face recognition algorithm that combines multiple losses. The overall network structure is shown in Figure 5 [Figure 5: see original paper].

Figure 5 Overall structure of network

In Figure 5, the input low-resolution face image (LR Face) passes through SRNet to obtain SR Face. The high-resolution face image (HR Face) is downsampled to the size of SR Face to produce MR Face. SR Face and MR Face form the image content loss L_{image}, which supervises SRNet to reconstruct more realistic high-resolution face images. SR Face is upsampled to HR Face size to obtain ISR Face. ISR Face and HR Face are fed into FRNet to extract facial features, and the distance between the two feature vectors forms the identity loss L_{id}, which supervises SRNet to preserve more identity information.

Consistent with [10], this paper adopts L_1 loss for image content loss:

$$L_{image} = \frac{1}{n} \sum_{i=1}^{n} \|SR_i - MR_i\|_1$$

where SR_i represents SR Face, MR_i represents MR Face, and n is the number of training samples.

The identity loss function maintains identity information during face super-resolution reconstruction. This algorithm uses the cosine distance between features of reconstructed and high-resolution face images as the identity loss, as shown in Equation (4):

$$L_{id} = \frac{1}{n}\sum_{i=1}^{n} g(f_{ISR}^{i}, f_{HR}^{i})$$

where f_{ISR}^{i} and f_{HR}^{i} represent feature vectors of ISR Face and HR Face, respectively, and $g(\cdot)$ denotes cosine distance between features, calculated as:

$$g(f_{ISR}, f_{HR}) = 1 - \frac{f_{ISR} \cdot f_{HR}}{\|f_{ISR}\|\|f_{HR}\|}$$

During training of the proposed joint-loss network, the pre-trained FRNet model parameters are fixed while only SRNet parameters are updated. The FRNet component uses a MobileFaceNet model trained on the CASIA-WebFace dataset [13], while SRNet is trained from scratch on the CelebA dataset [14]. Preprocessing involves: (1) face detection and alignment using MTCNN [15], (2) resizing images to $112\$\times 96 as high - resolution images, and (3) random down sampling by factors of 4, 8, and 16 to obtain low - resolution images of sizes 28\times 24, 14\times 12, and 7\times\$6$, respectively. During training, SRNet is supervised by the joint loss of image content loss and identity loss to simultaneously achieve face super-resolution reconstruction and identity preservation. The joint loss is:

$$L = L_{image} + \alpha L_{id}$$

where α is the weight factor for identity loss.

2 Experiments

2.1 Experimental Environment

Both model training and network architecture implementation utilize the PyTorch framework. PyTorch is an open-source Python-based scientific computing package commonly used in machine learning, featuring powerful GPU-accelerated tensor computation (similar to NumPy) and an automatic differentiation system. Due to its simplicity, efficiency, minimal encapsulation, and ease of use, PyTorch has gained widespread adoption. The specific hardware and software configuration includes: Intel(R) Core(TM) i7-6700 CPU @ 3.4GHz (8 cores, 8GB RAM), 64-bit Ubuntu 16.04 OS, NVIDIA TITAN X GPU with 12GB memory, and Anaconda software with Python 3.5 environment and relevant scientific computing libraries.

2.2 Super-Resolution Reconstruction Results and Analysis

To demonstrate the effectiveness of the proposed SRNet for low-resolution face image super-resolution reconstruction, selected faces from the LFW dataset [16]

are tested by downsampling them as input low-resolution images. While super-resolution performance is typically reflected through image pixels, face image reconstruction must also consider preservation of identity-discriminative details. Therefore, this paper primarily evaluates SRNet reconstruction performance through subjective visual quality.

The proposed method is compared against bicubic interpolation, the super-resolution algorithm from [10] (EDSR), and the algorithm from [17] (ESRGAN) from a subjective visual perspective. Bicubic interpolation is a widely used traditional method, while EDSR and ESRGAN represent state-of-the-art deep learning-based super-resolution algorithms using deep convolutional neural networks and generative adversarial networks, respectively. The upscaling factor is set to 4, with comparison results shown in Figure 6 [Figure 6: see original paper], where HR denotes the original high-resolution image.

Figure 6 Comparison of four times super-resolution reconstruction results

The results show that Bicubic reconstruction only preserves rough facial contours with extremely blurred facial information. The deep CNN-based EDSR algorithm produces relatively clear facial feature positions but suffers from artifacts that make individual identities difficult to discern. The GAN-based ESRGAN algorithm yields locally clear images but still exhibits blurred details around mouths in some cases. In contrast, the proposed SRNet reconstructs images with sharper overall edge contours, more complete preservation of detail information in facial features (eyes, eyebrows, etc.), and richer textures in identity-representative local regions, providing higher identity discriminability and better facilitating low-resolution face recognition.

2.3 Low-Resolution Face Recognition Results and Analysis

To validate the effectiveness of the proposed algorithm for low-resolution face recognition, the LFW dataset [16] is downsampled by factors of 4, 8, and 16 to obtain low-resolution images of sizes 28$\times$24, 14$\times$12, $and 7\times 6, respectively, with original 112\times$96 images serving as the high-resolution gallery. Face recognition broadly includes 1:1 verification and 1:N identification. This paper follows the unrestricted protocol of LFW [16], selecting 6,000 face pairs to test verification accuracy through 10-fold cross-validation, using average accuracy as the evaluation metric.

The proposed method is compared against bicubic interpolation and the face hallucination algorithm TDAE from [18]. Face verification accuracy results on LFW [16] are presented in Table 1 .

Table 1 Face verification accuracy of LFW dataset

Approach	14$\times$12\|16$\times$16\|28$\times$24\|112$\times$$96			
Bicubic	...	...	...	...
ESRGAN	...	...	...	...

Approach	14$×12|16×16|28×24|112×$96			
Ours(0)	···	···	···	···
Ours(0.5)	···	···	···	···

Since TDAE and ESRGAN require 16$×16 *input images, test images are down sampled to* 16×$16 for fair comparison. In Table 1, Ours(0) and Ours(0.5) denote the proposed algorithm, where Ours(0) represents the network trained with only image content loss, and Ours(0.5) represents the network trained with joint loss (weight of identity loss is 0.5). The data demonstrate that the proposed algorithm achieves significant accuracy improvements over traditional super-resolution methods (Bicubic) and deep learning-based face hallucination algorithms (TDAE, ESRGAN), particularly at lower input resolutions. The network trained with joint loss further improves accuracy compared to using only identity loss, with more pronounced improvements at lower resolutions. This indicates that under the supervision of the combined image content and identity loss, SRNet preserves not only image details but also identity-relevant information during low-resolution face reconstruction, enabling FRNet to extract more discriminative features and thereby enhancing low-resolution face recognition performance.

To verify the practical utility of the proposed algorithm in security surveillance scenarios, the challenging surveillance face dataset QMUL-SurFace [19] is used as the test set. Table 2 compares the proposed algorithm with several super-resolution methods (VDSR [20], SRResNet [21], FSRNet [7], FSRGAN [7]) on QMUL-SurFace [19] to evaluate performance on real surveillance data.

Table 2 Face verification results of QMUL-SurFace dataset

Approach	TAR(%)@FAR	AUC	Mean Acc/%
SRResNet	···	···	···
FSRNet	···	···	···
FSRGAN	···	···	···
Ours(0)	···	···	···
Ours(0.5)	···	···	···

Following the QMUL-SurFace [19] protocol, three metrics are used: TAR@FAR, AUC, and Mean Acc. TAR (True Acceptance Rate) and FAR (False Acceptance Rate) are used in face verification, where positive samples are two images of the same person and negative samples are from different persons. TAR represents the proportion of correctly accepted positive samples, while FAR represents the proportion of incorrectly accepted negative samples. TAR@FAR=0.001 indicates TAR value when FAR=0.001. AUC (Area Under the ROC Curve) measures classifier performance, with larger values indicating better face recognition performance. Mean Acc denotes average verification accuracy.

Table 2 shows that the proposed algorithms Ours(0) and Ours(0.5) generally outperform previous methods across TAR, AUC, and Mean Acc metrics. The model trained with joint loss (identity loss weight = 0.5) achieves higher Mean Acc and AUC than the model trained with only image content loss (identity loss weight = 0), confirming that introducing identity loss alongside image content loss preserves more identity information during super-resolution reconstruction of low-resolution surveillance faces, thereby validating the effectiveness of the joint loss training approach.

To illustrate the algorithm's effectiveness in actual surveillance scenarios, sample images from the UCCS dataset [22] are processed following the flowchart in Figure 1, with results shown in Figure 7 [Figure 7: see original paper]. Two pairs are selected: two different images of the same person and two images from different persons. After face detection, low-resolution faces undergo 4× super-resolution reconstruction and face recognition, yielding cosine similarity scores. The same-identity pair shows high similarity (0.7134), while the different-identity pair shows low similarity (0.2698), further demonstrating the practical utility of the proposed algorithm in real surveillance environments.

Figure 7 Low-resolution face recognition renderings in actual surveillance scene

2.4 Algorithm Complexity Analysis

In addition to evaluating super-resolution reconstruction quality and low-resolution face recognition accuracy, algorithm complexity is another important performance indicator. This paper analyzes complexity from both time and space perspectives.

For deep learning algorithms, time complexity directly determines model testing time, which affects face verification efficiency in practical applications. In the experimental environment described in Section 2.1, the proposed low-resolution face recognition algorithm requires 12.5ms to test a pair of 28$×$24 face images, including super-resolution reconstruction, feature extraction, and feature comparison.

Space complexity is reflected by model parameter count, which represents the total weight parameters of all parameterized layers and can be measured by model storage size. The proposed algorithm comprises two models: SRNet (198.8MB) and FRNet (4.1MB).

These results demonstrate that the proposed algorithm achieves low-resolution face verification with short processing time and small storage requirements, offering advantages of fast computation and compact model size that facilitate deployment in practical low-resolution face recognition applications.

3 Conclusion

This paper proposes a low-resolution face recognition algorithm for security surveillance scenarios comprising two sub-networks. The algorithm simulta-

neously considers image content loss for super-resolution reconstruction and identity loss for face recognition, enabling preservation of both image details and identity information during low-resolution face reconstruction. During super-resolution reconstruction, the wide activation concept is applied by increasing feature map count before activation functions to learn more image details. Experimental results show that the algorithm reconstructs images with realistic, rich details and clear edges, achieving good performance on both standard public datasets and surveillance face datasets. Compared with existing super-resolution methods, the proposed algorithm significantly improves low-resolution face verification accuracy.

Future work will extend the algorithm from single-image super-resolution to video sequences, leveraging spatio-temporal correlations between video frames for video super-resolution reconstruction and low-resolution face recognition. Multi-frame video fusion will be implemented to obtain higher-quality reconstructed face images and further improve low-resolution face recognition accuracy.

References

[1] Liu Wei. Research on face recognition under unconstrained condition [D]. Chengdu: University of Electronic Science and Technology of China, 2019.

[2] Zhang Kaibing, Zheng Dongdong, Jing Junfeng. Survey of low-resolution face recognition [J]. Computer Engineering and Applications, 2019, 55(22): 14-24.

[3] Baker S, Kanade T. Hallucinating faces [C]// Proc of the 4th IEEE International Conference on Automatic Face and Gesture Recognition. Piscataway, NJ: IEEE Press, 2000: 83-88.

[4] Wang Xiaogang, Tang Xiaoou. Hallucinating face by eigentransformation [J]. Systems Man and Cybernetics, 2005, 35(3): 425-434.

[5] Tappen M F, Liu C. A Bayesian approach to alignment-based image hallucination [C]// Proc of European Conference on Computer Vision. Berlin: Springer, 2012: 236-249.

[6] Song Yibing, Zhang Jiawei, He Shengfeng, et al. Learning to hallucinate face images via component generation and enhancement [C]// Proc of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4537-4543.

[7] Chen Yu, Tai Ying, Liu Xiaoming, et al. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 2492-2501.

[8] Liu Caogang. Low-resolution face recognition based on deep learning [D]. Harbin: Harbin Engineering University, 2018.

[9] Xu Yuanyuan, Yan Wan, Sun Haixin, et al. CenterFace: Joint Face Detection and Alignment Using Face as Point [J]. ArXiv Preprint, arXiv: 2019.

[10] Lim B, Son S, Kim H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE Press, 2017: 136-144.

[11] Salimans T, Kingma D P. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks [C]// In Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016: 901-909.

[12] Chen Sheng, Liu Yang, Gao Xiang, et al. MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices [C]// Proc of Chinese Conference on Biometric Recognition. Berlin: Springer, 2018: 428-438.

[13] Yi Dong, Lei Zhen, Liao Shengcai, et al. Learning Face Representation from Scratch [J]. ArXiv Preprint, arXiv: 1411.7923, 2014.

[14] Liu Ziwei, Luo Ping, Wang Xiaogang, et al. Deep Learning Face Attributes in the Wild [C]// Proc of the IEEE international conference on computer vision. Piscataway, NJ: IEEE Press, 2015: 3730-3738.

[15] Zhang Kaipeng, Zhang Zhanpeng, Li Zhifeng, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[16] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Technical Report 07-49 [R]. Amherst: University of Massachusetts, 2007.

[17] Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [C]// Proc of the European Conference on Computer Vision. Berlin: Springer, 2018: 63-79.

[18] Yu X, Porikli F. Hallucinating Very Low-Resolution Unaligned and Noisy Face Images by Transformative Discriminative Autoencoders [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 5367-5375.

[19] Cheng Z, Zhu X, Gong S, et al. Surveillance Face Recognition Challenge [J]. arXiv: Computer Vision and Pattern Recognition, 2018.

[20] Kim J, Lee J K, Lee K M, et al. Accurate Image Super-Resolution Using Very Deep Convolutional Networks [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 1646-1654.

[21] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [C]// Proc of the IEEE Con-

ference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 4681-4690.

[22] Gyunther M., Hu, P., Herrmann, C., et al. Unconstrained face detection and open-set face recognition challenge [C]// Proc of the IEEE International Joint Conference on Biometrics (IJCB). Piscataway, NJ: IEEE Press, 2017: 697-706.

Note: Figure translations are in progress. See original paper for figures.