Spectral Bayesian Uncertainty for Image Super-resolution

Tao Liu, Jun Cheng, Shan Tan*
Huazhong University of Science and Technology, Wuhan, China
{hust_liutao, jcheng24, shantan}@hust.edu.cn

Abstract

*Recently deep learning techniques have significantly advanced image super-resolution (SR). Due to the black-box nature, quantifying reconstruction uncertainty is crucial when employing these deep SR networks. Previous approaches for SR uncertainty estimation mostly focus on capturing pixel-wise uncertainty in the spatial domain. SR uncertainty in the frequency domain which is highly related to image SR is seldom explored. In this paper, we propose to quantify spectral Bayesian uncertainty in image SR. To achieve this, a **D**ual-**D**omain **L**earning (DDL) framework is first proposed. Combined with Bayesian approaches, the DDL model is able to estimate spectral uncertainty accurately, enabling a reliability assessment for high frequencies reasoning from the frequency domain perspective. Extensive experiments under non-ideal premises are conducted and demonstrate the effectiveness of the proposed spectral uncertainty. Furthermore, we propose a novel **S**pectral **U**ncertainty based **D**ecoupled **F**requency (SUDF) training scheme for perceptual SR. Experimental results show the proposed SUDF can evidently boost perceptual quality of SR results without sacrificing much pixel accuracy.*

1. Introduction

Image super-resolution (SR) is a basic computer vision task that aims to recover an underlying high-resolution (HR) image from its degraded low-resolution (LR) observation. Image SR is widely used in many applications where high-frequency (HF) information is required, such as medical imaging [38], microscopy imaging [36], surveillance [46], etc. In recent years, learning-based approaches with convolutional neural networks (CNN) have become the primary workhorse for image SR. Starting from the pioneering work SRCNN [9], various CNN-based SR models [7,21,24,31,43,49] have been proposed and significantly pushed the frontier of image SR research.

Despite the impressive success in image SR benchmarks, most of these CNN-based SR models tend to overfit the training data so that their reliability and generalizability may not be guaranteed in practice. A well-trained SR model often makes inaccurate reasoning for HF details when it receives LR images away from its training distribution, thereby making the downstream processing unreliable. Therefore, it is quite crucial to quantify reconstruction uncertainty when employing these SR models, especially in some high risk applications (e.g. medical imaging) or when under some harmful adversarial attacks. Bayesian neural networks (BNNs) which combine deep neural networks with Bayesian learning open up the possibility to capture model uncertainty, by placing distributions over the network weights and then obtaining the predictive distribution through marginalization over posterior. Since the exact Bayesian inference is usually intractable for deep networks, various stochastic techniques that are compatible with modern deep learning are widely used for posterior approximation, such as dropout [11], batch normalization [41], weight initialization [22], etc.

However, existing Bayesian models for image SR are mostly developed in the spatial domain to capture pixel-wise uncertainty [40, 41]. The uncertainty in the frequency domain which is highly related to image SR is seldom explored. From the frequency domain perspective, image SR is essentially a task of recovering HF components given low-frequency (LF) ones. Thus the uncertainty of HF components directly characterizes the reliability of the SR results. Besides, the common pixel-wise uncertainty is sensitive to local mismatch of spatial structures, where a slight pixel shift among Monte Carlo (MC) samples may result in high uncertainty. So it is also desirable to quantify the reconstruction uncertainty in a global way. Moreover, image HF components in the frequency domain usually play an important role in some specific areas. For instance, the calculation of imaging resolution in optical imaging heavily depends on the HF components of objects [8]. The uncertainty of HF components directly reveals the credibility of the imaging resolution. Therefore, estimating frequency spectral uncertainty for image SR is valuable.

*Corresponding author.

To fill this research gap, we aims to quantify SR uncertainty not only in the spatial domain but also in the frequency domain. Concretely, we first propose a dual-domain learning (DDL) framework for image SR. The proposed DDL introduces explicit frequency learning within networks and learns to reconstruct SR images and spectra simultaneously. Then combined with Bayesian approaches (MC-dropout [11] in this paper), the DDL model is able to estimate both spatial and spectral uncertainty of SR results. To the best of our knowledge, we are the first to quantify SR uncertainty in the frequency domain. Extensive experiments on different non-ideal premises are conducted to show the effectiveness of the spectral uncertainty. Lastly, we further propose a spectral uncertainty based decoupled frequency (SUDF) training scheme for perceptual SR. The SUDF decouple the training of different image frequencies with the guidance of estimated spectral uncertainty map, thereby boosting perceptual quality of SR results significantly without sacrificing much pixel accuracy.

In summary, the contributions of this paper are:

- We propose to quantify the frequency spectral uncertainty for deep SR networks. Experiments under several non-ideal premises demonstrate the effectiveness. To the best of our knowledge, it is the first work to estimate SR uncertainty in the frequency domain.
- A DDL method is proposed for image SR. By performing explicit frequency domain learning in feature space, DDL can restore more HF information and thus provide more accurate uncertainty estimation when combined with Bayesian approaches.
- Based on the estimated spectral uncertainty, a novel SUDF training scheme is proposed, helping enhance perceptual quality of SR results while maintaining reconstruction faithfulness.

2. Related work

2.1. Image Super-resolution

Recently, image SR solutions have been dominated by learning-based methods with deep neural networks, which aims to learn general image priors automatically from given exemplar LR-HR pairs. Among these works, SRCNN [9] makes the first attempt to adopt CNN for image SR with only three convolution layers. Inspired by SRCNN, a variety of CNN architectures are developed to improve SR performance. These improvements primarily arise from increase of model depth [21], more flexible information flow [24, 49, 50], and various efficient attention techniques [7, 26, 31, 48, 49]. Another research line of image SR is to devise better loss functions. Pixel-wise L1 or L2 loss is typically used in most works to ensure accuracy in pixel domain [9, 21, 24, 26, 49]. However, such pixel-wise losses are demonstrated to produce over-smooth results owing to their limited ability in capturing perceptually relevant similarity [23]. To enhance the visual quality, some perception-oriented SR methods are also proposed, by introducing perceptual loss [19, 35] or adversarial loss [23, 33, 37, 43]. These perceptual-driven losses can help restore more fine details but also lead to much higher distortion.

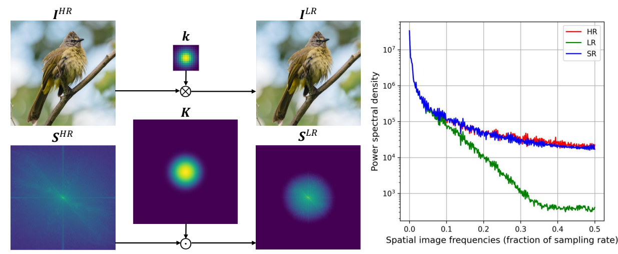

Figure 1. **Left**: The SR degradation visualization in both spatial and frequency domains. **Right**: The power spectral density of HR, LR and SR images. Image SR is to restore the high frequencies given low frequencies in LR images.

2.2. Applications of Frequency Domain Knowledge

Frequency domain knowledge has been widely applied in computer vision. CNNs can be understood in the frequency domain [32], and have been proved to be biased towards fitting low frequencies, i.e. the so-called F-principle or spectral bias [44]. To promote the ability in capturing frequency discrepancy, several studies attempt to introduce frequency domain knowledge to deep models, by designing frequency-based loss [10, 16, 18], or exploring information interaction between spatial and frequency domains [26, 27, 34]. For image SR, the training of SR networks can be viewed from the standpoint of frequency domain as an implicit conditional learning of HF components given LF ones [14]. So understanding the faithfulness of HF components is the core to assess credibility of SR results.

2.3. Uncertainty in Bayesian Deep Learning

Bayesian uncertainty has drawn much attention in recent years. BNN assign a prior distribution over the weights instead of deterministic weights as in non-Bayesian models. However, the optimization of BNNs is intractable since there is no conjugate prior postierior pairs for complex deep networks. Hence, approaches of approximate Bayesian inference are required to calculate posterior distribution of weights, such as variational inference [12, 20] and Markov Chain Monte Carlo [13]. Recently, some more efficient techniques for capturing model uncertainty are explored. For instance, Gal et al. [11] prove that applying dropout [39] in deep networks which utilizes Bernoulli variational distribution is mathematically equivalent to approximate varia-

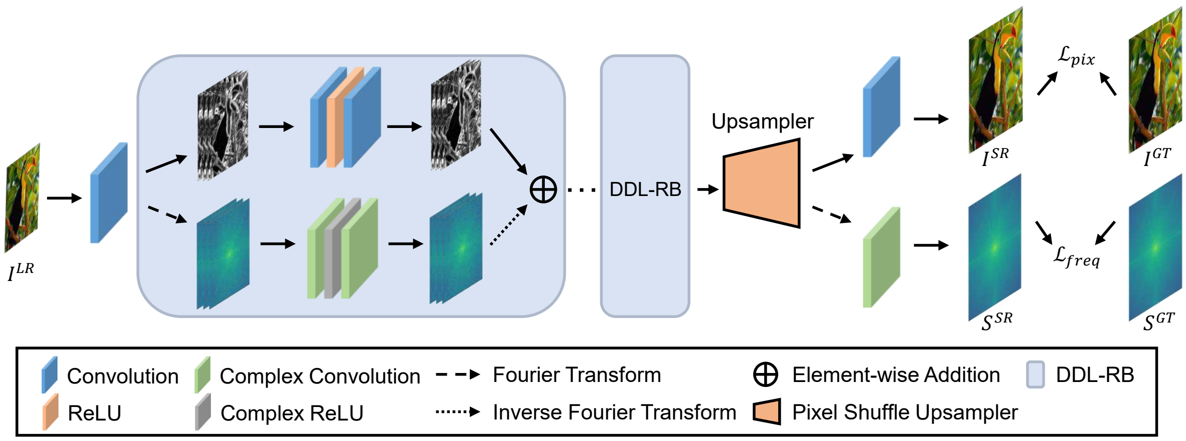

Figure 2. The architecture of the proposed DDL-EDSR.

tional inference in the deep Gaussian process. Likewise, other stochastic techniques like DropConnect [30], batch normalization [41], and weight initialization [22] are also widely used for quantifying Bayesian uncertainty. These methods are also widely applied for image reconstruction [2, 3, 15, 40, 41].

3. Methodology

3.1. Frequency Perspective of Image SR

In image SR, the degradation process is typically modeled as:

$$\boldsymbol{I}^{LR} = (\boldsymbol{I}^{HR} \otimes \boldsymbol{k}) \downarrow_s, \tag{1}$$

where the HR image $\boldsymbol{I}^{HR}$ is convolved with blur kernel $\boldsymbol{k}$, followed by a s-fold downsampler. Then the LR image $\boldsymbol{I}^{LR}$ is generated. Image SR aims to find an inverse mapping of the degradation process: $\mathcal{M}: \boldsymbol{I}^{LR} \rightarrow \boldsymbol{I}^{HR}$.

From the frequency domain perspective, Eq. (1) can be re-written as:

$$\boldsymbol{S}^{LR} = \mathcal{F}(\boldsymbol{I}^{LR}) = \sum_{n_\mu}\sum_{n_\upsilon}[\boldsymbol{S}^{HR} \cdot \boldsymbol{K}](\mu - n_\mu \mu_s, \upsilon - n_\upsilon \upsilon_s), \tag{2}$$

where $\mathcal{F}$ denotes Fourier transform (FT). $\boldsymbol{S}$ and $\boldsymbol{K}$ denote the frequency spectra of $\boldsymbol{I}$ and $\boldsymbol{k}$, respectively. (u, v) are the coordinates of the frequency domain and μ_s and υ_s are the sampling rates along these two dimensions. To avoid multiple replicas of $\boldsymbol{S}^{HR}$ overlapping their HF components (i.e. the so-called aliasing), the $\boldsymbol{K}$ are typically modeled as low-pass filters to attenuate HF in $\boldsymbol{S}^{HR}$. Hence, only low frequencies are preserved in the $\boldsymbol{S}^{LR}$. Assuming $\boldsymbol{K}$ an ideal low-pass filter, the frequency understanding of SR is essentially an implicit conditional learning of erased HF content according to the remained LF information [14]. An example of using common isotropic Gaussian filter as $\boldsymbol{k}$ is shown in Fig. 1. One can see HF components are attenuated during degradation process and then well restored in SR images.

3.2. Dual-domain Learning for Image SR

Existing SR networks are primarily developed in the spatial domain, where the recovery of HF components is taking place in an implicit manner. Inspired by [27], we propose a DDL method, which combines complex CNN layers [42] with the existing SR models, thereby achieving explicit frequency domain learning within models. Below, We first briefly introduce two basic complex-valued layers [42] (i.e. complex conlution and activation) and then present the DDL method.

Complex layers. Complex layers is used for dealing with complex-valued signals, which treat the real and imaginary part of a complex number as logically distinct real-valued entities and then achieve complex arithmetic through real-valued arithmetic [42]. Given complex-valued input feature $F = F^{real} + iF^{imag}$ where $real$ and $imag$ respectively represent the real and imaginary part, complex convolution adopts a complex filter $w = w^{real} + iw^{imag}$ and convolve with F in the form of:

$$\begin{bmatrix} F^{real}_{out} \\ F^{imag}_{out} \end{bmatrix} = \begin{bmatrix} w^{real} & -w^{imag} \\ w^{imag} & w^{real} \end{bmatrix} * \begin{bmatrix} F^{real} \\ F^{imag} \end{bmatrix}, \tag{3}$$

in which $F_{out} = F^{real}_{out} + iF^{imag}_{out}$ is the output feature. For complex activation, we use $\mathbb{C}$ReLU [42] which applies ReLU on the real and imaginary parts separately:

$$\mathbb{C}ReLU(F) = ReLU(F^{real}) + iReLU(F^{imag}). \tag{4}$$

DDL. We make use of above complex layers to extend the existing models to DDL models. In this paper, we utilize the classic EDSR [24] (the baseline version) as a base architecture and term its DDL derivative as DDL-EDSR. As displayed in Fig. 2, there are two modifications in DDL-EDSR: 1) all ResBlocks (RB) in EDSR are replaced by our proposed DDL-RB. Beside a normal spatial branch as in RB, DDL-RB adds an parallel frequency branch which starts with a FT and then employs two complex convolution layers and one ℂReLU in between. Lastly the output features of frequency branch are transformed back to spatial domain and added with those of spatial branch. 2) In the tail of the model, DDL-EDSR adopts two heads to output SR images and SR spectra simultaneously. For training of DDL-EDSR, dual-domain restrictions are imposed:

$$\mathcal{L}_{DDL} = \underbrace{\|\boldsymbol{I}^{SR} - \boldsymbol{I}^{GT}\|_1}_{\mathcal{L}_{pix}} + \lambda \underbrace{\|\boldsymbol{S}^{SR} - \boldsymbol{S}^{GT}\|_1}_{\mathcal{L}_{freq}}, \tag{5}$$

where the superscript "SR" and "GT" denote the image or spectra of SR results and ground truth, respectively. λ is a balancing parameter and fixed as 0.01. Note that the mean of the two heads is taken as our final result.

3.3. Spectral Bayesian Uncertainty Estimation

In this section, we further extend the DDL-EDSR to a BNN model. Given training data D, our goal is to find a Bayesian model $G(\boldsymbol{I}^{\boldsymbol{LR}}; \Theta) : \boldsymbol{I}^{LR} \rightarrow \boldsymbol{I}^{SR}, \boldsymbol{S}^{SR}$, where parameter Θ follows posterior distribution $p(\Theta|D)$. Then for a new input $\boldsymbol{I}^{LR*}$, the predictive distribution of $\boldsymbol{I}^{SR*}$ and $\boldsymbol{S}^{SR*}$ can be obtained by integrating

$$\begin{aligned} p(\boldsymbol{I}^{SR*}|\boldsymbol{I}^{LR*}, D) &= \int_{\Theta} p(\boldsymbol{I}^{SR*}|\boldsymbol{I}^{LR*}, \Theta) p(\Theta|D) d\Theta, \\ p(\boldsymbol{S}^{SR*}|\boldsymbol{I}^{LR*}, D) &= \int_{\Theta} p(\boldsymbol{S}^{SR*}|\boldsymbol{I}^{LR*}, \Theta) p(\Theta|D) d\Theta. \end{aligned} \tag{6}$$

However, Eq. (6) is intractable since no conjugate prior postierior pairs exist for deep networks so that approximations are typically required to achieve Bayesian posterior inference. In this paper, we utilize MC-dropout [11] for its simplicity. We open dropout in both training and inference phases. In this way, MC samples of $\boldsymbol{I}^{SR*}$ and $\boldsymbol{S}^{SR*}$ can be generated through multiple stochastic forward passes: $\left\{\boldsymbol{I}_1^{SR*}, ..., \boldsymbol{I}_T^{SR*}\right\}, \left\{\boldsymbol{S}_1^{SR*}, ..., \boldsymbol{S}_T^{SR*}\right\}$. Then the Bayesian uncertainty can be induced by measuring the prediction dispersion of these MC samples. For uncertainty in the spatial domain, pixel-wise variance is typically used as the uncertainty. However, using variance becomes inappropriate for quantifying frequency-wise uncertainty since the dynamic range of different frequencies varies a lot. In this paper, we propose that the coefficient of Variation (CV) is a proper statistic for measuring spectral uncertainty:

$$U_{\boldsymbol{S}} = \frac{\sum_{t=1}^{T}\left(\boldsymbol{S}_t^{SR*}\right)^2 - \left(\sum_{t=1}^{T}\boldsymbol{S}_t^{SR*}\right)^2}{\sum_{t=1}^{T}\boldsymbol{S}_t^{SR*}}, \tag{7}$$

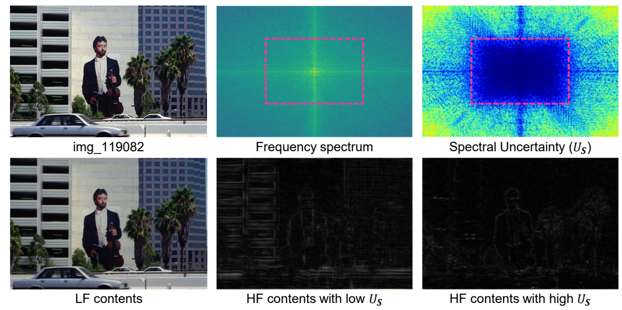

Figure 3. Identify the corresponding image contents of LF components, HF components with low spectral uncertainty, and HF components with high spectral uncertainty. The red dashed rectangle denotes the frequency support of the ideal bicubic kernel.

where $U_{\boldsymbol{S}}$ is the quantified spectral uncertainty of SR results.

3.4. Spectral Uncertainty based Perceptual SR

For SR networks training, loss function is a pivotal ingredient that significantly affects SR performance. The commonly-used loss functions can be mainly classified into two categories: PSNR-oriented (e.g. pixel-wise L1, L2) and perceptual-driven (e.g. adversarial loss). Training SR networks with the former can obtain results with high PSNR value but poor perceptual quality, since such losses are apt to learning LF components but struggle in capturing similarity of HF information. On the other hand, perceptual-driven losses can help generate results with rich HF details, but leads to much higher image distortion. Considering that different image frequencies encode different image contents and thus have different effects on SR results, an intuitive idea is to distinguish different image frequencies and adopt different loss functions to guide their training towards their suitable objectives. So how to distinguish different image frequencies properly becomes the key issue.

In this paper we find the spectral uncertainty could be a good indicator to separate image frequencies. As shown in Fig. 3, we separate frequencies into three parts: LF components, HF components with low $U_{\boldsymbol{S}}$, and HF components with high $U_{\boldsymbol{S}}$. The LF components are well preserved during degradation and thus can be restored in a well-posed way. HF components with low $U_{\boldsymbol{S}}$ correspond to the simple or periodic structures (e.g. building) while the HF ones with high $U_{\boldsymbol{S}}$ encode more complex textures (e.g. tree). We find the former highly relies on context information and can also be well resolved by PSNR-oriented losses. In contrast, the restoration of the latter one is more difficult and needs to resort to perceptual-driven methods. To conclude, the learning of frequencies with low $U_{\boldsymbol{S}}$ (the former two parts in Fig. 3) can be guided by PSNR-oriented losses and training with perceptual-driven losses is a better option for

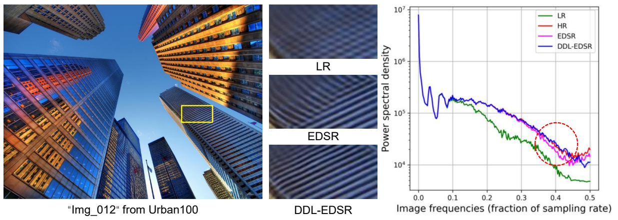

Figure 4. Evaluation of the proposed DDL framework based on EDSR model [24]. **Left**: The qualitative comparison between DDL-EDSR and EDSR (×4). **Right**: The power spectral density of SR results are shown. The example image is from Urban100 [17].

frequencies with high $U_{\boldsymbol{S}}$ (the latter one part in Fig. 3).

Based on above considerations, we propose a **S**pectral **U**ncertainty guided **D**ecoupled **F**requency (SUDF) training for perceptual SR. SUDF is a three-step method. In the first step, we train a Bayesian DDL-EDSR model to get the spectral uncertainty map $U_{\boldsymbol{S}}$ given input LR images. To aviod discontinuity, we smooth the $U_{\boldsymbol{S}}$ by a gaussian filter. The result is served as a frequency mask (denoted as M) in the latter steps. In the second step, a PSNR-oriented SR model is trained with $\mathcal{L}_{DDL}$ defined in Eq. (5). This model is denoted as G_{PSNR} parameterized by Θ_{PSNR}, which can provide accurate restoration for frequencies with low $U_{\boldsymbol{S}}$. In the third step, we employ the well-trained G_{PSNR} as an initialization and obtain another GAN-based SR model (denoted as G_{GAN} parameterized by Θ_{GAN}) by fine-tuning. Note that G_{GAN} is responsible to perform perceptual learning for high frequencies with high $U_{\boldsymbol{S}}$. In order to improve reconstruction faithfulness of G_{GAN}, another term that directly measures discrepancies of corresponding frequencies are also exerted. That is:

$$\begin{aligned}\mathcal{L}_{GAN} = \mathcal{L}_{adv}(\mathcal{F}^{-1}(\boldsymbol{S}^{SR} \odot M), \mathcal{F}^{-1}(\boldsymbol{S}^{GT} \odot M))\\ +\gamma\|M \odot (\boldsymbol{S}^{SR} - \boldsymbol{S}^{GT})\|,\end{aligned} \tag{8}$$

where $\mathcal{F}^{-1}$ is the inverse FT (iFT). $\mathcal{L}_{adv}$ is the loss provided by relativistic discriminator as in [43]. γ is set to 50.

In model inference, the final SR result is obatined by frequency spectrum fusion between results of G_{PSNR} and G_{GAN}:

$$\begin{aligned}\boldsymbol{S}^{SR} = G_{PSNR}(\boldsymbol{I}^{LR}, \Theta_{PSNR}) \cdot (1 - M)\\ +G_{GAN}(\boldsymbol{I}^{LR}, \Theta_{GAN}) \cdot M.\end{aligned} \tag{9}$$

This way, the advantage of both PSNR-oriented and perceptual-driven methods are inherited. The spatial results can be obtained easily by iFT: $\boldsymbol{I}^{SR} = \mathcal{F}^{-1}(\boldsymbol{S}^{SR})$.

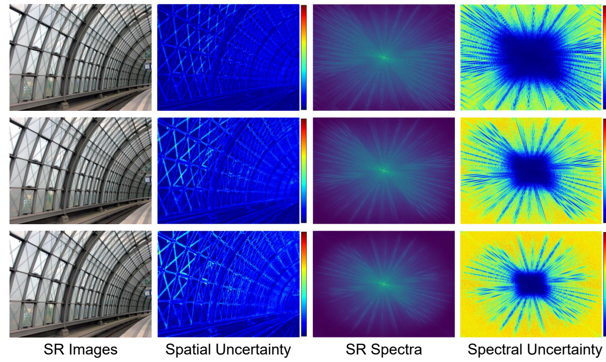

Figure 5. Visualizations of the reconstructed results and the corresponding uncertainty in both spatial and frequency domain. From top to bottom: ×2, ×3, and ×4 SR. The example image is from Urban100 [17].

4. Experiments

4.1. Experimental Settings

Datasets and Evaluation. Following prior arts [7, 24, 49], we use 800 training images of DIV2K [1] as the training set. LR images are obtained by downsampling HR images using MATLAB bicubic kernel. For testing, five standard benchmark datasets including Set5 [4], Set14 [45], B100 [28], Urban100 [17], and Manga109 [29] are used. As for evulation metric, the SR results are evaluated by PSNR and SSIM on Y channel of image YCbCr space. LPIPS [47] metric is also reported when involving perceptual SR.

Implementation details. In this paper, we choose the classic EDSR [24] (baseline version) as the base model. In DDL-EDSR, all complex convolution layers adopt 1×1 kernel to introduce negligible extra parameters. We use DDL-EDSR to analyze subsequent spectral uncertainty and SUDF training scheme. For MC-dropout, we place dropout with $p = 10\%$ after each DDL-RB. In testing phase, we use 40 MC samples.

To train our models, a batch of 16 LR images of size 48 × 48 are randomly cropped as model input. The training patches are further augmented by random horizontal flips and 90° rotations. Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ is employed for training. Learning rate is set to 1×10^{-4} initially and decays with a factor of 0.5 every 2×10^5 iterations of back-propagation. All our experiments are conducted on a server equipped with four NVIDIA RTX 2080Ti GPUs.

4.2. Evaluation of DDL models

We first demonstrate the effectiveness of the proposed DDL method for image SR. The classic EDSR [24] is chosen as the base model which is then extended to DDL-EDSR. The quantitative comparison between EDSR and DDL-EDSR for ×2, ×3, and ×4 SR are listed in Tab. 1.

Table 1. Quantitative evaluation of the proposed DDL-EDSR.

Model	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EDSR	$\times 2$	37.914	0.9602	33.512	0.9172	32.125	0.8995	31.881	0.9263	38.362	0.9766
DDL-EDSR		38.048	0.9607	33.754	0.9188	32.209	0.9003	32.412	0.9308	38.897	0.9775
EDSR	$\times 3$	34.288	0.9261	30.278	0.8414	29.055	0.8044	27.988	0.8493	33.412	0.9434
DDL-EDSR		34.500	0.9277	30.453	0.8447	29.159	0.8072	28.454	0.8574	33.898	0.9460
EDSR	$\times 4$	32.036	0.8920	28.540	0.7811	27.539	0.7357	25.965	0.7825	30.339	0.9066
DDL-EDSR		32.302	0.8951	28.703	0.7845	27.644	0.7388	26.342	0.7921	30.764	0.9100

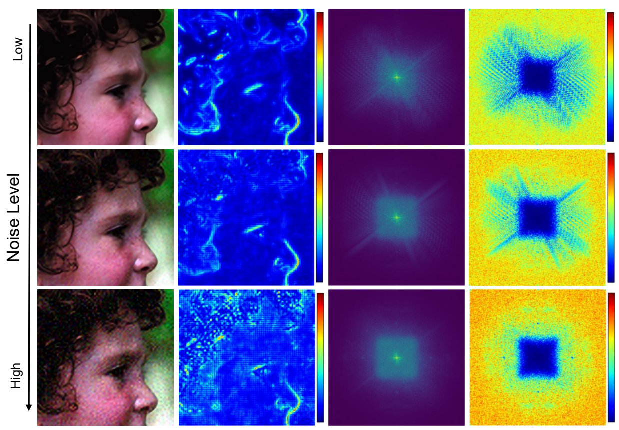

Figure 6. Dual-domain $\times 4$ SR results and the corresponding uncertainty estimation when input LR images are contaminated by Gaussian noise of different variances. Noise variance from top to bottom: 0, 5, 10. The example image is from Set5 [4].

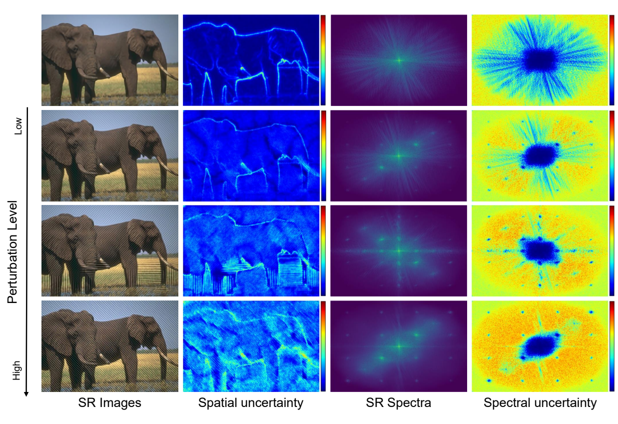

Figure 7. Dual-domain $\times 4$ SR results and corresponding uncertainty under adversarial attacks of different perturbation levels. From top to bottom: the perturbation level is 0/255 (without attack), 1/255, 2/255, 3/255. The example image is from B100 [28].

It can be observed that DDL-EDSR achieves remarkable performance gains across scales and benchmarks, indicating the promising potential of utilizing explicit frequency domain learning for image SR. The visual comparison of an example of $\times 4$ SR from Urban100 is presented in Fig. 4. One can see that the DDL-EDSR can produce SR results with more faithful HF structures. The power spectral density visualizations clearly show the DDL method can help models to restore more HF components (highlighted by the red dashed circle). Note that the proposed DDL method is general to various CNN backbones. We also conduct experiments based on RCAN [49], which can be seen in Appendix. Due to the greater ability in terms of HF restoration, DDL-EDSR can estimate more accurate spectral uncertainty when combined with Bayesian approaches.

4.3. Analysis of Spectral Uncertainty

4.3.1 Spectral Uncertainty in SR

By combining MC-dropout [11] with our DDL-EDSR for MC samples generation, spatial and spectral uncertainty can be obtained. In this paper we focus on the latter one. Fig. 5 displays the spatial and frequency visualizations of SR results and the corresponding uncertainty. It is clear that frequencies with low spectral uncertainty are primarily situated in LF region since LF information is well preserved in LR images and easy to recover. In HF region, there are HF components with low uncertainty and others with high uncertainty. In the earlier section, we have identified their corresponding spatial contents in Fig. 3, i.e. the HF components with low spectral uncertainty mostly correspond to simple structures while the ones with high spectral uncertainty primarily encode complex structures or textures. Besides, the spectral uncertainty increases with the increase of SR scale factor, similar as the behavior of the spatial uncertainty.

4.3.2 Spectral Uncertainty for Input Noise

Most existing SR models are trained under ideal noise-free condition, thereby lacking the ability in dealing with input noise. Condidering noise is a common degradation in real-world applications, identifying the model limitation in

Table 2. Quantitative evaluation of the proposed SUDF training for $\times 4$ SR. The results are based on DDL-EDSR model.

Scale	Loss	Set5		Set14		B100		Urban100		Manga109	
		PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓
$\times 4$	$\mathcal{L}_{DDL}$	32.302	0.1377	28.703	0.2336	27.644	0.3122	26.342	0.1899	30.764	0.0824
	$+\mathcal{L}_{adv}$	30.177	0.0768	26.271	0.1413	25.103	0.1661	24.110	0.1273	27.673	0.0688
	SUDF	31.883	0.0995	28.243	0.1574	27.232	0.2093	26.032	0.1375	30.217	0.0622

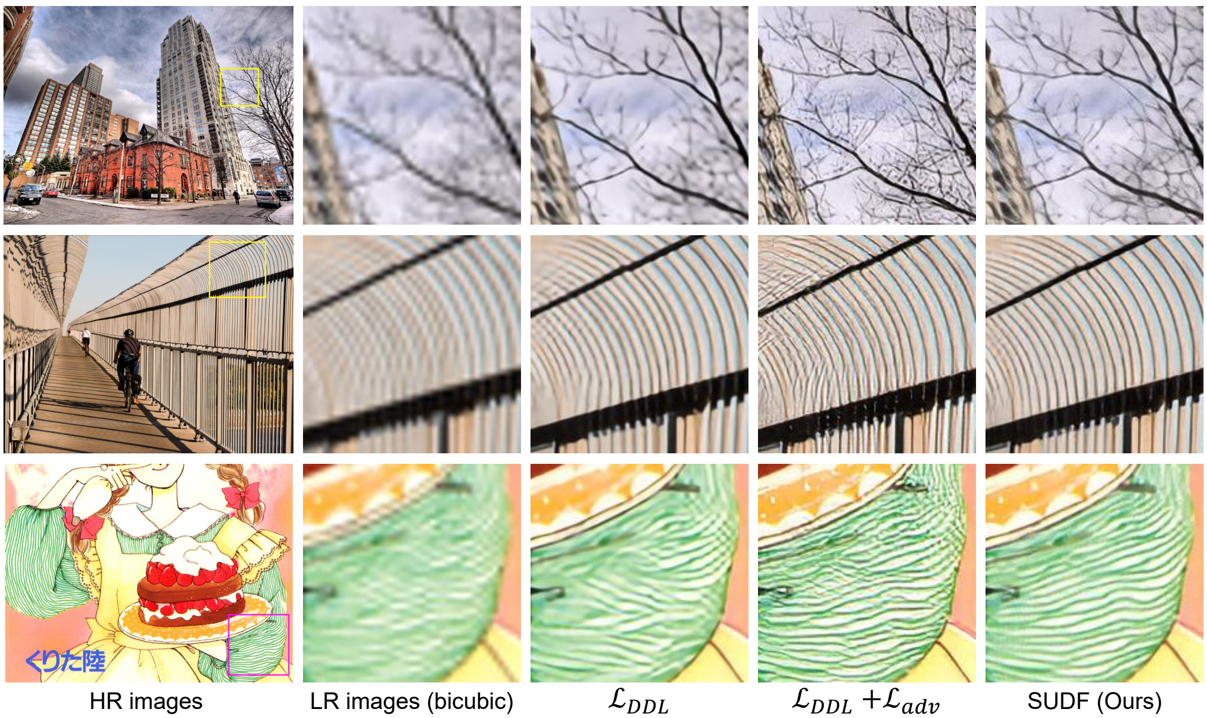

Figure 8. Visual evaluation of the proposed SUDF results for $\times 4$ SR. The example images from top to bottom are: img20 from Urban100, img024 from Urban100 [17], and YumeiroCooking from Manga109 [29].

terms of dealing with input noise is critical in practice. In this section, we investigate the effect of input noise, for not only spatial uncertainty, but also spectral uncertainty. Fig. 6 illustrates the SR results and uncertainty in dual domains, with escalating noise levels. Gaussian white noise of different variances (i.e. $\sigma = 0, 5, 10$) are added in input LR images for its simplicity. As can be seen, SR uncertainty in both domains increases with the increase of noise level. However, when input noise is not very obvious (e.g. the case where variance is 5 in Fig. 6), spatial pixel-wise uncertainty cannot well reflect the SR performance deteriorate. In contrast, spectral uncertainty is more sensitive, where fewer high frequencies with high certainty are inferred, indicating the advantage of spectral uncertainty over the common spatial uncertainty in this scenario. We suggest that spatial and spectral uncertainty can be complementary to measure the reliability of SR results locally and globally.

4.3.3 Spectral Uncertainty under Attacks

Robustness of deep SR networks against adversarial attacks is a key issue in practice. Unfortunately, previous studies [6, 41] have shown the vulnerability of existing SR models. Well-trained SR networks could be confused by a very slight perturbation in the input LR image, and produce unpleasant artifacts in SR results. Hence, it is desirable to adopt Bayesian models for characterizing imperfections caused by these harmful adversarial attacks.

In this section, we apply PGD algorithm [25] to perturb LR images with different levels (the maximum perturbation pixel intensity is 0, 1/255, 2/255, and 3/255). The implementation details can refer to Appendix. SR results and corresponding uncertainty are shown in Fig. 7. Under adversarial attacks, plenty of fake fringe patterns arise in the reconstructed SR images, which corresponds to the impulse-like regions in SR spectra. As seen in Fig. 7, such fake HF rea-

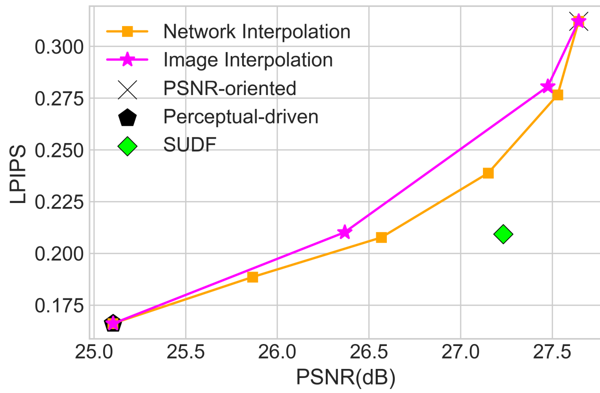

Figure 9. Comparison of perception-distortion trade-off between SUDF and other methods on B100 [28].

soning of adversarial LR images can be detected by both spatial uncertainty and our proposed spectral uncertainty. But when perturbation level is relative small, spectral uncertainty map could be the better indicator that help to be aware of these failure cases caused by attacks. Besides, we can witness the spectral uncertainty increase with the increase of perturbation level, especially in HF regions. More experimental results about spectral uncertainty quantification can be seen in Appendix.

4.4. Results with SUDF Training

4.4.1 Comparison with Common Losses

In this section, we compare the results of our proposed SUDF method with two commonly-used losses, one of which is the PSNR-oriented $\mathcal{L}_{DDL}$ and the other is the perceptual-driven $\mathcal{L}_{DDL}+\mathcal{L}_{adv}$. Tab. 2 shows the quantitative comparison for $\times 4$ SR. Compared with PSNR-oriented $\mathcal{L}_{DDL}$ loss, the LPIPS metrics of our SUDF results are significantly reduced, implying a much better perceptual quality. The PSNR is slightly lower since our SUDF relaxes the training and allows the mismatch of complex textures between SR results and ground truth. Compared with the GAN-based model, SUDF is able to yield very close LPIPS but much higher PSNR. The visual results are displayed in Fig. 8. We observe that the results of PSNR-oriented $\mathcal{L}_{DDL}$ are blurry and perceptual unpleasant. GAN-based perceptual SR help infer more fine-grained details but suffer from reconstruction distortion artifacts. In contrast, our SUDF can recover more faithful HF details and alleviate distortion artifacts.

4.4.2 Comparison with Other Trade-off Methods

Previous work [5] has shown that image perceptual quality and distortion are at odds with each other in image restoration tasks including image SR. A simple method for balancing the perception-distortion trade-off is to train a PSNR-oriented network and obtain another perceptual-driven one by fine-tuning, then interpolate the SR results of the two model in the pixel domain. Wang et al. [43] proposes another alternative to the trade-off by directly interpolating all the corresponding parameters of the two models. In this part, we demonstrate our proposed SUDF training scheme can help approach a better perception-distortion trader-off. We use the two competing models in Tab. 2 as the PSNR-oriented and perceptual-driven models, and draw the perception-distortion trade-off curves by employing pixel interpolation and network interpolation, respectively. Experiments are conducted on B100 dataset. As presented in Fig. 9, both image interpolation and network interpolation methods achieve a compromise between the contradictory image distortion (measured by PSNR) and perceptual quality (measured by LPIPS). Our SUDF is beyond the two trade-off curves, indicating a better perception-distortion trade-off performance for image SR.

5. Conclusion

In this paper, we propose a DDL method for image SR which enables the quantification of spectral uncertainty in the frequency domain when combined with Bayesian approaches. Our experiments show that the spectral uncertainty can characterize the reliability of image HF components in a global way. Image SR under several nonideal input LR premises demonstrate the better sensitiveness of spectral uncertainty against noise and adversarial attacks. Furthermore, we treat the spectral uncertainty map as a indicator for distinguishing frequencies that encode different image contents, and then propose SUDF training scheme for perceptual SR. Experimental results reveal the SUDF method can evidently enhance image perceptual quality while maintaining excellent faithfulness. The proposed spectral uncertainty is a valuable supplement to commonly-used pixel-wise uncertainty. We hope our work can enlighten researchers to explore the potential of reconstruction uncertainty in other domains.

6. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NNSFC), under Grant Nos. 61672253 and 62071197.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1122–1131, 2017. 5

[2] Riccardo Barbano, Chen Zhang, Simon Robert Arridge, and Bangti Jin. Quantifying model uncertainty in inverse prob-

lems via bayesian deep gradient descent. *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 1392–1399, 2021. 3

[3] Riccardo Barbano, Željko Kereta, Chen Zhang, Andreas Hauptmann, Simon Robert Arridge, and Bangti Jin. Quantifying sources of uncertainty in deep learning-based image reconstruction. *ArXiv*, abs/2011.08413, 2020. 3

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012. 5, 6

[5] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *CVPR*, 2018. 8

[6] Jun-Ho Choi, Huan Zhang, Jun-Hyuk Kim, Cho-Jui Hsieh, and Jong-Seok Lee. Evaluating robustness of deep image super-resolution against adversarial attacks. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 303–311, 2019. 7

[7] Tao Dai, Jianrui Cai, Yongbing Zhang, Shutao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11057–11066, 2019. 1, 2, 5

[8] Adrien C. Descloux, Kristin S. Grussmayer, and Aleksandra Radenović. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods*, 16:918–924, 2019. 1

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:295–307, 2016. 1, 2

[10] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2340–2349, 2021. 2

[11] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *ArXiv*, abs/1506.02142, 2016. 1, 2, 4, 6

[12] Alex Graves. Practical variational inference for neural networks. In *NIPS*, 2011. 2

[13] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970. 2

[14] Majed El Helou, Ruofan Zhou, and Sabine Süsstrunk. Stochastic frequency masking to improve super-resolution and denoising networks. In *ECCV*, 2020. 2, 3

[15] Ming Hong, Jianzhuang Liu, Cuihua Li, and Yanyun Qu. Uncertainty-driven dehazing network. In *AAAI*, 2022. 3

[16] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1698–1706, 2017. 2

[17] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. 5, 7

[18] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13899–13909, 2021. 2

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *ArXiv*, abs/1603.08155, 2016. 2

[20] Michael I. Jordan, Zoubin Ghahramani, T. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233, 2004. 2

[21] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016. 1, 2

[22] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS*, 2017. 1, 3

[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. 2

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. 1, 2, 3, 5

[25] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *ArXiv*, abs/1706.06083, 2018. 7

[26] Salma Abdel Magid, Yulun Zhang, D. Wei, Won-Dong Jang, Zudi Lin, Yun Raymond Fu, and Hanspeter Pfister. Dynamic high-pass filtering and multi-spectral attention for image super-resolution. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4268–4277, 2021. 2

[27] Xintian Mao, Yiming Liu, Wei Shen, Qingli Li, and Yan Wang. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745*, 2021. 2, 3

[28] David R. Martin, Charless C. Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2:416–423 vol.2, 2001. 5, 6, 8

[29] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, T. Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2016. 5, 7

[30] Aryan Mobiny, Hien Van Nguyen, Supratik Moulik, Naveen Garg, and Carol C Wu. Dropconnect is effective in modeling uncertainty of bayesian deep networks. *Scientific Reports*, 11, 2021. 3

[31] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *ECCV*, 2020. 1, 2

[32] Hengyue Pan, Yixin Chen, Xin Niu, and Wenbo Zhou. Learning convolutional neural networks in the frequency domain. *ArXiv*, abs/2204.06718, 2022. 2

[33] Seong-Jin Park, Hyeongseok Son, Sunghyun Cho, K. S. Hong, and Seungyong Lee. Srfeat: Single image super-resolution with feature discrimination. In *ECCV*, 2018. 2

[34] Zequn Qin, Pengyi Zhang, Fei Wu, and Xi Li. Fcanet: Frequency channel attention networks. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 763–772, 2021. 2

[35] Mohammad Saeed Rad, Behzad Bozorgtabar, Urs-Viktor Marti, Max Basler, Hazim Kemal Ekenel, and Jean-Philippe Thiran. Srobb: Targeted perceptual loss for single image super-resolution. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2710–2719, 2019. 2

[36] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017. 1

[37] Mehdi S. M. Sajjadi, Bernhard Schölkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4501–4510, 2017. 2

[38] Wenzhe Shi, Jose Caballero, Christian Ledig, Xiahai Zhuang, Wenjia Bai, Kanwal K. Bhatia, Antonio de Marvao, Timothy J. W. Dawes, Declan P. O'Regan, and Daniel Rueckert. Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16 Pt 3:9–16, 2013. 1

[39] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1929–1958, 2014. 2

[40] Ryutaro Tanno, Daniel E. Worrall, Aurobrata Ghosh, Enrico Kaden, Stamatios N. Sotiropoulos, Antonio Criminisi, and Daniel C. Alexander. Bayesian image quality transfer with cnns: Exploring uncertainty in dmri super-resolution. *ArXiv*, abs/1705.00664, 2017. 1, 3

[41] Mattias Teye, Hossein Azizpour, and Kevin Smith. Bayesian uncertainty estimation for batch normalized deep networks. In *ICML*, 2018. 1, 3, 7

[42] Chiheb Trabelsi, Olexa Bilaniuk, Dmitriy Serdyuk, Sandeep Subramanian, João Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher Joseph Pal. Deep complex networks. *ArXiv*, abs/1705.09792, 2018. 3

[43] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCV Workshops*, 2018. 1, 2, 5, 8

[44] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yan Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *ArXiv*, abs/1901.06523, 2020. 2

[45] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, 2010. 5

[46] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859, 2010. 1

[47] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 5

[48] Xindong Zhang, Huiyu Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. *ArXiv*, abs/2203.06697, 2022. 2

[49] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Raymond Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 1, 2, 5, 6

[50] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Raymond Fu. Residual dense network for image super-resolution. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 2