Smart Structures and Systems, *Vol. 30, No. 3 (2022) 287-301*
https://doi.org/10.12989/sss.2022.30.3.287

Investigation of the super-resolution methods for vision based structural measurement

Lijun Wu, Zhouwei Cai, Chenghao Lin, Zhicong Chen*, Shuying Cheng and Peijie Lin

College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China

(*Received June 26, 2022, Revised July 6, 2022, Accepted July 6, 2022*)

Abstract. The machine-vision based structural displacement measurement methods are widely used due to its flexible deployment and non-contact measurement characteristics. The accuracy of vision measurement is directly related to the image resolution. In the field of computer vision, super-resolution reconstruction is an emerging method to improve image resolution. Particularly, the deep-learning based image super-resolution methods have shown great potential for improving image resolution and thus the machine-vision based measurement. In this article, we firstly review the latest progress of several deep learning based super-resolution models, together with the public benchmark datasets and the performance evaluation index. Secondly, we construct a binocular visual measurement platform to measure the distances of the adjacent corners on a chessboard that is universally used as a target when measuring the structure displacement via machine-vision based approaches. And then, several typical deep learning based super resolution algorithms are employed to improve the visual measurement performance. Experimental results show that super-resolution reconstruction technology can improve the accuracy of distance measurement of adjacent corners. According to the experimental results, one can find that the measurement accuracy improvement of the super resolution algorithms is not consistent with the existing quantitative performance evaluation index. Lastly, the current challenges and future trends of super resolution algorithms for visual measurement applications are pointed out.

Keywords: deep learning; machine vision; structural measurement; super-resolution reconstruction

1. Introduction

Machine vision techniques are broadly employed in the structure displacement measurement (Ye *et al*. 2016a, Wu *et al*. 2014, Lee *et al*. 2017, Ye *et al*. 2016b, Xu and Brownjohn 2018), thanks to the merits of non-contact, simple installation, low cost and the flexibility of increasing the measuring points (Wu *et al*. 2014). Machine-vision based systems can be used in structural condition assessment, especially in system identification (Yoon *et al*. 2016, Caetano *et al*. 2007, Oh *et al*. 2015). The measured deformation information can be used for finite element model calibration (Feng and Feng 2015), damage detection (Cha *et al*. 2017), weigh-in-motion system of bridges with camera assistance for traffic monitoring (Ojino *et al*. 2016), and so on. Basically, the procedure of machine-vision based displacement measurement consist of feature points detection and the 3D coordinators reconstruction based on feature point pixel coordinators and the projection mapping principle or binocular measurement principle (Xu and Brownjohn 2018, Wu *et al*. 2020a). Even though various image acquisition hardware and video processing improvements have been made in the previous works, the accuracy of vision-based displacement measurement is directly related to the resolution of the camera. Given the field of view to be measured, the higher the resolution, the smaller the physical size corresponding to each pixel, and thus the higher the measurement accuracy. Instead of upgrading the hardware system, upgrading the software can save costs and can be compatible with existing systems.

Image super-resolution (SR) is a software method for recovering high-resolution images from low-resolution images (Wang *et al*. 2020a), which has the advantages of low cost and good effect. It has a wide range of applications in practical remote sensing images (Nguyen *et al*. 2005), medical images (Finkelstein *et al*. 2003) and surveillance (Rasti *et al*. 2016). Recently, SR is introduced into machine-vision based measurement. The Super Resolution Digital Image Correlation (SR-DIC) (Hansen *et al*. 2021) demonstrates the potential application of super resolution imaging to improve high-magnification DIC measurements using open source SR software. However, the SR-DIC adopts the traditional SR, and requires capture several low-resolution images in order to reconstruct a SR image. Thanks to the development of deep learning algorithms, it is possible to reconstruct the SR image from single image via a trained model.

Image super-resolution reconstruction is a classic ill-posed problem (Tian and Ma 2011), because a given low resolution (LR) image always has multiple corresponding high resolution (HR) images. At present, the SR algorithms can be divided into three categories: interpolation-based methods (Mecheri *et al*. 2007), reconstruction-based methods (Chantas *et al*. 2007), and learning-based methods (Chen 2011). The interpolation-based methods (Mecheri *et al*. 2007) use the gray value of original pixels around the

*Corresponding author, Ph.D., Professor,
E-mail: lijun.wu@fzu.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

interpolation point to estimate the gray value of the point, which is simple and fast, suitable for real-time applications; Typical interpolation-based algorithms includes nearest neighbor interpolation (Nearest), bilinear interpolation (Bilinear) (Blu *et al*. 2004), bicubic interpolation (Bicubic) (Keys 2003). The reconstruction-based methods (Chantas *et al*. 2007) try to model the image degradation process in order to reconstruct the high-resolution image that satisfies the constraint provided by the low-resolution image (Yang *et al*. 2019a). Commonly used reconstruction algorithms include iterative back projection (IBP) algorithm (Irani and Peleg 1991), maximum a posterior probability (MAP) algorithm (Schultz and Stevenson 1996), projection onto convex set (POCS) algorithm (Stark and Oskoui 1989), and maximum a posterior probability- projection onto convex set (MAP-POCS) hybrid algorithm (Elad and Feuer 2002).

The learning-based methods (Chen 2011) learn the mapping of high-resolution and low-resolution images in the training dataset, and then apply this mapping relationship to the super-resolution reconstruction of other images. At the beginning, the machine-learning based methods use manual or semi-automatic methods to select features and adjust parameters, which is time-consuming and requires expert knowledge of the corresponding field (Lin *et al*. 2008). The learned mapping relationship is often a simple mapping among convolution kernels, comparison tables or image blocks (Freeman *et al*. 2002). Traditional learning methods mainly include neighbor embedding (Chang *et al*. 2004), sparse representation (Yang *et al*. 2010), and regression mapping (Ni and Nguyen 2007).

The SR algorithm based on Convolutional Neural Network (SRCNN) (Dong *et al*. 2014) firstly introduced deep learning network (Wang *et al*. 2020b) into the field of image SR reconstruction, greatly improving the performance of SR algorithm. Since then, various deep-learning based SR models have been proposed and the performance are greatly improved.

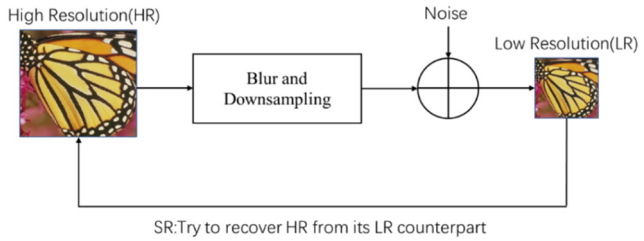


Fig. 1 Typical frame of image super-resolution

In this study, we focus on the investigation of the deep learning based image SR reconstruction for machine-vision based measurement. Firstly, the existing deep-learning-based SR algorithms are introduced in details, together with some public benchmark datasets and commonly used performance evaluation index for the sake of validation and evaluation. Then, we will construct a binocular visual measurement system to measure the distance of the adjacent corners on chessboard that is universally used as a target in the machine-vision based structure displacement measurement system. It is worth noting that the chessboard is selected as the measurement object since the distance between its adjacent points (ground truth) is known. However, SR algorithm is not limited to the visual measurement scheme with target. After that, several classical SR algorithms are employed to improve the image quality of the measurement system to verify the potential of SR in the displacement measurement accuracy improvement.

2. Basic principles and classification

As shown in Fig. 1, in a typical SR framework (Park *et al*. 2003), the low resolution (LR) image y is transformed from the high resolution (HR) image x by Eq. (1).

$$y = (x \otimes k) \downarrow_s + n \tag{1}$$

Among them, x⊗k is the convolution between the fuzzy kernel k and the HR image x, ↓s is the down-sampling operation with the scale factor s, and n is the independent noise term. The SR algorithm focuses on how to recover the HR image from its LR image.

With the development of deep learning in the image field, image SR reconstruction has also been inspired by neural network models and has made significant progress. The deep learning-based SR method is based on the standard convolutional neural network (CNN) (Krizhevsky *et al*. 2012), residual network (ResNet) (He *et al*. 2016), densely connected network (DenseNet) (Huang *et al*. 2017) and generative adversarial network (GAN) (Goodfellow *et al*. 2014), as shown in Fig. 2.

2.1 SR network based on standard CNN

The first deep-learning based SR algorithm is the

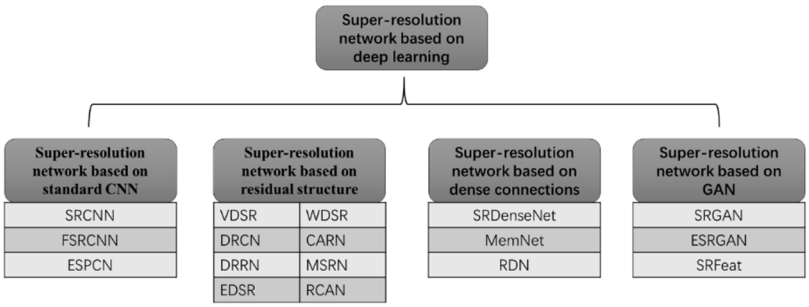


Fig. 2 Classification of SR networks based on deep learning

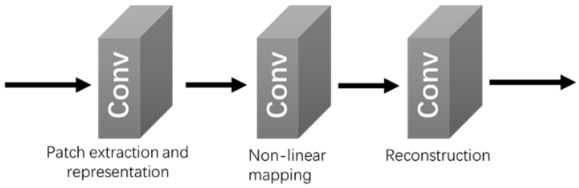


Fig. 3 SRCNN structure

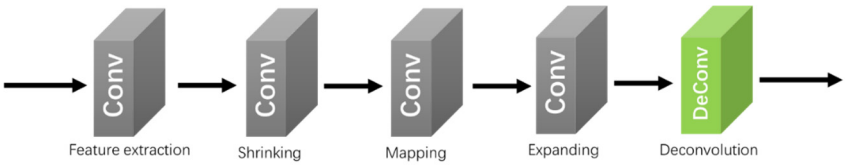


Fig. 4 Structure of FSRCNN

SRCNN (Dong *et al*. 2014), as shown in Fig. 3. It simulates the sparse coding reconstruction algorithm, and learns how to map low-resolution images to high-resolution images. The input image is enlarged by bicubic interpolation to the required size, and then the CNN is used to perform three steps of feature extraction, nonlinear mapping and reconstruction, so as to achieve SR reconstruction.

The structure of SRCNN is simple, and the accuracy is better than other non-deep learning algorithms, which fully reflects the prospects of deep learning in the field of single-image SR reconstruction. But the input is an interpolated LR image, which not only increases the amount of calculation, but also introduces unnecessary noise that affects the reconstruction effect. Besides, one should train a model separately for each different scale magnification factor.

To solve the problems of the SRCNN, the Fast Super-Resolution Convolutional Neural Networks (FSRCNN) (Dong *et al*. 2016) is proposed, as shown in Fig. 4. Instead of bicubic interpolation, the FSRCNN introduces a deconvolution layer to upsample the input image, which greatly reduces the amount of calculation and can provide different scale amplification by changing the deconvolution layer. Meanwhile, the FSRCNN deepens the network to 8 layers to improve the network performance.

In the Efficient Sub-Pixel Convolutional Neural Network (ESPCN) proposed by Shi *et al*. (2016), a new image reconstruction strategy, i.e., the pixel rearrangement strategy, is given, which is very similar to image interpolation. At the same time, by setting the number of feature channels, different scale factors can be obtained. The structure of ESCPN is shown in Fig. 5.

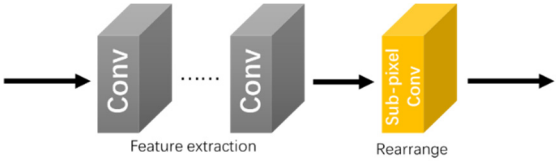


Fig. 5 ESCPN structure

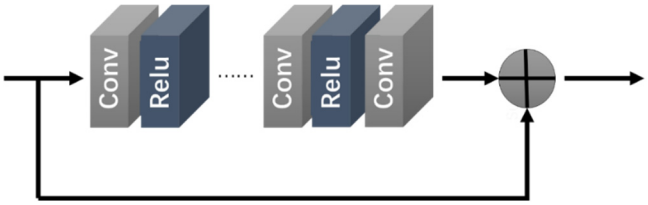


Fig. 6 VDSR structure

2.2 SR network based on residual structure

Very Deep Convolutional Networks for Super-Resolution (VDSR) (Kim *et al*. 2016a) introduced the residual structure and constructed a network layer with 20 layers. It states that the low-frequency information carried by the LR image is similar to the low-frequency information of the HR image. And thus, it focuses on learning the high-frequency residual information. It inputs the LR image with the target size into the network, then adds this image and the residual error learned by the network to obtain the final HR reconstruction image, as shown in Fig. 6.

The deeper network in VDSR brought better expressive capabilities, as well as better performance. However, because only a skip connection is introduced in the network, the vanishing gradient problem has not been well alleviated. Hence, an adaptive gradient clipping strategy is used to solve the vanishing gradient problem in the training process.

The Deeply-Recursive Convolutional Network (DRCN) (Kim *et al*. 2016b) adopts the embedded network to extract shallow features of low-resolution pictures as the input of the recursive network. Then, the inference network is composed of 16 recursive units to strengthen the information transfer between layers and make full use of the connection of contextual information. Therefore, it can better restore the high-frequency parts of images. At last, the final reconstruction output is the weighted sum of the different convolution layer's output. Due to the parameter sharing, the recursive layers in DRCN does not increase the number of parameters. However, the recurrent neural network is prone to overfitting or gradient disappearance/explosion, so a skip connection is added between the recursive unit and the reconstructed network, while

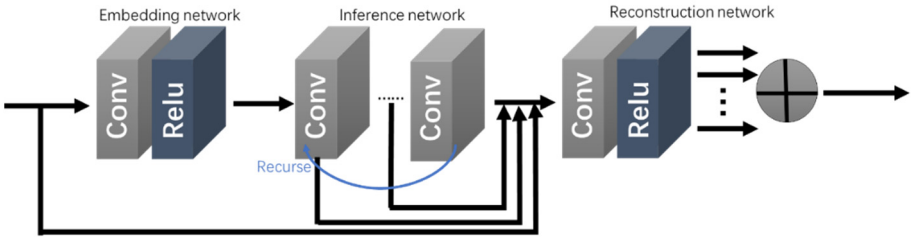


Fig. 7 DRCN structure

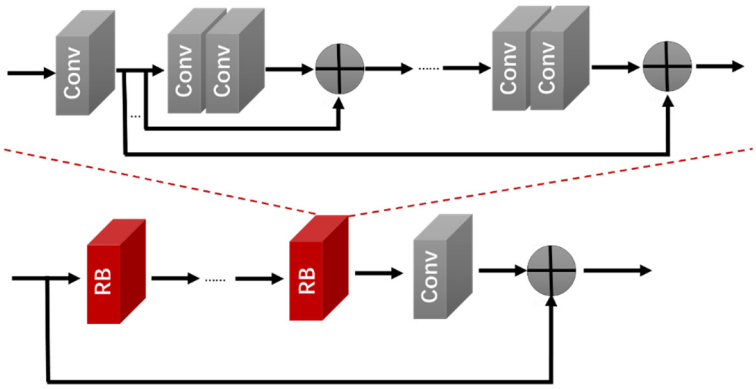


Fig. 8 DRRN structure. Here, "RB" layer refers to a recursive block

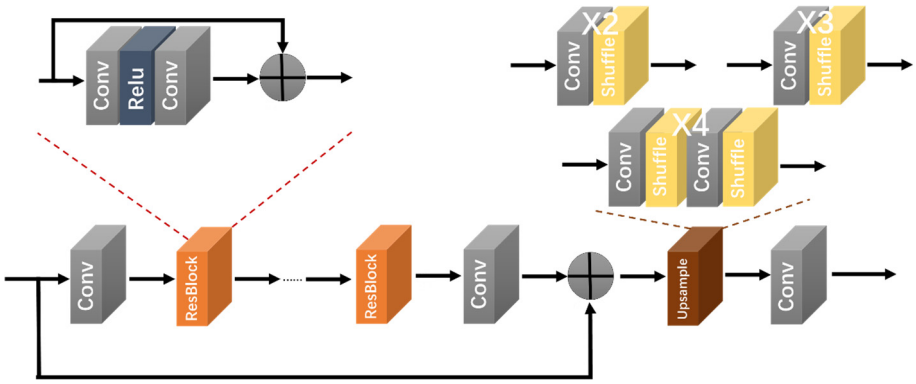


Fig. 9 EDSR structure

a recursive supervision strategy is used to alleviate the gradient disappearance/explosion problem. The structure of DRCN is shown in Fig. 7.

Inspired by VDSR and DRCN, the Deep Recursive Residual Network (DRRN) (Tai *et al*. 2017b) is proposed, which combines local residual learning, global residual learning, and multi-weight recursive learning. As shown in Fig. 8, the whole network is composed of an adjustable number of recursive blocks. Because the parameters of the sub-recursive units are shared, and local skip connections are added, the network depth can achieve 52 layers.

The Enhanced/Multi-scale Deep Super-Resolution (EDSR/MDSR) (Lim *et al*. 2017) removes the Batch Normalization (BN) layer from the deep residual network, and thus can stack more network layers and achieve better performance, as shown in Fig. 9. To learn the mapping relationship between the high-resolution images and the low resolution images from different "perspectives", the EDSR perform multiple geometric transformations on each LR image, and then the reconstructed HR image are inversely transformed and averaged as the final reconstruction HR image.

Based on the EDSR, the Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR) (Yu *et al*. 2018) is proposed, which wins the championship of the NTRIE 2018 super-resolution competition. In WDSR, the redundant convolutional layer outside the residual block is removed, and the batch normalization is replaced by weight normalization. Hence, the network occupies less memory and can be trained faster. Fig. 10 demonstrates the simplified WDSR network.

In order to reduce the size of the network and enhance its practicability, a cascade mechanism (Ahn *et al*. 2018) is used to propose a lightweight SR reconstruction model Cascading Residual Network (CARN), as shown in Fig. 11. The CARN contains multiple cascaded blocks (CB) with

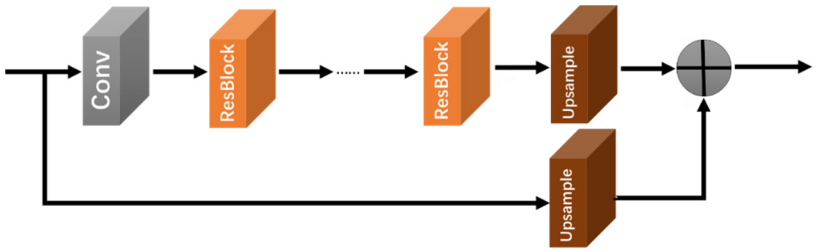


Fig. 10 The simplified WDSR network

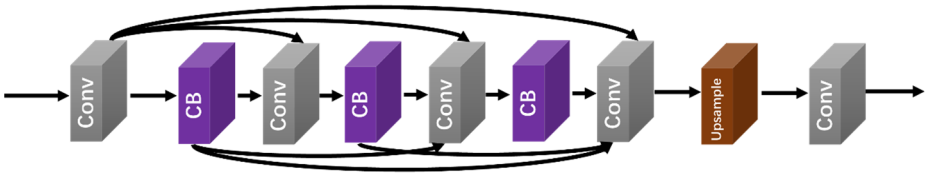


Fig. 11 CARN structure

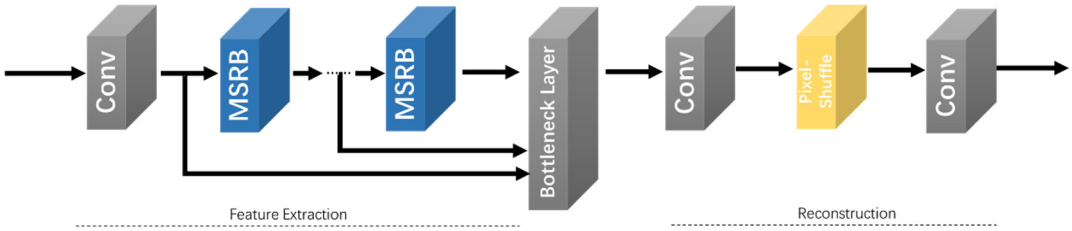


Fig. 12 MSRN structure. Here, "MSRB" layer refers to a multi-scale residual block

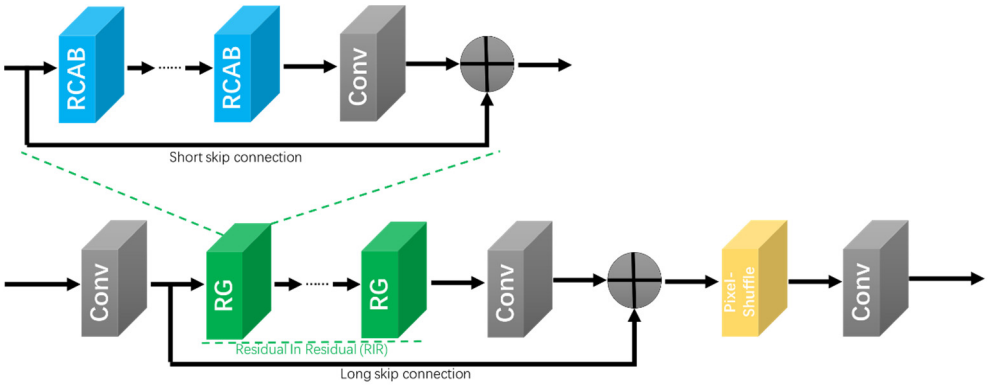


Fig. 13 RCAN structure

multiple skip connections, so that the cascaded residual network can mix multi-level features locally and globally, and make information transfer more efficient.

The Multiscale Residual Network (MSRN) (Li *et al.* 2018) is a SR network that can make use of multi-scale features for arbitrary scale amplification, as shown in Fig. 12. The network is divided into a feature extraction part and a reconstruction part. The feature extraction part is composed of the multi-scale residual block (MSRB) that completes the multi-scale feature fusion. The task of the reconstruction part is to enlarge the image. By adopting the channel pixel rearrangement method in the ESPCN model, the reconstructed images with different scales are realized.

In the previous SR models, the information in each channel was treated indiscriminately, which limits the expressive ability of the network. The Residual Channel Attention Network (RCAN) (Zhang *et al.* 2018a) adds a channel attention mechanism, as shown in Fig. 13. In order to alleviate the difficulty of deep network training, RCAN adopts Residual In Residual (RIR) structure. The RIR is connected by a series of residual groups (RG) through a long-skip connection, which relieves the burden of information transmission and stabilizes the training process. Each RG structure was connected by the residual channel attention block (RCAB) through short-skip connection, which made the network mainly learn high frequency features. The structure of RCAB is shown in Fig. 14, in which a channel attention mechanism is added, so that the

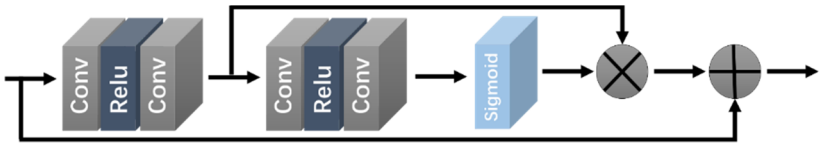


Fig. 14 RCAB structure

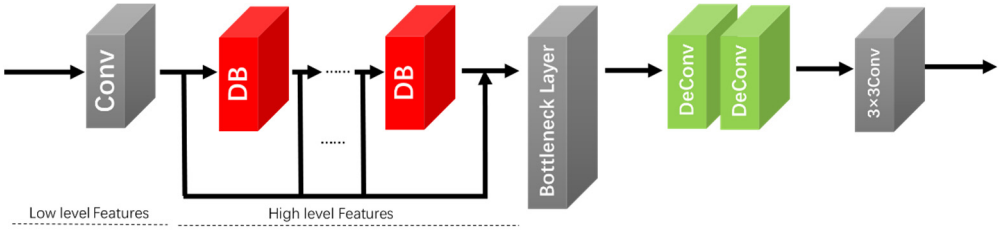


Fig. 15 Structure of SRDenseNet. Here, "DB" refers to a dense block

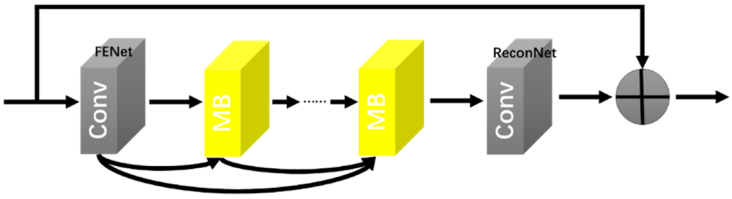


Fig. 16 MemNet structure. Here, "MB" refers to a memory block

RCAN can adaptively learn more useful channel functions.

2.3 SR network based on dense connections

Compared with ResNet, the Dense convolutional network (DenseNet) (Huang *et al*. 2017) proposes a more radical dense connection mechanism, which connects all layers to each other. Specifically, each layer will accept all the previous layers as its additional input.

The SRDenseNet (Tong *et al*. 2017) applied the DenseNet into the SR problem, as shown in Fig. 15. Multiple dense blocks are used to learn high-level features, which enables the sharing of complementary features at different depth levels, and achieves better performance with fewer parameters and calculation.

The Very Deep Persistent Memory Network (MemNet) (Tai *et al*. 2017a) further introduces a Memory Block (MB) containing a recursive unit and a gate unit, as shown in Fig. 16. In each MB, the recursive unit learns the multi-level representations of the current state in different receptive fields, which can be regarded as short-term memory. Long-term memory is generated by the previous MB. These short-term and long-term memory are combined and sent to the gate unit. Compared with the SRDenseNet, the MemNet considers the influence of the previous state on the subsequent state.

The Residual Dense Network (RDN) (Zhang *et al*. 2018b) combines DenseNet and ResNet, and makes full use of the hierarchical features extracted by the convolutional neural network. It contains 4 modules, namely, the shallow feature extraction network (SFENet), the residual dense block (RDB), the dense feature fusion (DFF) and the up-sampling network (UPNet). The network structure is shown in Figs. 17 and 18. The network uses a series of concatenated residual dense blocks to extract and fuse local features, establish a continuous memory mechanism, retain the information of the previous layer to the greatest extent, and perform global fusion of the obtained local features.

2.4 SR network based on GAN

The Super-Resolution Using a Generative Adversarial Network (SRGAN) (Ledig *et al*. 2017) introduces the GAN (Goodfellow *et al*. 2014) into the field of image SR, and obtains images more consistent with human visual perception. The generator and discriminator are shown in Figs. 19(a) and (b) respectively. The generator is responsible for generating high-resolution images, and the discriminator is responsible for identifying whether the input image comes from generated data or real data. When the discriminator fails to discriminate, it can be considered that the generator has the ability to synthesize high-resolution images.

The enlarged details of HR reconstructed by SRGAN are usually accompanied by artifacts. The Enhanced SRGAN (ESRGAN) (Wang *et al*. 2018) replaces the basic residual unit of the network by a Residual-in-Residual Dense Block (RRDB) in order to restore more realistic texture details, as shown in Fig. 20. Specifically, the RRDB removes the BN layer in order to utilize more features and reduce the artifact problem. Besides, the discriminator uses the relativistic average GAN (RaGAN) to judge 'whether an image is more real than another' instead of 'an image is true or false'.

The Super-Resolution with Feature Discrimination (SRFeat) (Park *et al*. 2018) adds a feature discriminator in

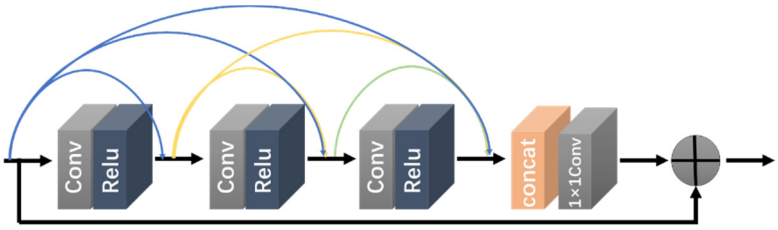


Fig. 17 RDB structure

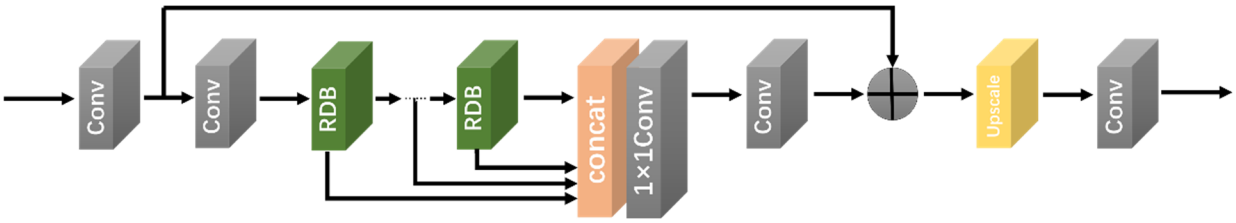
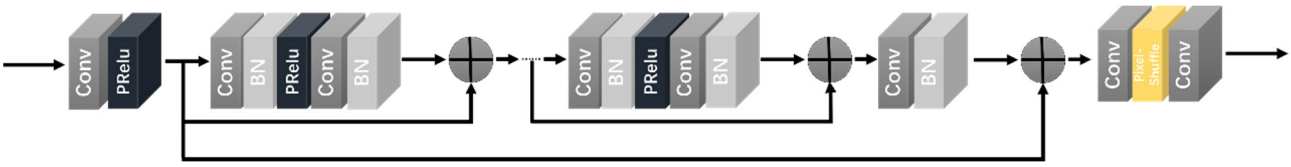
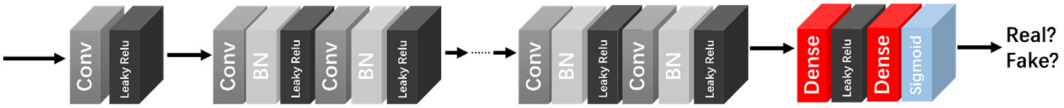


Fig. 18 RDN structure. Here, "RDB" refers to a residual dense block



(a) The generator



(b) The discriminator

Fig. 19 SRGAN structure

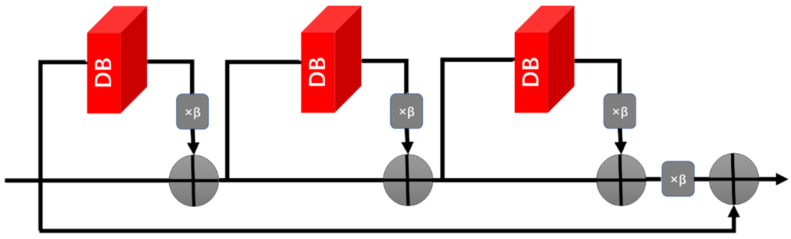
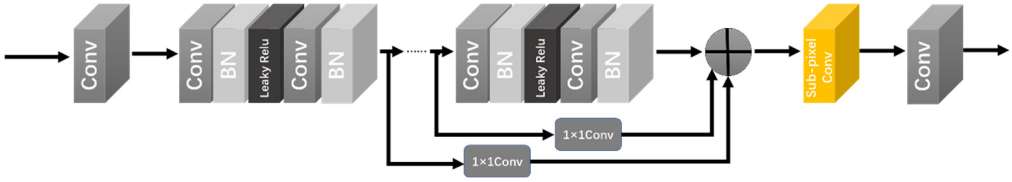
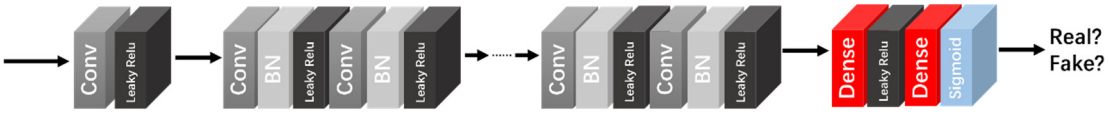


Fig. 20 RRDB structure. Here, "DB" refers to a dense block



(a) The generator



(b) The discriminator

Fig. 21 SRFeat structure

the discriminant network, so that not only the reconstructed image can be trained against the high-resolution image, but also the extracted feature can be trained against the high-resolution image, so as to produce more realistic results. The generator and discriminator are shown in Figs. 21(a) and (b) respectively.

3. SR public benchmark datasets

A list of public benchmarks image datasets has been constructed for SR task, as listed in Table 1. Some provide LR-HR image pairs, while others only provide HR images and a degradation operation is required to obtain LR images.

4. Evaluation Index and the performance of previous models

How to evaluate the quality of SR images, including both subjective and objective index, has always been a very active topic. Subjective quality evaluation is the most direct and effective method to evaluate the image quality based on the subjective feelings of the observer. One of the commonly used methods is subjective mean opinion of score (MOS) (Ledig *et al*. 2017). However, due to the influence of individual differences and environmental factors, it is difficult to form a standard evaluation result. In addition, in order to reduce the influence of subjective factors on the evaluation results, multiple test images are generally required for repeated experiments in actual evaluation, which requires a lot of time and energy. The objective evaluation method is to quantitatively evaluate the image quality by establishing a mathematical model, mainly including mean square error (MSE), peak signal-to-noise ratio (PSNR) (Hore and Ziou 2010), structural similarity (SSIM) (Wang *et al*. 2004) and information fidelity criterion (IFC) (Sheikh *et al*. 2006).

4.1 Subjective quality evaluation

The MOS (Ledig *et al*. 2017) is a commonly used subjective image quality evaluation method, requiring many experimenters to score the perceived quality of the test images. Usually the score ranges from 1 (bad) to 5 (good), and the final average opinion score is the average of all the scores.

4.2 Objective quality evaluation

The PSNR (Hore and Ziou 2010) is one of the most commonly used evaluation indicators in image processing. In SR work, the PSNR is defined by MSE and the maximum pixel value L which is generally equal to 255. Given a real image I with N pixels and a reconstructed image $\hat{I}$, the PSNR is defined as Eq. (2).

$$\begin{aligned} \mathrm{PSNR} &= 10 \cdot log_{10}\left(\frac{L^2}{MSE}\right) \\ &= 10 \cdot log_{10}\left(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\hat{I}(i)\right)^2}\right) \in (0,+\infty) \end{aligned} \tag{2}$$

The higher the PSNR value between two images, the less distortion of the reconstructed image relative to the high-resolution image. The PNSR only pays attention to the difference between pixels, and the reconstruction of real scenes pay more attention to human perception, so high PNSR does not guarantee high reconstruction quality of work in real scenes. However, due to the need for performance comparison and the lack of accurate perceptual metrics, the PSNR is still the most widely used evaluation standard for SR models.

The SSIM (Wang *et al*. 2004) compares the brightness, contrast and structure information between the reconstructed image and the original image to obtain the image quality value. The mean value is used as the brightness estimation, the standard deviation as the contrast

Table 1 Super-resolution public benchmark datasets

Datasets	Usage	Format	Number	Average resolution	Remarks information
Set5 (Bevilacqua *et al*. 2012)	TEST	PNG	5	313×336	baby, bird, butterfly, head, women
Set14 (Zeyde *et al*. 2010)		PNG	14	492×446	humans, animals, insects, flowers, vegetables, etc.
BSD100 (Martin *et al*. 2001)		JPG	100	481×321	animal, building, food, landscape, people, plant, etc.
Urban100 (Huang *et al*. 2015)		PNG	100	984×797	architecture, city, structure, urban
Manga109 (Fujimoto *et al*. 2016)		PNG	109	827×1170	a variety of 109 Japanese comic books
T91 (Yang *et al*. 2010)	TRAIN	PNG	91	264×204	car, flower, fruit, human face, etc.
BSDS300 (Martin *et al*. 2001)		JPG	300	435×367	animal, building, food, landscape, people, plant, etc.
BSDS500 (Arbelaez *et al*. 2011)		JPG	500	432×370	animal, building, food, landscape, people, plant, etc.
General-100 (Dong *et al*. 2016)		BMP	100	435×381	animal, food, people, plant, texture, etc.
DIV2K (Agustsson and Timofte 2017)	TRAIN/ TEST	PNG	1000	1972×1437	environments, flora, fauna, handmade object, people, etc.
Flickr2K (Wang *et al*. 2018)		PNG	2650	2K	2K images from Flickr
DF2K		PNG	3650	2K	a merged training dataset of DIV2K and Flickr2K

Table 2 Quantitative results of the SR models with scale factor ×2 (Unit: dB/-)

Methods	Scale factor	Set5 (Bevilacqua *et al*. 2012) PSNR/SSIM	Set14 (Zeyde *et al*. 2010) PSNR/SSIM	BSD100 (Martin *et al*. 2001) PSNR/SSIM	Urban100 (Huang *et al*. 2015) PSNR/SSIM
Bicubic		33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403
SRCNN (Dong *et al*. 2014)		36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946
FSRCNN (Dong *et al*. 2016)		37.05/0.9560	32.66/0.9090	31.53/0.8920	29.88/0.9020
ESPCN (Shi *et al*. 2016)		37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065
VDSR (Kim *et al*. 2016a)		37.53/0.9588	33.03/0.9124	31.90/0.8960	30.76/0.9140
DRCN (Kim *et al*. 2016b)		37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133
DRRN (Tai *et al*. 2017b)		37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188
EDSR (Lim *et al*. 2017)	×2	38.11/0.9602	33.92/0.9195	32.32/0.9013	32.93/0.9351
WDSR (Yu *et al*. 2018)		38.10/0.9608	33.72/0.9182	32.25/0.9004	32.37/0.9302
CARN-M (Ahn *et al*. 2018)		37.53/0.9583	33.26/0.9141	31.92/0.8960	30.83/0.9233
MSRN (Li *et al*. 2018)		38.08/0.9605	33.74/0.9170	32.23/0.9013	32.22/0.9326
RCAN (Zhang *et al*. 2018a)		38.33/0.9617	34.23/0.9225	32.46/0.9031	33.54/0.9399
SRDenseNet (Tong *et al*. 2017)		—	—	—	—
MemNet (Tai *et al*. 2017a)		37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195
RDN (Zhang *et al*. 2018b)		38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353
Bicubic		30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349
SRCNN (Dong *et al*. 2014)		32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989
FSRCNN (Dong *et al*. 2016)		33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080
ESPCN (Shi *et al*. 2016)		33.02/0.9135	29.49/0.8271	28.50/0.7937	26.41/0.8161
VDSR (Kim *et al*. 2016a)		33.68/0.9201	29.86/0.8312	28.83/0.7966	27.15/0.8315
DRCN (Kim *et al*. 2016b)		33.85/0.9215	29.89/0.8317	28.81/0.7954	27.16/0.8311
DRRN (Tai *et al*. 2017b)		34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378
EDSR (Lim *et al*. 2017)	×3	34.65/0.9280	30.52/0.8462	29.25/0.8093	28.80/0.8653
WDSR (Yu *et al*. 2018)		34.48/0.9279	30.39/0.8434	29.16/0.8067	28.38/0.8567
CARN-M (Ahn *et al*. 2018)		33.99/0.9236	30.08/0.8367	28.91/0.8000	26.86/0.8263
MSRN (Li *et al*. 2018)		34.38/0.9262	30.34/0.8395	29.08/0.8041	28.08/0.5554
RCAN (Zhang *et al*. 2018a)		34.85/0.9305	30.76/0.8494	29.39/0.8122	29.31/0.8736
SRDenseNet (Tong *et al*. 2017)		—	—	—	—
MemNet (Tai *et al*. 2017a)		34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376
RDN (Zhang *et al*. 2018b)		34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653
Bicubic		28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577
SRCNN (Dong *et al*. 2014)		30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
FSRCNN (Dong *et al*. 2016)		30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280
ESPCN (Shi *et al*. 2016)		30.66/0.8646	27.71/0.7562	26.98/0.7124	24.60/0.7360
VDSR (Kim *et al*. 2016a)		31.35/0.8830	28.02/0.7680	27.29/0.7251	25.18/0.7540
DRCN (Kim *et al*. 2016b)		31.56/0.8810	28.15/0.7627	27.24/0.7150	25.15/0.7530
DRRN (Tai *et al*. 2017b)		31.68/0.8888	28.21/0.7721	27.38/0.7284	25.44/0.7638
EDSR (Lim *et al*. 2017)	×4	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033
WDSR (Yu *et al*. 2018)		32.27/0.8963	28.67/0.7838	27.64/0.7383	26.26/0.7911
CARN-M (Ahn *et al*. 2018)		31.92/0.8903	28.42/0.7762	27.44/0.7304	25.63/0.7688
MSRN (Li *et al*. 2018)		32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896
RCAN (Zhang *et al*. 2018a)		32.73/0.9013	28.98/0.7910	27.85/0.7455	27.10/0.8142
SRDenseNet (Tong *et al*. 2017)		--	--	--	--
MemNet (Tai *et al*. 2017a)		31.74/0.8893	29.26/0.7723	27.40/0.7281	25.50/0.7630
RDN (Zhang *et al*. 2018b)		32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028

estimation, and the covariance as the structural similarity measure. Given two images x and y, the structural similarity of the two images can be calculated according to Eq. (3).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \in [-1,1] \quad (3)$$

Among them, μ_x and μ_y are the average values of x and y respectively, σ_y^2 and σ_y^2 are the variances of x and y respectively, σ_{xy} is the covariance of x and y. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are constants used to maintain stability, and L is the dynamic range of the pixel value, generally k_1 = 0.01, k_2 = 0.03. The larger the SSIM, the smaller the gap between the output image and the undistorted image, i.e., the better image quality. The SSIM is widely used because it evaluates the reconstruction quality from the perspective of the human visual system.

The IFC (Sheikh *et al*. 2006) uses high-frequency components to calculate the similarity between images by wavelet transform, which is used to measure the information of the distorted image and retains the reference image. This measurement method can better match the human visual system which is more sensitive to high-frequency information. However, the calculation is more complicated.

4.3 Performance of previous models

Table 2 show the PSNR/SSIM of the ×2, ×3, and ×4 amplification factor reconstruction results of the above mention network models on the benchmark datasets Set5 (Bevilacqua *et al*. 2012), Set14 (Zeyde *et al*. 2010), BSD100 (Martin *et al*. 2001), and Urban100 (Huang *et al*. 2015). The PSNR/SSIM values based on the generated adversarial network model are not given since they pay attention to perception details but not the PSNR/SSIM.

5. Experimental verification

In order to compare the performance of various SR algorithms, we construct a testing platform. Based on two ALLIED GIGE GT1910C cameras whose resolution is 1920×1080, image pairs of a black and white chessboard are collected. The SR algorithms are used to obtain the high-resolution images. It is worth noting that the low resolution images are acquired via Bicubic with 4×down sampling and the real images are used as the ground truth. The corners are detected and matched both on the low-resolution image pairs and the high-resolution image pairs. The 3D coordinators of the corners are respectively reconstructed according to the principle of binocular vision (Wu *et al*. 2020a). Then, we calculate the distance of the reconstructed adjacent corners and compared it with the real distance in order to access the distance measurement accuracy.

In this case, the shooting distance is about 4 m, and the distance between adjacent corner points is designed to be 15 mm. Fig. 22 shows the binocular vision measurement platform.

The Fig. 23 shows the enlarged details of the chessboard image. Fig. 23(a) is the original captured image and is used as the ground truth. Fig. 23(b) is the image acquired by ↓4 down sampled by Bicubic and is used as the low resolution image. Fig. 23(c)-(f) are the ×4 reconstructed images acquired by SRCNN, VDSR, EDSR and RCAN. As one can see from Fig. 23, the image SR algorithms can improve the image quality in different degrees from the perception of

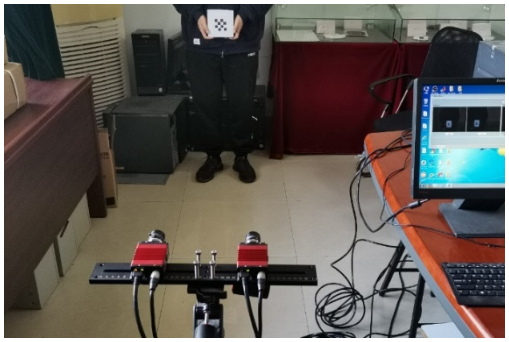
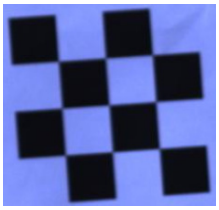
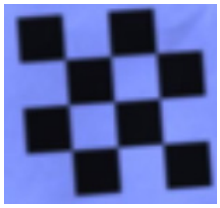


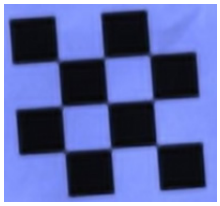
Fig. 22 Binocular vision measurement platformsystem



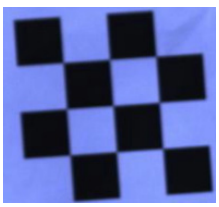
(a) GT



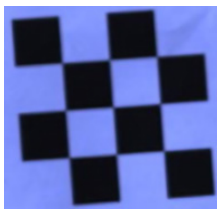
(b) 4×down sampling



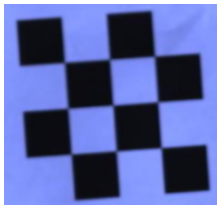
(c) Reconstructed by SRCNN



(d) Reconstructed by VDSR

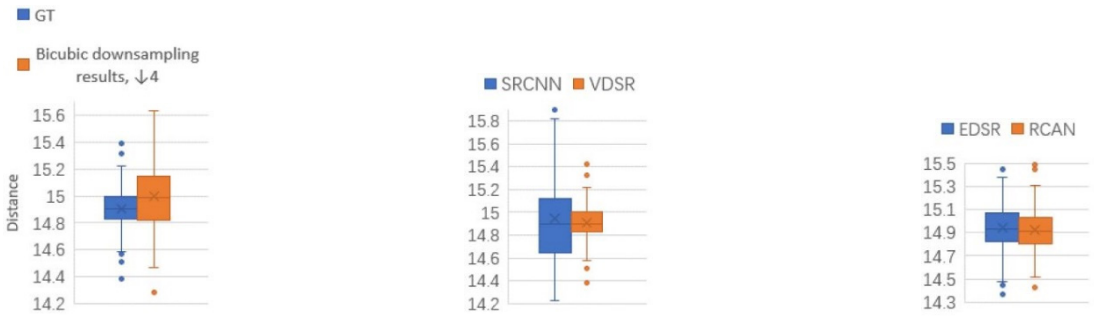


(e) Reconstructed by EDSR



(f) Reconstructed by RCAN

Fig. 23 The enlarged detail of the 4×reconstructed chessboard images



(a) (b) (c)

Fig. 24 Box plots: (a) Original image (ground truth, GT) and Bicubic down sampling ↓4 results; (b) SRCNN, VDSR results, ×4; (c) EDSR, RCAN results, ×4

human eyes.

Then, the distances of the adjacent corner points are measured based on the principle of binocular vision measurement, and the accuracy of the measurement results can be seen from the box plots in Fig. 24. One can see from Fig. 24(a) that the corner detection accuracy is directly related to image resolution, and the measurement accuracy is decreased after the down sampling. From Figs. 24(b) and (c), one can see that most of the SR reconstruction algorithms to be tested can improve the measurement accuracy to different degrees, except the SRCNN. The visual quality of the pictures after SRCNN has been improved, but the edge areas, especially the corner points are blurred. Corner detection results are worse than the input image, with more outliers and a wider error range. The VDSR has better visual effects than SRCNN, and the corner detection results are also better. If the results in Fig. 24 are compared with the ones listed in Table 3, one can find the results are inconsistent. For example, in Table 3, the PSNR and SSIM of images reconstructed by RCAN and EDSR are better than the ones of VDSR. However, the distance measurement accuracy based on VDSR SR images is better than the one of EDSR and RCAN. Therefore, if we want to improve the measurement accuracy through SR reconstruction algorithm, we need to select the appropriate algorithm. The existing PSNR and SSIM evaluation indexes cannot perfectly evaluate the performance of the algorithm in visual measurement applications.

6. Future directions

Currently, the deep learning based image SR has achieved great success, but there are still a lot of issues for the application of machine-vision based measurement.

The improvement of network structure is always important in order to improve the performance and reduce excessive space and computational redundancy. Currently, the image SR algorithms usually have too many parameters, which lead to long training time and large memory consumption, and it is difficult to be deployed on the embedded platforms. Hence, the design of lightweight models is also important.

6.1 Towards real-world image super-resolution

SR algorithms aim to learn the map between the LR and HR image pairs. However, it is difficult to collect image pairs with different resolutions in site. Instead of constructing a paired dataset in real life (Cai *et al*. 2019), the collected images are often treated as the HR image and are down sampled by bicubic algorithm to obtain a simulation of the LR image (Zhang *et al*. 2019) in order to train the model.

However, the reasons for the degradation of low-resolution images in the real world are very complex, and the low-resolution images in the public dataset are obtained by performing a predefined degradation (usually bicubic degradation) on high-resolution images, so the quality of models derived from supervised learning are often very poor when applied to real-world images (Bulat *et al*. 2018). How to better fit the real-world image degradation process and improve the reconstruction effect of image SR reconstruction technology in practical applications, is one of the remaining issues for practical applications (Zhang *et al*. 2017).

Finally, the SR reconstruction technology has a wide range of applications in areas where it is difficult to extract useful information from low-resolution images, such as target tracking (Luo *et al*. 2017), medical imaging (Luo *et al*. 2020b), and video surveillance (Yan *et al*. 2015). However, the current SR reconstruction technology mostly stays in the research stage. In practical applications, for different application fields, the requirements for SR reconstruction technology are different, and the network structure needs to be adjusted according to different application scenarios to meet different needs.

6.2 Selection of loss function and evaluation index

The loss function can be seen as a constraint between the high-resolution image and the reconstructed image, which can guide the model to optimize (Seif and Androutsos 2018). The evaluation index has a certain connection with the loss function (Seif and Androutsos 2018). A good evaluation index can effectively evaluate the quality of the reconstructed image, so as to optimize the model. Most of the loss functions used in current image SR reconstruction work are inherited from previous work in

other image fields. For example, at the beginning, people simply adopted L2 loss in order to pursue higher PSNR. But as the PSNR gets higher and higher, the researchers found that the image quality has not achieved good visual effects. Therefore, it is very important to find a set of loss function (Wu *et al*. 2020b) and a more accurate evaluation index system for image SR reconstruction in the future.

7. Conclusions

This article investigates the application of deep-learning based image SR reconstruction algorithms for improving the accuracy of machine-vision structural measurement. Firstly, the CNN-based and GAN-based SR models are introduced, together with the discussion of their advantages and disadvantages. Secondly, several datasets and evaluation index commonly used in image SR reconstruction work are introduced for the validation of the SR methods. Then, several experiments are conducted in order to test the potential of applying SR algorithms into visual measurement. Experiment results show that SR algorithms can improve the accuracy of corner detection and have great development prospects in visual measurement. However, the currently existing model performance evaluation indexes cannot perfectly access the algorithm performance in visual measurement. Lastly, the future research directions and challenges of image SR reconstruction are proposed from the aspects of model design, performance optimization and practical application.

Acknowledgments

This work was supported in part by the Fujian Provincial Department of Science and Technology of China under Grant 2019H0006, 2021J01580 and 2022H0008; in part by the National Natural Science Foundation of China under Grant 51508105.

References

Agustsson, E. and Timofte, R. (2017), "Ntire 2017 challenge on single image super-resolution: Dataset and study", *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, July, pp. 1122-1131.

Ahn, N., Kang, B. and Sohn, K.A. (2018), "Fast, accurate, and lightweight super-resolution with cascading residual network", *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, September, pp. 252-268.

Arbelaez, P., Maire, M., Fowlkes, C. and Malik, J. (2011), "Contour detection and hierarchical image segmentation", *IEEE Transact*. *Pattern Anal*. *Mach*. *Intell*., **33**(5), 898-916. https://doi.org/10.1109/TPAMI.2010.161

Bevilacqua, M., Roumy, A., Guillemot, C. and Alberi-Morel, M.L. (2012), "Low-complexity single-image super-resolution based on nonnegative neighbor embedding", *Proceedings of the 23rd British Machine Vision Conference* (*BMVC*), Surrey, UK, September, pp. 1-10. http://dx.doi.org/10.5244/C.26.135

Blu, T., Thévenaz, P. and Unser, M. (2004), "Linear interpolation revitalized", *IEEE Transact*. *Image Process*., **13**(5), 710-719. https://doi.org/10.1109/TIP.2004.826093

Bouguet, J.Y. (2013), Camera Calibration Toolbox for Matlab, Wwwvisioncaltechedu/bouguetj

Bulat, A., Yang, J. and Tzimiropoulos, G. (2018), "To learn image super-resolution, use a GAN to learn how to do image degradation first", *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, September, pp. 185-200.

Caetano, E., Silva, S. and Bateira, J. (2007), "Application of a vision system to the monitoring of cable structures", *Proceedings of the 7th International Symposium on Cable Dynamics*, Vienna, Austria, December, pp. 225-236.

Cai, J., Zeng, H., Yong, H., Cao, Z. and Zhang, L. (2019), "Toward real-world single image super-resolution: A new benchmark and a new model", *Proceedings of the IEEE*/*CVF International Conference on Computer Vision*, Seoul, Korea, October-November, pp. 3086-3095.

Cha, Y.J., Chen, J.G. and Büyüköztürk, O. (2017), "Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters", *Eng*. *Struct*., **132**, 300-313. https://doi.org/10.1016/j.engstruct.2016.11.038

Chang, H., Yeung, D.Y. and Xiong, Y. (2004), "Super-resolution through neighbor embedding", *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *CVPR 2004*, Washington, DC, USA, June-July. https://doi.org/10.1109/CVPR.2004.1315043

Chantas, G.K., Galatsanos, N.P. and Woods, N.A. (2007), "Super-resolution based on fast registration and maximum a posteriori reconstruction", *IEEE Transact*. *Image Process*., **16**(7), 1821-1830. https://doi.org/10.1109/TIP.2007.896664

Chen, Y.W. (2011), "Learning-Based Super-Resolution", IEICE Technical Report; **111**, 61-66.

Dong, C., Loy, C.C., He, K. and Tang, X. (2014), "Learning a deep convolutional network for image super-resolution", *Proceedings of the 13th European Conference on Computer Vision*, Zurich, Switzerland, September, pp. 184-199.

Dong, C., Loy, C.C. and Tang, X. (2016), "Accelerating the super-resolution convolutional neural network", *Proceedings of European Conference on Computer Vision*, Amsterdam, The Netherlands, October, pp. 391-407. https://doi.org/10.1007/978-3-319-46475-6_25

Elad, M. and Feuer, A. (2002), "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images", *IEEE Transact*. *Image Process*., **6**(12), 1646-1658. https://doi.org/10.1109/83.650118

Feng, D. and Feng, M.Q. (2015), "Model updating of railway bridge using in situ dynamic displacement measurement under trainloads", *J*. *Bridge Eng*., **20**(12), 4015019. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000765

Finkelstein, S.E., Schrump, D.S., Nguyen, D.M., Hewitt, S.M., Kunst, T.F. and Summers, R.M. (2003), "Comparative evaluation of super high-resolution CT scan and virtual bronchoscopy for the detection of tracheobronchial malignancies", *Chest*, **124**(5), 1834-1840. https://doi.org/10.1378/chest.124.5.1834

Freeman, W.T., Jones, T.R. and Pasztor, E.C. (2002), "Example-based super-resolution", *Comput*. *Graphics Applicat*. *IEEE*, **22**(2), 56-65. https://doi.org/10.1109/38.988747

Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T. and Aizawa, K. (2016), "Manga109 dataset and creation of metadata", *Proceedings of the 1st International Workshop on Comics Analysis*, *Processing and Understanding*, Cancun, Mexico, December, pp. 1-5. https://doi.org/10.1145/3011549.3011551

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014),

"Generative adversarial nets", (Ghahramani, Z., Welling, M., Cortes, C., *et al*.), In: *Advances in Neural Information Processing Systems*, Vol. 27, pp. 2672-2680.

Hansen, R.S., Waldram, D.W., Thai, T.Q. and Berke, R.B. (2021), "Super Resolution Digital Image Correlation (SR-DIC): An Alternative to Image Stitching at High Magnifications", *Experim*. *Mech*., **61**(9), 1351-1368. https://doi.org/10.1007/s11340-021-00729-2

He, K., Zhang, X., Ren, S. and Sun, J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June, pp. 770-778.

Hore, A. and Ziou, D. (2010), "Image quality metrics: PSNR vs. SSIM", *Proceedings of 2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, August, pp. 2366-2369. https://doi.org/10.1109/ICPR.2010.579

Huang, J.B., Singh, A. and Ahuja, N. (2015), "Single image super-resolution from transformed self-exemplars", *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June, pp. 5197-5206.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017), "Densely connected convolutional networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July, pp. 2261-2269.

Irani, M. and Peleg, S. (1991), "Improving resolution by image registration", *CVGIP*: *Graphical models and image processing*, **53**(3), 231-239. https://doi.org/10.1016/1049-9652(91)90045-L

Keys, R.G. (2003), "Cubic convolution interpolation for digital image processing", *IEEE Transact*. *Acoust*. *Speech Signal Process*., **29**(6), 1153-1160. https://doi.org/10.1109/TASSP.1981.1163711

Kim, J., Lee, J.K. and Lee, K.M. (2016a), "Accurate image super-resolution using very deep convolutional networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June, pp. 1646-1654.

Kim, J., Lee, J.K. and Lee, K.M. (2016b), "Deeply-recursive convolutional network for image super-resolution", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June, pp. 1637-1645.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012), "Imagenet classification with deep convolutional neural networks", *Adv*. *Neural Inform*. *Process*. *Syst*., **25**.

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", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July, pp. 105-114.

Lee, J.H., Jung, C.Y., Choi, E. and Cheung, J.H. (2017), "Vision-based multipoint measurement systems for structural in-plane and out-of-plane movements including twisting rotation", *Smart Struct*. *Syst*., *Int*. *J*., **20**(5), 563-572. https://doi.org/10.12989/sss.2017.20.5.563

Li, J., Fang, F., Mei, K. and Zhang, G. (2018), "Multi-scale residual network for image super-resolution", *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, September, pp. 527-532.

Li, B., Wang, B., Liu, J., Qi, Z. and Shi, Y. (2020), "s-LWSR: Super lightweight super-resolution network", *IEEE Transact*. *Image Process*., **29**, 8368-8380. https://doi.org/10.1109/TIP.2020.3014953

Lim, B., Son, S., Kim, H., Nah, S. and Mu Lee, K. (2017), "Enhanced deep residual networks for single image super-resolution", *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, USA, July, pp. 136-144.

Lin, Z., He, J., Tang, X. and Tang, C.K. (2008), "Limits of learning-based superresolution algorithms", *Int*. *J*. *Comput*. *Vis*., **80**(3), 406-420. https://doi.org/10.1007/s11263-008-0148-2

Liu, Y., Dong, Z., Lim, K.P. and Ling, N. (2020a), "A densely connected face super-resolution network based on attention mechanism", *Proceedings of 2020 15th IEEE Conference on Industrial Electronics and Applications* (*ICIEA*), Kristiansand, Norway, November, pp. 148-152. https://doi.org/10.1109/ICIEA48937.2020.9248111

Liu, K., Ma, Y., Xiong, H., Yan, Z., Zhou, Z., Fang, P. and Liu, C. (2020b), "Medical image super-resolution method based on dense blended attention network", *Laser & Optoelectronics Progress*, **57**(2), 021014. https://doi.org/10.48550/arXiv.1905.05084

Luo, Y., Zhou, L., Wang, S. and Wang, Z. (2017), "Video satellite imagery super resolution via convolutional neural networks", *IEEE Geosci*. *Remote Sens*. *Lett*., 14(12), 2398-2402. https://doi.org/10.1109/LGRS.2017.2766204

Martin, D., Fowlkes, C., Tal, D. and Malik, J. (2001), "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", *Proceedings of the 8th IEEE International Conference on Computer Vision* (*ICCV 2001*), Vol. 2, Vancouver, BC, Canada, July, pp. 416-423. https://doi.org/10.1109/ICCV.2001.937655

Mecheri, K., Ziou, D. and Deschenes, F. (2007), "Super-resolution based on interpolation and global sub pixel translation", *Proceedings of the 4th Canadian Conference on Computer and Robot Vision* (*CRV'07*), Montreal, QC, Canada, May, pp. 448-458. https://doi.org/10.1109/CRV.2007.62

Nguyen, M.Q., Atkinson, P.M. and Lewis, H.G. (2005), "Superresolution mapping using a Hopfield neural network with LIDAR data", *IEEE Geosci*. *Remote Sens*. *Lett*., **2**(3), 366-370. https://doi.org/10.1109/LGRS.2005.851551

Ni, K.S. and Nguyen, T.Q. (2007), "Image superresolution using support vector regression", *IEEE Trans*. *Image Process*., **16**(6), 1596-1610. https://doi.org/10.1109/TIP.2007.896644

Oh, B.K., Hwang, J.W., Kim, Y., Cho, T. and Park, H.S. (2015), "Vision-based system identification technique for building structures using a motion capture system", *J*. *Sound Vib*., **356**, 72-85. https://doi.org/10.1016/j.jsv.2015.07.011

Ojio, T., Carey, C., O'Brien, E.J., Doherty, C. and Taylor, S.E. (2016), "Contactless bridge weigh-in-motion", *J*. *Bridge Eng*., **21**, 4016032. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000776

Park, S.C., Park, M.K. and Kang, M.G. (2003), "Super-resolution image reconstruction: a technical overview", *IEEE Signal Process*. *Magaz*., **20**(3), 21-36. https://doi.org/10.1109/MSP.2003.1203207

Park, S.J., Son, H., Cho, S., Hong, K.S. and Lee, S. (2018), "Srfeat: Single image super-resolution with feature discrimination", *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, September, pp. 439-455.

Rasti, P., Uiboupin, T., Escalera, S. and Anbarjafari, G. (2016), "Convolutional neural network super resolution for face recognition in surveillance monitoring", *Proceedings of International Conference on Articulated Motion and Deformable Objects*, Palma de Mallorca, Spain, July, pp. 175-184. https://doi.org/10.1007/978-3-319-41778-3_18

Schultz, R.R. and Stevenson, R.L. (1996), "Extraction of high-resolution frames from video sequences", *IEEE Transact*. *Image Process*. *A Publicat*. *IEEE Signal Process*. *Soc*., **5**(6), 996-1011. https://doi.org/10.1109/83.503915

Seif, G. and Androutsos, D (2018), "Edge-based loss function for single image super-resolution", *Proceedings of 2018 IEEE International Conference on Acoustics*, *Speech and Signal*

Processing (*ICASSP*), Calgary, Alberta, Canada, April, pp. 1468-1472. https://doi.org/10.1109/ICASSP.2018.8461664

Sheikh, H.R., Bovik, A.C. and De Veciana, G. (2006), "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Transact*. *Image Process*., **14**(12), 2117-2128. https://doi.org/10.1109/TIP.2005.859389

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z. (2016), "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June, pp. 1874-1883.

Song, X., Dai, Y. and Qin, X. (2016), "Deep depth super-resolution: Learning depth super-resolution using deep convolutional neural network", *Proceedings of the 13th Asian Conference on Computer Vision*, Taipei, Taiwan, November, pp. 360-376. https://doi.org/10.1007/978-3-319-54190-7_22

Song, X., Liu, W., Liu, J., Liu, C., Lu, C. and Gao, H. (2019), "Deep CNN jointing low-high level feature for image super-resolution", *Proceedings of the 10th International Conference on Graphics and Image Processing* (*ICGIP 2018*), Vol. 11069, Chengdu, China, December, pp. 1059-1065. https://doi.org/10.1117/12.2524412

Stark, H. and Oskoui, P. (1989), "High-resolution image recovery from image-plane arrays, using convex projections", *J. Optical Soc*. *Am*. *A Optics Image Sci*., **6**(11), 1715-1726.

Tai, Y., Yang, J., Liu, X. and Xu, C. (2017a), "Memnet: A persistent memory network for image restoration", *Proceedings of the IEEE International Conference on Computer Vision*, Salt Lake City, UT, USA, June, pp. 4539-4547.

Tai, Y., Yang, J. and Liu, X. (2017b), "Image super-resolution via deep recursive residual network", *Proceedings of ~~the 30th~~ 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, USA, July, pp. 2790-2798.

Tian, J. and Ma, K.K. (2011), "A survey on super-resolution imaging", *Signal Image Video Process*., **5**(3), 329-342. https://doi.org/10.1007/s11760-010-0204-6

Tong, T., Li, G., Liu, X. and Gao, Q. (2017), "Image super-resolution using dense skip connections", *Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, Italy, October, pp. 4809-4817.

Trajković, M. and Hedley, M (1998), "Fast corner detection", *Image Vision Comput*., **16**(2), 75-87. https://doi.org/10.1016/S0262-8856(97)00056-5

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004), "Image quality assessment: from error visibility to structural similarity", *IEEE Trans*. *Image Process*., **13**(4), 600-612. https://doi.org/10.1109/TIP.2003.819861

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C. (2018), "Esrgan: Enhanced super-resolution generative adversarial networks", *Proceedings of the European Conference on Computer Vision* (*ECCV*) *Workshops*, Munich, Germany, September.

Wang, Z., Chen, J. and Hoi, S.C. (2020a), "Deep learning for image super-resolution: A survey", *IEEE Transact*. *Pattern Anal*. *Mach*. *Intell*., **43**(10), 3365-3387. https://doi.org/10.1109/TPAMI.2020.2982166

Wang, W., Hu, Y., Luo, Y. and Zhang, T. (2020b), "Brief survey of single image super-resolution reconstruction based on deep learning approaches", *Sens*. *Imag*., **21**(1), 1-20. https://doi.org/10.1007/s11220-020-00285-4

Wu, L.J., Casciati, F. and Casciati, S. (2014), "Dynamic testing of a laboratory model via vision-based sensing", *Eng*. *Struct*., **60**, 113-125. https://doi.org/10.1016/j.engstruct.2013.12.002

Wu, L., Su, Y., Chen, Z., Chen, S., Cheng, S. and Lin, P. (2020a), "Six-degree-of-freedom generalized displacements measurement based on binocular vision", *Struct*. *Control Health Monitor*., **27**(2), e2458. https://doi.org/10.1002/stc.2458

Wu, Q., Fan, C., Li, Y., Li, Y. and Hu, J. (2020b), "A novel perceptual loss function for single image super-resolution", *Multimedia Tools Applicat*., **79**(29), 21265-21278. https://doi.org/10.1007/s11042-020-08878-7

Xu, Y. and Brownjohn, J.M. (2018), "Review of machine-vision based methodologies for displacement measurement in civil structures", *J*. *Civil Struct*. *Health Monitor*., **8**, 91-110. https://doi.org/10.1007/s13349-017-0261-4

Yan, X., Shen, Q. and Liu, X. (2015), "Super-resolution reconstruction for license plate image in video surveillance system", *Proceedings of 2015 10th International Conference on Communications and Networking in China* (ChinaCom), Shanghai, China, August, pp. 643-647. https://doi.org/10.1109/CHINACOM.2015.7498016

Yang, J., Wright, J., Huang, T.S. and Ma, Y. (2010), "Image super-resolution via sparse representation", *IEEE Transact*. *Image Process*., **19**(11), 2861-2873. https://doi.org/10.1109/TIP.2010.2050625

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H. and Liao, Q. (2019a), "Deep learning for single image super-resolution: A brief review", *IEEE Transact*. *Multimedia*, **21**(12), 3106-3121. https://doi.org/10.1109/TMM.2019.2919431

Yang, W., Wang, W., Zhang, X., Sun, S. and Liao, Q. (2019b), "Lightweight feature fusion network for single image super-resolution", *IEEE Signal Process*. *Lett*., **26**(4), 538-542. https://doi.org/10.1109/LSP.2018.2890770

Ye, X.W., Yi, T.H., Dong, C.Z. and Liu, T. (2016a), "Vision-based structural displacement measurement: System performance evaluation and influence factor analysis", *Measurement*, **88**, 372-384. https://doi.org/10.1016/j.measurement.2016.01.024

Ye, X.W., Dong, C.Z. and Liu, T. (2016b), "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms", *Smart Struct*. *Syst*., *Int*. *J*., **17**(6), 935-956. https://doi.org/10.12989/sss.2016.17.6.935

Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M. and Spencer Jr, B.F. (2016), "Target-free approach for vision-based structural system identification using consumergrade cameras", *Struct*. *Control Health Monitor*., **23**(12), 1405-1416. https://doi.org/10.1002/stc.1850

Yu, J., Fan, Y., Yang, J., Xu, N., Wang, Z., Wang, X. and Huang, T. (2018), "Wide activation for efficient and accurate image super-resolution", arXiv preprint arXiv:1808.08718. https://doi.org/10.48550/arXiv.1808.08718

Zeyde, R., Elad, M. and Protter, M. (2010), "On single image scale-up using sparse-representations", *Proceedings of International Conference on Curves and Surfaces*, Avignon, France, June, pp. 711-730. https://doi.org/10.1007/978-3-642-27413-8_47

Zhang, Z. (2000), "A flexible new technique for camera calibration", *IEEE Transact*. *Pattern Anal*. *Mach*. *Intell*., **22**(11), 1330-1334. https://doi.org/10.1109/34.888718

Zhang, K., Zuo, W. and Zhang, L. (2017), "Learning a single convolutional super-resolution network for multiple degradations", *Proceedings of Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July, pp. 33662-3271.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y. (2018a), "Image super-resolution using very deep residual channel attention networks", *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, September, pp. 286-301.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B. and Fu, Y. (2018b), "Residual dense network for image super-resolution", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June, pp. 2472-2481.

Zhang, Y., Liu, S., Dong, C., Zhang, X. and Yuan, Y. (2019), "Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution", *IEEE Transact. Image Process*., **29**, 1101-1112. https://doi.org/10.1109/TIP.2019.2938347