Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19292-8

Enhanced collapsible linear blocks for arbitrary sized image super-resolution

Prathap Soma[1] · Xiaosong Yang[2] · Jian Chang[2] · Jian Jun Zhang[2]

Received: 30 November 2023 / Revised: 23 March 2024 / Accepted: 22 April 2024

Abstract
Image up-scaling and super-resolution (SR) techniques have been a hot research topic for many years due to its large impact in the field of medical imaging, surveillance etc. Especially single image super-resolution (SISR) become very popular because of the fast development of deep convolution neural network (DCNN) and the low requirement on the input. They are achieving outstanding performance. However, there are still problems in the state-of-the-art works, especially from two perspectives: 1. failed at exploiting the hierarchical characteristics from the input, resulting in loss of information and artifacts in the final high resolution (HR) image; 2. failed to handle arbitrary-sized images; the existing research works are focused on fixed size input images. To address these challenges, this paper proposed a residual dense network (RDN) and multi-scale sub-pixel convolution network (MSSPCN) which are integrated into a Collapsible Linear Block Super Efficient Super-Resolution (SESR) network. The RDNs aims to tackle the first challenge, carrying the hierarchical features from end-to-end. An adaptive cropping strategy (ACS) technique is introduced before feature extraction targeting at the image size challenge. The novelty of this work is extracting the hierarchical features and integrating RDNs with MSSPCNs. The proposed network can upscale any arbitrary-sized image (1080p) to $\times 2$ (4K) and $\times 4$ (8K). To secure ground truth for evaluation, this paper follows the opposite flow, generating the input LR images by down-sampling the given HR images (ground truth). To evaluate the performance, the proposed algorithm is compared with eight state-of-the-art algorithms, both quantitatively and qualitatively. The results are verified on six benchmark datasets. The extensive experiments justify that the proposed architecture performs better than other methods and upscales the images satisfactorily.

✉ Prathap Soma
psoma@bournemouth.ac.uk

✉ Xiaosong Yang
xyang@bournemouth.ac.uk

Jian Chang
jchang@bournemouth.ac.uk

Jian Jun Zhang
jzhang@bournemouth.ac.uk

[1] NCCA Faculty of Media and Communications, Bournemouth University, Bournemouth, UK

[2] National Center for Computer Animation, Bournemouth University, Bournemouth, UK

Published online: 21 May 2024

Keywords Super resolution · Residual dense network · Collapsible linear blocks · Adaptive cropping strategy · Sub-pixel convolutional layer

1 Introduction

Generating the high resolution image (HR) from its equivalent low resolution (LR) is a dominant research topic in computer vision and image processing applications. This piece of work is known as image up-scaling or super-resolution (SR). The major application includes HDTV, satellite imaging, face recognition, medical imaging, surveillance and mobile applications. In the context of SR, LR data is previewed as a low pass filtered or downscaled version of HR data. During this process, there is a chance of loosing high-frequency data makes upscaling as an ill-posed problem. In addition, the SR operation are basically one-to-many mapping that can have many solutions. The major key assumption in all SR technique is that the high-frequency data is redundant and can be generated from low frequency components. As a result SR is an inference problem.

A few SR methods have used multiple images of the same scene to generate the HR images. It can be called as multi-image SR method. They take advantage of explicit redundancy (due to overlapping) by attempting to invert the downsampling process. Moreover, these methods required complicate computation such as image restoration and fusion. If the user's focus is on quality, the network will require more computations. Less computations will lead to degraded image. Therefore, most of the time, the user needs to find a trade-off between quality and computational complexity.

Another side, single-image super-resolution techniques have been proposed. They learn to generate the HR information from implicit redundant data of LR image. These methods requires prior information and the generating time is inversely proportional to the image redundant information. This technique does not require multiple images. In practice, it is difficult to capture the multiple images of the same scene, especially when there are moving objects. Therefore most of researchers preferred single image super resolution techniques. Hence we employed single-image strategy. We adopted the collapsible linear blocks model integrated with residual dense networks, multi-scale sub-pixel convolutional layers and an adaptive cropping strategy. The key contributions and advantages of our model is as follows:

- A residual dense network with feed forwarding the input is adopted in our design. It will exploit the hierarchical features from LR to HR space, resulting in producing the artifact-free image.
- Multi-scale sub-pixel convolutional layer is used as an up-sampler in our model. It provides multi-range contextual information for image super-resolution.
- The adaptive cropping strategy is employed for up-scaling any arbitrary sized image based on the scaling factor.

Compared to earlier SISR approaches, our method can achieve a superior result.

2 State-of-the-art

To solve the problem of SR, numerous approaches based on deep neural architectures [1, 7] have been developed and produced satisfactory results. Firstly, Dong et al. [8, 9] proposed a 3-layer convolutional neural network (SRCNN) to generate HR images from LR but failed at generating the different-sized up-scale images. Motivated by the VGG (ImageNet classi-

fication) model [3], Kim et al. [7, 11] developed a very deep convolutional neural network. It enhances the quality but difficult to train. They have tackled this problem using deeply-recursive convolutional network (DRCN) [6]. The parameters of this model are reduced significantly compared with existing algorithms. However, it failed at preserving sensitive image information such as depth, texture, and edges. Later, Wang et al. [10] proposed a deep network framework based on the new technique called sparse prior coding to achieve satisfactory results at higher scaling factor ($\times 4$, $\times 8$..). The network architecture is also deep and complex.

To overcome the complexity issue, Tai et al. [12] developed a deep recursive residual network (DRRN) with memory networks (MemNet) [14]. The MemNet eliminates the problem of complexity. This technique is designed based on recursive learning to optimize the model parameters. But, MemNets suffered from long-term dependency problems because of the huge number of memory blocks utilized. Furthermore, this algorithm is only suitable for LR-HR image pairs and not suitable for generating new images. It also required a longer running time, heavy computational cost, and large graphics memory during the training and testing phase.

Later, based on mapping techniques Shi et al. [15] proposed sub-pixel convolutional network. They upsample the features at the end of the main architecture. But it take more time for generation. Dong et al. [16] proposed FSRCNN that used learnable upsampling layer to achieve post-upsampling super resolution and suffered from long-term dependency. The Laplacian pyramid SR network (LapSRN) was proposed by Lai et al. [17]. Based on this network, MS-LapSRN [20] and ProSR [21] networks are proposed and achieved better results than LapSRN, but failed at preserving texture information. Next, EDSR [18] achieved a considerable improvement in SR and won the NTIRE 2017 competition [19]. To get better outcomes, the authors have eliminated some of the redundant modules in the SRResNet [22]. Such framework required high-computational cost because the most of CNN operations are performed in the HR space. In addition to residual block in EDSR, Zhang et al. [23] added densely connected block and constructed residual dense network called RDN. Later, they proposed Channel attention mechanism to implement very deep residual attention networks (RCAN). Recently, Zhang et al. [23] developed the residual non-local attention network (RNAN) [24] for various image restoration tasks by adding spatial attention (non-local module) to the residual block. Even though the residual blocks performed well, the upscaled image did not enhance up to a satisfactory level.

Later, some researchers have focused on developing the algorithms on resource-constrained devices for real-time applications. Unfortunately, the above methods are difficult to deploy due to their operating speed (training and testing) and computational complexity. In [25]-[28], they suggested that linearly over-parametrization techniques are helpful for fast processing with less computational cost. Hence, such techniques are suitable for implementation on resource-constrained devices. Arora et al. [25] demonstrated theoretically that the linear over-parameterization with fully connected layers can accelerate the training of deep linear networks by acting as a time-varying momentum and adaptive learning rate. Recent work on ExpandNets [28] and ACNet [27] propose to overparameterize a convolutional layer and show that it accelerates the training of various CNNs and enhances image quality. Since the real-time SR algorithms are working with fewer computations, the generated HR image will become degraded. Therefore, it is necessary to enhance the quality of upscaled image. Hence, we proposed a model to generate HR data from LR feature maps and enhance the resolution from LR to HR at the very end of the network. To achieve this, we adopted a multi-scale sub-pixel convolution layer.

The rest of this paper is structured as follows. Section 3 describing the motivation from the literature. Section 4 presenting the detailed explanation of proposed model. The experimental setup and results analysis briefly demonstrated in Section 5 and at last, conclusions & future scope are drawn in Section 6.

3 Motivation

The majority of deep CNN-based SR models [1, 8, 9, 13] perform poorly because they do not fully exploit the hierarchical features such as edges, corners, and structure of the image from the original low-resolution (LR) images. Therefore, we provide a unifying framework for SR on high-quality images using residual dense networks (RDNs). The RDNs are made to enable direct connections from the state of the previous RDN to every layer of the current RDN and to extract an enormous number of local features through densely connected convolutional layers [29]. All of the hierarchical characteristics from the initial LR images are fully used by the network. This RDNs required a special memory to store the read state from the previous RDN called Contiguous memory (CM). Later, this hierarchical features from all RDNs in the LR space will be adaptively fused using the global feature fusion mechanism. This mechanism is used to produce the global dense residual features from the original LR images, by combining the shallow and deep features. In this way, the hierarchical features can be exploit from the LR to HR images.

According to Osendorfer et al. [31], the size of the image gradually increases at the middle of the network. Increasing the resolution at the network's first layer or earlier is another strategy [9, 32]. But this strategy has a number of shortcomings. Firstly, the computational complexity is increased by increasing the resolution of the LR pictures before the image enhancing process. It not only affects the training time of convolutional networks but also extremely degrade the picture quality. Moreover, this strategy required interpolation techniques and did not provide any new information to generate the HR image.

Contrary to other research, we proposed a new model to generate HR data from LR feature maps and enhance the resolution from LR to HR at the very end of the network. To achieve this novelty work, we adopted a multi-scale sub-pixel convolution layer [34]. We directly takes the hierarchical features in the form of depth maps from residual dense network (RDN) and feed into the multi-scale sub-pixel convolution layer with a varying factor 2. This way we preserved the detailed information of the image.

The major advantages of using multi-scale sub-pixel convolution layer in our method are

- The final network layer in the proposed model is responsible for upscaling. Firstly, LR image's (1080x1920) information is supplied directly to the network, and depth-maps extraction takes place using 32-residual dense networks (RDNs) in LR space. We employed a smaller sized filters (3x3) to combine the same information while preserving a specific contextual region due to the reduced input resolution. The major advantage of employing small sized filters is extracting the detailed, smaller complex information of the image.
- Instead of learning from a single small sized upscaling filter with L layers, we have trained n_{L-1} upscaling filters for the n_{L-1} depth-maps. The typical values of L=32 and n_{L-1}=31. We did not use any explicit interpolation filters. However, the network implicitly picks up the processing information required for SR. In contrast to a single fixed filter upscaling at the first layer, the network can learn a better and more intricate LR to HR mapping. As a result, generation accuracy of the model will be improved further.

4 Enhanced collapsible linear block super resolution (ECLB-SR)

Figure 1 depicts the proposed network (ECLB-SR) architecture. It consists of a Collapsible Linear Blocks (CLBs) with a novel multi stage residual dense connections and an up-sampler with multi-scale sub-pixel convolution layer. Adaptive cropping strategy(ACS) is another technique used for handling the arbitrary sized images, which will be discussed in detailed in Section 4.2. Initially, ACS blocks takes the input LR image. If the image dimensions are divisible by four the proposed model works normally. If it is not divisible by four can be called an arbitrary image. When it happens the entire image is divided into four equal parts, and then process each patch independently in the main network. The main network consist of sequence of linear convolutional layer blocks (LB) and multi Scale Sub-pixel Convolution Layer (MSSPCL). The linear convolutional layer blocks are responsible for exploiting the hierarchical features from LR to HR space, resulting in producing the artifact-free image. At last, multi scale sub-pixel convolution layers upscaling the LR image to a desired level resulting in HR image.

The collapsible linear blocks are sequence of linear convolutional layers that can be mathematically reduced at inference time to single or narrow convolutional layer (in terms of input/output channels). For example, with x input channels, y output channels can be produced with the help of k×k and 1×1 linear blocks as shown in Fig. 2. It can be done as, a k×k linear block with x input channels firstly expand to p intermediate channels using a k×k convolution ($p >> x$). Then, a 1×1 convolution is used to project the p intermediate channels to y final output channels. Since no non-linearity is used between these two convolutions blocks, they can be analytically collapsed into a single narrow convolution layer at inference time, hence, the name Collapsible Linear Blocks (CLB). All these block connected together with residual dense networks (RDNs).

The structure of a residual dense network (RDN) is shown in Fig. 3. The RDN have dense connected layers and local feature fusion (LFF) with local residual learning (LRL). The LFF extracts the features like patterns, shapes and depth information of the images. The LR input image I, which is passed through conv and ReLU layers resulting in produce the F_{-1} features. The output of each stage is passed to another stage along with input image (I) through connected layers. This process is called as contiguous state pass. This process repeats until the deep hierarchical features are extracted. At the final stage, all the RDN states are concatenated. In parallel to this process, the LRL extracts multi-level local dense features by adaptively preserving the information. Moreover, LRL allows a very high growth rate by stabilizing the training of a wider network. After extracting multi-level local dense features, we have conducted a global feature fusion (GFF) technique for capturing the region of interest (ROI) features. The mathematical analysis of this approach is given below:

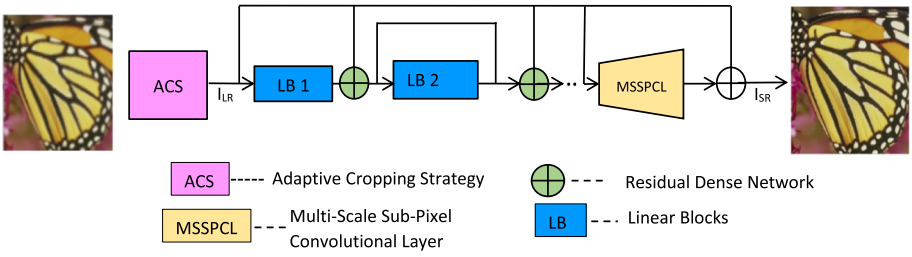

Fig. 1 Architecture of proposed method

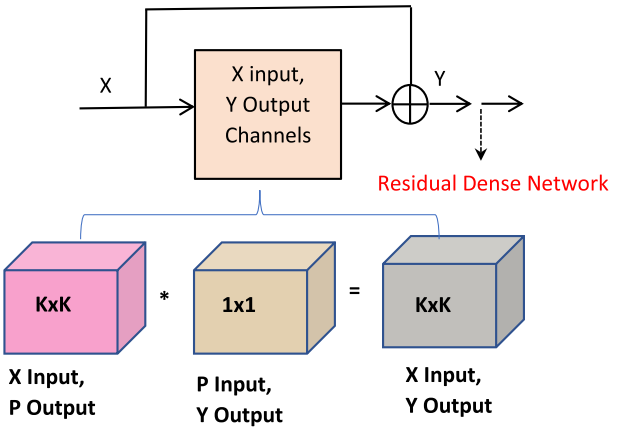

Fig. 2 Collapsible linear blocks

The *f=64* Conv layers are used to extract the features from input low resolution image(I_{LR}) and denoted as

$$I = 5 \times 5 \times f \times 1 \times I_{LR} \tag{1}$$

The I is used for further shallow feature extraction and global residual learning. It can be denoted as

$$F_{-1} = \sigma\left(H_{SFE1}\left(I\right)\right) \tag{2}$$

Where σ is an activation function typically ReLU and $H_{SFE1}(.)$ denotes convolution operation. F_{-1} is the output of preceding residual dense network. Similarly we have $n = 32$ residual dense network with outputs of $F_0 ..., F_n$ respectively is given by

$$F_0 = \sigma\left(H_{SFE2}\left(F_{-1}, I\right)\right) \tag{3}$$

$$F_n = \sigma\left(H_{SFEn}\left[I, F_{-1}, F_0, F_1 ... F_{n-1}\right]\right) \tag{4}$$

After extracting hierarchical features with a set of RDNs, we further conduct local feature fusion (LFF) and global feature fusion (GFF). LFF makes full use of features from all the preceding layers and can be represented as

$$F_{d,LF} = H^d_{LFF}\left(\left[I, F_{-1}, F_0, ... F_n\right]\right) \tag{5}$$

where $F_{d,LF}$ output feature maps in the LR space. These features together with LR image are fed to the up-sampler resulting in I_{SR}.

$$I_{SR} = upsampler\left(F_D\right) \tag{6}$$

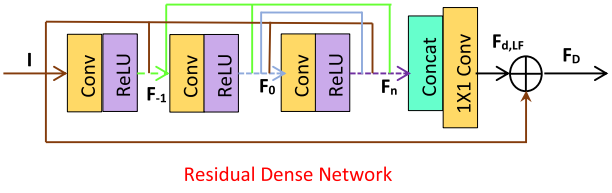

Fig. 3 Residual Dense Network

where, $F_D = I + F_{d,LF}$.

This up-sampler works based on the multi-scale sub-pixel Convolution layer technique. The

4.1 Multi scale sub-pixel convolution layer

Existing techniques have adopted interpolation technique to upscale the LR images within the LR space. They used a fractional stride of $\frac{1}{r}$ (r=1,2,3...), followed by a convolution with stride 1 in the HR space. This procedure involves a lot of convolution operations and computational complexity, r^2. An innovative technique is proposed to tackle this problem by employing a convolution stride $\frac{1}{r}$ in the LR space with a filter W_s having a size k_s and weight spacing $\frac{1}{r}$. It will be activated in different parts of W_s for convolution. The weights between these pixels are not activated therefore no need to compute them and the total number of activated parts are r^2. During this time $\left(\frac{k_s}{r}\right)^2$ number of weights are activated. These activated parts are periodically changed across the image depending upon sub-pixel location during the convolution operation. Eventually, the number of computations are reduced.

The super resolution of the image (I) using sub-pixel convolution layer can be defined as

$$I^{SR} = f^L\left(I^{LR}\right) = PS\left(W_L * f^{L-1}\left(I^{LR}\right) + b_L\right) \tag{7}$$

where I^{SR}, I^{LR} are the high resolution, low resolution images respectively. f^{L-1} is the neural network with $L-1$ layers, where L=64. $PS(.)$ is a periodic shuffling operator that rearranges the pixels of $H \times W \times C.r^2$ image (LR) to $rH \times rW \times C$ image (HR). W_L represents the convolution operator has a shape $n_{L-1} \times r^2C \times k_L \times k_L$. The value of $k_L = \frac{k_s}{r}$ and mod (ks, r) = 0 it is equivalent to sub-pixel convolution in the LR space with the filter W_s.

In practical, the training set consisting of HR images I_n^{HR} $n = 1....N$, and corresponding LR images I_n^{LR} $n = 1....N$. We have used pixel-wise mean square error (MSE) as an objective function to train the network:

$$\iota\left(W_{1:L}, b_{1:L}\right) = \frac{1}{r^2HW}\sum_{x=1}^{rH}\sum_{x=1}^{rW}\left(I_{x,y}^{HR} - f_{x,y}^{L}\left(I^{LR}\right)\right)^2 \tag{8}$$

This network is $\log_2 r^2$ times faster than deconvolution layer and r^2 times faster compared to other upscaling before convolution. We further enhanced the speed of operation of this network by applying multiple depth maps with up-scaling factor 2 at each stage. These depth maps are directly taken from the residual dense blocks hence they carried the detailed information of the input image. Thus, we can easily preserve the detailed information in the upscaled image.

An input LR image having a height (H), width (W), and multiple feature maps then its multi-scale sub-pixel convolution feature map can be defined as,

$$P_{n\times} = PS\left(H\left(\left[P_{2n\times}, B_{2n\times}^{2_{n\times}}, B_{2n\times}^{2_{n\times}^3}, ..., B_{2n\times}^{32_{n\times}}\right]\right)\right) \quad n\epsilon\{1, 2, 4, 8\} \tag{9}$$

Here, $\left[P_{2n\times}, B_{2n\times}^{2^2_{n\times}}, B_{2n\times}^{2^3_{n\times}}, ..., B_{2n\times}^{32_{\times}}\right]$ are the input feature maps. $H(\cdot)$ represents the concatenation of all the feature maps through convolutional layers to make the reconstruction of upscaled high quality image.

4.2 Adaptive cropping strategy

The adaptive cropping strategy (ACS) is another novel technique proposed to upscale any arbitrary sized image. If the image dimensions are divisible by four the proposed model works normally. If it is not divisible by four can be called an arbitrary image. When it happens the entire image is divided into four equal parts, and then process each patch independently in the main network. The overlapped patches data will be eliminated at the final stage.

For example the Fig. 4 depicts the arbitrary sized image with four equal patches. Consider the first patch in the top left corner having a height and width $\left(\lfloor \frac{H}{2} \rfloor + \Delta I_H\right), \left(\lfloor \frac{W}{2} \rfloor + \Delta I_W\right)$ respectively. The ΔI_H and ΔI_W are the overlapped height and width to make the patch size divisible by 4. The below (10) represents fundamental principle of the image patch dimensions of the adaptive cropping technique

$$\begin{aligned} \left(\left\lfloor \frac{H}{2} \right\rfloor + \Delta I_H\right) \%4 = 0 \\ \left(\left\lfloor \frac{W}{2} \right\rfloor + \Delta I_W\right) \%4 = 0 \end{aligned} \tag{10}$$

The amount of extra added patch $(\Delta lH, \Delta lW)$ sizes can be calculated as

$$\begin{aligned} \Delta I_H = pad_H - \left(\left\lfloor \frac{H}{2} \right\rfloor + pad_H\right) \%4 \\ \Delta I_W = pad_W - \left(\left\lfloor \frac{W}{2} \right\rfloor + pad_W\right) \%4 \end{aligned} \tag{11}$$

where pad_H, pad_W are used for presetting the additional lengths. In general, these values can be set by

$$pad_H = pad_W = 4k \quad k \geq 1 \tag{12}$$

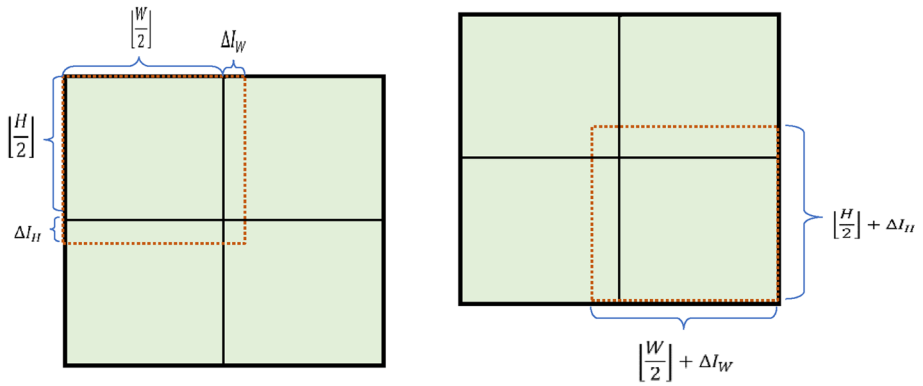

Fig. 4 Adaptive cropping strategy

where k is an integer. Once these patches are processed from the main network, the extra increments (ΔI_H and ΔI_W) are eliminated.

5 Experimental setup and results

We validate of our algorithm with six different datasets namely Set5 [36], Set14 [37], BSD100 [38], Urban100 [39], Manga 109 [40] and DIV2K [41]. We also compared with state-of-the-art algorithms in qualitatively and quantitatively. This framework is implemented on a desktop computer with Intel i7-11800H @ 2.30GHz CPU, 32GB RAM with NVIDIA GeForce RTX 3070 GPU (8GB RAM) for the any arbitrary image resolution to 4K ($\times 2$) and to 8K ($\times 4$). The following techniques are proposed: an adaptive cropping strategy to handle arbitrary-sized images, instead of PReLU a residual dense block, and a highly efficient multi-scale sub-pixel convolutional network before feature extraction during the upscaling. We used mean square error (MSE) as loss function between high quality (Ground Truth) and generated image. The network is trained by ADAM optimizer with a parameter $\beta_1 = 0.9$. For training efficiency, we take a random crops of size 64×64 and arbitrary size from each image; The reason behind choosing an image size 64×64 is that it can divisible by 4 and it is a minimum-sized image that humans can see and analyze easily. We considered 300 epochs and each epoch conducts 1600 training steps.

5.1 Preparing dataset

We train our models on the DIV2K training dataset with 800 HR images. The validation sets are Set5, Set14, BSD100, Urban100, and DIV2K. All these datasets are commonly used in super resolution techniques. The LR images are obtained using the MATLAB R2023a and bicubic kernel function of HR images. Based on [42], humans are more sensitive to luminance changes. We also considered the luminance channel(Y) in YCbCr colour space to evaluate performance of our model. For each upscaling factor ($\times 2$ and $\times 4$), we train a specific network.

5.2 Quantitative analysis

Table 1 present the comparison of state-of-the-art algorithms for $\times 2$ super resolution. We compared PSNR, SSIM and the total number parameters utilized by the networks. Based on the parameters the networks are divided into three types. A network with 25K or less parameters is small network, between 25K to 100K is medium network and having more than 100K parameters are called large network.

In the case of the small network category, we compared our proposed network with six state-of-the-art techniques Bicubic, FSRCNN [43], MOREMNAS-C [44], SESR-M3, M5 and M7 [42]. The proposed ECLB-M3 network achieved significantly better PSNR and SSIM values for all the datasets except Set5. But, the number of parameters utilized by our model are little higher. When compared with SESR-M3, our method utilized around 1.67K extra parametres. Since the goal of this research is to further improve the quality of SESR architecture [42]. It can achieve by ACS, multi-scale sub-pixel convolution layers, and replacing the pooling layers with RDN's blocks. These extra added layers have contributed more parameters.

Table 1 Comparison of PSNR/SSIM metrics on ×2 Super Resolution

Complexity	Model	Parameters	Set5	Set14	BSD100	Urban100	Manga109	DIV2K
Small	Bicubic	–	33.68/0.9307	30.24/0.8693	29.56/0.8439	26.88/0.8408	30.82/0.9349	32.45/0.9043
	FSRCNN	12.46K	36.98/0.9556	32.62/0.9087	31.50/0.8904	29.85/0.9009	36.62/0.9710	34.74/0.9340
	MOREMNAS-C	25K	37.06/0.9561	32.75/0.9094	31.50/0.8904	29.92/0.9023	–/–	–/–
	SESR-M3	8.91K	37.21/0.9577	32.70/0.9100	31.56/0.8920	29.92/0.9034	36.47/0.9717	35.03/0.9373
	SESR-M5	13.52K	37.39/0.9585	32.84/0.9115	31.70/0.8938	30.33/0.9087	37.07/0.9734	35.24/0.9389
	SESR-M7	18.12K	37.47/0.9588	32.91/0.9118	31.77/0.8946	30.49/0.9105	37.14/0.9738	35.32/0.9395
	ECLB-M3	10.58K	38.15/0.9650	34.84/0.9781	32.72/0.9153	31.42/0.9178	37.54/0.9843	36.12/0.9473
	ECLB-M5	15.74K	38.14/0.9617	33.12/ 0.9240	32.21/0.9056	30.74/0.9147	37.13/0.9755	35.51/0.9217
	ECLB-M7	19.57K	38.53/0.9708	33.76/0.9047	32.11/0.9043	31.41/0.9158	37.16/0.9778	35.98/0.9410
Medium	TPSR-NoGAN	60K	37.38/0.9583	33.00/0.9123	31.75/0.8942	30.61/0.9119	–/–	–/–
	SESR-M11	27.34K	37.58/0.9593	33.03/0.9128	31.85/0.8956	30.72/0.9136	37.40/0.9746	35.45/0.9404
	ECLB-M11	28.12K	37.84/0.9601	34.01/0.9208	32.41/0.8992	31.54/0.9178	38.13/0.9874	36.75/0.9412
Large	VDSR	665K	37.53/0.9587	33.05/0.9127	31.90/0.8960	30.77/0.9141	37.16/0.9740	35.43/0.9410
	LapSRN	813K	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.53/0.9740	35.31/0.9400
	BTSRN	410K	37.75/–	33.20/–	32.05/–	31.63/–	–/–	–/–
	CARN-M	412K	37.53/0.9583	33.26/0.9141	31.92/0.8960	31.23/ 0.9193	–/–	–/–
	MOREMNAS-B	1118K	37.58/0.9584	33.22/0.9135	31.91/0.8959	31.14/0.9175	–/–	–/–
	SESR-XL	105.37K	37.77/0.9601	33.24/ 0.9145	31.99/ 0.8976	31.16/0.9184	38.01/0.9759	35.67/0.9420
	ECLB-XL	105.98K	37.92/0.9612	34.19/0.9276	32.41/0.9147	31.54/0.9210	38.61/0.9880	36.61/0.9575

(Blue/ Red colored numbers indicates first/ second best values (PSNR/SSIM) among the methods) Input image size=1080x1920, Output image size=2160x3840

Table 2 Comparison of PSNR/SSIM metrics on ×4 Super Resolution

Complexity	Model	Parameters	Set5	Set14	BSD100	Urban100	Manga109	DIV2K
Small	Bicubic	–	28.43/0.8113	26.00/0.7025	25.96/0.6682	23.14/0.6577	24.90/0.7855	28.10/0.7745
	FSRCNN	12.46K	30.70/0.8657	27.59/0.7535	26.96/0.7128	24.60/0.7258	27.89/0.8590	29.36/0.8110
	SESR-M3	13.71K	30.75/0.8714	27.62/0.7579	27.00/0.7166	24.61/0.7304	27.90/0.8644	29.52/0.8155
	SESR-M5	18.32K	30.99/0.8764	27.81/0.7624	27.11/0.7199	24.80/0.7389	28.29/0.8734	29.65/0.8189
	SESR-M7	22.92K	31.14/0.8787	27.88/0.7641	27.13/0.7209	24.90/0.7436	28.53/0.8778	29.72/0.8204
	ECLB-M3	13.98K	31.54/0.8876	28.15/0.7674	28.12/0.7247	25.24/0.7458	29.54/0.8732	29.98/0.8257
	ECLB-M5	19.76K	31.58/0.8819	28.23/0.7748	28.61/0.7278	25.09/0.7410	29.01/0.8842	30.76/0.8241
	ECLB-M7	23.02K	31.79/0.8891	28.40/0.7748	28.65/0.7332	29.17/0.7578	29.76/0.8992	30.92/0.8374
Medium	TPSR-No GAN	61K	31.10/0.8779	27.95/0.7663	27.15/0.7214	24.97/0.7356	–/–	–/–
	SESR-M11	32.14K	31.27/0.8810	27.94/0.7660	27.20/0.7225	25.00/0.7466	28.73/0.8815	29.81/0.8221
	ECLB-M11	33.78K	32.40/0.8973	28.44/0.7751	28.47/0.7373	26.13/0.7709	29.78/0.8872	30.12/0.8401
Large	VDSR	665K	31.35/0.8838	28.02/0.7678	27.29/0.7252	25.18/0.7525	28.82/0.8860	29.82/0.8240
	LapSRN	813K	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8900	29.88/0.8250
	BTSRN	410K	31.85/–	28.20/–	27.47/–	25.74/–	–/–	–/–
	CARN-M	412K	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.62/0.7694	–/–	–/–
	SESR-XL	114.97K	31.54/0.8866	28.12/0.7712	27.31/0.7277	25.31/0.7604	29.04/0.8901	29.94/0.8266
	ECLB-XL	115.73K	32.11/0.8990	28.08/0.7701	28.22/0.7379	28.15/0.7591	29.29/0.8898	30.54/0.8416

(Blue/ Red colored numbers indicates first/ second best values (PSNR/SSIM) among the methods) Input image size=1080x1920, Output image size=4320x7680

Table 3 Training time efficiency of the ECLB networks

Model	ECLB-M3	ECLB-M5	ECLB-M7	ECLB-M11	ECLB-XL
Avg. training time (in Sec) of 300 epochs	28	37	45	63	90

In the case of medium network, we compare against the most recent super resolution technique called TPSR-NoGAN [45] network and SESR-M11. It is clearly indicates that the M11-proposed network performs better in all datasets.

For the large network category, we compared against VDSR [46], LapSRN [47], BTSRN [48], CARN-M [19], MOREMNAS-B [50] and SESR-XL [42] networks. All the network's performances are closer to each other. But, the proposed network achieved better PSNR/SSIM values than others. 0.61K excess parameters are required in the proposed network compared with SESR-XL and can be negligibly small.

Similarly, Table 2 presents the performance of the proposed network at $\times 4$ resolution. In case of small complexity network, FSRCNN model utilizes fewer parameters than other methods. The proposed M3 method and SESR-M3 method required the equal number of parameters. Whereas, the outstanding performance of the ECLB-M7 method is indicated with PSNR/SSIM values. In the case of medium complex networks, the proposed model is compared with SESR-M11 and TPSR-No GAN. The proposed model performs well and 1.64K excess parameters are required compared with the SESR-M11 method. Similarly, for large complexity network category 0.76K extra parameters are needed compared with SESR-XL. Even though, the proposed network achieved better PSNR/SSIM values in the datasets.

To further analyze the performance of the proposed method we have performed a visual comparison analysis (Qualitative Analysis) as follows:

5.3 Qualitative analysis

Figures 5 and 6 shows the qualitative analysis of the proposed method at various scales $\times 2$, and $\times 4$ respectively. The comparisons are done on small, medium, and large networks, with the state-of-the-art networks of FSRCNN and SESR's three methods. We directly took the results of SESR and FSRCNN from their papers and compared with our results.

From the each dataset we have randomly selected one image to show the efficiency of our method. From Set14 dataset PPT3 and lenna images, From Set5 dataset baby image, from DIV2K dataset 0808 image, from urban100 dataset img_096 are selected for visualization purpose. It is clearly showing that our algorithm outperforms existing algorithms (FSRCNN, SESR-M5,M11 and XL). The quality of the baby and img_096 images in FSRCNN is degraded and blurred. But, ECLB-XL and ECLB-M11 tackle these images easily and produce enhanced images. The minor details (eyelid and pupil) are preserved satisfactorily. At sharper edges, the SESR methods failed, and images 210088 and PPT3 became blurred. Whereas, ECLB-M5, M11, and XL images are close to ground truth. Additionally, if LR images are corrupted with noise it is hard to recover back. However, our algorithm can not only handle such cases but also recover the noiseless details. Lenna image is an example for such cases, it can be observed that the quality of the proposed network (ECLB-XL) is more

Fig. 5 Qualitative comparison on $\times 2$ SISR

closer to the ground truth. Since, the hierarchical features are generated from the residual dense block, the network can easily recover sharper, clearer edges and more faithful to the ground truth.

This results indicating that collapsible linear blocks with additional residual dense block network and multi scale sub-pixel convolution layer outperforms existing methods.

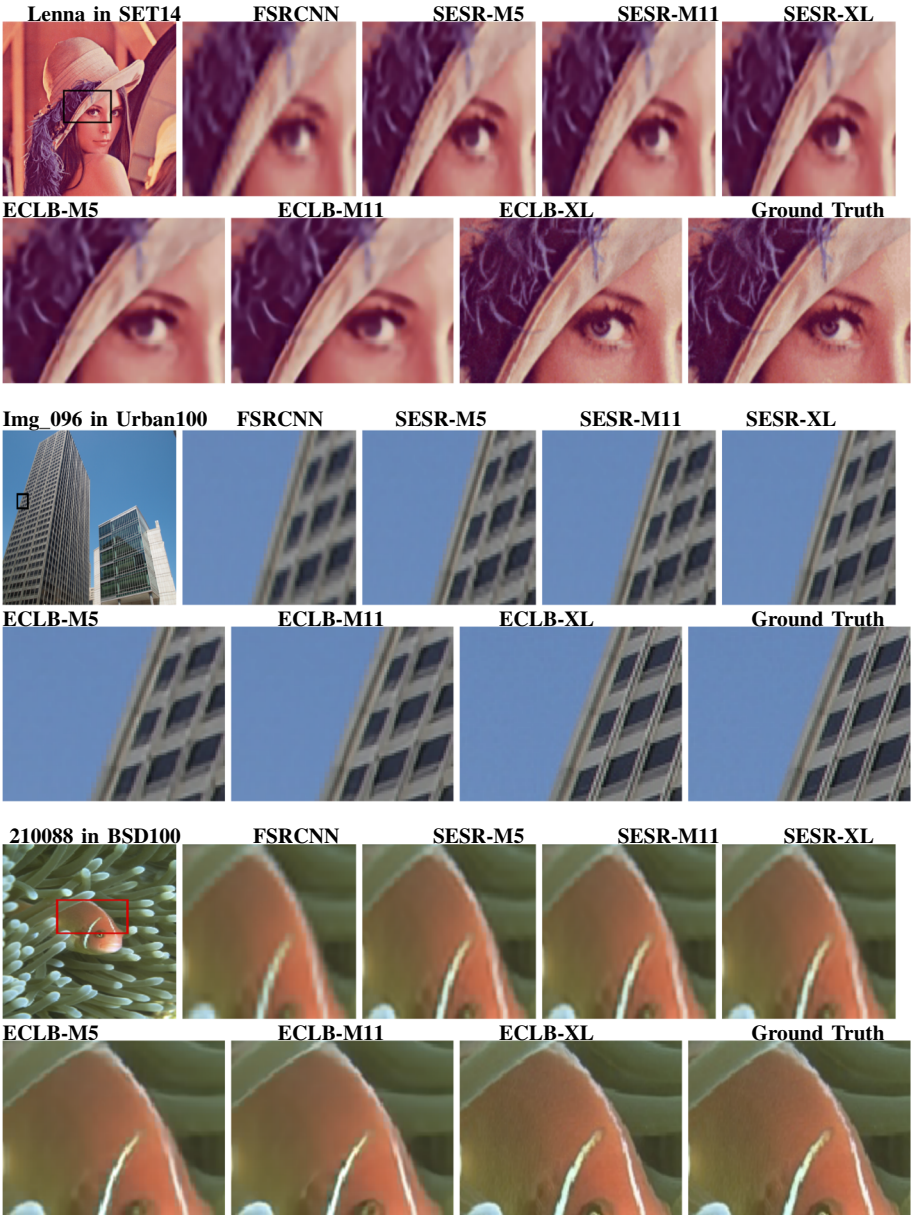

Fig. 6 Qualitative comparison on $\times 4$ SISR

Benefits of residual dense block network In general RDNs are specially designed for carrying hierarchical features from LR to HR space. The architecture of original dense layer is modified to achieve better results. We removed the pooling layers since they discard some pixel-level information and degrades the final features. Instead, we employed a multi-scale setup (combination of Conv, ReLU and input image) for producing better features, i.e., the output of the $(d-1)^{th}$ RDN layer has direct connections with d^{th} and $(d+1)^{th}$ layers

along with input image. This way the sensitive information like edges, textures and hair can preserved. The Fig. 5, baby image clearly proven that edge and sensitive information (eyelid and pupil) restored successfully.

Benefits of multi-scale sub-pixel convolution layer The existing methods used fractional stride $\frac{1}{r}$. Hence, the complexity increased. But, our method takes multiple depth maps and which contains hierarchical depth maps. The consecutive depth maps are separated with a scaling factor 2. The depth maps are combined at each stage and removes the redundant information at final stage. Resulting in, producing the high-quality image at higher resolution. Eventually, The computational complexity significantly reduced. Image 0808 in Fig. 5 and img_096 in Fig. 6 are the best examples to show the qualitative efficiency of this module; the finer details, sharp edges and depth information are recovered satisfactorily.

5.4 Training efficiency of ECLB

Table 3 presents the training cost of proposed ECLB networks. We trained our four networks with standard input image size 1080p resolution. The average of 300 epoch's training time is calculated in Seconds. This timing efficiency is verified on GeForce RTX 3070 GPU (8GB RAM). It is observed that the network ECLB-M3 have less number of parameters hence it took less training time of 28s. The parameters such as number of convolution layers and residual dense networks are utilized by this network are 64 and 32 respectively. These 32 hierarchical features are transferred from LR to HR space. The total number of multi-scale sub-pixel convolutional layers are utilized by this network are 64. The number of parameters are increasing with the technology used to generate the HR image. The less number of parameters are utilized for ECLB-M5, M7 and M11 networks and training time of 37,45 and 63 seconds required respectively. During 8K image generation the linear blocks were collapsed into sub blocks and then perform the upscaling operation, resulting in the model ECLB-XL's an average training time is 90 Sec. It is showing that the model is well trained. Hence, it produces satisfactory results during the validation.

Even though the proposed ECLB network consumed more parameters, with the advantage of collapsible linear blocks the training and validation efficiency is satisfactorily better.

6 Conclusions and future scope

We have successfully developed a quality enhanced image super resolution technique based on collapsible linear blocks with RDNs, multi-scale sub-pixel network and adaptive cropping technique. The RDN's successfully forwarded the hierarchical features to preceding blocks. All the depth-maps and features were trained with n_{L-1} up-scaling filters. In addition, adaptive cropping strategy (ACS) helped up-scaling any arbitrary sized image to a desired quality. The multi-scale sub-pixel convolutional layers successfully upscaled the given LR image into HR. The proposed algorithm is verified across six benchmark datasets quantitatively and qualitatively. The results demonstrated that proposed algorithm outperforms all the baselines on previous state-of-the-art algorithms.

We did not intend to design this algorithm for resource-constraint devices. We will be done it in the future.

Acknowledgements The research leading to these findings has received funding from the EU H2020 Marie Skłodowska-Curie COFUND Scheme (Grant No. 900025), CfACTs.

Author Contributions The corresponding author has made a substantial contribution to the concept of this article, the acquisition, analysis, and interpretation of data. The remaining authors have drafted the article or revised it critically for important intellectual content.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 900025 (CfACTs)

Data availability statement Data sharing not applicable to this article as no datasets were generated during the current research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition pp 770–778
2. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 4700–4708
3. Simonyan KZA (2015) Very deep convolutional networks for largescale image recognition, p 1409
4. Hui Z, Wang X, Gao X (2018) Fast and accurate single image super-resolution via information distillation network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 723–731
5. Tong T, Li G, Liu X, Gao Q (2017) Image super-resolution using dense skip connections. In: Proceedings of the IEEE international conference on computer vision pp 4799–4807
6. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y (2018) Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV) pp 286–301
7. Kim J, Lee JK, Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1646–1654
8. Dong C, Loy CC, He K, Tang X (2014) Learning a deep convolutional network for image super-resolution. In: Computer vision-ECCV 2014: 13th European conference, Zurich, Switzerland, Proceedings, Springer International Publishing, Part IV 13 pp 184–199. Accessed 6–12 Sept 2014
9. Dong C, Loy CC, He K, Tang X (2015) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295–307
10. Wang Z, Liu D, Yang J, Han W, Huang T (2015) Deep networks for image super-resolution with sparse prior. In: Proceedings of the IEEE international conference on computer vision pp 370–378
11. Kim J, Lee JK, Lee KM (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1637–1645
12. Tai Y, Yang J, Liu X (2017) Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 3147–3155
13. Duanmu C, Zhu J (2020) The image super-resolution algorithm based on the dense space attention network. IEEE Access 8:140599–140606
14. Tai Y, Yang J, Liu X, Xu C (2017) Memnet: A persistent memory network for image restoration. In: Proceedings of the IEEE international conference on computer vision pp 4539–4547

15. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1874–1883
16. Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z, Shi W (2017) Real-time video super-resolution with spatio-temporal networks and motion compensation. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 4778–4787
17. Lai WS, Huang JB, Ahuja N, Yang MH (2017) Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 624–632
18. Lim B, Son S, Kim H, Nah S, Mu Lee K (2017) Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops pp 136–144
19. Hui Z, Gao X, Yang Y, Wang X (2019) Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th acm international conference on multimedia pp 2024–2032
20. Lai WS, Huang JB, Ahuja N, Yang MH (2018) Fast and accurate image super-resolution with deep laplacian pyramid networks. IEEE Trans Pattern Anal Mach Intell 41(11):2599–2613
21. Wang Y, Perazzi F, McWilliams B, Sorkine-Hornung A, Sorkine-Hornung O, Schroers C (2018) A fully progressive approach to single-image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops pp 864–873
22. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 4681–4690
23. Zhang Y, Li K, Li K, Zhong B, Fu Y (2019) Residual non-local attention networks for image restoration. arXiv preprint arXiv:1903.10082
24. Zhang Y Li K, Li K, Wang L, Zhong B, Fu Y (2018) Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV) pp 286–301
25. Arora S, Cohen N, Hazan E (2018) On the optimization of deep networks: Implicit acceleration by overparameterization. In: International conference on machine learning, PMLR, pp 244–253
26. Wu F, Souza A, Zhang T, Fifty C, Yu T, Weinberger K (2019) May. Simplifying graph convolutional networks. In, International conference on machine learning, PMLR, pp 6861–6871
27. Ding X, Guo Y, Ding G, Han J (2019) Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In: Proceedings of the IEEE/CVF international conference on computer vision pp 1911–1920
28. Guo S, Alvarez JM, Salzmann M (2020) Expandnets: Linear over-parameterization to train compact convolutional networks. Adv Neural Inform Process Syst 33:1298–310
29. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y (2018) Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 2472–2481
30. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1–9
31. Osendorfer C, Soyer H, Van Der Smagt P (2014) Image super-resolution with fast approximate convolutional sparse coding. In: Neural information processing: 21st international conference, ICONIP 2014, Kuching, Malaysia, Proceedings, Springer International Publishing, Part III 21 pp 250–257. Accessed 3–6 Nov 2014
32. Wang Z, Liu D, Yang J, Han W, Huang T (2015) Deep networks for image super-resolution with sparse prior. In: Proceedings of the IEEE international conference on computer vision pp 370–378
33. Schulter S, Leistner C, Bischof H (2015) Fast and accurate image upscaling with super-resolution forests. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 3791–3799
34. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1874–1883
35. Hui Z, Gao X, Yang Y, Wang X (2019) Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th acm international conference on multimedia pp 2024–2032
36. Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel ML (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding

37. Zeyde R, Elad M, Protter M (2012) On single image scale-up using sparse-representations. In: Curves and surfaces: 7th international conference, Avignon, France, Revised Selected Papers 7, Springer Berlin Heidelberg, pp 711–730. Accessed 24–30 June 2010
38. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings eighth IEEE international conference on computer vision. ICCV 2001, IEEE, vol 2, pp 416–423
39. Huang JB, Singh A, Ahuja N (2015) Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 5197–5206
40. Matsui Y, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T, Aizawa K (2017) Sketch-based manga retrieval using manga109 dataset. Multimed Tool Appl 76:21811–21838
41. Agustsson E, Timofte R (2017) Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops pp 126–135
42. Bhardwaj K, Milosavljevic M, O'Neil L, Gope D, Matas R, Chalfin A, Suda N, Meng L, Loh D (2022) Collapsible linear blocks for super-efficient super resolution. Proceedings of machine learning and systems 4:529–547
43. Dong C, Loy CC, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Springer International Publishing, Part II 14 pp 391–407. Accessed 11–14 Oct 2016
44. Chu X, Zhang B, Xu R (2021) Multi-objective reinforced evolution in mobile neural architecture search. In: Computer Vision-ECCV 2020 Workshops: Glasgow, UK, August 23-28, 2020, Proceedings, Cham: Springer International Publishing, Part IV pp 99–113
45. Lee R, Dudziak Ł, Abdelfattah M, Venieris SI, Kim H, Wen H, Lane ND (2020) Journey towards tiny perceptual super-resolution. In: Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, Proceedings, Cham: Springer International Publishing, Part XXVI pp 85–102. Accessed 23–28 Aug 2020
46. Kim J, Lee JK, Lee KM (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 1637–1645
47. Lai WS, Huang JB, Ahuja N, Yang MH (2017) Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 624–632
48. Fan Y, Shi H, Yu J, Liu D, Han W, Yu H, Wang Z, Wang X, Huang TS (2017) Balanced two-stage residual networks for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops pp 161–168
49. Ahn N, Kang B, Sohn KA (2018) Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European conference on computer vision (ECCV) pp 252–268
50. Chu X, Zhang B, Xu R (2021) Multi-objective reinforced evolution in mobile neural architecture search. In: Computer vision-ECCV 2020 workshops: glasgow, UK, Proceedings, Cham: Springer International Publishing, Part IV pp 99–113. Accessed 23–28 Aug 2020