Article

Deep Residual Dense Network for Single Image Super-Resolution

Yogendra Rao Musunuri ¹ and Oh-Seol Kwon ^{2,*}

¹ Department of Control & Instrumentation Engineering, Changwon National University, Changwon 51140, Korea; musunuri3@gmail.com

² School of Electrical, Electronics & Control Instrumentation Eng., Changwon National University, Changwon 51140, Korea

* Correspondence: osk1@changwon.ac.kr; Tel.: +82-55-213-3669

Abstract: In this paper, we propose a deep residual dense network (DRDN) for single image super-resolution. Based on human perceptual characteristics, the residual in residual dense block strategy (RRDB) is exploited to implement various depths in network architectures. The proposed model exhibits a simple sequential structure comprising residual and dense blocks with skip connections. It improves the stability and computational complexity of the network, as well as the perceptual quality. We adopt a perceptual metric to learn and assess the quality of the reconstructed images. The proposed model is trained with the Diverse2k dataset, and the performance is evaluated using standard datasets. The experimental results confirm that the proposed model exhibits superior performance, with better reconstruction results and perceptual quality than conventional methods.

Keywords: super-resolution; perceptual; DRDN; residual; dense blocks

Citation: Musunuri, Y.R.; Kwon, O.-S. Deep Residual Dense Network for Single Image Super-Resolution. *Electronics* **2021**, *10*, 555. <https://doi.org/10.3390/electronics10050555>

Academic Editor: Jihoon Yang

Received: 6 January 2021

Accepted: 22 February 2021

Published: 26 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Single image super-resolution (SISR) is used to reconstruct a high-resolution image (HR) from a single low-resolution (LR) input image with better visual quality [1,2]. For instance, in the sharing and research of multimedia content, access to the original data is non-existent, and the quality of the received image cannot be estimated. If the image quality is poor, then it becomes difficult to restore the information. Therefore, it is important to restore the original image in super-resolution. This is currently being implemented in many applications, such as closed-circuit television surveillance [3] and security systems [4], satellite remote sensing [5], medical imaging [6,7] atmospheric monitoring [8], and robotics [9].

Super-resolution methods can be broadly classified into two main categories: conventional [10] and deep learning methods [11]. Conventional approaches in computer vision for super-resolution are interpolation-based methods such as bicubic, nearest-neighbor [12], and bilinear interpolation [13–15]. Deep learning methods yield better performances than conventional methods. We categorized the network architecture as based on Saeed et al. [11]. In this reference, they are categorized as single image super resolution models based on their structures. Furthermore, the deep learning methods of single image super-resolution can be categorized into two types: peak signal-to-noise ratio (PSNR) and perception-oriented methods. In PSNR-oriented methods, deep neural networks [16–18] provide significantly improved performance measures in terms of the PSNR and structural similarity index (SSIM) [19]. In SISR problems, deep learning models are implemented using basic linear networks [20], which are simple structures involving one path and the signal flows sequentially from the start layer to the end layer. By contrast, residual networks [21] use dense and skip connections [22] as well as multiple branches for residual learning. Residual learning with deep networks yields better performance. We will briefly discuss recent deep residual-based [23–25] methods used to improve visual quality that are relevant to our current study. One of them includes the enhanced deep residual network (EDSR) [26]. This network was derived from a super-resolution residual network

(SRResNet) [27]. In an EDSR, the network is designed to improve the high-frequency details that can improve the visual quality. Other residual methods, such as the very deep super-resolution network (VDSR) [28,29] and cascading residual network (CARN) [30], enhance the performance using different types of architectures and parameters.

Additionally, deep learning methods involving single image super-resolution include perceptual-oriented methods. This study focuses on enhancing perception-based [31,32] image quality. Methods related to perceptual image quality include generative adversarial networks (GANs) [33], where the deep spatial feature transforms GANs [34] and focuses on recovering realistic textures using spatial feature transformation in GAN architectures. Enhanced super-resolution GANs [35] are designed to enhance visual quality in a realistic and natural manner. In general, GAN models pose generalization problems, such as training unstable and complex structures that consume excessive memory. To mitigate issues of training instability, computational complexity, and memory usage, it replaces the residual in residual dense block (RRDB) instead of that in the residual block. The proposed model is a simple sequential structure for improving the perceptual quality of reconstructed images using various depth models.

We propose a deep residual dense network (DRDN) for super-resolution in this paper. The proposed model consists of the residual network and parts of the RRDB based on the EDSR training parameters. Therefore, the proposed network is designed with local and global residual learning which includes a deep residual layers along with the dense and skip connections. Finally, it is implemented with a new structure as a DRDN, which is designed for perceptual quality, fewer parameters, and low computational cost. In the proposed network, the experimental results show that the model yielded better reconstruction results and perceptual quality when evaluated using perceptual metrics. The remainder of this paper is organized as follows. Section 2 provides a brief background of studies pertaining to single image super-resolution. Section 3 provides the details of the proposed approach for a perceptual visual quality network. In Section 4, the experimental process and the results of the proposed method on various datasets are presented. Additionally, comparisons of the proposed method with reference methods in terms of perceptual quality, universal image quality index (UQI), PSNR, and SSIM are reported. Section 5 presents the conclusions.

2. Related Works

The era of deep learning began with cybernetics [36] based on biological learning. Subsequent research advances were realized using the perceptron network [37] for image recognition, and artificial neural networks [38] were developed to achieve intelligent behavior. Finally, researchers performed deeper training in networks, and named it as deep learning [39]. Deep learning is a leading technology in the field of single image super-resolution for reconstructing high-resolution images. The evolution of single image super-resolution for deep learning can be categorized based on the network architecture, such as linear networks, residual networks, recursive networks, progressive reconstruction designs, densely connected networks, multi-branch designs, attention-based networks, and GAN models.

CARN: The cascading residual network (CARN) is a residual network that learns locally and globally with cascading connections. Strategy learning from global cascading to local cascading is efficient for high-frequency details.

VDSR: This is a deep neural network for residual-based learning that solves the problems of super-resolution in convolution neural networks [40]. In addition, it provides solutions for context, convergence, and scale factor issues. This network generates residual maps and combines them with a HR image to form a super-resolution image.

SRResNet: This is one of the residual networks that exhibits better performance than conventional computer vision problems in low-level resolution. It is implemented using basic residual networks.

EDSR: The EDSR is a recent state-of-the-art super-resolution method that is a modified network based on SRResNet, whose batch normalization (BN) process consumes excessive

memory while training the model using the graphical processing unit. In the EDSR network, BN layers are removed, thereby improving image reconstruction.

AGTM: The aggregated residual transformation method (AGTM) is a modified version of the EDSR for reducing the number of parameters and time complexity. The aggregated residual transformation method [41–43] achieves the same level of performance as the EDSR.

PM-DAN: The perceptual-metric-guided deep attention network (PM-DAN) [44] is an attention-based decoder–encoder network that focuses on the visual quality of newly constructed images. In this case, the residual spatial attention unit captures key information.

ESRGAN: This network is a family of GANs that comprises two main building components: a discriminator and a generator. ESRGAN is a modified version that improves the visual quality of the reconstructed image. In this study, the *RRDB* was introduced instead of the residual block with a perceptual loss function.

3. Proposed Methods

The EDSR performed better results in single image super-resolution, in which the perception quality improved when it is compared with other models. To improve the perceptual quality, the modified residual block was replaced with the *RRDB* in the EDSR structure. The *RRDB* was designed for the discriminator of the GAN model. Finally, the *RRDB* was built in the residual network of the proposed model. A comparison between the residual block of the EDSR and the proposed model is shown in Figure 1. The original EDSR comprised two convolutional layers and a rectified linear unit (ReLU), as shown in Figure 1a. The convolution layer was used for feature extraction, whereas the ReLU was used to activate the network. The main difference between the EDSR and the proposed model is evident in the *RRDB* layers. Depending on the depth, we implemented two models, *RRDB_20* and *RRDB_28*.

Figure 1. Comparison of residual blocks in the enhanced deep residual network (EDSR) and our proposed model. (a) EDSR residual block, (b) Proposed model with residual block (residual in residual dense block strategy).

If the model has a depth of 20, then the number of *RRDB* layers in the model is 20. If the model has a depth of 28, then the number of *RRDB* layers in the model is 28. The proposed model comprised three residual dense blocks (*RDBs*) in the *RRDB*, as shown in Figure 1b. In addition, it combined multiple *RDBs* that were concatenated with each other. The *RDB* comprised a dense block with its own convolutional layers, dense connections that are stacked together (as shown in Figure 2), and a leaky rectified linear unit (LReLU) to rectify the network. After modifying the residual block, the appearance of the proposed model is as illustrated in Figure 3.

Figure 2. Inner layers and dense connections of *RDB*.

Figure 3. Structure of proposed model.

The process of this model begins from the low-resolution image as the input x_{lr} and y_{sr} as the output of the proposed model. The first convolutional layer extracts the features F_0 from the low-resolution input, as shown in (1) [45,46].

$$F_0 = H_{fe}(x) \quad (1)$$

where $H_{fe}(\cdot)$ denotes the convolution operation for the first convolutional layer, and F_0 is the feature extraction used as input to the *RRDB*. Suppose we have rd residual in residual dense blocks, and the output F_{RD} the rd th *RRDB* can be obtained by (2).

$$F_{RD} = H_{RRDB,rd}(F_{rrdb-1}) \quad (2)$$

where $H_{RRDB,rd}$ denotes the operations of the rd th *RRDB*. The inner layers of the *RRDB* comprised three *RDBs*, and the features of the *RRDB* can be calculated using (3).

$$F_{RRDB} = H_{RDB,rd}((H_{RDB,rd-1}(F_{rd-1}))(\dots)(H_{RDB,rd2}(F_{rd2}))) + F_{rrl} \quad (3)$$

where $H_{RDB,rd}$ represents the operations of the rd th *RDB*. As $H_{RDB,rd}$ is a combined operation of the convolution and LReLU. As F_{RRDB} is obtained using all *RDBs* and residual

in residual learning (F_{rri}). The output of the d th RDB can be obtained using the input of F_{d-1} as shown in (4).

$$F_{rd} = H_{DB,d}(F_{d-1}) \quad (4)$$

where, F_{rd} represents the local features of the RDB. As F_{rd} is obtained using all convolutional layers and the LReLU, the inner layers of the dense block are formulated using (5).

$$F_{d,c} = \sigma(W_{d,c}[F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]) \quad (5)$$

where σ denotes the LReLU activation function. $W_{d,c}$ the weights of the c -th convolutional layer, and $F_{d-1}, F_{d,1}, \dots, F_{d,c-1}$ the concatenation of the feature-maps yielded by the $(d-1)$ th RDB, convolutional layers $1, \dots, (c-1)$ in the d th RDB. Furthermore, the residual in residual dense fusion (RRDF) is shown in (6).

$$F_{RRDF} = F_0 + F_{GRL} \quad (6)$$

where F_{GRL} is the global residual learning, and F_0 is the feature extraction of the first convolution layer. Final output of the DRDN can be obtained by the (7).

$$y_{sr} = H_{DRDN}(x_{lr}) \quad (7)$$

The residual blocks of the EDSR and the proposed model are shown in Figure 4. Figure 4a shows the inner convolutional layers (as shown in Figure 1a) of the EDSR residual block as a building block. In the EDSR, the residual block consists of two convolutional layers, which are the same size as the 256-d input and output with a 3×3 kernel. Figure 4b represents the inner convolutional layers (as shown in Figure 2) of the proposed RRDB as a building block. Figure 4b shows the image transformation steps on one block. In the proposed model, the residual block (RRDB) has five convolutional layers to extract the features that begin with the 64-d input to the first convolutional layer, and its output is 32-d with the 3×3 kernel.

Figure 4. Equivalent building blocks, (a) EDSR residual block, (b) proposed model residual block (RRDB).

Likewise, the second to fifth convolutional layers are followed by different input and output image sizes. Finally, reconstruction of the super-resolution image was performed. In addition, the proposed models were implemented at various depths, such as RRDB_20 and RRDB_28. These two models were distinguished by their depths, and the number of parameters involved is shown in the Table 1. Furthermore, Table 1 provides the model type, number of residual blocks, total number of parameters, residual scaling, and loss function. Based on to the performances of the three models, RRDB_28 and RRDB_20

exhibited performances comparable to the *EDSR* in terms of the *PSNR* and *SSIM*. In terms of perceptual quality, the *RRDB_20* model outperformed the *EDSR*.

Table 1. Parameters of *EDSR* and proposed models.

Model		No. of Residual Blocks	Total no. of Parameters	Residual Scaling	Loss Function
<i>EDSR</i>		32	~45,450 K	–	L1
<i>DRDN</i>	<i>RRDB_28</i>	28	~20,627 K	0.2	L1
(ours)	<i>RRDB_20</i>	20	~14,872 K	0.2	L1

4. Experimental Results

In this section, we present the experimental procedures and results of the proposed model. A comparison with a state-of-the-art method is presented as well. For the experimental evaluation, the Diverse2k resolution (Div2K) training dataset [47] was used, and a quantitative evaluation was performed using a public benchmark dataset. Finally, perceptual quality was evaluated using the perception-based image quality evaluator (PIQE).

4.1. Training Datasets

In the training, the Div2K dataset [47] was used, which is a 2k-resolution high-image-quality dataset comprising 800 training images, 100 validation images, and 100 test images. In addition, low-resolution (LR) bicubic images were available in $\times 2$, $\times 3$, $\times 4$, $\times 8$ factors for training and evaluating the proposed model.

Evaluation on Benchmark Datasets

To evaluate the performance of the model in terms of quantitative measures, we compared our model with publicly available benchmark datasets, such as set 5 [48], set 14 [13], BSD100 [49], and Urban100 [50]. The main purpose of these datasets are testing and predicting the performance of the newly designed network architectures. It is also easy to compare with the existing conventional models.

Set 5 [48]: This is a standard dataset with five test images of a baby, bird, butterfly, head, and woman.

Set 14 [13]: More categories were compared with set 5. However, the number of images used was 14, which included a baboon, Barbara, bridge, coastguard, comic, face, flowers, foreman, lenna, man, monarch, pepper, ppt3, and zebra.

BSD100 [49]: It is one of the classical datasets, consisting of 100 test images. The dataset consists of images ranging from nature to individual objects such as plants, people, food, animals, and devices, etc.

Urban100 [50]: It is also a classical dataset composed of the same 100 images as BSD100. The dataset focuses on artificial structures which are made by humans, such as urban scenes.

4.2. PIQE

The *PIQE* [51] is a metric used for human perceptual quality assessment in the field of super-resolution image reconstruction. The mathematical expression for the *PIQE* is as follows:

$$PIQE = \frac{\left(\sum_{k=1}^{N_{SA}} D_{sk} \right) + C_1}{(N_{SA} + C_1)} \quad (8)$$

Here, N_{SA} represents the number of spatially active blocks in an image, C_1 a positive constant, and D_{SK} the distorted block. The *PIQE* metric returns a positive scalar in the range of 0–100. The *PIQE* score is the individual image quality score, which is inversely correlated with the perceived quality of an image. A low score indicates good perceptual

quality, whereas a high score indicates otherwise. The *PIQE* metric was evaluated based on benchmark datasets, such as set 5 and set 14, with scale factors $\times 2$, $\times 3$, $\times 4$, and $\times 8$ presented in Table 2. As shown, $\times 2$ of the *RRDB_20*, set 5 and set 14 *PIQE* exhibited a lower value compared with that of the EDSR and *RRDB_28*, which were 56.3124 and 48.1648, respectively. Hence, *RRDB_20* demonstrated better perceptual quality. Similarly, other scale models of set 5 and set 14, such as $\times 3$, $\times 4$, and $\times 8$, yielded values of 66.4755, 75.7457, 76.2772 and 71.9378, 77.4722, 79.5802 respectively. In addition, *RRDB_28* of the set 14 ($\times 2$), BSD100 ($\times 3$), and urban100 ($\times 2$, $\times 3$, $\times 4$, $\times 8$) are higher *PIQE* values than conventional methods. It is evident that the proposed models exhibited better perceptual quality than the EDSR.

Table 2. Perceptual evaluation based on benchmark datasets (perception-based image quality evaluator (*PIQE*)).

Dataset	Scale	VDSR	EDSR	DRDN (ours)	
				<i>RRDB_28</i>	<i>RRDB_20</i>
Set 5	$\times 2$	86.8745	56.3733	56.7676	56.3124
	$\times 3$	84.7418	74.1060	73.2497	66.4755
	$\times 4$	80.2138	82.0745	79.9208	75.7457
	$\times 8$	95.8925	82.2949	80.9344	76.2772
Set 14	$\times 2$	86.6052	48.5839	48.1912	48.1648
	$\times 3$	84.7468	75.5955	67.7200	71.9378
	$\times 4$	80.7410	79.5800	79.7969	77.4722
	$\times 8$	96.4151	83.9043	80.3520	79.8502
BSD100	$\times 2$	41.9925	40.5223	37.8143	37.7747
	$\times 3$	65.0338	64.3396	63.6483	63.2722
	$\times 4$	81.0542	80.1188	74.1877	78.6450
	$\times 8$	89.3395	87.5180	85.6520	84.6193
Urban100	$\times 2$	53.3344	50.4255	50.2638	50.4026
	$\times 3$	67.8616	65.6758	63.8491	65.2255
	$\times 4$	77.0920	74.5159	67.6317	68.6164
	$\times 8$	83.4780	78.1662	72.8245	73.1051

4.2.1. UQI

The image quality measures for the human visual system (HVS) were calculated using the universal image quality index (*UQI*) [52,53] metrics. In particular, *UQI* models the correlation loss, luminance distortion, and contrast distortion. The *UQI* metric is expressed as in (9).

$$UQI = \frac{1}{M} \sum_{j=1}^M Q_j \quad (9)$$

where *UQI* is universal image quality index, Q_j is the local quality index, and M is the total number of steps. The *UQI* metric was used to evaluate the image quality for the HVS based on benchmark datasets such as set 5, set 14, and BSD100, and urban100 with scale factors of $\times 2$, $\times 3$, $\times 4$, and $\times 8$ are presented in Table 3. As shown, $\times 2$ of the *RRDB_20* set 5, *UQI* exhibited a higher value compared with that of VDSR, *RRDB_28*, and EDSR was 0.9951. Additionally, set 14 was 0.9920. Hence, *RRDB_20* shows improved HVS image quality. Similarly, other scale models of set 5, such as $\times 3$, $\times 4$, and $\times 8$, yielded values of 0.9929, 0.9870, 0.9666, and set 14, yielded 0.9889, 0.9820, and 0.9673, respectively. In addition, *RRDB_28* of urban100 ($\times 3$) is higher value than the conventional methods in the Table 3. Hence, the proposed model shows improved image quality when compared to other models.

Table 3. Universal image quality index (UQI) evaluation based on benchmark datasets.

Dataset	Scale	VDSR	EDSR	DRDN (ours)	
				RRDB_28	RRDB_20
Set 5	$\times 2$	0.9936	0.9949	0.9950	0.9951
	$\times 3$	0.9814	0.9844	0.9928	0.9929
	$\times 4$	0.9819	0.9868	0.9869	0.9870
	$\times 8$	0.9262	0.9640	0.9632	0.9666
Set 14	$\times 2$	0.9886	0.9906	0.9919	0.9920
	$\times 3$	0.9754	0.9792	0.9888	0.9889
	$\times 4$	0.9736	0.9753	0.9819	0.9820
	$\times 8$	0.9432	0.9667	0.9662	0.9673
BSD100	$\times 2$	0.9840	0.9942	0.9944	0.9947
	$\times 3$	0.9882	0.9919	0.9923	0.9925
	$\times 4$	0.9794	0.9807	0.9842	0.9845
	$\times 8$	0.9714	0.9722	0.9717	0.9731
Urban100	$\times 2$	0.9779	0.9897	0.9898	0.9902
	$\times 3$	0.9656	0.9825	0.9833	0.9834
	$\times 4$	0.9687	0.9736	0.9744	0.9743
	$\times 8$	0.9467	0.9469	0.9475	0.9479

4.2.2. Training Details

For training, we followed the training parameters given in Bee et al. [26]. We also used the RGB input patches of size 48×48 from the LR image using the corresponding HR patches. We trained our model using the Adam optimizer by setting $\beta_1 = 0.9$ $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. We set the mini-batch to 16. The learning rate was initialized as 10^{-4} , and the learning decay varying at boundaries (70,000, 100,000) were 5×10^{-5} and 1×10^{-5} respectively. We implemented the proposed model using the tensor flow framework and trained it using NVIDIA GeForce GTX GPUs device. The proposed model architectures required 10 days to train.

4.3. PSNR (dB)/SSIM Evaluation

The mean PSNR and SSIM of the super-resolution methods evaluated on benchmark datasets such as set 5, set 14, BSD100, and urban100 for super-resolution factors $\times 2$, $\times 3$, $\times 4$, and $\times 8$ of the proposed models were compared with those of the simulated EDSR, as shown in Table 4. The simulated performance of the EDSR indicated better PSNR and SSIM scores when compared with those of the reference EDSR [26]. By comparison, the PSNR and SSIM scores were measured on the y-channel. The validation of the PSNR and SSIM scores of the EDSR and proposed models on the Div2K dataset are shown in Figure 5. In addition, visual comparisons of the super-resolution images from set 5 ($\times 2$ and $\times 4$) and set 14 ($\times 3$ and $\times 8$) are shown in Figures 6 and 7, respectively. As shown by the results in Table 3, the proposed model performed comparably with the EDSR in terms of reduced parameters and better perceptual quality during the reconstruction of super-resolution images. In addition, RRDB_20 yielded better average perceptual quality than RRDB_28, and its performance varied with individual images, as shown in Figures 6 and 7.

Table 4. Public benchmark test results (peak signal-to-noise ratio (PSNR) (dB)/structural similarity index (SSIM)).

Dataset	Scale	VDSR [54]	EDSR	DRDN (ours)	
				RRDB_28	RRDB_20
Set 5	$\times 2$	37.53/0.9559	38.16/0.9550	38.03/0.9546	38.08/0.9548
	$\times 3$	33.67/0.9210	35.29/0.9332	35.19/0.9320	35.22/0.9326
	$\times 4$	31.35/0.8830	32.31/0.8829	32.35/0.8835	32.30/0.8828
	$\times 8$	25.93/0.7240	26.94/0.7461	26.80/0.7388	26.84/0.7402
Set 14	$\times 2$	33.05/0.9130	33.93/0.9122	33.69/0.9100	33.74/0.9101
	$\times 3$	29.78/0.8320	31.13/0.8487	31.09/0.8480	31.11/0.8483
	$\times 4$	28.02/0.7680	28.80/0.7693	28.83/0.7704	28.79/0.7690
	$\times 8$	24.26/0.6140	25.16/0.6200	25.07/0.6163	25.05/0.6164
BSD100	$\times 2$	31.90/0.8960	33.85/0.9196	33.76/0.9188	33.77/0.9190
	$\times 3$	28.83/0.7976	29.74/0.8075	29.72/0.8075	29.73/0.8074
	$\times 4$	27.29/0.7252	28.60/0.7480	28.59/0.7482	28.60/0.7483
	$\times 8$	24.49/0.5830	25.44/0.5916	25.38/0.5867	25.40/0.5903
Urban100	$\times 2$	30.77/0.9141	32.59/0.9263	32.19/0.9225	32.29/0.9237
	$\times 3$	27.14/0.8279	29.73/0.8719	29.67/0.8716	29.61/0.8702
	$\times 4$	25.18/0.7525	26.40/0.7805	26.34/0.7796	26.33/0.7790
	$\times 8$	21.70/0.5710	22.63/0.5993	22.52/0.5926	22.51/0.5928

(a)

(b)

(c)

(d)

Figure 5. (a) Validation of PSNR of EDSR $\times 2$ model and proposed $\times 2$ models (RRDB_20 and RRDB_28), (b) validation of PSNR of EDSR $\times 3$ model and proposed $\times 3$ models (RRDB_20 and RRDB_28), (c) validation of PSNR of EDSR $\times 4$ model and proposed $\times 4$ models (RRDB_20 and RRDB_28), and (d) validation of PSNR of EDSR $\times 8$ model and proposed $\times 8$ models (RRDB_20 and RRDB_28).

Ground truth

PIQE Score

VDSR

88.223316

EDSR

55.650376

RRDB20

49.474508

RRDB28

48.915827

Ground truth

PIQE Score

VDSR

49.177059

EDSR

42.126118

RRDB20

40.585472

RRDB28

39.573356

Figure 6. Perceptual evaluation of proposed models on set 5 ($\times 2$) and set 14 ($\times 3$) datasets.

Figure 7. Perceptual evaluation of proposed models on set 5 ($\times 4$) and set 14 ($\times 8$) datasets.

5. Conclusions

In this paper, we proposed a *DRDN* for single image super-resolution. The proposed models were designed based on *RRDBs* and it required fewer parameters and demonstrated better stability and perceptual quality. The performance of the proposed model is limited with the increased number of layers, and to train comfortably, the network needs to be optimized. Furthermore, we need to investigate the cascaded multiscale deep network to improve the visual textures with a lower computational cost. The experimental results confirmed that the proposed model achieved better reconstruction results and perceptual quality than the conventional methods.

Author Contributions: Conceptualization, Y.R.M. and O.-S.K.; methodology, Y.R.M. and O.-S.K.; investigation, Y.R.M. and O.-S.K.; writing—original draft preparation, Y.R.M.; writing—review and editing, Y.R.M. and O.-S.K.; supervision, O.-S.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2019R1F11058489).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wujie, Z.; Jingsheng, L.; Qiuping, J.; Lu, Y.; Ting, L. Blind Binocular Visual Quality Predictor Using Deep Fusion Network. *IEEE Trans. Comput. Imaging* **2020**, *6*, 883–893.
- Sheikh, H.R.; Bovik, A.C.; de Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [\[CrossRef\]](#)
- Turchini, F.; Seidenari, L.; Uricchio, T.; Del Bimbo, A. Deep Learning Based Surveillance System for Open Critical Areas. *Inventions* **2018**, *3*, 69. [\[CrossRef\]](#)
- Debapriya, H.; Yung-Cheol, B. Upsampling Real-Time, Low-Resolution CCTV Videos Using Generative Adversarial Networks. *Electronics* **2020**, *9*, 1312. [\[CrossRef\]](#)
- Das, M.; Ghosh, S.K. Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1984–1988. [\[CrossRef\]](#)
- Greenspan, H. Super-Resolution in Medical Imaging. *Comput. J.* **2009**, *52*, 43–63. [\[CrossRef\]](#)
- Venkateswararao, C.; Tiantong, G.; Steven, J.S.; Vishal, M. Deep MR Brain Image Super-Resolution Using Spatio Structural Priors. *IEEE Trans. Image Process.* **2020**, *29*, 1368–1383.
- Dudczyk, J. A method of feature selection in the aspect of specific identification of radar signals. *Bull. Pol. Acad. Sci.-Tech.* **2017**, *65*, 113–119. [\[CrossRef\]](#)
- Harry A, P. Deep Learning in Robotics: A Review of Recent Research. *Adv. Robot.* **2017**, *31*, 821–835.
- Yunfeng, Z.; Qinglan, F.; Fangxun, B.; Yifang, L.; Caiming, Z. Single-Image Super-Resolution Based on Rational Fractal Interpolation. *IEEE Trans. Image Process.* **2018**, *27*, 3782–3797.
- Saeed, A.; Salman, K.; Nick, B. A Deep Journey into Super-resolution: A Survey. *ACM Comput. Surv.* **2020**, *53*, 1–21.
- Gao, X.; Zhang, K.; Tao, D. Image super-resolution with sparse neighbor embedding. *IEEE Trans. Image Process.* **2012**, *21*, 3194–3205.
- Timofte, R.; De Smet, V.; Van Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, 1–2 November 2014.
- Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T. Coupled dictionary training for image super-resolution. *IEEE Trans. Image Process.* **2012**, *21*, 3467–3478. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, J.; Wright, J.; Thomas, S.H.; Ma, Y. Image Super-Resolution via Sparse Representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [\[CrossRef\]](#)
- Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [\[CrossRef\]](#)
- Viet, K.H.; Ren, J.; Xu, X.; Zhao, S.; Xie, G.; Masero, V.; Hussain, A. Deep Learning Based Single Image Super-resolution: A Survey. *Int. J. Autom. Comput.* **2019**, *16*, 413–426.
- Gao, X.; Zhang, L.; Mou, X. Single Image Super-Resolution Using Dual-Branch Convolutional Neural Network. *IEEE Access* **2018**, *7*, 15767–15778. [\[CrossRef\]](#)
- Wang, Z.; Bovik, A.C.; Sheikh, H.R. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
- Dong, C.; Loy, C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014.
- Kangfu, M.; Aiwen, J.; Juncheng, L.; Bo, L.; Jihua, Y.; Mingwen, W. Deep residual refining based pseudo-multi-frame network for effective single image super-resolution. *IET Image Process.* **2019**, *13*, 591–599.
- Shamsolmoali, P.; Zhang, J.; Yang, J. Image super resolution by dilated dense progressive network. *Image Vision Comput.* **2019**, *88*, 9–18. [\[CrossRef\]](#)
- Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [\[CrossRef\]](#)
- Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
- Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

26. Bee, L.; Sanghyun, S.; Heewon, K.; Seungjun, N.; kyoung Mu, L. Enhanced Deep Residual Networks for Single Image super-Resolution. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the CVPR 2015, Boston, MA, USA, 7–12 June 2015.
29. Jiwon, K.; Jung, K.L.; Kyoung, M.L. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
30. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and, light weight super-resolution with cascading residual network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
31. Luo, X.; Chen, R.; Xie, Y.; Qu, Y.; Li, C. Bi-GANs-ST for perceptual image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016.
33. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.
34. Xintao, W.; Ke, Y.; Chao, D.; Chen, C. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In Proceedings of the CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018.
35. Xintao, W.; Ke, Y.; Shixiang, W.; Jinjin, G.; Yihao, L.; Chao, D.; Chen, C.L.; Yu, Q.; Xiaoou, T. ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018.
36. Wiener, N.; Schade, J.P. *Cybernetics of the Nervous System*; Elsevier: Amsterdam, The Netherlands, 2008; Volume 17, pp. 1–423.
37. Martin, T.H.; Howard, B.D.; Mark, H.B.; Orlando, D.J. *Neural Network Design*, 2nd ed.; eBook; MTH Publications: Oklahoma, OK, USA, 2014; pp. 1–120.
38. Joao Luis, G.R. Artificial Neural Networks—Models and Applications. *IN-TECH* **2016**, *1*, 1–412.
39. Saeed, A.; Salman, K.; Nick, B. A Deep Journey into Super-resolution: A Survey. In *Computer Vision and Pattern Recognition (cs.CV)*; DBLP-CS Publications: Ithaca, NY, USA, 2020; pp. 1–21.
40. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]
41. Yan, L.; Guangrui, Z.; Hai, W.; Wei, Z.; Min, Z.; Hongbo, Q. An efficient super-resolution network based on aggregated residual transformations. *Electronics* **2019**, *8*, 339. [[CrossRef](#)]
42. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.
43. Wazir, M.; Supavadee, A. Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach. *Electronics* **2019**, *8*, 892. [[CrossRef](#)]
44. Yubao, S.; Yuyang, S.; Ying, Y.; Wangping, Z. Perceptual Metric Guided Deep Attention Network for Single Image Super-Resolution. *Electronics* **2020**, *9*, 1145. [[CrossRef](#)]
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016.
46. Yulun, Z.; Yapeng, T.; Yu, K.; Bineng, Z.; Yun, F. Residual Dense Network for Image Super-Resolution. In Proceedings of the CVPR 2018, Salt Lake City, Utah, USA, 18–22 June 2018.
47. Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Ntire 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.
48. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the 23rd British Machine Vision Conference Location (BMVC), Guildford, UK, 3–7 September 2012.
49. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the 8th international Conference of Computer Vision (ICCV), Vancouver, Canada, 7–14 July 2001.
50. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015.
51. Venkatanath, N.; Praneeth, D.; Chandrasekhar, B.M.; Channappayya, S.S.; Medasani, S.S. Blind Image Quality Evaluation Using Perception Based Features. In Proceedings of the 21st National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015.
52. Zhou, W.; Alan Conrad, B.; Eero P, S. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *4*, 600–612.

-
53. Zhou, W.; Alan C, B. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *3*, 81–84. [[CrossRef](#)]
 54. Wei-Sheng, L.; Jia-bin, H.; Narendra, A.; Ming-Hsusan, Y. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2599–2613.