Weighted Multi-Kernel Prediction Network for Burst Image Super-Resolution

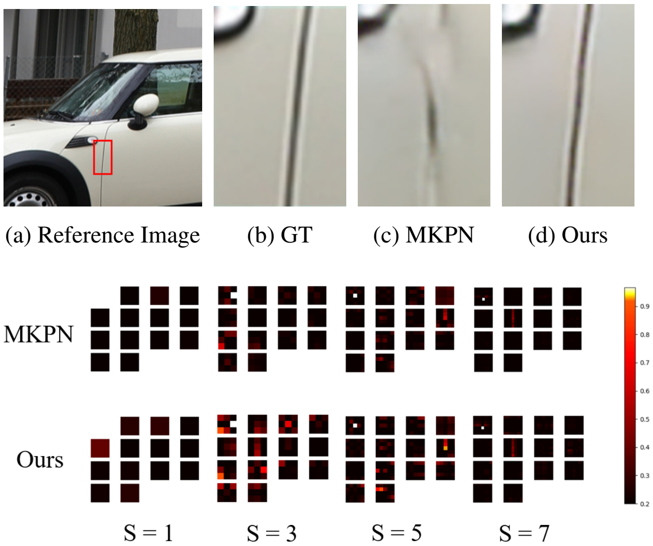
Wooyeong Cho Sanghyeok Son Dae-Shik Kim
KAIST
{chowy333, ssh816, daeshik} @kaist.ac.kr

Abstract

Burst image super-resolution is an ill-posed problem that aims to restore a high-resolution (HR) image from a sequence of low-resolution (LR) burst images. To restore a photo-realistic HR image using their abundant information, it is essential to align each burst of frames containing random hand-held motion. Some kernel prediction networks (KPNs) that are operated without external motion compensation such as optical flow estimation have been applied to burst image processing as implicit image alignment modules. However, the existing methods do not consider the interdependencies among the kernels of different sizes that have a significant effect on each pixel. In this paper, we propose a novel weighted multi-kernel prediction network (WMKPN) that can learn the discriminative features on each pixel for burst image super-resolution. Our experimental results demonstrate that WMKPN improves the visual quality of super-resolved images. To the best of our knowledge, it outperforms the state-of-the-art within kernel prediction methods and multiple frame super-resolution (MFSR) on both the Zurich RAW to RGB and BurstSR datasets.

1. Introduction

Recently, burst image processing algorithms [3, 1, 21, 10, 25, 29, 30, 48, 42] have been introduced to be utilized for various applications in computer vision tasks. Burst images captured with the fast shutter speed of the cameras are generally used to express the continuous motion of the objects. The image devices such as mobile phone cameras for capturing burst images become more portable and smaller, the spatial resolution of the image sensor also goes limited. Burst image super-resolution [9, 6, 3, 2] attempts to overcome this limitation by estimating HR images in certain applications such as surveillance [33, 46, 23] and computational photography [28, 45]. It takes advantage of being able to utilize abundant information from multiple frames, but it is necessary to consider that each multi-frames contains shot noise from short exposure of the camera and random hand-held motion [43] among a sequence of burst images.



(a) Reference Image (b) GT (c) MKPN (d) Ours

(e) Heatmaps of multiple kernels on a sequence of burst frames

Figure 1: (*top*) Qualitative results of MKPN [29] and our model. Weighted multi-kernel prediction helps our model to restore the high-frequency details compared to [29]. (*bottom*) Visualization of activated values for different sizes of kernels $S \in \{1, 3, 5, 7\}$ (*each column*) on the center pixel of the red box in (a) using [29] and ours, respectively (*first row and second row*). We find that several kernel values, which are implicitly important for image restoration, are emphasized with higher activation values.

The recent advances of Multi-Frame Super-Resolution (MFSR) make it possible to utilize abundant information of multiple images through several alignment modules. [26, 4, 38, 40, 3] show outstanding performance with the methods based on optical flow estimation as image alignment modules. Despite these approaches, the potential problem which can lead to artifacts for image restoration still exists in real-world applications [16, 39, 5] due to the errors in estimated optical flow. In contrast, the kernel prediction network (KPN) [30] alleviates this issue by utilizing the implicit alignment module for burst image processing.

KPN predicts the kernels existing for each spatial location corresponding to each temporal image and these kernels are used to be convolved with the input of burst images. As the further improvement of KPN, it is followed by multiple kernel prediction network (MKPN [29]) to acquire various receptive field sizes [24].

However, the existing MKPN method does not take into account the interdependencies between multiple kernels. MKPN utilizes the kernels which are averaged based on the largest size of them. It implies that the different sizes of kernels are treated equally. In previous works [13, 50], they pointed out that lack of discriminative learning ability hinders the representation power of deep networks. To solve this problem, each of the weights corresponding with the channel of the feature maps is utilized as explicit constraints to learn discriminatively the global distribution of channel-wise feature response and increase the performance.

Therefore, we address this issue with a novel weighted multi-kernel prediction network (WMKPN), which can lead to restoring the high-quality images, while considering the interdependencies between multiple kernels by assigning the global weights for each of them.

In this paper, we propose a weighted multi-kernel prediction to address this issue and enhance the existing kernel prediction method. We developed a dynamic and discriminative mechanism, which helps multiple kernels to learn the interdependencies among the different sizes of kernels. Predicted weights for each kernel induce the model to discriminatively aggregate the features from each kernel and improve the image quality of the reconstructed HR image. The contributions of this paper are summarized as follows:

- We introduce a weighted multi-kernel prediction that can allow the model to induce to learn the interdependencies among the different sizes of kerenl using global weights for each of them and align the LR features without external motion compensation such as optical flow estimation.
- We propose a novel burst image super-resolution framework that is an end-to-end network utilizing the WMKPN as a powerful alignment module.
- Extensive experiments with various KPN methods are conducted in both Zurich RAW to RGB [15] and BurstSR dataset [3]. We improve the performance of the existing method while outperforming the state-of-the-art MFSR architecture with public metrics on image restoration.

2. Related Work

Multi-frame Super Resolution MFSR aims to reconstruct HR images from multiple LR frames containing the same scene. Farsiu et al. [11] design a hand-crafted MFSR algorithm that is robust to motion blur and noise by using the L1 norm. [6] concentrates pre-processing step that is composed of blur image filtering, denoising, and alignment rather than the SR step that is implemented by bicubic upsampling and merged with the L2 norm [11]. Reddy et al. [34] also use a two-step approach to solve the MFSR problem in the Ocular Biometrics domain. It uses a discrete cosine transform interpolation filter to perform upscaling and denoising, followed by deep learning-based deblurring. Kawulok et al.[18] propose a method of mapping multi-frame images to multiple HR images using SRResNet [22] before the EvoIM [17] process. HighRes-net [7] extracts the features of LR frames into the latent representations using an encoder that is a recursive scheme. The single global encoding is fed to the decoder that is followed by ShiftNet and Lanczos resampling to reconstruct. DeepSUM [31] leverages the registration filter that is the output of RegNet to fuse HR images from multiple feature maps. RAMS [36] harnesses attention mechanisms which are temporal and feature attention in the architecture. The model is composed of RFAB and RTAB that contain 3D and 2D convolution respectively. Bhat et al. [3] propose a novel deep learning-based BurstSR method that aligns burst frames using optical flow vectors and fuses images with an attention mechanism.

In this paper, we focus on burst image super-resolution covering RAW burst LR images acquired from hand-held cameras such as Wronski et al. [43]. Our method is different from the above methods in terms of alignment module that is inspired by the kernel prediction network. [30]

Kernel Prediction Network Given a sequence of burst images, the model generates per-pixel kernels that are convolved with the input frames to produce high quality images. It can be used as efficent image alignment modules that can implicitly capture the motion from burst images without external motion compensation. Since the KPN achieved improvements, several KPN based methods have been proposed. MKPN [29] uses the different sizes of the kernel with separable convolution and kernel fusion for computational efficiency. AME-KPNs [48] predict spatially adaptive kernels and weight maps to consider spatial-temporal components. In the Xia et al [44], inspired by self-similarity, model extracts 3D kernel basis and coefficient maps to reduce the computational costs.

We exploit KPN that can be used as the alignment module before the SR phase which reconstructs the HR image from the feature maps.

3. Method

The proposed model takes multiple LR RAW burst images with noise $\{b_i\}_{i=1}^{T}$ where T is the burst length and

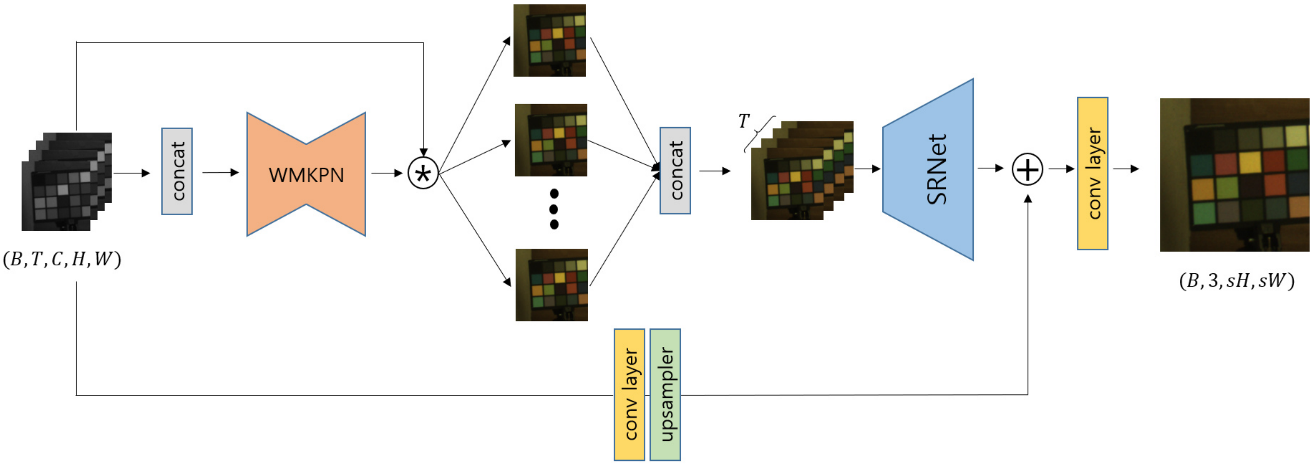


Figure 2: The total architecture of the proposed framework with WMKPN. Predicted images from WMKPN are concatenated in channel dimension for exploiting abundant information to be provided for the SR network

predicts denoised single HR RGB image I_{HR}. Inspired by the success of KPN [30] for burst image processing, we employ it as the backbone of our architecture for the image alignment. Our overall model as shown in Figure 2, consists of two significant parts: A module that aligns the burst of image and a network that increases the resolution of images while fusing the output of the align module.

Input of burst images are fed into the modified U-net as shown in Figure 3, which can be formulated as:

$$I(x,y) = H_{WMKPN}(F_{concat}(\{b_i\}_{i=1}^{T})(x,y)) \quad (1)$$

where $F_{concat}(\cdot)$ denotes function of concatenation in channel dimension and $I(x,y)$ where (x,y) means spatial locations is the output of the modified U-net before splitting into two branches as illustrated in Figure 3. The accumulated kernels are estimated by using outputs from each of two branches as shown in Eq. 9. The estimated kernels are then directly convolved with the input to generate T images, which is described in Eq. 2:

$$\hat{I}_i(x,y) = \tilde{K}_i(x,y) * P_i(x,y) \quad (2)$$

Here $\hat{I}_i(x,y)$ is the output of WMKPN. $\tilde{K}_i(x,y)$ is accumulated kernel and $P_i(x,y)$ is patch of the input image. $\hat{I}_i(x,y)$ are provided for the SR reconstruction network after being merged into channel dimension, which can be formulated as:

$$M = F_{SR}(F_{concat}(\hat{I}_i(x,y))) \quad (3)$$

Before the prediction of the single HR RGB image, global residual skip connection [19] is utilized to help the model to restore the high frequency of image, as shown in Eq. 4:

$$M_{res} = H_{res}(F_{concat}(\{b_i\}_{i=1}^{T})) \quad (4)$$

The last convolution layer is used to predict the final RGB output image from integrated information. as shown in Eq 5.

$$I_{SR} = Conv(M + M_{res}) \quad (5)$$

where $+$ means the element-wise addition for global residual learning.

3.1. Separable Kernel Estimation

KPN estimates the s^2 parameters for each pixel in the image where s is kernel size. It can be a burdensome task in terms of computation and memory consumption. The separable kernel estimation [29, 32] is leveraged to alleviate this issue and It can be formalized as follows:

$$K_i^s(x,y) = k_{i_1}^s(x,y) \otimes k_{i_2}^s(x,y) \quad (6)$$

where $K_i^s(x,y)$ is estimated kernel from separable kernel estimation. $k_{i_1}^s(x,y), k_{i_2}^s(x,y)$ are the pairs of one by s dimensional kernels that are predicted from the image. $s \times s$ kernel can be obtained from the outer product operation. This method helps model to reduce s^2 memory consumption costs to 2s.

3.2. Weighted Multi-Kernel Prediction

Inspired by comprehensive experiments, [24] show that multi-scale information is crucial to restoring high image quality. For this reason, the different sizes of multiple kernels are also widely used for burst image processing. However, existing multi-kernel prediction methods [29] treat

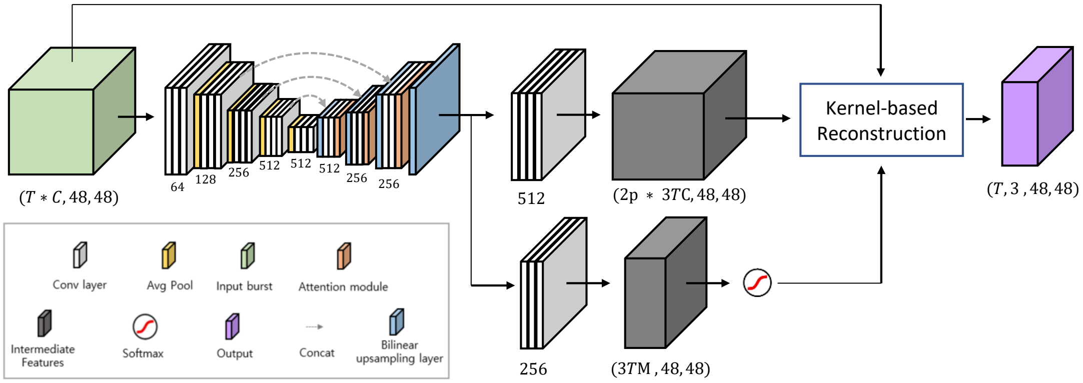


Figure 3: Overview of the WMKPN architecture. The input of our network is a sequence of burst that has T burst length (T = 14) and C channels (C = 4) per single burst image. The network is split into two branches *i.e*. kernel prediction branch and kernel weight branch.

multiple kernels equally, which does not consider the interdependencies among the different sizes of multiple kernels that are computed by each pixel. We propose our WMKPN which is an extension of [29] to deal with this issue.

The architecture of WMKPN is illustrated in Figure 3. In the encoder part, the spatial sizes of the feature maps are reduced by the average pooling layer. On the other side, the decoder increases the spatial sizes of the feature maps by a bilinear upsampling layer. In addition, we exploit the attention module, proposed in [48], which is composed of a series of the channel attention (CA) [50] and the spatial attention (SA) [14]. Also, encoded feature maps are concatenated to the decoder side that have the same spatial sizes like U-net architecture [35]. This modified U-net is split into two branches *i.e*. kernel prediction branch, kernel weight branch which predicts kernels and weights, respectively and it can be formulated as:

$$K_i^s(x,y) = B_k(I_i(x,y)),\ w_i^s(x,y) = B_w(I_i(x,y)) \quad (7)$$

where B_k is kernel prediction branch and B_w is kernel weight branch. B_k extracts 2p * 3TC channels where p, T, C are the sum of different kernel sizes (*e.g*. p = 1 + 3 + 5 + 7 = 16), length of the sequence of burst and channel of the single input burst image, respectively. On the other branch, kernel weights that have 3TM output channels are extracted where M means the number of different sizes of the kernel (*e.g*. M = 4, S $\in \{1, 3, 5, 7\}$) where S is different sizes of kernels and we additionally use 3 sets of filters to make output have 3 channels.

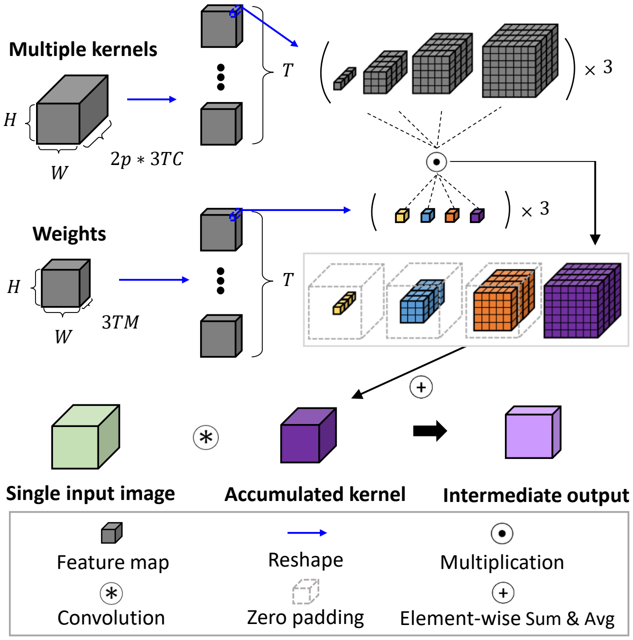


Figure 4: The procedure of weighted multi-kernel prediction. Feature maps from each of two branches are utilized to estimate the accumulated kernel. In the case of the process for a spatial pixel location, reshaped feature maps that come from the kernel prediction branch are multiplicated by corresponding to those of reshaped feature maps that are estimated by the other branch. The accumulated kernels are calculated by summation and average of zero-padded multiple kernels. After that, the predicted kernels are convolved with the input of burst images.

Next is the weighted multi-kernel prediction which is conducted on a pixel by pixel as shown in Figure 4. First, extracted feature maps from each branch are reshaped in the temporal dimension. In the case of one spatial pixel that is from a temporally reshaped tensor, multiple kernels which have C channel dimension can be generated by separable convolution. Similarly, kernel weights that correspond to each kernel also can be obtained from the kernel weight branch. To provide a discriminative mechanism, kernel weights are normalized with the softmax operator where it is applied in M dimension, which can be formulated as:

$$\tilde{W}_i = \frac{e^{w_i(x,y)}}{\sum_j e^{w_j(x,y)}} \tag{8}$$

The multiple kernels obtained in the kernel prediction branch are multiplied by the corresponding kernel weights. The weighted kernels are then added and averaged after zero padding for element-wise calculation, which can be formulated as:

$$\tilde{K}_i(x,y) = \frac{1}{|S|}\sum_{s\in S} K_i^s(x,y)\cdot \tilde{W}_i^s(x,y) \tag{9}$$

After that, The accumulated kernel is convolved with the input of burst image, as shown in Eq. 2. This kernel allows the model to learn the interdependencies among the multiple sizes of kernels, therefore, increases the performance of the network.

3.3. SR Reconstruction Network

We leverage the residual blocks (RBs) with local skip connections which are exploited in enhanced deep SR (EDSR) network [27]. For utilizing abundant information of features, this network takes T temporarily aligned and concatenated images that come from the output of WMKPN. SR network extracts deep features without upsampling the spatial size for efficient computation and speeding up the SR process [8]. This network is composed of tens of RBs and three sub-pixel convolution layers [37] to upsample the resolution of the image.

3.4. Loss Fucuntion

In several super-resolution literature, they use a simple, L_1 or L_2 loss for better HR reconstruction. Some comprehensive experiments [51] in image restoration tasks show that L_2 loss tends to give strong penalty to outliers, which leads to poor image quality. Therefore, we utilize the L_1 loss for reconstruction loss, which can be formulated as:

$$L_{SR} = \frac{1}{N}\sum_{i=1}^{N} \left|\left| H_{SR}(\{b_j\}_{j=1}^{T}) - I_{HR}^{i} \right|\right|_1 \tag{10}$$

where H_{SR} is our total network for burst image super-resolution.

Perceptual loss functions such as VGG loss [22] or the structural similarity index measure (SSIM) [41] loss are also utilized to train the model for visual qualitative results. SSIM loss is motivated by SSIM that takes into account contrast, structure and luminance of the images, which can be expressed as:

$$L_{SSIM}(P) = \frac{1}{N}\sum_{p\in P} 1 - SSIM(p) \tag{11}$$

defined via the dependence of means and standard deviations on pixel p that is in some boundary region P.

In this paper, we use two loss functions that are composed of the mean absolute error (MAE) and the SSIM loss to train our model. Utilizing these two loss functions, the total loss function is computed as follows:

$$L_{Total} = \lambda_{SR} * L_{SR} + \lambda_{SSIM} * L_{SSIM} \tag{12}$$

where λ_{SR} and λ_{SSIM} are coefficients as hyper-parameters which are assigned to each loss function for a balanced learning scheme of consistency and visual quality.

4. Experiments

4.1. Dataset

We evaluate our model on both the synthesized Zurich RAW to RGB [15] and BurstSR datasets [3]. The synthetic datasets have 46839 HR RGB images for training, 1204 images for testing. To obtain the pair of synthesized LR burst images and HR RGB images, ground-truth images are warped by affine transform that includes rotation and translation, utilizing several camera pipeline parameters *i.e.* color correction matrix, color gains, gamma expansion, *etc*. Those HR images are downsampled with bilinear downscale degradation models.

Meanwhile, the BurstSR datasets have 200 RAW burst sequences with their corresponding HR RGB images. Configuration of datasets is 160 images for training, 20 for validation, 20 for the test, respectively. LR burst sequences are acquired in identical settings of devices such as Samsung Galaxy S8 mobile phone camera. HR RGB images as ground-truth are collected by Canon 5D Mark 4 DSLR camera.

4.2. Training settings

We implement our SR framework based on WMKPN in PyTorch and use an Nvidia Titan V to train the model with a batch size of 8. Each batch has 14 LR noisy burst frames and HR RGB images that are flipped and randomly cropped to make the sizes of patches (48 × 2) s × (48 × 2) s where s is the upsampling scale factor. Our network is trained with Adam optimizer [20] where β_1 = 0.9, β_2 = 0.999, $\epsilon = 10^{-8}$. The learning rate is initialized as 1×10^{-4} and reduced to half for every 30 epochs.

4.3. Evaluation metrics

All downsampled images including validation and test datasets are super-resolved to a scale of 4. The reconstructed images are evaluated by the Pixel Signal to Noise

Ratio (PSNR), the Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [49]. For the PSNR metric, we ignore the boundary forty pixels to measure the image quality. SSIM score is calculated for each channel and then averaged. LPIPS is also used to measure the perceptual similarity between output images and ground truth.

5. Experimental Results

5.1. Ablation Study

Alignment Loss Inspired by KPN as powerful alignment tools, we utilize the KPN as a baseline of our model for the alignment module of our super-resolution framework. Several networks such as MKPN and AWE-KPNs derived from KPN exploit the annealed loss to prevent kernels existing in each temporal frame from being biased to the reference frame. We experiment on the annealed loss as alignment loss to investigate the impact of it in our case. The loss is calculated for the difference between the predicted intermediate output from WMKPN and LR ground-truth images from the bilinear downsampled from the HR ground truth images. To match the number of channel dimensions, the intermediate feature maps from WMKPN are summated and averaged over the temporal axis. We report the quantitative results on the test set of Zurich RAW to RGB datasets. We find that alignment loss has no remarkable changes and even reduces the performance of the architecture as shown in Table 1. These results allow us to design the final loss function for our framework.

Impact of Abundant Features In our burst super-resolution framework, the intermediate outputs from WMKPN are concatenated in channel dimension and then fed to the SR reconstruction network that is composed of several residual blocks. Aligned features that are calculated by kernels for each temporal frame are provided after concatenation in channel-wise to allow the SR network to fully utilize the aligned features. We observe that utilizing the aggregated feature maps before being provided for the SR reconstruction network increases the performance of our model.

Weighted Kernel Prediction We analyze the impact of the weighted kernel prediction module on our super-resolution framework. We compare our WMKPN with the method that does not use the weighted kernel. As the results are shown in Table 1, we find that the model with PSNR of 36.5641dB, considering the weights for multiple kernels outperforms the comparison model with PSNR of 36.3066dB. These experiment results imply that the interdependencies among the multiple kernels affect the performance of our architecture in our task.

Method	PNSR	SSIM	LPIPS
Alignment Loss	36.4742	0.9105	0.1185
No Concat	36.3811	0.9095	0.1200
WMKPN - W	36.3066	0.9059	0.1288
WMKPN	**36.5641**	**0.9120**	**0.1172**

Table 1: Ablation study on a synthetic dataset, investigating the model that can lead to the best performance of our results. All values are reported in terms of averaged score on test set. **Bold** indicates the best score on this table.

Method	PSNR	SSIM	LPIPS
Synthetic dataset			
KPN	36.2047	0.9046	0.1319
EfDeRain	36.1701	0.9044	0.1303
MKPN	36.2895	0.9056	0.1302
AWE-KPNs	36.3054	0.9063	0.1286
WMKPN	36.5641	0.9120	0.1172
WMKPN*	**36.8959**	**0.9170**	**0.1095**
BurstSR			
KPN	41.6928	0.9562	0.0785
EfDeRain	41.6091	0.9548	0.0856
MKPN	41.6972	0.9562	0.0780
AWE-KPNs	41.7260	0.9564	0.0777
WMKPN	**41.8740**	**0.9585**	**0.0746**
WMKPN*	41.8472	0.9584	0.0745

Table 2: Comparison of other KPN approaches with our proposed model. WMKPN* means the other version of our model that uses the concatenation of different sizes of the kernels without exploiting the kernel fusion that is proposed in [29]. **Bold** and underline indicate the best and the second-best score respectively.

Method	PSNR	SSIM	LPIPS
Bilinear	26.4749	0.7337	0.1410
HighRes-net	41.3086	0.9520	0.0881
RAMS	41.5722	0.9555	0.0916
WMKPN	**41.8740**	**0.9585**	**0.0746**

Table 3: Comparison of the existing MFSR architecture with our proposed model in BurstSR dataset.

5.2. Comparison with other KPN approaches

Next, we report the results of different KPN architectures on the test set of both the synthetic dataset and the BurstSR dataset. We compare with the other four models: i) **KPN** that introduces the kernel prediction network for burst image denoising, ii) **EfDeRain** [12] that propose the enhanced version of KPN with dilated convolution [47] for efficient deraining, iii) **MKPN** using the multiple kernels to

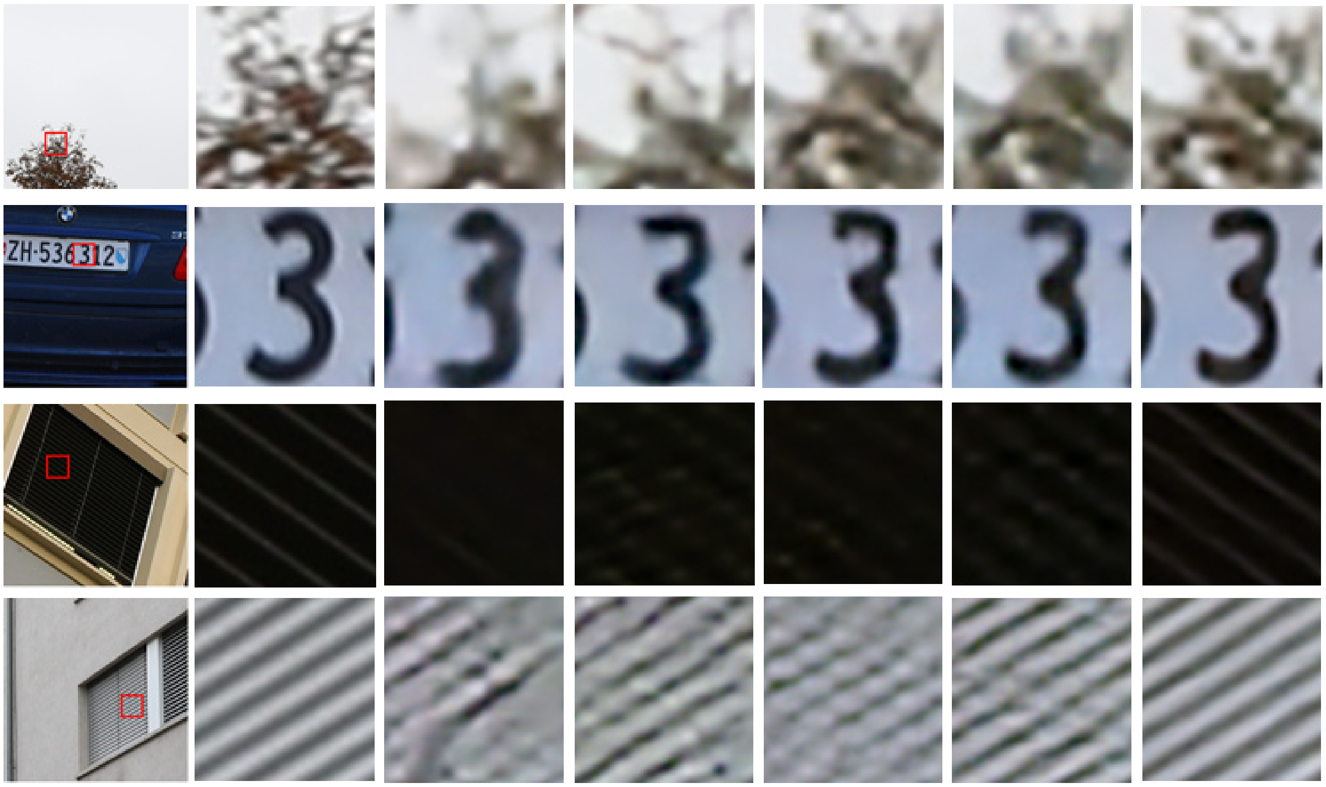

(a) Reference (b) GT (c) KPN (d) EfDeRain (e) MKPN (f) AWE-KPNs (g) Ours

Figure 5: Qualitative comparison of different KPN methods on synthetic dataset.

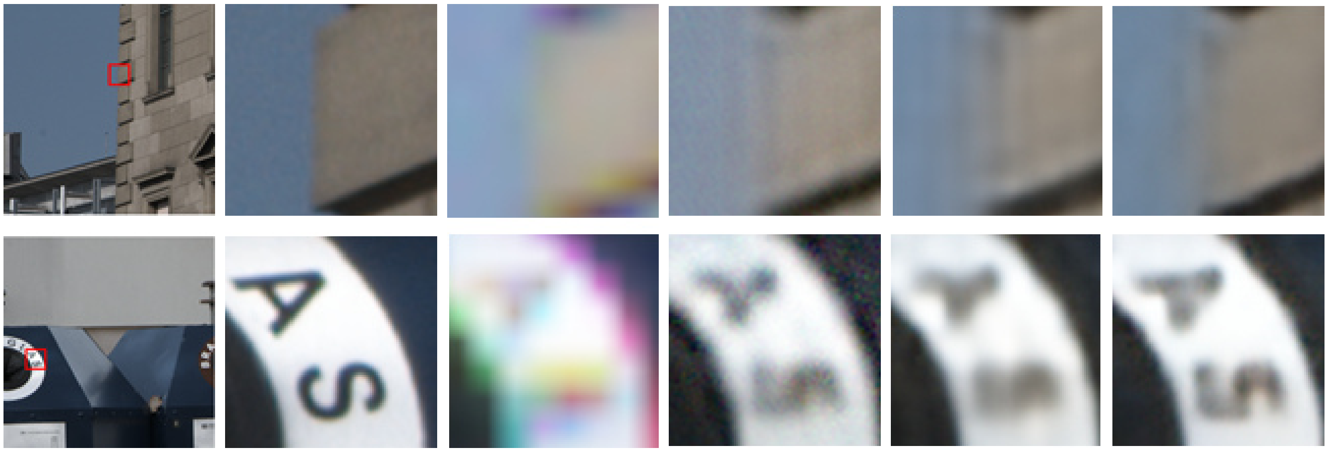

(a) Reference (b) GT (c) Bilinear (d) HighRes-net (e) RAMS (f) Ours

Figure 6: Qualitative comparison of MFSR model on BurstSR dataset.

have various receptive field. iv) **AWE-KPNs** utilizing the attention module on the modified U-net and allocating the weights for the predicted output image. As a module for alignment before being provided for the SR reconstruction network, the performances of each model are compared in the same environment based on our SR framework. We reported the results with averaged PSNR, SSIM and LPIPS scores on the test set as shown in Table 2. In the synthetic dataset, we find that concatenating the multiple kernels increases the model capacity which helps the model enhance

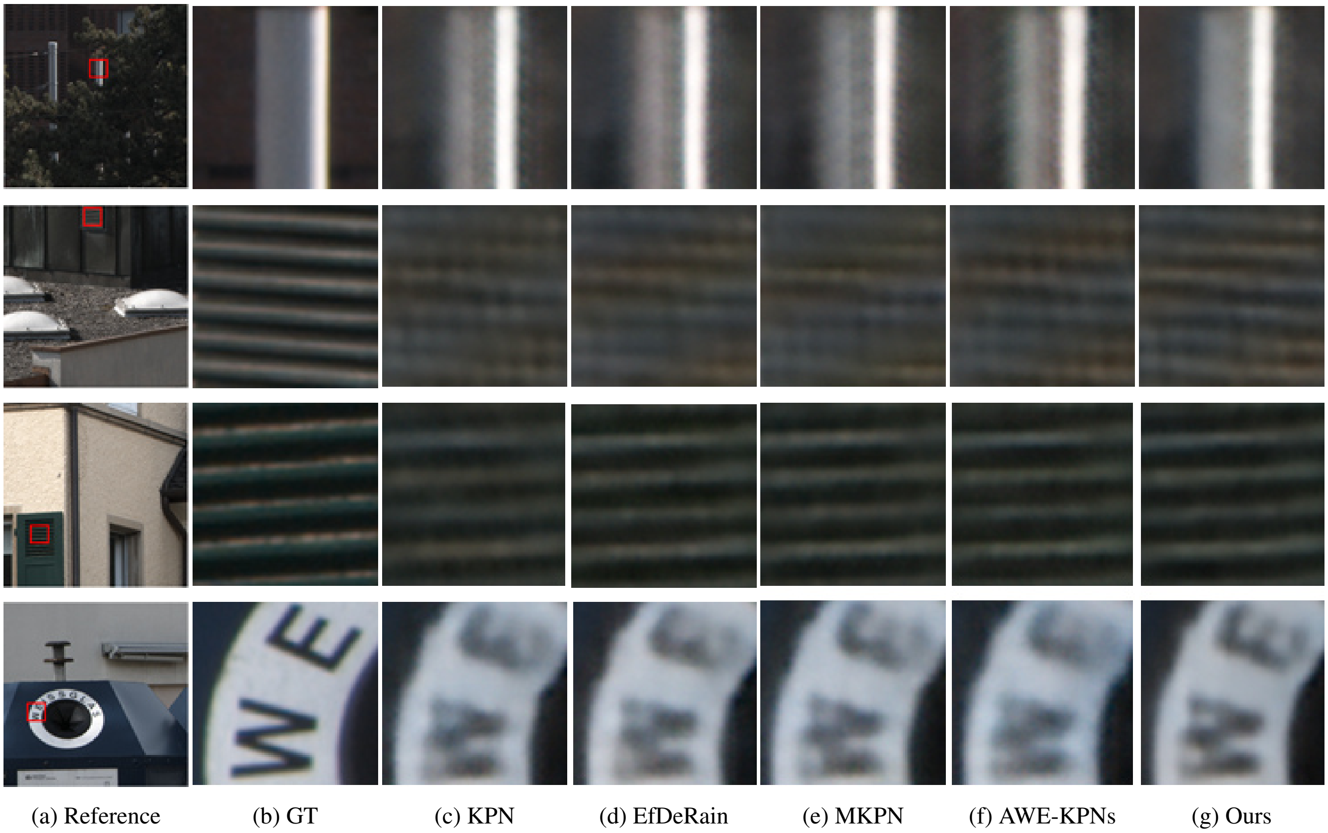


Figure 7: Qualitative comparison of different KPN methods on BurstSR dataset.

the quantitative results, but training speed slows down to converge and the number of model parameters increases inefficiently.

5.3. Comparison with other MFSR architectures

To compare the existing MFSR architectures with our model, we conduct a comparison study in the BurstSR dataset. We compare our model with two architectures that are from remote sensing applications: i) **HighRes-net** [7] which consists of an encoder with recursive fusion, decoder and ShiftNet for image registration, and ii) **RAMS** [36] that has RFAB and RTAB with 3D convolution layer. To evaluate in RAW burst image super-resolution application, we adapt the model to fit the BurstSR dataset. For a fair comparison, each of these architectures is employed on the default settings, respectively. The results of the comparison studies are shown in Table 3. We observe that our WMKPN method outperforms the recent MFSR architecture on the test set of BurstSR for real-world applications.

6. Conclusion

In this paper, we propose a novel framework: WMKPN for burst image super-resolution. WMKPN is motivated by KPN as an image alignment module that captures implicitly the random hand-held motions among temporal frames and improved by taking into account the interdependencies between the multiples sizes of kernels. We present that WMKPN helps ours framework to improve the performance for burst image super-resolution. The extensive experiments demonstrate that our model outperforms other super-resolution frameworks including the existing KPN methods and the state-of-the-art MFSR method. In future work, we would like to test our model in broad multiple frame image restoration tasks, such as burst image denoising, video deblurring, multiple frame image super-resolution.

Acknowledgements

This research has been supported by the Samsung Electronics Corporation. (IO201209-07893-01)

References

[1] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 731–747, 2018. 4321

[2] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 challenge on burst super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 4321

[3] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. *arXiv preprint arXiv:2101.10997*, 2021. 4321, 4322, 4325

[4] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4778–4787, 2017. 4321

[5] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 4321

[6] Neil Patrick Del Gallego and Joel Ilao. Multiple-image super-resolution on mobile devices: an image warping approach. *EURASIP Journal on Image and Video Processing*, 2017(1):1–15, 2017. 4321, 4322

[7] Michel Deudon, Alfredo Kalaitzis, Md Rifat Arefin, Israel Goytom, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. Highres-net: Multi-frame super-resolution by recursive fusion. 2019. 4322, 4328

[8] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 4325

[9] Juan Du, Wenlan Wei, Cien Fan, Lian Zou, Jiawei Shen, Ziyu Zhou, and Zezong Chen. Lightweight image super-resolution with mobile share-source network. *IEEE Access*, 8:60008–60018, 2020. 4321

[10] Thibaud Ehret, Axel Davy, Pablo Arias, and Gabriele Facciolo. Joint demosaicking and denoising by fine-tuning of bursts of raw images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8868–8877, 2019. 4321

[11] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multiframe super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004. 4322

[12] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. *arXiv preprint arXiv:2009.09238*, 2020. 4326

[13] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 4322

[14] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3911–3927, 2019. 4324

[15] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 536–537, 2020. 4322, 4325

[16] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 466–467, 2020. 4321

[17] Michal Kawulok, Pawel Benecki, Daniel Kostrzewa, and Lukasz Skonieczny. Evolving imaging model for super-resolution reconstruction. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 284–285, 2018. 4322

[18] Michal Kawulok, Pawel Benecki, Szymon Piechaczek, Krzysztof Hrynczenko, Daniel Kostrzewa, and Jakub Nalepa. Deep learning for multiple-image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 17(6):1062–1066, 2019. 4322

[19] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 4323

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4325

[21] Filippos Kokkinos and Stamatis Lefkimmiatis. Iterative residual cnns for burst photography applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5929–5938, 2019. 4321

[22] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 4322, 4325

[23] Feng Li, Xiuping Jia, and Donald Fraser. Universal hmt based super resolution for remote sensing images. In *2008 15th IEEE International Conference on Image Processing*, pages 333–336. IEEE, 2008. 4321

[24] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 517–532, 2018. 4322, 4323

[25] Zhetong Liang, Shi Guo, Hong Gu, Huaqi Zhang, and Lei Zhang. A decoupled learning scheme for real-world burst denoising from raw images. In *European Conference on Computer Vision*, pages 150–166. Springer, 2020. 4321

[26] Renjie Liao, Xin Tao, Ruiyu Li, Ziyang Ma, and Jiaya Jia. Video super-resolution via deep draft-ensemble learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 531–539, 2015. 4321

[27] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single

image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 4325

[28] Tairan Liu, Kevin De Haan, Yair Rivenson, Zhensong Wei, Xin Zeng, Yibo Zhang, and Aydogan Ozcan. Deep learning-based super-resolution in coherent imaging systems. *Scientific reports*, 9(1):1–13, 2019. 4321

[29] Talmaj Marinč, Vignesh Srinivasan, Serhan Gül, Cornelius Hellge, and Wojciech Samek. Multi-kernel prediction networks for denoising of burst images. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2404–2408. IEEE, 2019. 4321, 4322, 4323, 4324, 4326

[30] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018. 4321, 4322, 4323

[31] Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Deepsum: Deep neural network for super-resolution of unregistered multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3644–3656, 2019. 4322

[32] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 261–270, 2017. 4323

[33] Zongxu Pan, Jing Yu, Huijuan Huang, Shaoxing Hu, Aiwu Zhang, Hongbing Ma, and Weidong Sun. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9):4864–4876, 2013. 4321

[34] Narsi Reddy, Dewan Fahim Noor, Zhu Li, and Reza Derakhshani. Multi-frame super resolution for ocular biometrics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 453–461, 2018. 4322

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4324

[36] Francesco Salvetti, Vittorio Mazzia, Aleem Khaliq, and Marcello Chiaberge. Multi-image super resolution of remotely sensed images using residual attention deep neural networks. *Remote Sensing*, 12(14):2207, 2020. 4322, 4328

[37] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 4325

[38] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4472–4480, 2017. 4321

[39] Rao Muhammad Umer, Gian Luca Foresti, and Christian Micheloni. Deep generative adversarial residual convolutional networks for real-world super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 438–439, 2020. 4321

[40] Longguang Wang, Yulan Guo, Li Liu, Zaiping Lin, Xinpu Deng, and Wei An. Deep video super-resolution using hr optical flow estimation. *IEEE Transactions on Image Processing*, 29:4323–4336, 2020. 4321

[41] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4325

[42] Patrick Wieschollek, Bernhard Schölkopf, Hendrik PA Lensch, and Michael Hirsch. End-to-end learning for image burst deblurring. In *asian conference on computer vision*, pages 35–51. Springer, 2016. 4321

[43] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics (TOG)*, 38(4):1–18, 2019. 4321, 4322

[44] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11844–11853, 2020. 4322

[45] Xiangyu Xu, Yongrui Ma, and Wenxiu Sun. Towards real scene super-resolution with raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1723–1731, 2019. 4321

[46] Daiqin Yang, Zimeng Li, Yatong Xia, and Zhenzhong Chen. Remote sensing image super-resolution: Challenges and approaches. In *2015 IEEE international conference on digital signal processing (DSP)*, pages 196–200. IEEE, 2015. 4321

[47] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 4326

[48] Bin Zhang, Shenyao Jin, Yili Xia, Yongming Huang, and Zixiang Xiong. Attention mechanism enhanced kernel prediction networks for denoising of burst images. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2083–2087. IEEE, 2020. 4321, 4322, 4324

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 4326

[50] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4322, 4324

[51] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 4325