Learning Omni-frequency Region-adaptive Representations for Real Image Super-Resolution

Xin Li*, Xin Jin†, Tao Yu, Simeng Sun, Yingxue Pang, Zhizheng Zhang, Zhibo Chen†
CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System,
University of Science and Technology of China
Hefei 230027, China
{lixin666, jinxustc, yutao666, smsun20, pangyx, zhizheng}@mail.ustc.edu.cn, chenzhibo@ustc.edu.cn

Abstract

Traditional single image super-resolution (SISR) methods that focus on solving single and uniform degradation (i.e., bicubic down-sampling), typically suffer from poor performance when applied into real-world low-resolution (LR) images due to the complicated realistic degradations. The key to solving this more challenging real image super-resolution (RealSR) problem lies in learning feature representations that are both informative and content-aware. In this paper, we propose an Omni-frequency Region-adaptive Network (OR-Net) to address both challenges, here we call features of all low, middle and high frequencies omni-frequency features. Specifically, we start from the frequency perspective and design a Frequency Decomposition (FD) module to separate different frequency components to comprehensively compensate the information lost for real LR image. Then, considering the different regions of real LR image have different frequency information lost, we further design a Region-adaptive Frequency Aggregation (RFA) module by leveraging dynamic convolution and spatial attention to adaptively restore frequency components for different regions. The extensive experiments endorse the effective, and scenario-agnostic nature of our OR-Net for RealSR.

Introduction

With the development of deep learning, single image super-resolution (SISR) has achieved great success either on PSNR values (Dong et al. 2015; Haris, Shakhnarovich, and Ukita 2018; Kim, Kwon Lee, and Mu Lee 2016; Lim et al. 2017; Zhang et al. 2018a; Dai et al. 2019; Mei et al. 2020; Pan et al. 2020) or on visual quality (Ledig et al. 2017; Sajjadi, Scholkopf, and Hirsch 2017). In general, these traditional SISR methods typically focus on restoring the low-resolution (LR) image with single and uniform synthetic degradation, such as bicubic down-sampling and Gaussian down-sampling. However, the degradations in real-world LR images are usually far more complicated, which makes most SISR models become less effective when directly applied to practical scenarios.

Figure 1: Difference analysis in frequency domain between RealSR and the conventional SISR: (a) the overall frequency distribution comparison between RealSR and SISR. (b) the frequency components distribution comparison between different regions of the same real LR image. Δ means the absolute difference of frequency degradation.

In recent years, some studies that focus on solving the real image super-resolution (RealSR) problem have attracted more and more attention (Cai et al. 2019; Wei et al. 2020). Unlike the single and uniform synthetic degradation in SISR, the LR and HR images of RealSR are captured with digital single lens reflex (DSLR) cameras, which typically contains various/complex non-uniform real-world degradations, including blur, noise, and down-sampling. That is why those classic conventional SISR methods (RCAN (Zhang et al. 2018a), EDSR (Lim et al. 2017), SAN (Dai et al. 2019) etc.) cannot handle the RealSR problem well.

Among recent representative RealSR methods, Cai et al. (Cai et al. 2019) first construct a RealSR dataset by capturing the LR-HR image pairs on the same scene with different focal lengths of a digital camera. And then, they propose a Laplacian pyramid based kernel prediction network (LP-KPN) to handle such non-uniform degradation. Concur-
Based on our analysis, we propose an
DRealSR. Considering the targets of RealSR vary with image regions, they further design a component divide-and-conquer (CDC) model to adaptively restore the LR image. But, LP-KPN and CDC both only focus on super-resolving LR images by collaborating different pixel-wise local restorations (e.g., flat regions, edges and corners), they ignore to make full use of hierarchical features across different frequency domains to comprehensively enhance texture details for real LR images.

In this paper, we first analyse the latent and essential challenges of RealSR from a completely different perspective of frequency distributions. In particular, we perform an experimental comparison between RealSR and conventional SISR in the frequency domain to explore their difference. Specifically, we visualize two LR-HR pairs in frequency domain through the wavelet transform tool, where one pair belongs to SISR and the other belongs to RealSR. As shown in Figure 1(a), we observe that the degradation (from HR image to LR image) of the general SISR mainly exists in the high-frequency component. In contrast, the degradation of RealSR exists in all frequency components. Besides, in Figure 1(b), we also see that the degradation of different regions of an image are usually distributed in the different frequency components.

Based on the above analysis, we argue that an effective RealSR model should learn feature representations that are both informative and content-aware. Informative property promises to restore sufficient realistic texture details across multiple frequency domains for the degraded LR, and content-aware property satisfies the varied targets of RealSR for different image regions (i.e., smoothing for flat regions and sharpening for edges). In this paper, we propose a Omni-frequency Region-adaptive Network (OR-Net) to efficiently solve the RealSR problem. Specifically, we first design a Frequency Decomposition (FD) module to decompose LR image into low-frequency, middle frequency, and high frequency components. Then, we employ multiple interactive branches to enhance the corresponding frequency factors. Second, to achieve the content-aware super-resolution for real LR images, we further design a Region-adaptive Frequency Aggregation (RFA) module by combining the dynamic convolution and spatial attention to selectively restore different frequency components for the different positions of HR images. The contributions of this paper can be summarized as follows:

- We analyse the essential difference between general SR and RealSR from the frequency perspective, to answer the question that why the classic SISR methods cannot handle RealSR problem well.
- Based on our analysis, we propose an Omni-frequency Region-adaptive Network (OR-Net) for RealSR, which contains two technical novelties–1) Frequency Decomposition (FD) module that aims to achieve the LR image content separation in frequency domain and enhance texture details across all frequency components, 2) Region-adaptive Frequency Aggregation (RFA) module that aims to appropriately restore different frequency components for real HR images in different positions.
- Extensive experiments on multiple RealSR benchmarks have validated the effectiveness and superiority of our OR-Net. Sufficient intuitive visualization results/analysis are also provided to support/verify the expected functions of the proposed FD and RFA modules.

Related Work

Conventional Single Image Super-Resolution

In the last decade, the traditional SISR has achieved great progress, especially for deep learning based approaches (Dong et al. 2015; Dong, Loy, and Tang 2016; Lim et al. 2017; Haris, Shakhnarovich, and Ukita 2018; Zhang et al. 2018a; Dai et al. 2019). These methods usually perform well on the the synthetic degradation (e.g., bicubic down-sampling) but generalize poorly to realistic complicated distortions in real-world scenarios. This is problematic especially in practical applications, where the target scenes typically have hybrid/complex non-uniform degradations (e.g., blur, noise, and down-sampling), and also, there is always no readily available paired data for training.

Real Image Super-Resolution

RealSR has drawn more and more attention in recent years. Different from the general SISR that typically focuses on the simple and uniform synthetic degradation, RealSR mainly aims to solve these realistic complicated degradations in real-world scenarios. To capture the distortion of real scene, Chen et al. (Chen et al. 2019a) design two novel data acquisition strategies. Cai et al. (Cai et al. 2019) build a real-world super-resolution (RealSR) dataset by adjusting the focal length of a digital camera, and introduce the Laplacian pyramid based kernel prediction network (LP-KPN) to solve such non-uniform distortions. Recently, Wei et al. (Wei et al. 2020) present a large-scale diverse real-world image super-resolution dataset (DRealSR) and a component divide-and-conquer (CDC) model with gradient weighted (GW) loss, achieving great performance in RealSR.

However, above methods ignore to consider and study the difference between general SISR and RealSR. They didn’t design solutions based on the essential difference analysis, which is sub-optimal and un-targeted. In this paper, we start from analysing the difference between two kinds of super-resolution tasks in frequency domain in detail. Besides, we also analyse the degradation of different regions in the same LR image. Based on these analysis, we study how to design a generalizable and efficient RealSR framework that can exploit the merits of previous works while more targeting on the specific characteristics of RealSR itself.

Frequency Decomposition and Content Adaptation

Lots of excellent low-level restoration studies explore to enhance/reconstruct content details from the frequency decomposition perspective, including image denoising/deraining (Fu et al. 2017), rescaling (Xiao et al. 2020) and super-resolution (Fritsche, Gu, and Timofte 2019; Pang et al. 2020). Particularly, Chen et al. introduce an octave
convolution (Chen et al. 2019b) to decompose features in the frequency domain to reduce spatial redundancy in CNNs. Akbari et al. (Akbari et al. 2020) further extend the octave convolution and frequency decomposition idea to the image compression field. These methods typically decompose the degraded image into low-/high-frequency factors, and individually deal with different components to achieve divide-and-conquer. However, they usually ignore the interaction and aggregation between multiple frequency factors. In addition, with the development of some content-adaptation techniques (e.g., attention mechanism (Wang et al. 2018; Woo et al. 2018; Hou et al. 2020; Mei et al. 2020; Zhang et al. 2020), region normalization (Yu et al. 2020) and dynamic convolution (Jia et al. 2016; Mildenhall et al. 2018; Lin et al. 2019; Chen et al. 2020), some low-level restoration algorithms (e.g., image interpolation (Niklaus, Mai, and Liu 2017), hybrid-distorted image restoration (Li et al. 2020) and denoising (Mildenhall et al. 2018)) also achieve the purpose of ‘divide-and-conquer’, where the processing operations can be adjusted adaptively according to the changing image content in the reference. In this paper, considering that the different regions of real LR image inevitably have different frequency information lost, we explore to simultaneously utilize dynamic convolution and spatial attention to achieve adaptive frequency information compensation.

Omni-frequency Region-adaptive Network
We aim at designing a generalized and efficient framework for RealSR. During the training, we have access to the annotated real-world SR datasets with LR-HR pairs. The trained model is expected to work well with high generalization capability for the real LR images. Figure 2 shows the overall flowchart of our framework. Particularly, we propose an Omni-frequency Region-adaptive Network (OR-Net) to boost the SR performance on the real low-resolution images. OR-Net is designed from a frequency perspective, and contains two novel technical designs: Frequency Decomposition (FD) module and Region-adaptive Frequency Aggregation (RFA) module. As shown in Figure 2(a), FD module first employs an architecture with three branches, which decomposes the input real LR image into low-/middle-/high-frequency components (i.e., omni-frequency) in frequency domain and enhance them in an interactive manner to achieve the comprehensive information compensation. Moreover, the RFA module (see Figure 2(c), 4) adaptively fuses the enhanced omni-frequency for different regions to achieve the content-aware super-resolution.

**Frequency Decomposition (FD) Module**
To obtain the informative omni-frequency representation for RealSR, we propose the Frequency Decomposition (FD) module. As shown in Figure 2(a), FD module is consists of two phases: frequency decomposition phase and frequency enhancement phase. The first frequency decomposition phase aims to separate the low-/middle-/high- frequency components from the LR input, and the frequency enhancement phase is dedicated to enhance the different frequency representations. To encourage the interaction between different frequency components and enhance frequency components in a coarse-to-fine (easy-to-hard) manner, we also
progressively utilize the enhanced lower frequency representation to help the enhancement of higher frequency components by concatenating two frequency representations.

For the FD module, we denote the input (which is an RGB LR image) by $I \in \mathbb{R}^{h \times w \times 3}$ and the output includes three kinds of enhanced frequency features: low-frequency enhanced features $\tilde{f}_l$, mid-frequency enhanced features $\tilde{f}_m$, and high-frequency enhanced features $\tilde{f}_h$.

**Frequency Decomposition Phase.** In FD module, we first decompose the input LR image $I$ into different frequency components. Such frequency components separation could be achieved in wavelet transform (??) or discrete cosine transform (Ahmed, Natarajan, and Rao 1974) in the traditional signal-processing methods. However, with the mathematical operations being deterministic and task-irrelevant, such transforms inevitably discard some critical/detailed information for low-level restoration tasks. To imitate the wavelet transform while avoiding key information lost, we propose to factorizes the mixed feature representations through the learnable latent-wise spatial down-sampling, the similar operation can be found in the recently-proposed octave convolution (OctConv) (Akbari et al. 2020) but OctConv aims to reduce channel-wise redundancy like group or depth-wise convolutions. Specifically, we first utilize the convolution layer with the larger stride (e.g., stride= 2) to downsample the feature to extract the coarse features, i.e., the low-frequency components. Then, we remove such relatively low frequency components from the original feature (before downsample) to obtain the rest relatively high frequency feature representations.

In formula, and as shown in Figure 2(a), we downsample the feature space by using convolution layer with stride= 4 to get the corresponding low-frequency components $f_l$. Then we get the middle frequency components $f_m$ by removing the $f_l$ from the corresponding original features, which is also downsampled with stride= 2. Similarly, to get the high frequency components $f_h$, we remove the downsampled features with stride= 2 from the features without down-sampling, which has same spatial size with original LR image. The whole process can be denoted as follows:

$$
\begin{align*}
  f_l &= \text{Conv}_2 \downarrow 2(\text{Conv}_2 \downarrow 2(I)), \\
  f_m &= \text{Conv}_2 \downarrow 2(I) - \text{Conv}_2 \downarrow 2(\text{Conv}_2 \downarrow 2(I)) \uparrow 2, \\
  f_h &= \text{Conv}(I) - \text{Conv}_2 \downarrow 2(I) \uparrow 2,
\end{align*}
$$

where $\text{Conv}_2 \downarrow 2$ denotes the convolution layer with stride= 2 and $\text{Conv}$ denotes the convolution layer without down-sampling. $\uparrow$ means the bilinear up-sampling operation. The corresponding interpretive/analysis experimental results that support the reasonableness of such frequency decomposition design can be found in Figure 6.

**Frequency Enhancement Phase.** After extracting the low-/middle-/high-frequency components from LR image, we enhance these representations by a well-designed Frequency Enhancement Unit (FEU), to make up for low-/middle-/high-frequency components lost. Specifically, the FEU is designed based on the popular GRDB module (Kim, Ryun Chung, and Jung 2019), which can be regarded as an dense connection block. But, as shown in the Figure 2(b) and Figure 3, the features on the different stages of FEU are usually different and contain different information (some features focus on object structure but others focus on texture details). Hence, the common fusion/concatenate (directly sum up all features) that used in traditional dense connection is not consistent with our purpose (i.e., divide-and-conquer). Intuitively, the simple sum-up of all features cannot promise that the low-frequency branch only focuses on the low-frequency features, same issues also exist in other two branches (middle- and high-freq. branch). To address this problem, except of the regular operations (e.g., non-linear transformation achieved by Conv+ReLU, dense connection), we further integrate the channel-wise attention to adaptively adjust the residual information aggregation in FEU as shown in Figure 2(b), which helps each branch selectively fuse the corresponding frequency components at different stages, and thus the representation capability in frequency domain of each branch is significantly improved.

Moreover, considering that the high-frequency feature components are relatively difficult to restore/enhance (Fritsche, Gu, and Timofte 2019; Wei et al. 2020), we propose to enhance such challenging frequency component in a coarse-to-fine/easy-to-hard manner. In detail, we encourage the interaction between different frequency components, and progressively utilize the enhanced lower frequency representations to help the enhancement of higher frequency components by concatenating them together. We define this process as:

$$
\begin{align*}
  \tilde{f}_l &= \text{Enhance}(f_l), \\
  \tilde{f}_m &= \text{Enhance}(f_m, \tilde{f}_l), \\
  \tilde{f}_h &= \text{Enhance}(f_h, \tilde{f}_m, \tilde{f}_l),
\end{align*}
$$

where $\text{Enhance}()$ denotes a set of several frequency enhancement unit (FEU), the specific number of FEU of each branch can be found in Figure 2(a).

**Region-adaptive Frequency Aggregation (RFA) Module**

Different from the conventional SISR, the degradation (from HR→LR image) of RealSR generally exists in all frequency components. Besides, the frequency information loss of different regions in a real LR image is different. Therefore, it is necessary to adaptively aggregate omni-frequency components for different regions to restore a more realistic HR image with the rich texture details. In this section, we introduce the Region-adaptive Frequency Aggregation (RFA) module in detail (shown in Figure 4), which achieves the
Table 1: Quantitative results on the DRealSR dataset. We compare our OR-Net to the general SISR methods, including Bicubic, VDSR, EDSR, RDN, DDBPN, RCAN, and RealSR methods, including LP-KPN and CDC. We use PSNR, SSIM and LPIPS as evaluation metrics.

<table>
<thead>
<tr>
<th>Method</th>
<th>Category</th>
<th>Scale</th>
<th>DRealSR PSNR</th>
<th>DRealSR SSIM</th>
<th>DRealSR LPIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicubic</td>
<td></td>
<td></td>
<td>32.67</td>
<td>0.877</td>
<td>0.201</td>
</tr>
<tr>
<td>VDSR</td>
<td>SISR</td>
<td>x2</td>
<td>34.02</td>
<td>0.901</td>
<td>0.154</td>
</tr>
<tr>
<td>EDSR</td>
<td></td>
<td></td>
<td>34.24</td>
<td>0.908</td>
<td>0.155</td>
</tr>
<tr>
<td>RDN</td>
<td></td>
<td></td>
<td>34.46</td>
<td>0.910</td>
<td>0.151</td>
</tr>
<tr>
<td>DDBPN</td>
<td></td>
<td></td>
<td>34.26</td>
<td>0.906</td>
<td>0.157</td>
</tr>
<tr>
<td>RCAN</td>
<td></td>
<td></td>
<td>34.34</td>
<td>0.908</td>
<td>0.158</td>
</tr>
<tr>
<td>LP-KPN</td>
<td></td>
<td></td>
<td>33.88</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CDC</td>
<td>RealSR</td>
<td>x2</td>
<td>34.45</td>
<td>0.910</td>
<td>0.146</td>
</tr>
<tr>
<td>OR-Net(Ours)</td>
<td></td>
<td></td>
<td>34.56</td>
<td>0.910</td>
<td>0.145</td>
</tr>
</tbody>
</table>

Figure 4: Region-adaptive map synthesis in RFA module. * is matrix multiplication and $\odot$ is element multiplication.

content-aware super-resolution by combining the dynamic convolution and spatial attention mechanism.

To achieve the region-adaptive aggregation, a straightforward solution is to utilize spatial attention in (Woo et al. 2018) to fuse the low-/mid-/high-frequency components. However, the general spatial attention only leverages the spatial contextual information but lacks of considering the relationship among low-/mid-/high frequency features, since the attention map for each region is generated through the same convolution filter. Therefore, in our RFA module, to achieve better omni-frequency feature adaptive aggregation, we combine the dynamic omni-frequency feature (Zhao et al. 2018) and spatial attention (Woo et al. 2018) to achieve adaptive frequency aggregation. Specifically, we utilize dynamic convolution kernels to synthesize the region-adaptive attention map from omni-frequency features. With such attention map, we achieve flexible and accurate frequency component fusion for different image regions.

As shown in Figure 4, we first concatenate the enhanced low-/mid-/high-frequency $\bar{f}_l$, $\bar{f}_m$ and $\bar{f}_h$ to get $f_{omni}$:

$$f_{omni} = [(\bar{f}_l)^T, (\bar{f}_m)^T, \bar{f}_h]$$

where $[\cdot]$ represents the ‘concat’ operation. Then we set up a learnable basis kernel pool $K \in \mathbb{R}^{m \times n \times c \times k}$, where $m$, $n$, $c$, and $k$ represent the number of filters, input channel, output channel and kernel size (see Figure 4). After that, we pass $f_{omni}$ through an embedding layer to obtain a coefficient tensor $\alpha \in \mathbb{R}^{2 \times w \times m}$, where $h$ and $w$ represent the same height and width of omni-frequency $f_{omni}$. Finally, we re-weight $m$ filters in $K$ with $\alpha(i, j, :)$ for region $(i, j, :)$ to get region-adaptive dynamic convolution filter $\tilde{K}$:

$$\tilde{K} = \sum_{n=1}^{m} \alpha_n(i, j, :) \times K_n. \quad (4)$$

With dynamic convolution kernel $\tilde{K}$, we can get the region-adaptive attention map $A(i, j, :)$ for region $(i, j, :)$ as:

$$A(i, j, :) = f_{omni}(i, j, :) \ast \tilde{K}, \quad (5)$$

where $\ast$ represents the convolution operation. Finally, the aggregated omni-frequency feature $f$ can be obtained by:

$$f(i, j, :) = A(i, j, :) \bullet f_{omni}, \quad (6)$$

where $\bullet$ represent the dot multiplication.

Experiments

In this section, we first describe the datasets of RealSR and our implementation details in Section 4. And then, to verify the superiority of our method, we compare the proposed OR-Net with the current state-of-the-art RealSR methods and conventional SISR methods in Section 5. To validate the effectiveness of the proposed FD and RFA modules, we show the visualization analysis in Section 6 and present a series of ablation studies for OR-Net in Section 7.

Dataset and Implementation Details

RealSR is now under-explored and few works focus on such new challenge. Hence, these are few datasets can be used for
Evaluation. Here we evaluate our OR-Net on DRealSR (Wei et al. 2020). DRealSR dataset is collected by (Wei et al. 2020). The training dataset contains 35,065, 26,118, and 30,502 image patches for scales of ×2, ×3 and ×4, respectively. The size for patches of scale ×2, ×3 and ×4 are 380×380, 272×272 and 192×192. The testing dataset contains 83, 84, and 93 images for ×2~×4, respectively.

The implementation of OR-Net is based on PyTorch framework. In the training process, we utilize Adam optimizer with an initial learning rate of 0.0001 and the learning rate decay by a factor of 0.5 each epoch. Batch size is 8 and we leverage random flip, random rotation and random cropping to achieve data augmentation. We randomly crop the training image as 192×192. For FD module, we set the channels of three frequency branches as 128, 128 and 64 from low-frequency to high-frequency components. For RFA module, we set the number of basis kernels $K$ as 5.

$L_1$ loss has been verified effective and been widely used in many super-resolution works (Lim et al. 2017; Zhang et al. 2018a). In this paper, we also utilize the $L_1$ loss to optimize our OR-Net.

Comparison with State-of-the-Arts

We compare our OR-Net with the state-of-the-art traditional SISR models (including Bicubic, VDSR (Kim, Kwon Lee, and Mu Lee 2016), EDSR (Lim et al. 2017), RDN (Zhang et al. 2018b), DDBPN (Haris, Shakhnarovich, and Ukita 2018), RCAN (Zhang et al. 2018a)), and the RealSR models (including LP-KPN (Cai et al. 2019) and CDC (Wei et al. 2020)). As shown in Table. 1, our OR-Net achieves the best performance in terms of PSNR, SSIM and LPIPS compared to other general SISR and RealSR methods. We analyse that the general learning-based SISR methods that focus on solving synthetic degradation usually ignore the restoration of the full-frequency components of real LR, and cannot handle non-uniform distortions well. Besides, CDC and LP-KPN both ignore to leverage rich hierarchical information/features in frequency domains to comprehensively enhance texture details for real LR, which limits their practicality and scalability. In contrast, with the proposed frequency decomposition (FD) module and region-adaptive frequency aggregation (RFA) module, our OR-Net could enhance texture details for real LR images across all frequency scope while achieving content-aware super-resolution.

We provide the qualitative comparison of our OR-Net with state-of-the-art conventional SISR methods (including Bicubic, VDSR, EDSR, DBPN, and RCAN) and RealSR method (including CDC). We can see that those traditional SISR methods (including VDSR, EDSR, DBPN) fail to restore well some corrupted details of real LR, e.g., the letters
in the second rows. RealSR method CDC cannot effectively remove blur artifacts. However, our OR-Net can achieve better restoration of textures and details for all regions.

We also evaluate our OR-Net on RealSR dataset (Cai et al. 2019) and traditional SISR datasets in Supplementary.

**Visualization**

To study the influence of each module in our OR-Net, we visualize their inner features to understand how they work. For FD module, we first visualize the features for three frequency scales in Figure 6(a), we see that the feature in high-frequency branch contains more details and texture information. We then analyse the low-/mid-/high-frequency features according to wavelet transform in Figure 6(b). From left to right, with the frequency scale increases from 1 to 6, we find that the energy of low-frequency features $f_l$ is almost concentrated in the low frequency domain, and the energy of mid-/high-frequency features $f_m/f_h$ are concentrated in the relatively higher frequency domains, which validate the effectiveness of our frequency decomposition module.

For RFA module, we visualize the generated attention maps that used to adaptively aggregate frequency features in Figure 7. We see that the attention map $A_h$ for high-frequency features (Note that $A_h \in A$) tends to enhance the high-frequency details at the object edges, and the low-frequency attention map $A_l$ ($A_l \in A$) tends to select the low-frequency features at the flat region.

**Ablation Studies**

**Effectiveness of omni-frequency.** To study the influence of omni-frequency, we set the number of decomposed frequency branches in OR-Net as 1, 2, 3, 4 to get several schemes (denoted as bran. 1, 2, 3, 4). As shown in Table 2, we observe that the performance of OR-Net is improved as the number of frequency branch increases, but begins to decay when the number of branch exceeds 3. To balance the complexity and performance, we employ 3 frequency branches (corresponding to the low-/mid-/high-frequency components) in this paper.

**Effectiveness of our RFA module and FEU design.** As shown in Table. 3, the proposed RFA module could bring 0.25dB gain in PSNR, and the FEU design also could bring 0.36dB gain in PSNR. Moreover, when we replace our RFA module with a general spatial attention (SA) to implement frequency aggregation, we see that our RFA module outperforms SA by 0.13dB in PSNR.

**Conclusion**

In this paper, we propose an Omni-frequency Region-adaptive Network (OR-Net) to enable effective real image super-resolution. To efficiently promise the learned feature representations of OR-Net are both informative and content-aware, we first start from the frequency perspective and design a Frequency Decomposition (FD) module to fully leverage the omni-frequency features to comprehensively enhance texture details for LR images. Then, to adaptively aggregate such omni-frequency features to achieve the content-aware super-resolution, we further introduce a Region-adaptive Frequency Aggregation (RFA) module. Extensive experiments on several benchmarks demonstrate the superiority of OR-Net, and comprehensive ablation analysis verify the effectiveness of FD and RFA modules.
Acknowledgments

This work was supported in part by the National Key Research and Development Program of China 2018AAA0101400, and NSFC under Grant U1908209, 61632001.

References


