Authors' Responses to the Reviews of Paper #507

We sincerely thank all the reviewers for their recognition on the constructed dataset. Please find what below our itemized responses to all the reviewers' comments.

R1.Q1: Image registration. We used global registration. The reviewer actually mentioned two misalignment problems: the change of image center and the perspective change. The first one is caused by the slight shift of optical center at different focal lengths, which is basically a global shift for modern camera lens. To reduce the effect of perspective change caused by object depths at different focal lengths, we selected to capture scenes with large depth-of-field at a long distance, and then conducted careful screening to remove those image pairs with noticeable misalignment. We originally captured more than 500 scenes but only 234 were remained in the dataset. As can be seen from Table 2, the PSNR of simple Bicubic interpolator on our dataset can be 27.99dB for scaling factor $\times 4$, which validates the effectiveness of our registration scheme.

R1.Q2: PSNR/SSIM. We agree that PSNR/SSIM are sometimes inconsistent with human perception. But they are still the most popular metrics for image restoration tasks. As suggested by this reviewer, we will add the LPIPS metric and arrange perceptual user-study in the revision.

R1.Q3: Noise in real-world SR. We didn't explicitly consider noise in the SR model training for 2 reasons. First, this is to fairly compare with most of the SR works in literature, which are trained on simulated clean image pairs. Second, noise reduction is an important module in camera ISP and it is usually performed before SR. Joint denoising and SR is becoming a promising direction for new ISP design, while in this work we focus on the SR part. However, we will release the raw data of our collected dataset, on which researchers can study joint denoising and SR models.

R1.Q4: Discussion on the concurrent zoom paper. We have noticed this very interesting paper released recently, which also zooms lens to collect images. A novel loss function was proposed to learn SR models from mild misaligned image pairs. Compared with this work, we made more efforts on carefully choosing the scenes and designed an image registration algorithm to precisely align the image pairs. With our dataset, researcher can more conveniently train their models and evaluate their results. We believe both the two datasets will greatly facilitate the research of real-world SR. We will add discussions in the revision.

R2.Q1: Center crops. Yes we only keep the registered center crops in the dataset. However, once an SR model is trained, it can be well applied to whole images since SR aims to enhance the image local details. Actually, Fig. 6 in the main paper shows the SR results of our model on full images captured by iPhone X and Google Pixel2.

R2.Q2: Real-world blur kernels. The blur kernels in real-world images are unknown and spatially variant. We totally agree with this reviewer that it is worth investigating how to recover and visualize these kernels using our dataset. We will surely work on this in the next step.

R2.Q3: Registration by SIFT/SURF. Sorry that we may not explain this clearly enough. Due to the large difference in resolution and small changes in luminance, the sparse key points based SIFT/SURF methods are not accurate enough for our registration task. Note that our registration algorithm is a dense solution which involves all image pixels in optimization, aiming at subpixel-wise alignment. It has 0.5dB and 1.0dB higher PSNR than SURF for scaling factor 2 and 4, respectively. We will clarify this in the revision.

R2.Q4: Distortions of MD. The MD paper simulates real-world degradations by combining different types/sizes of synthetic kernels, noise and downsampling strategies. Unfortunately, such simulations cannot generalize well to the real-world images. This is actually one of the reasons motivating us to construct the RealSR dataset.

R2.Q5: Lens distortion correction. The lens manufacturer has provided the distortion calibration parameters. The Photoshop reads them to correct the lens distortion.

R3.Q1: Difference from [R1-R3]. The image pairs in [R1] are not precisely aligned and some have moving objects. The degraded raw data in [R2] are simulated which can be different from real-world degradations. The dataset in [R3] only contains images of postcards but not natural scenes and there is only one scaling factor. Our dataset consists of pixel-wise aligned high-quality image pairs, which provides a good benchmark for real-world SR study. Please also refer to our response to Q4 of R1.

R3.Q2: Comparison between LP-KPN and KPN. LP-KPN is designed to reduce the computational cost of KPN at the last kernel prediction layer (KPL). For one image with 12M pixels, the FLOPs of KPL in KPN ($k = 20$) and LP-KPN ($k = 5$) are 14.4G and 1.18G, respectively. Since the evaluated SR models have very deep convolutional layers which cost a lot of computation, the efficiency improvement by LP-KPN over KPN on the whole model is not obvious. But if a lightweight network (e.g., 10 layers) is used, the speed-up will be significant.

R3.Q3: Comparisons in table 2. Table 2 is to demonstrate the advantages of our RealSR dataset over those simulated ones to show the necessity of real-world SR datasets.

R3.Q4: Training and testing split. We followed the practice in previous SR datasets to use one random training and testing split. For sure we will release the split used in our experiments. We will also add results on k-fold cross-validation in the revision.