Super Resolution with Kernel Estimation and Dual Attention Mechanism

Huan Liang, Youdong Ding *, Fei Wang, Yuzhen Gao and Xiaofeng Qiu

Shanghai Engineering Research Center of Motion Picture Special Effects, Shanghai University, Shanghai 200072, China; lianghuan@shu.edu.cn (H.L.); s852989587@shu.edu.cn (F.W.); gyz18723239@shu.edu.cn (Y.G.); xiaofengQ@shu.edu.cn (X.Q.)

* Correspondence: ydding@shu.edu.cn

Received: 13 September 2020; Accepted: 24 October 2020; Published: 29 October 2020

Abstract: Convolutional Neural Networks (CNN) have led to promising performance in super-resolution (SR). Most SR methods are trained and evaluated on predefined blur kernel datasets (e.g., bicubic). However, the blur kernel of real-world LR image is much more complex. Therefore, the SR model trained on simulated data becomes less effective when applied to real scenarios. In this paper, we propose a novel super resolution framework based on blur kernel estimation and dual attention mechanism. Our network learns the internal relations from the input image itself, thus the network can quickly adapt to any input image. We add the blur kernel estimation structure into the network, correcting the inaccurate blur kernel to generate high quality images. Meanwhile, we propose a dual attention mechanism to restore the texture details of the image, adaptively adjusting the features of the image by considering the interdependencies both in channel and spatial. The combination of blur kernel estimation and attention mechanism makes our network perform well for complex blur images in practice. Extensive experiments show that our method (KASR) achieves promising accuracy and visual improvements against most existing methods.

Keywords: super resolution; blur kernel; attention mechanism

1. Introduction

Super-resolution (SR) [1] aims to generate a high-resolution (HR) image from its low-resolution(LR) image. The key of SR is to add some additional information in the image reconstruction process to compensate for the loss of detail information caused by image degradation, so as to reconstruct the clear HR image from the LR image. Obtaining HR Images is significant in security surveillance [2], medical [3], object recognition [4], and other fields. To solve the SR problem, many learning-based approaches have been proposed to learn the mapping between LR and HR image pairs.

Since the rapid development of deep convolutional neural networks (CNN), various architectures of SR methods have been designed to improve the performance of SR models. SRCNN [5] is the first work to use a three-layered convolutional neural network for SR. The depth of neural network is critical to deep learning. With the emergence of residual net (ResNet) [6], many methods [7–10] try to deepen the network in order to obtain high quality images. However, hundreds of layers make the network hard to train and adjust. Furthermore, the great deep network would lose low-frequency information. In addition, simply building deeper networks is difficult to obtain better improvements, and Peak Signal-to-Noise Ratio (PSNR) [11] values have reached certain limits.

On the one hand, most deep learning SR methods use external predefined downsampling blur kernels (e.g., bicubic downsampling) images to train the network. The learning relationships from this datasets are limited. The real world LR images blur kernels are unknown and complex, thus, when the predefined blur kernel is different from the real world, these methods suffer from severe performance...
degradation. Recently, more scholars pay attention to SR for real world images. Some methods [12,13] establish new datasets to study SR from the perspective of camera images. Some methods [14,15] attempt to add blur kernels as additional inputs but cannot predict every image accurately. On the other hand, the majority of SR methods [5,7–9,16,17] treat features among channel and spatial equally, which lacks flexibility in processing different information among low and high frequency. Image SR can be seen as a process in which we try to recover as much high frequency information as possible. Treating features equally would lack identification learning ability across feature maps, and finally hinder the representation ability of network.

In this paper, we focus on using deep learning methods to solve the SR problem in different cases. In view of the shortcomings of existing methods, we propose a new SR method which combine blur kernels’ estimation and attention mechanism together based on ZSSR [15]. Our method reconstructs an HR image by learning the recurrence information from the input LR image itself. We add the blur kernels estimation structure into network. With the predicting image blur kernel, our network performs well in practice. Furthermore, we propose a new dual attention mechanism into our network. Such a dual attention mechanism enables our method to focus on more useful information and enhance the ability of recognition and learning.

Overall, our main contributions are summarized as follows:

- We propose a new SR method based on blur kernels estimation and attention mechanism to improve the performance for Super-Resolution. Our method is more suitable for solving the problem of SR reconstruction in real life.
- We propose a new dual attention module to consider the interdependence between features both in channels and spatial.
- We test the performance on real world images, isotropic Gaussian blur kernels, and specific blur kernels images. The experimental results show that our method achieves better SR performance than most existing methods.

2. Related Work

In the past few years, many SR methods have been studied. Since SRCNN [5] was proposed as the first SR method using CNN network, more and more CNN architectures [10,16–18] have been proposed in image SR. Based on residual architecture [6], the majority of existing CNN-based SR models focus on designing a deeper or broader network to achieve better performance. VDSR [16] is the pioneer work in introducing residual blocks into SR networks and they achieved significant improvements in accuracy. EDSR [7] improves it by removing unnecessary batch normalization layers to simplify the network architecture. DenseSR [19] uses effective residual dense block in the SR network.

However, the above SR networks are trained by external datasets with predefined blur kernel (e.g., bicubic). They exhibit poor performance in non-bicubic downsampling scenarios. In order to solve the problem, SRMD [18] was proposed for multiple blur kernels to achieve better results than other SR methods in a non-bicubic condition. CAB [14] proposes the conditional regression model. They generate SR images by effectively utilizing additional kernel information in the training and inference. ZSSR [15] trains the network with examples extracted from the input image itself. To get a higher-quality SR image, these methods still need to add extra true blur kernel information.

Extracting features from the original LR inputs and improving the resolution at the end of the network has become the main choice for deep architecture. These methods must first interpolate the LR input to the required size, which inevitably loses some of the detail and greatly increases the computation. Attention in human perception generally means that the human visual system focuses on the most informative components [20].

In recent years, attention mechanisms have been widely used in super resolution reconstruction to improve performance, but it mainly focuses on channel or spatial attention mechanism, respectively. NLRN [21] takes into account the feature correlation of spatial dimension. RCAN [10] use squeeze-and-excitation (SE) block [22] to improve the representational ability of the SR network. Inspired by RCAN,
SCA [23] proposes a deep second-order attention to further improve SR performance by exploring second-order statistics of features. To investigate spatial and channel mechanism further, we propose a novel dual attention mechanism. Our network pays more attention to both channel and spatial features to obtain improvements.

3. Proposed Method

We constructed a new super resolution framework (KASR) which combines blur kernels estimation and dual attention mechanism to improve SR performance. Our method generates a super resolution image from the input LR image itself. The details of our proposed framework are described below.

3.1. ZSSR Method Introduction

ZSSR utilizes the internal information of the image and the generalization ability of deep learning. As shown in Figure 1, given a LR image $I_{LR}$, ZSSR network can construct a high quality HR image without external examples available to train on. Since their network is trained to infer the complex LR–HR relationships from test LR image $I_{LR}$. The test image $I_{LR}$ is downscaled to many smaller versions of itself ($I'_{LR} = I_0, I_1, I_2, \ldots, I_n$). With these smaller versions, LR–HR sample image pairs are formed to construct the training set. The network can then randomly train on these pairs. In other words, they train the network to extract examples from the test image itself. These learning relationships are then applied to the LR input image $I_{LR}$ to generate the HR output $I_{HR}$.

3.2. Proposed Network

Based on ZSSR, we propose a progressive model for image SR. Our method can obtain an HR image from the input LR image itself like ZSSR without pre-training. The network is trained to infer the complex image-specific HR–LR relationships from LR image $I_{LR}$ and its downscaled versions $I'_{LR}$. As shown in Figure 2, our network mainly consists of three parts: Kernel estimation network is designed to predict the blur kernel of an input LR image, feature extraction network part whose purpose is reconstructing the HR image, and the dual attention network parts which aims to extract texture details from input images.
3.2.1. Kernel Estimation Network

The kernel estimation network designs are shown in the left part of Figure 2. This structure takes the image \( I_{LR} \) as input. This part contains four \( 3 \times 3 \) convolution layers with ReLU activations and a global average pooling layer. The convolution layers give the estimation of the kernel spatially and form the estimation maps. Then, the global average pooling layer gives the global estimation by taking the mean value spatially. Then, the kernel estimation feature map and the LR image are input together into a convolution layer.

3.2.2. Feature Extraction Network

The feature extraction network mainly captures the relationships from the LR–HR pairs which are downscaled by the LR input image \( I_{LR} \). The network learns the high-frequency details and reconstructs the high-resolution image. In this architecture of feature extraction, the network takes the concatenated LR image and kernel maps as input. We use full convolution operation with eight layers before features fusion. The convs use the same filter size \( 3 \times 3 \) each has 64 channels. The 8th layer output will merge with the features extracted from the attention blocks later. The \( i \)-th layer feature is represented as:

\[
F_i = \sigma (W_i * F_{i-1}) + F_0
\]

where \( W_i \) is the weight of the convolution layer, \( \sigma \) is the ReLU activation function, \( F_{i-1} \) denotes the output of the feature extract from first convolution layer, and we omit the bias for simplicity. In order to prevent the network from losing the shallow information, the skip connection is used in our network. The rich low frequency information from the LR input images can be transmitted to the network behind layer. It makes our network deeper and uses more pixel information to achieve better SR performance.

3.2.3. Attention Network

As illustrated in Figure 3, we proposed a new dual attention mechanism architecture to improve SR performance. This module extracts the interdependencies statistic both among channel and spatial dimensions, respectively. Channel attention and spatial attention are complementary.

\[ \text{Channel SE} \]
\[ \text{Spatial SE} \]

**Figure 3.** Architecture of the dual attention network.

**Channel Attention:** The channel attention(CA) structure we proposed is shown in the upper part of the Figure 3. The channel attention is built to leverage the interdependencies between channel maps, it strengthens the feature mapping of interdependence, and improves the feature representation of specific semantics. We input the shallow feature \( F_0 \) to calculate the channel attention map \( CA^{C \times C} \). We add a SE [22] operation after the convolution layer to further enhance the channel weight. The feature map \( F^{C \times H \times W} \) is reshaped to \( F_R \), \( F_R \) transposed to \( F_T \), and then we multiply these two matrices. The result is applied a softmax layer to get the channel attention map CA:

\[
ca_{ij} = \frac{\exp (F_i \cdot F_j)}{\sum_{i=1}^{C} \exp (F_i \cdot F_j)}
\]
where \(c_{aij}\) represents the \(i\)th channel’s influence on the \(j\)th channel. Furthermore, we restore the feature map result \(F_{ji}\) to the original shape by multiplying the transpose matrices of \(CA\) and \(F\). Finally, the output \(F_{CA}\) can be computed as:

\[
F_{CA} = \alpha \sum_{i=1}^{C} (c_{aij}F_{i}) + F_{j}
\]  

(3)

where \(\alpha\) is a scale parameter.

**Spatial Attention:** As shown in the lower part of the Figure 3, the spatial attention (SA) is generated by using the spatial relation of features, which is a supplement to channel attention. To compute spatial attention, we first apply SE [22] operation with the features output from the previous layer. Then, average pool and max pool operations aggregate the information of the feature map to generate two 2D maps: average-pooled features \(F_{avg}\) and max-pooled features \(F_{max}\). Pooling operation effectively highlights the information area. Those are then concatenated and input to a convolution layer to generate the spatial attention map, which encodes where emphasis or suppression is needed. The spatial attention \(F_{SA}\) can be computed as:

\[
F_{SA} = \sigma \left( f^{3\times3}\left( [\text{Avg Pool}(F); \text{Max Pool}(F)] \right) \right)
\]

(4)

where \(f^{3\times3}\) is the convolution layer with the kernel size of \(3 \times 3\), \(\sigma\) here is the sigmoid function.

To take full advantage of the context information, we put together the features of the two attention modules. Specifically, we transform the output of the two attention modules through the convolution layer, and carry out an element-wise sum to achieve the feature fusion. Finally, the final prediction map is generated after the convolution layer. Our attention module is simple and can be plugged directly into an existing Fully Convolutional Network (FCN) pipeline. They do not add too many parameters, but effectively enhance the feature representation.

4. Experiments and Discussion

We both consider the blur kernel estimation and attention mechanism to solve the SR problem. Our method (KASR) generates the HR image without external pre-defined datasets, and the network utilizes the internal information of input LR images. In order to verify our experimental performance, we ran several experiments in a variety of settings. Experimental results demonstrate that our method KASR produces competitive results in application. Our method performs better than most existing SR methods in cases of real world LR images with unknown blur kernel, without massive external datasets. In addition, we also show superior results for images with complex blur kernels. Furthermore, our method also achieves comparable results on the benchmark datasets with predefined blur kernel (bicubic).

4.1. Implementation Details

For our experiments, we use \(L_1\) loss with ADAM optimizer to optimize our networks. We start the learning rate of 0.001 start. We periodically do a linear fitting of the reconstruction error and, if the standard deviation is one factor higher than the slope of the linear fitting, we divide the learning rate by 10. It stops when the learning rate reaches \(10^{-6}\). Because our training set is generated from the one input test image, data augmentation is performed on input images to take more LR–HR sample pairs for learning. The test image is downscaled to many smaller versions of itself by the desired scale-factor \(s\). We further enriched the training set LR–HR pairs, which are randomly rotated by using four rotations \((0^\circ, 90^\circ, 180^\circ, 270^\circ)\) and flipped in both vertical and horizontal directions.
4.2. Real World Cases

Most of the SR methods are thoroughly trained and optimized for a specific predefined blur kernel, but real LR images are often not ideally generated. These specialized SR networks perform terribly in practice. We compare our method with ZSSR [15], SRMD [18], state of the art (SotA) deep trained SR method RCAN [10], and SCA [23]. Quantities’ experiments show that our method works well for real-world images. Since the real images don’t have the ground-truth images, we only provide visual comparison. As shown in Figure 4, the visual results prove that our method (KASR) performs significantly better performance on different real world cases: Internet image, old movie images, and phone images.

Figure 4. SR results (×3) of real world images (old movie image, phone image) by different methods.
With the self-learning ability of the internal information in LR image, the performance of our method (KASR) is more robust for real-world LR images with unknown blur kernels. Our method (KASR) trains at test time on examples extracted from the test image, and it is better to provide more clear reconstruct images in unconstrained and unknown settings.

### 4.3. KASR with Complex SR Kernel

In this section, we conduct further experiments to verify the ability of the proposed KASR to handle the complex SR kernel images. The purpose of this experiment is to test the ability of a more realistic blur kernel to evaluate the results numerically.

We use isotropic Gaussian blur kernels to downscale the HR images on the numerous dataset. The kernel width ranges are set to [0.2], [1.3] and [2.6] for SR factors 2–4, respectively. We uniformly sample the width of kernel within the above ranges.

Table 1 compares our results of PSNR value with that of the leading SR approach CAB [14] and ZSSR [15]. When compared with CAB [14] and ZSSR [15], our method KASR performs better values than other methods on most datasets with all scaling factors and kernels. Our KASR’s PSNR obtains ~3 dB improvement over CAB [14] on Set5 (SR × 2 for kernel set to [0.2], [1.3]). In other cases, KASR also exceeds CAB by about 1 dB on average. This further demonstrated the effect of our network. Figure 5 shows the SR visual result comparison of different methods with SR factor 4 and kernel width 1.6 on image “Img_97” from Urban100. Our method reconstructs the texture and edge of the images better than other methods. The results prove that KASR can still be greatly improve the performance of complex blur kernel images.

### Table 1. Comparison of SR results (PSNR) for isotropic Gaussian blur kernel cases.

| Method       | Kernel | Set5   | Set14  | B100   |
|--------------|--------|--------|--------|--------|
|              |        | ×2     | ×3     | ×4     | ×2     | ×3     | ×4     |
| CAB [14]     | 0.2    | 33.27  | 31.03  | 29.31  | 30.29  | 28.29  | 26.91  |
| ZSSR [15]    | 0.2    | 36.13  | 33.70  | 31.27  | 32.91  | 29.95  | 27.88  |
| KASR (ours)  | 0.2    | 36.27  | 33.81  | 31.30  | 32.93  | 29.98  | 27.95  |
|              | 1.3    | 33.42  | 31.14  | 29.50  | 30.51  | 28.34  | 27.02  |
| CAB [14]     | 1.3    | 36.01  | 33.28  | 31.14  | 32.56  | 29.43  | 27.25  |
| ZSSR [15]    | 1.3    | 36.17  | 33.40  | 31.26  | 32.71  | 29.52  | 27.34  |
| KASR (ours)  | 1.3    | 33.21  | 30.82  | 28.81  | 29.74  | 27.83  | 26.15  |
|              | 2.6    | 32.97  | 31.62  | 30.21  | 29.96  | 27.83  | 26.47  |
| CAB [14]     | 2.6    | 33.01  | 31.63  | 30.25  | 29.97  | 27.88  | 26.46  |
| ZSSR [15]    | 2.6    | 32.11  | 30.82  | 28.81  | 29.74  | 27.83  | 26.15  |
| KASR (ours)  | 2.6    | 31.30  | 30.25  | 29.97  | 27.88  | 26.46  | 25.59  |

*Figure 5. SR visual result comparison of different methods with SR factor 4 and kernel width 1.6 on image “Img_97” from Urban100.*

### 4.4. KASR with Bicubic Blur SR Kernel

In addition, we also conduct experiments on images fixed bicubic SR kernel (scale factor × 2, × 3, × 4). Following [15,18], we use three standard benchmarks datasets: Set5, Set14, B100. The SR results are evaluated with PSNR and SSIM [11] on Y channel of transformed YCbCr space. Here, we compare SR results with four SotA methods: SRCNN [5], VDSR [16], SelfExS [24], and ZSSR [15]. As shown in Table 2, our method KASR achieves competitive results with the external-supervised methods like VDSR [16] which are carefully trained with massive external datasets. In fact, KASR is significantly
superior to the earlier method SRCNN [5] over ~1 dB. KASR has a great advantage over SelfExS [24] ~0.5 dB on average and, in some cases, achieves comparable or better results than VDSR [16] for SR ×2, ×3.

Table 2. Comparison of SR results for Bicubic blur kernel cases.

| Method    | Scale | Set5 PSNR | Set14 PSNR | B100 PSNR |
|-----------|-------|-----------|------------|-----------|
| SRCNN [5] | ×2    | 36.66     | 32.42      | 31.36     |
| VDSR [16] | ×2    | 37.53     | 33.03      | 31.90     |
| SelfExSR [24] | ×2   | 36.49     | 32.22      | 31.18     |
| ZSSR [15] | ×2    | 37.37     | 33.00      | 31.65     |
| KASR (ours) | ×2  | 37.54     | 33.09      | 31.78     |

Table 2. Comparison of SR results for Bicubic blur kernel cases.

| Method    | Scale | Set5 SSIM | Set14 SSIM | B100 SSIM |
|-----------|-------|-----------|------------|-----------|
| SRCNN [5] | ×2    | 0.9542    | 0.9063     | 0.8879    |
| VDSR [16] | ×2    | 0.9587    | 0.9124     | 0.8960    |
| SelfExSR [24] | ×2   | 0.9537    | 0.9034     | 0.8855    |
| ZSSR [15] | ×2    | 0.9570    | 0.9108     | 0.8920    |
| KASR (ours) | ×2  | 0.9586    | 0.9135     | 0.8951    |

Notably, experiments show that our method (KASR) performs well in images with strong internal repeating structures. The visual results on Urban100 are significantly better than on other datasets, as urban scenes in the dataset mainly include structured scenes with lots of patch-redundancy. More qualitative comparison visual results are shown in Figure 6, and we choose images from Urban100 with bicubic downscaling. The result images prove that our method KASR tends to surpass ZSSR [15]. Our method generates better visual results than RCAN [10], although the PSNR value of our method is lower. This further proves that the preference of SR problems for internal learning and attention mechanism.

Figure 6. SR results (×3) of BICUBIC blur kernel images by different methods.
4.5. Ablation Study for Attention

In addition, to demonstrate the effect of our proposed dual attention mechanism structure in SR, we remove channel attention or/and spatial attention from our network. First, we evaluate the PSNR on Urban100 (2×) with bicubic downsampling. As shown in Table 3, when both CA and SA are removed, the PSNR value is the lowest. When SA is added, the performance can be improved from 31.12 dB to 31.14 dB. Comparing the results of the second and the third columns, we find that networks with CA would perform better than those with SA. The performance would further obtain better results by using both of them, and the PSNR is increased to 31.22 dB. Furthermore, in Figure 7, we also choose visual results from B100 with redefined bicubic blur kernel and real world natural image. The addition of SA and CA makes the network construct higher quality images.

| CA Block | SA Block | PSNR on Urban100 (2×) |
|----------|----------|-----------------------|
|          |          | 31.12                 |
| X        |          | 31.14                 |
|          | ✓        | 31.19                 |
| ✓        |          | 31.22                 |

Table 3. Investigations of Attention (CA and SA).

![Figure 7. SR results (×3) of different attention.](image)

When CA and SA are both added into the network, we can get images with more details. The results confirm that the dual attention mechanism we proposed plays a vitally important role in our network.

4.6. Runtime

Our method’s training is done at test time, and the average runtime per image for SR (×2) is 2.5 min (on a NVIDIA TITAN X(12 GB) GPU). For training iterations, the final test run time is negligible. As a comparison, the training-time of the leading RCAN [10] (on the same platform) is more than a week for one specific blur kernel. Although RCAN’s test-time is fast, it only works well on the specific blur kernel used in training.

5. Conclusions

In this paper, we proposed a novel super resolution framework based on the blur kernel estimation and dual attention mechanism. Compared with existing SR methods, our method extracts the internal relations from the input LR image itself and estimates the blur kernel, which is suitable for practical application in Internet image clearing, old movie restoration, natural images sharpening, etc. Moreover,
we propose a new parallel dual attention mechanism which is crucial for restoring image texture details to enhance super-resolution performance. The experiments’ results show that our method achieves comparable performance to the state-of-the-art deep model. We believe that the proposed method can be applied not only to specific downscaling blur kernel but also to various types of complex blur kernel and real world LR images.

**Author Contributions:** Conceptualization, H.L.; Data curation, H.L. and F.W.; Investigation, X.Q.; Project administration, Y.G.; Resources, Y.D.; Software, Y.D.; Writing—original draft, H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

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

**References**

1. Freeman, T.; Pasztor, C.; Carmichael, T. Learning Low-Level Vision. *Int. J. Comput. Vis.* **2000**, *40*, 25–47. [CrossRef]
2. Wilman, W.; Zou, P.W.; Yuen, C. Very Low Resolution Face Recognition Problem. *IEEE Trans. Image Process.* **2002**, *21*, 327–340.
3. Shi, W.; Jose, C.; Christian, L.; Zhuang, X.; Bai, W.; Bhatia, K.; Simoes, A.; Dawes, T.; Declan, P.; Cardiac, D. Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 9–16.
4. Sajjadi, S.; Ikop, B.; Hirsch, M. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4501–4510.
5. Dong, C.; Loy, C.; He, K.M.; Tang, X.O. Learning a Deep Convolutional Network for Image Super-Resolution. *Eur. Conf. Comput. Vis.* **2014**, *8692*, 184–199.
6. He, K.M.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
7. Lim, B.; Son, S.; Kim, H.; Nah, S., Lee, M.K. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
8. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2790–2798.
9. Tai, Y.; Yang, J.; Liu, X.; Xu, C. MemNet: A Persistent Memory Network for Image Restoration. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4549–4557.
10. Zhang, Y.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018; pp. 294–310.
11. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2014**, *13*, 600–612. [CrossRef] [PubMed]
12. Chen, C.; Xiong, Z.; Tian, X.; Zha, Z.-J.; Wu, F. Camera Lens Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 1652–1660.
13. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward Real-World Single Image Super-Resolution: A New Benchmark and a New Model. In Proceedings of the International Conference on Computer Vision, Thessaloniki, Greece, 23–25 September 2019; pp. 3086–3095.
14. Riegler, G.; Schulte, S.; Ruther, M.; Bischof, H. Conditioned regression models for non-blind single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 522–530.
15. Shocher, A.; Cohen, N.; Irani, M. “Zero-Shot” Super-Resolution Using Deep Internal Learning. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 3118–3126.
16. Kim, J.; Lee, J.k.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
17. Kim, J.; Lee, J.k.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
18. Zhang, K.; Zuo, W.; Zhang, L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 3262–3271.
19. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.V.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2472–2481.
20. Itti, L.; Koch C.; Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 1254–1259. [CrossRef]
21. Liu, D.; Wen, B.; Fan, Y.; Change, C.; Huang, S. Non-Local Recurrent Network for Image Restoration. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 1680–1689.
22. Hu, J.; Shen, L.; Sun G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
23. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11065–11074.
24. Huang, J.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.

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

© 2020 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 (http://creativecommons.org/licenses/by/4.0/).