Toward Real-world Panoramic Image Enhancement

Yupeng Zhang[1] Hengzhi Zhang[1] Daojing Li[1] Liyan Liu[1] Hong Yi[1] Wei Wang[1]
Hiroshi Suitoh[2] Makoto Odamaki[2]
[1] Ricoh Software Research Center (Beijing) Co., Ltd. Haidian District, Beijing, China
[2] Ricoh Company, Ltd. Tokyo, Japan
{Yupeng.Zhang, Hengzhi.Zhang, Daojing.Li, Liyan.Liu, Hong.Yi, wei.wang}@srcb.ricoh.com
{hiroshi.suitoh, makoto.odamaki}@jp.ricoh.com

Abstract

Panoramic images captured by the fisheye lens cameras cover very wide field of view (FoV) ranging from 180° to 360°, but the image quality is very low compared to that of high-end cameras such as DSLR or compact cameras with APS-C or full frame sensors. In this paper, we aim to use deep neural network (DNN) based methods to improve panoramic image quality. Specifically, we enhance low quality panoramic images of 5K resolution (5376×2688) to high-end camera quality at the same resolution, which is good for applications that requires limited resources, low-cost but high image quality. We build a Panoramic-High-end dataset which is the first real world panoramic image dataset as far as we know. Based on the generative adversarial network (GAN) architecture, we also design a compact network employing multi-frequency structure with compressed Residual-in-Residual Dense Blocks (RRDBs) and convolution layers from each dense block. Experiments show that our method surpasses several state-of-the-art DNN based methods in both no-reference and full-reference evaluations as well as the processing speed. Our results show that it's practical to integrate DNN based image enhancer into optics design to achieve a balance between optical cost and image quality.

1. Introduction

Image enhancement plays an important role in digital image processing because the captured images are not perfect in terms of sharpness, texture detail, signal-to-noise ratio (SNR), dynamic range and the level of chromatic aberration etc. Tremendous amount of work have been proposed in this area by using traditional image processing methods or recently fast-developing deep learning based methods [1, 2].

Although image enhancement techniques are extensively studied for decades and have seen great achievement, few addresses the problem of improving panoramic image quality. Panoramic images can be obtained by a fisheye camera, a catadioptric system, a multi-camera system or a rotating camera [3, 4, 5]. The fisheye lens is a wide angle camera lens that can capture very large field of view (FoV), usually half of the full sphere. The catadioptric system uses reflecting mirrors and lenses for panoramic imaging. The multi-camera systems and the rotating cameras obtain 360° images by sewing images captured by multiple cameras or by one camera over time sequence. Among the four aforementioned devices, the fisheye panoramic camera is a low-cost consumer level camera which has poor image quality in terms of texture detail, image clarity, noise level and chromatic aberration.

Figure 1: Results of the proposed method. Top row: original 5K (5376×2688) panoramic image captured by a 360° consumer fisheye camera. Bottom left: image crop showing enlarged patch in the original image. Bottom right: enhanced by the proposed method.

In this work, we aim to enhance panoramic images captured by consumer fisheye lens cameras to high-end camera level at the same resolution by using a deep neural network (DNN) based method (Figure 1). Enhancement at the same resolution is good for applications that requires

limited resources, low-cost but high image quality. As far as we know, none has achieved this target by either traditional or DNN based methods. One reason lies in the dataset used for training. Most of the available datasets are designed for perspective image enhancement. Perspective images refer to distortion-free images usually captured by non-fisheye cameras [3], which do not produce large geometrical distortion. However, these images have low data similarity to panoramic ones captured by the fisheye lens cameras (panoramic image in short). Firstly, a panoramic image has severe geometrical distortions around poles. In the image matching phase, linear transformations such as homography for perspective low and high quality patch (LQ-HQ in short) pairs is straightforward and good matching result can be obtained [1]. However, panoramic image matching involves non-linear transformation which leads to mis-matched pairs (see Section 2.1 of the supplementary material). Secondly, the fisheye lens has large FoV, but the captured image is a circle which is smaller than the sensor size, so the effective resolution is less than that of perspective cameras. Thirdly, panoramic images taken at different angle of incidence may have different image quality too. For example, those taken around 170° may have much lower clarity and higher level chromatic aberration than those taken around 0°. Fourthly, commercial fisheye camera usually stitches two fisheye images to a single equirectangular image. This process leads to warping and stitching errors, results in poorer image quality than perspective camera images. Therefore, naively applying deep learning models trained by only datasets of perspective images leads to poor results (see Figure 6). It is necessary to train a network tailored to real world panoramic LQ-HQ pairs.

Existing datasets for enhancement purpose are either synthetic or real-world. The former creates low quality (LQ) image by degrading (e.g. downsample, Gaussian blur) the high quality (HQ) one [6, 11]. The latter collects LQ and HQ by using real world cameras [1, 2]. Recent works tend to use real world data rather than synthetic one because the former adapts well to real world complexity [30]. If we train a network with synthetic data, the network only learns the degradation that is artificially introduced but not from the real world. The real world degradation includes not only downsamping, Gaussian blur etc. but also optical defects such as lens distortion, chromatic aberration, and other noise components which cannot be modeled directly. For this reason, we decide to use real world data captured by a real panoramic camera and a high-end camera. We build a Panoramic-High-end (Pano-Hi) dataset consisting of panoramic equirectangular images and aligned APS-C camera image counterparts. A two-step patch matching algorithm is used to obtain small LQ-HQ image patches for training.

Among DNN based methods, perceptual-driven methods have proved superiority over PSNR/SSIM driven methods in terms of perceptual quality because the latter produces blurry images while the former produces photorealistic images with rich texture details, closer to human visual perception [8, 9]. Blau et al. [8] and [10] suggest that there is a trade-off between perceptual quality and reconstruction accuracy by full-reference metrics such as PSNR/SSIM. Since perceptual quality is more important in our case for panoramic image enhancement, we adopt perceptual-driven methods throughout the paper.

In this paper, we also propose a deep learning based method that enhances panoramic image quality at a high speed. Among the DNN based methods, ESRGAN [11] trained with our real world data meets the requirements of improved texture detail, image clarity, reduced noise and chromatic aberration. However, ESRGAN uses Residual-in-Residual Dense Blocks (RRDBs) which is not computationally efficient and takes too much time for a 5K image inference. Therefore, we design an efficient GAN architecture, which employs multi-frequency structure with compressed RRDBs and convolutional layers to reduce processing time while enhancing the image quality to the state-of-the-art level. Our method not only improves texture details and sharpness, but also removes color aberration and noise (due to optical defects) simultaneously (Figure 4).

The proposed network architecture is fast for 5K panoramic image enhancement without the need of high-spec hardwares (e.g. multi-camera system), and has impressive visual quality.

Our contributions are as follows:

(1) As far as we know, we are the first to enhance panoramic image captured by the fisheye lens cameras to high-end camera quality at the same resolution, which is good for applications that requires limited resources, low-cost but high image quality.

(2) We build the Pano-Hi dataset including panoramic equirectangular images and high quality counterparts. To the best of our knowledge, this is the first real world panoramic dataset for enhancement purpose. A two-step patch matching algorithm is proposed to obtain aligned small LQ-HQ image patches for training.

(3) We propose an efficient GAN architecture employing multi-frequency structure, which consists of high and low frequency representations process with compressed RRDBs and convolutional layers from each dense block.

2. Related work

2.1. Image enhancement

Image enhancement has many sub-tasks such as image deblurring [12], denoising [13, 14, 15], dehazing [16], super-resolution (SR) [17, 18, 19, 20], HDR imaging, color restoration and contrast enhancement etc.

Recent works have demonstrated advantages of using

DNN image enhancement over traditional methods. Generally, DNN based image enhancement can be seen as an image to image translation problem, which translates one representation of a scene into another [21]. In [21], Isola et al. demonstrated a general purpose framework that uses a GAN architecture to solve image translation problems such as image colorization, day to night, edge and label to photo. Similarly, Chen et al. [22] proposed a generalized image processing operator that can do ten different image processing tasks: denoising, dehazing, multiscale tone enhancement, style transfer etc. by using deep neural networks. Image translation is to learn an end to end mapping from an input to an output image. For example, in the area of image enhancement, a low quality (LQ) image is the input which is mapped to an enhanced (HQ_E) one by some mapping function $HQ_E = F(LQ)$, given a high quality (HQ) counterpart as the ground truth to compute the loss between HQ_E and HQ images. We find this rule is also effective to improve panoramic image quality: given a low quality panoramic image, which is mapped to a high-end quality image, despite the difference in the cameras' optical designs. As a result, it is straightforward for us to use DNN based enhancement techniques to improve panoramic image quality.

Many works obtain simulated low quality image by downsampling the ground truth [11, 17, 18, 23, 24, 25]. To improve adaptability to real world complexity, many propose to use real world images rather than simulated ones [1, 2, 26, 27, 28, 29]. Ignatov et al [2] used three phones and one high-end DSLR camera to capture low-high-quality pairs for image enhancement training. Zhang et al. [26] used optical zoom to obtain low resolution images and high resolution ground truths and propose a contextual bilateral loss for image alignment issue. Chen et al. [27] performed a low-light image enhancement which enhances short-exposure image captured at extremely dark environment. They employed short-exposure low-light images as their input and long-exposure counterparts as the ground truths. Cai et al. [30] also adopted real world low-high-resolution (LR-HR) data captured by a DSLR, the LR and HR were captured at a short and a long focal length, respectively. Ignatov et al. [57] used iPhone 3GS and Canon 70D DSLR to capture low and high quality images, respectively.

Although these works use real cameras to capture the real world low-high-quality image pairs for training, ours is the first to use a real panoramic camera to capture low quality panoramic image and a high-end camera to capture the high quality counterpart for training as far as we know. Our task of panoramic image enhancement is quite similar to Ignatov et al. [31], but our input is not phone images but panoramic images captured by a fisheye lens camera.

2.2. Perceptual-driven methods

As mentioned in Section 1, perceptual-driven image enhancement methods generate visually more pleasing results than PSNR/SSIM [58] driven methods. The former adopts perceptual loss with adversarial training by using Generative adversarial network (GAN)[32]. Blau et al. [8] provided a benchmark for perceptual driven SR. It ranked state-of-the-art algorithms which focus on perceptual quality of images and evaluated quantitatively by using no-reference metric such as Perceptual Index (PI) [8]. For example, ESRGAN [11] aimed to improve SRGAN [18] by introducing a deeper network with Residual-in-Residual Dense Blocks (RRDB) and removing batch normalization (BN) layer from the original SRGAN architecture. It redefined perceptual loss before the activation layer instead of after the activation layer adopted by SRGAN, and used a relativistic GAN loss instead of vanilla GAN loss. Michelini et al. [33] combined multi-scale loss and perceptual loss for adversarial training. Other perceptual-driven examples are [9, 34, 35, 36, 37, 38, 39]. All of them used adversarial loss and perceptual loss and demonstrated superior performance over PSNR/SSIM driven methods.

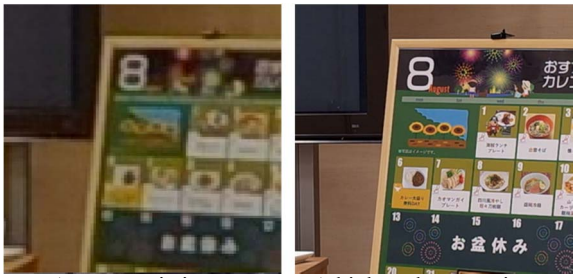
Although perceptual quality is of utmost importance in our case, we also refer to the state-of-the-art methods which are PSNR/SSIM driven. RCAN [41] proposed a residual channel attention network with many residual in residual (RIR) blocks, and evaluated the results using PSNR and SSIM. However, RCAN is not GAN based, which does not define perceptual loss and adversarial loss. It only uses $L1$ loss. Other works that favor high PSNR/SSIM scores are [6, 23, 30, 40, 42, 43, 44, 45, 46, 57].

Some works also tried to improve reconstruction accuracy like PSNR and SSIM scores and perceptual quality at the same time [47, 48, 49].

In this paper, we adopt perceptual-driven methods to enhance panoramic image quality, and we use no-reference and full-reference perceptual quality metrics to evaluate our results quantitatively.

2.3. Panoramic image enhancement

Deep learning based approaches for panoramic image enhancement are rare. Existing methods either enhance fisheye images or equirectangular ones using synthetic data, in which the LR images are created by downsampling the ground truth. Chang et al. [51] super-resolved fisheye images using a CNN based method. They applied different distortion coefficients to the original distortion-free images to produce distorted ones with barrel or pincushion distortion, then downsampled them with scaling factors 2×, 3× and 4× for training. Fakour-Sevom [52] used SRCNN for panoramic equirectangular image SR. They built their own panoramic equirectangular image dataset consisting of 34 images of different scenes. Instead of using real world



A panoramic image A high-end camera image

Figure 2: Visual comparison between a panoramic image (cropped, captured by a fisheye camera) and its high quality (cropped, captured by an APS-C camera) counterpart.

data, the LR samples were created by downscaling sub-images cropped from the original with a scaling factor 3×.

3. Method

3.1. Data collection

To the best of our knowledge, there is no public dataset for panoramic image enhancement with aligned image pairs prepared. To obtain best enhanced image, we collect low quality panoramic images and high quality perspective counterparts using different cameras instead of using synthetic data (by downsampling or Gaussian blur), then build a real world Pano-Hi dataset.

We use a consumer fisheye camera and a high-end APS-C camera. To obtain low quality panoramic image and high quality counterpart, we first set both cameras with reduced parallax according to [53] and [54]. Then, we take one photo using the fisheye camera and four or five photos using the APS-C camera by rotating 360 degrees along a tripod.

3.2. Patch matching for LQ-HQ pairs

The matching phase requires the same resolution and content for LQ and HQ patches. The content is confined by the field of view (FoV). In order to obtain the same FoV for LQ-HQ pairs, we have to downsample the APS-C camera image because panoramic cameras and APS-C cameras have different effective pixels per solid angle. The former's effective pixel is several times lower than that of the latter, resulting in poor image quality. For example, the effective pixel of a high-end APS-C camera: GR2 is 18 times that of a consumer panoramic fisheye Theta V camera. A visual comparison of images captured by these two cameras is shown in Figure 2.

The training requires precisely matched LQ-HQ pairs. In our work, we adopt a two-step strategy for matching. In the first step, we use feature matching and similarity transformation to match LQ and HQ images of a large size. In the second step, we extract small LQ and HQ patches from the matched LQ and HQ images, and use pixel mapping or pattern matching to align LQ and HQ patches precisely. The final matched patches are used for training. More details can be found in Section 2.2 of the supplementary material.

3.3. Network architecture

ESRGAN [11] improved SRGAN [18] by introducing the Residual-in-Residual Dense Blocks (RRDBs) with dense connections inside each block, which is more complex than SRGAN. Additionally, ESRGAN used perspective image datasets DIV2K [6] and Flickr2K [7] to train models, which are not suitable for real world panoramic image enhancement due to low data similarity as mentioned in Section 1.

We propose a compact network architecture which is based on ESRGAN but with several modifications. Figure 3 shows the network architecture of the generator.

Our main modification is motivated by [50], in which they decomposed feature maps into low and high frequency groups. We design a multi-frequency architecture employing RRDBs, which combines high and low frequency representation processes as depicted in Figure 3. The multi-frequency feature representation method stores the smooth-varying low-frequency feature maps in low-resolution tensors to reduce spatial redundancy. The multi-frequency structure consumes substantially less memory and computational resources than ESRGAN. We perform stride 2 convolutions to produce low frequency features. As a result, the receptive field is correspondingly enlarged by 2 times compared to RRDB.

The multi-frequency structure has two downsampling and two upsampling layers, as illustrated in Figure 3. The downsampling layer before the high-frequency representation process aims to reduce the amount of computation as well as the model inference time. In our case, we enhance 5K images, which is very computationally demanding. This layer helps us avoid the out of memory issue. The other downsampling layer aims to generate low-frequency feature maps. All two downsampling layers employ convolution with a stride of 2. The upsamping layer after the low-frequency representation process aims to restore the resolution of high-frequency feature maps. Similarly, the other upsamping layer aims to restore the resolution of the input image.

Another modification is the compression of RRDBs. We remove redundant RRDBs and some convolution layers from each dense block. Based on our observation, we find that simply stacking more RRDBs does not improve image quality significantly, while consumes much more memory and slows down the inference process (See results in Section 4). We also find experimentally that removing

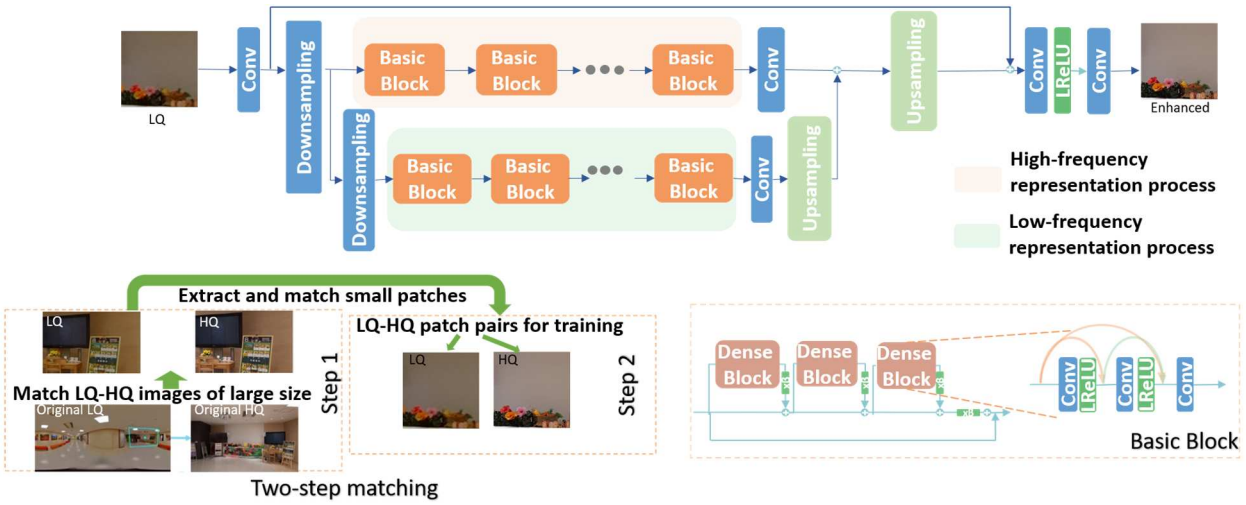


Figure 3: Generator of the proposed compact network. Our network is based on ESRGAN [11] with several modifications. The main modification is to design a multi-frequency structure, which combines high and low frequency representations. Another modification is the removal of redundant RRDBs and some convolution layers. These techniques reduce processing time significantly while maintaining similar image quality to ESRGAN. We adopt a two-step matching strategy to prepare perspective LQ-HQ patch pairs for training as described in section 3.2. Details about the matching algorithm can be found in Section 2.2 of the supplementary material.

some layers and blocks obtains similar performance at a very short time compared with the original ESRGAN. We reduce the number of RRDBs from 23 to 6, and the number of convolution layers with LeakeyReLU in each dense block from 5 to 3. The reduced blocks and layers result in reduced parameters and complexity in the network, thus require less processing time than ESRGAN. Finally, since the purpose of this work is to enhance panoramic image without changing the image resolution, we remove the last two upsampling layers from the original ESRGAN generator.

As to the discriminator, we only make a small modification. The original ESRGAN discriminator is composed of 9 basic blocks for input image size 128×128. In our discriminator, we delete two basic blocks to adapt to our input image size, which is 32×32.

3.4. Loss function

Similar to the original ESRGAN, we use content loss L1, perceptual loss defined by VGG features and adversarial loss defined by RaGAN [11].

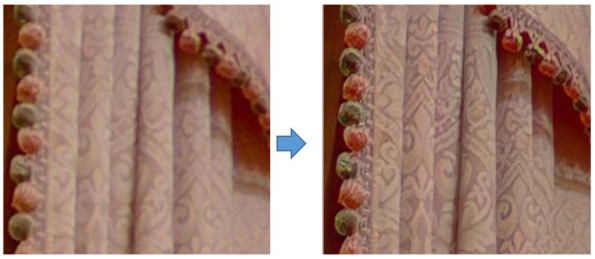
4. Experiments

4.1. Dataset

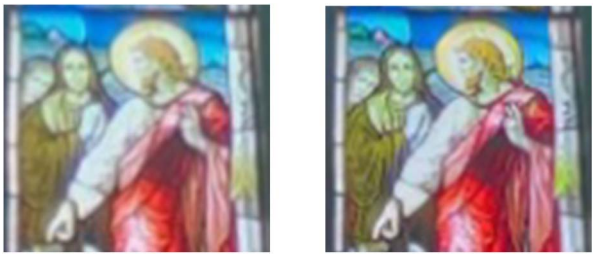
As mentioned in Section 1, datasets of perspective images do not work well for panoramic image enhancement due to low data similarity. Therefore we build a Pano-Hi dataset consisting of real world data instead. This dataset improves panoramic image quality effectively and solves the problems caused by low-cost optics: texture details, image clarity, noise and chromatic aberration (see Figure 4).

There is a color balance issue when using real world data taken by two different cameras. If one uses such images without any augmentation, stain artifacts may appear in the enhanced results (see supplementary material Section 3). Therefore, it's necessary to augment the real world data with synthetically created patches to overcome this issue. We use perspective image datasets to create synthetic data, which do not have color balance issue. Besides data augmentation, we can also solve this issue by conditioning our network (see Section 3 of the supplementary material for details). The final training data are a combination of patches from Pano-Hi and the perspective image datasets. The patch size is 32×32 and the numbers of low-high-quality pairs from these two datasets are 1,134k and 919k, respectively. We will use Pano-Hi to represent the combination of the aforementioned two datasets in the following paragraph for simplicity purpose.

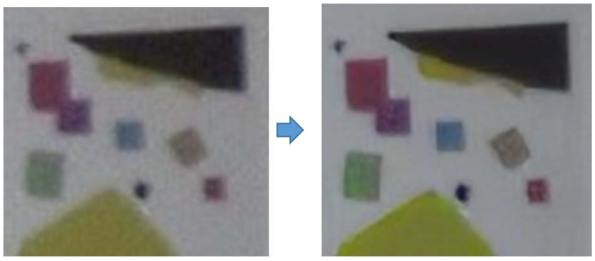
The testing set is selected from Pano-Hi dataset. To test our model, we select 39 panoramic images of different categories including indoor and outdoor, day and night, and different scenes including galleries, houses, museums, shopping malls, offices, gyms, exhibition centers, historical sites, bicycle parking lots and restaurants. Qualitative and quantitative results on the testing set are given in Section 4.3.



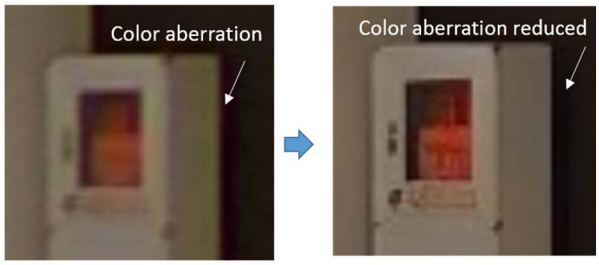
(a) Enhance texture details



(b) Enhance clarity



(c) Reduce noise



(d) Remove chromatic aberration (color fringe)

Figure 4: The proposed method is capable of (a) enhancing texture details (b)enhancing clarity (c) reducing noise and (d) removing chromatic aberration(color fringe). Please see this figure on a computer screen because it contains the effect of removing color fringes on the LQ image.

4.2. Training details

We use a small patch size 32×32 for both low and high-quality patches for enhancement purpose. 2053k pairs of LQ-HQ patches including augmented data are used for training. The minibatch size is set to 32, initial learning rate is 1×10^{-4} and decayed by a factor of 2 at 100k, 200k, 300k, 400k and 500k iterations. Adam [55] optimization was used with decay rates β_1= 0.9 and β_2= 0.999. The training stage takes 1.5 days using two Geforce GTX1080 Ti GPUs for the aforementioned setting. The network was trained from scratch without loading pre-trained models.

4.3. Results and discussion

In this sub section, we first show that our Pano-Hi dataset and the proposed compact network improves panoramic image quality from four aspects: (1) enhance texture details (2) enhance clarity (3) reduce noise and (4) remove chromatic aberration (color fringe). Then we visually compare results generated by our compact network and other state-of-the-art methods. It is followed by the comparison of results trained by Pano-Hi dataset and those by perspective dataset alone. Finally, we present quantitative evaluation results by using perceptual quality metrics, and the processing speed.

Figure 4 illustrates the quality improvement about the aforementioned four aspects. This suggests that we can overcome optical defects simply by using DNN based methods instead of using high-spec hardwares (e.g. multi-camera system).

Figure 5 compares enhanced images by different networks with some modifications. We show 3 images in Figure 5: one outdoor and two indoor scenes. We select SR networks RCAN [41], ESRGAN [11] and DBPN [56] and remove their upsampling layers for enhancement purpose (so as to maintain the same resolution between input and output images). Two well-known image enhancement methods based on GAN: DPED [1] and WESPE [2] are also selected for comparison. All models in Figure 5 are trained by using patches extracted from the Pano-Hi dataset. To show that datasets of perspective images do not work well for panoramic image enhancement due to low data similarity, we present our results trained only by perspective images in Figure 6. We use a combination of DIV2K [6] and Flickr2K [7], named as DIV2K_ Flickr2K datasets, and our network architecture to train a model, which generates images shown in the third column of Figure 6. Original and our network trained images using Pano-Hi dataset are also shown for comparison purpose.

We conduct both no-reference evaluation by perceptual index (PI) and full-reference evaluation by LPIPS [59]. PI consists of two no-reference image quality measures: Ma et al. [60] and NIQE et al. [61], expressed as $\text{PI} = \frac{1}{2}\big((10 - \text{Ma}) + \text{NIQE}\big)$. A lower PI means better image quality. The no-reference evaluation is conducted by using 39 panoramic images at 5K resolution (5376×2688) from the testing set. We compute average PI score and the average processing time for all 39 testing images for each method, and illustrate them in Table 1.

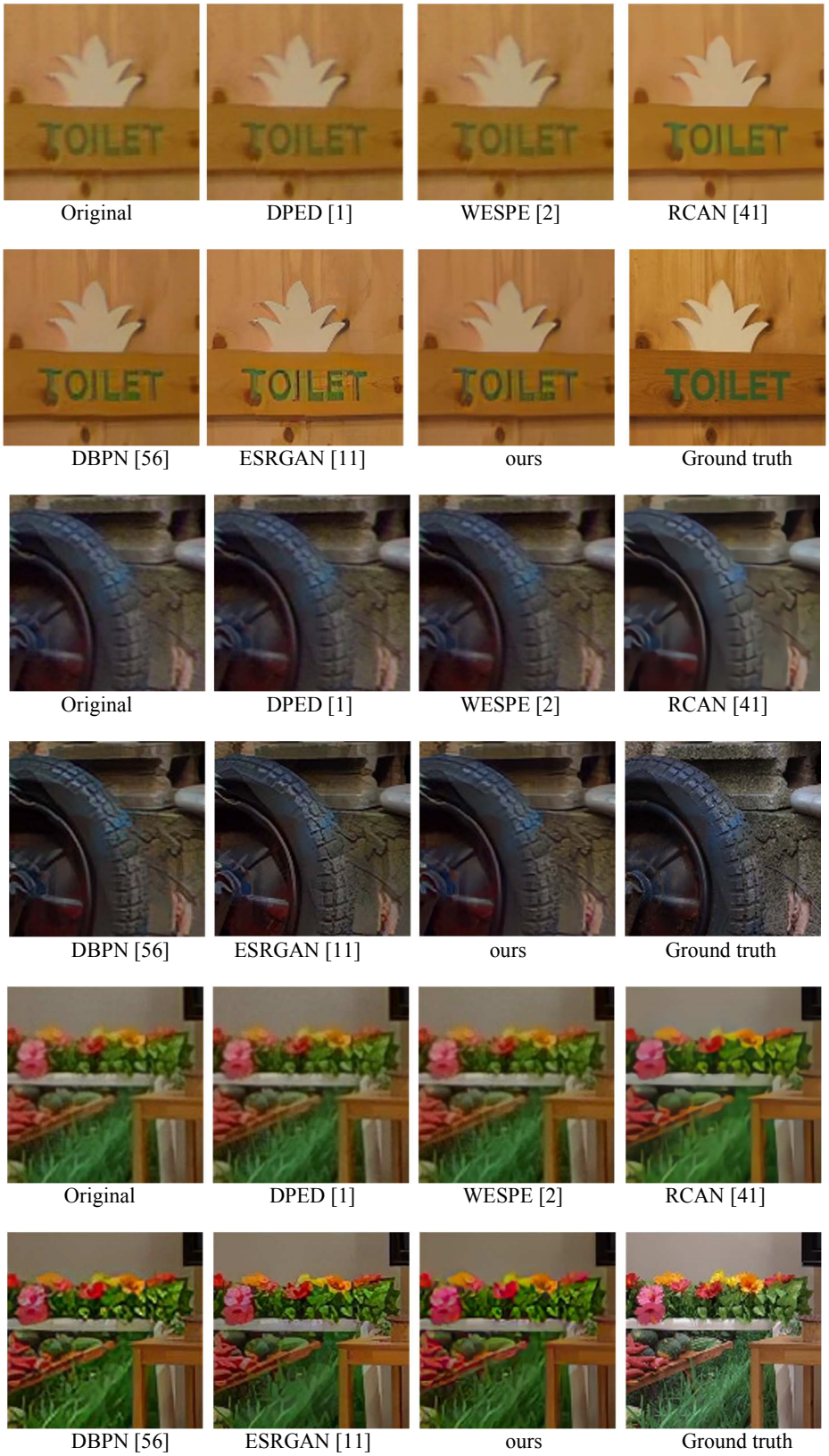


Figure 5: Visual comparison of ours and the state-of-the-art algorithms by using the Pano-Hi dataset.

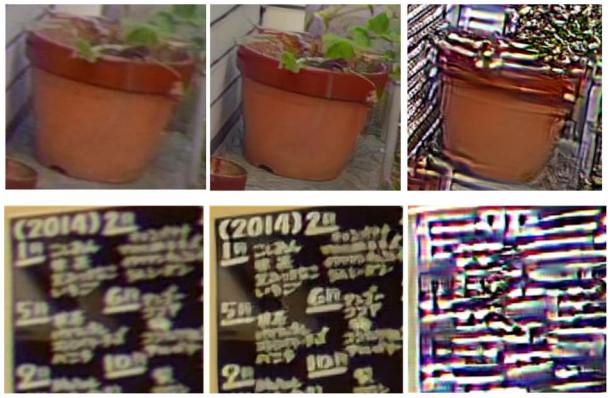


Figure 6: To demonstrate that the perspective image datasets do not work well for panoramic image enhancement due to low data similarity, we show the third column which is generated by using model trained only by DIV2K_ Flickr2K datasets of perspective images. The second column is generated by using Pano-Hi dataset. The original images are shown too in the first column.

As to full-reference evaluation, since there is no ground truth panoramic image, we extract more than 40k low-high-quality pairs of patch of resolution 128×128 from the testing set, and enhance the low quality patches. Then we compute LPIPS between the enhanced and high quality patches. A lower value indicates that a generated image is more similar to the ground truth. Results are shown in Table 2.

Models	PI	Time (sec)
DPED	4.51	2.92
WESPE	4.55	2.93
RCAN	5.19	192.53
ESRGAN	3.85	69.25
DBPN	4.07	9.56
ours	**3.74**	4.59

Table 1: No-reference evaluation. Perceptual index (PI) is the average score computed from 39 panoramic images of 5K (5376×2688) resolution. Processing time was run on one Geforce GTX 1080Ti GPU and averaged for 39 images.

Models	LPIPS
Original	0.782
DPED	0.782
WESPE	0.784
RCAN	0.774
ESRGAN	0.771
DBPN	0.775
ours	**0.769**

Table 2: Full-reference evaluation. The LPIPS of each row is the average score computed from more than 40k enhanced and high-quality pairs of patch of size 128 ×128.

As seen from Figure 5, ESRGAN, DBPN and our compact network generate images with richer and sharper textures than DPED, WESPE and RCAN. The visual quality of ours is almost indistinguishable from that of the ESRGAN. To demonstrate that we achieve the high-end image quality, we visually compare the enhanced image patches and the ground truth high-end (HQ) image in Figure 5. Note that we downsampled the ground truth image to obtain the same field of view (FoV) for LQ-HQ pairs for matching.

Table 1 indicates that our method obtains smaller PI score than ESRGAN (3.74 vs 3.85) while ours is 15.1 times faster. Although DPED and WESPE are 1.6 times faster than ours, their image quality in terms of PI scores is worse. Visual results in Figure 5 also suggest that DPED and WESPE generate blurrier images than ours. Since we do not have the ground truth panoramic image as HQ, we cannot compare it in Table 1.

Figure 6 shows that model trained with DIV2K_ Flickr2K dataset of perspective images possesses severe artifacts (third column) compared to that trained with Pano-Hi dataset (second column). This proves that low data similarity between training (perspective) and testing (panoramic) data leads to poor performance. We present more examples in Section 1 of the supplementary material.

The full-reference evaluation results in Table 2 also demonstrate that our method is more similar to the ground truth, obtains lowest score 0.769 among all 6 state-of-the-art methods for more than 40k small patches.

5. Conclusion and future work

In this paper, we propose a method which enhances low-quality 360° panoramic image to high-end camera (DSLR, compact camera with APS-C or full frame sensor) quality without changing image resolution. We build a Pano-Hi dataset consisting of panoramic equirectangular images and high quality counterparts and propose a two-step matching algorithm. We adopt an efficient GAN architecture modified from ESRGAN by employing multi-frequency structure with compressed RRDBs and convolutional layers within each dense block, and obtained good image quality visually and quantitatively at a high processing speed. Experiment shows that our method generates rich and sharp texture details, reduces noise and chromatic aberration at the same time. It also demonstrates superiority over several state-of-the-art DNN based methods in both no-reference and full-reference evaluations as well as the processing speed.

To achieve low cost and fast processing targets, we enhance image at the same resolution in this work. It is also possible to use SR techniques to further improve image quality.

References

[1] A.Ignatov, N.Kobyshev, R.Timofte, K.Vanhoey, and L .V. Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, pages 3277-3285, 2017.

[2] A.Ignatov, N.Kobyshev, R.Timofte, K.Vanhoey, and L .V. Gool. WESPE: weakly supervised photo enhancer for digital cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 691-700, 2018.

[3] S.K.Nayar, Catadioptric omnidirectional camera. In Proceedings of IEEE computer society conference on computer vision and pattern recognition, pages 482-488, 1997.

[4] T.E.Boult, X.Gao, R.Micheals, and M.Eckmann. Omni-directional visual surveillance. Image and Vision Computing, 22(7): 515-534, 2004.

[5] C.Jaramillo. Single-Image Omnidirectional Vision Systems: A Survey of Models, Calibration, and Applications. Doctoral dissertation, City University of New York, 2016.

[6] E.Agustsson and R.Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 126-135, 2017.

[7] R.Timofte et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 114-125, 2017.

[8] Y.Blau, R.Mechrez , R.Timofte, T.Michaeli, and L.Zelnik-Manor, The 2018 PIRM Challenge on Perceptual Image Super-Resolution. In European Conference on Computer Vision, pages 334-355, 2018

[9] M.S.Sajjadi, B.Scholkopf, and M.Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision, pages 4491-4500, 2017.

[10] Y.Blau and T.Michaeli. The perception-distortion tradeoff. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6228-6237, 2018.

[11] X.Wang , K.Yu , S.Wu, J.Gu, Y.Liu , C.Dong , C.C.Loy, Y.Qiao and X. Tang. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[12] L.Xu, J.S.Ren, C.Liu, and J.Jia. Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing Systems, pages 1790-1798, 2014.

[13] V.Jain and S.Seung. Natural image denoising with convolutional networks. In Advances in neural information processing systems, pages 769-776, 2009.

[14] H.C.Burger, C.J.Schuler, and S.Harmeling. Image denoising: Can plain neural networks compete with BM3D? In 2012 IEEE conference on computer vision and pattern recognition, pages 2392-2399, 2012.

[15] T.Plotz and S.Roth. Benchmarking denoising algorithms with real photographs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1586-1595, 2017.

[16] W.Ren, S.Liu, H.Zhang, J.Pan, X.Cao, and M.H.Yang. Single image dehazing via multi-scale convolutional neural networks. In European conference on computer vision, pages 154-169, 2016.

[17] C.Dong, C.C.Loy, K.He, and X. Tang. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38(2): 295-307, 2016.

[18] C.Ledig, L.Theis, F.Huszár, J.Caballero, A.Cunningham, A.Acosta, A.Aitken, A.Tejani, J.Totz, Z.Wang, and W.Shi. 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.

[19] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1637-1645, 2016.

[20] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision, pages 694-711, 2016.

[21] P.Isola, J.Y.Zhu, T.Zhou, and A.A.Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125-1134, 2017.

[22] Q.Chen, J.Xu, and V.Koltun. Fast image processing with fully-convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, pages 2497-2506, 2017.

[23] J.Kim, J.K.Lee, and K.M. 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.

[24] W.S.Lai, J.B.Huang, N.Ahuja, and M.H.Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 624-632, 2017.

[25] Y.Zhang, Y.Tian, Y.Kong, B.Zhong, and Y.Fu. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2472-2481, 2018.

[26] X.Zhang, Q.Chen, R.Ng, and V.Koltun. Zoom to Learn, Learn to Zoom. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3762-3770, 2019.

[27] C.Chen, Q.Chen, J.Xu, and V.Koltun. Learning to see in the dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3291-3300, 2018.

[28] C.Chen, Z.Xiong, X.Tian, ZZha, and F.Wu. Camera Lens Super-Resolution. arXiv preprint arXiv:1904.03378, 2019.

[29] J.Cai, H.Zeng, H.Yong, Z.Cao, and L.Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. arXiv preprint arXiv:1904.00523, 2019.

[30] J.Cai et al. Ntire 2019 challenge on real image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[31] A.Ignatov et al. PIRM challenge on perceptual image enhancement on smartphones: report. In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[32] I.Goodfellow, J.Pouget-Abadie, M.Mirza, B.Xu, D.Warde-Farley, S.Ozair, A.Courville, and Y.Bengio, Generative

adversarial nets. In Advances in neural information processing systems, pages 2672-2680, 2014.
[33] P.N.Michelini, D.Zhu, and H.Liu. Multi–scale Recursive and Perception–Distortion Controllable Image Super–Resolution. In European Conference on Computer Vision, pages 3-19, 2018,
[34] B.X.Chen, T.J.Liu, K.H.Liu, H.H.Liu, and S.C.Pei. Image Super-Resolution Using Complex Dense Block on Generative Adversarial Networks. In 2019 IEEE International Conference on Image Processing (ICIP), pages 2866-2870, 2019.
[35] J.H.Choi, J.H.Kim, M.Cheon, and J.S.Lee. Deep learning-based image super-resolution considering quantitative and perceptual quality. Neurocomputing. 2019.
[36] X.Wang, K.Yu, C.Dong, and C.C. Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 606-615, 2018.
[37] W.Zhang, Y.Liu, C.Dong, and Y.Qiao. RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 3096-3105, 2019.
[38] M.S.Rad, B.Bozorgtabar, U.V.Marti, M.Basler, H.K.Ekenel, and J.P.Thiran. SROBB: Targeted Perceptual Loss for Single Image Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 2710-2719, 2019.
[39] R.Mechrez, I. Talmi, F.Shama, and L.Zelnik-Manor. Maintaining natural image statistics with the contextual loss. In Asian Conference on Computer Vision, pages 427-443, 2018.
[40] R.Timofte et al. Ntire 2018 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 852-863, 2018.
[41] Y.Zhang, K.Li, K.Li, L.Wang, B.Zhong, and Y.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.
[42] W.Shi, J.Caballero, F.Huszár, J.Totz, A.P.Aitken, R.Bishop, D. Rueckert, and Z.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.
[43] B.Lim, S.Son, H.Kim, S.Nah, and K.M.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.
[44] H.Huang, R.He, Z. Sun, and T.Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 1689-1697, 2017.
[45] Z.Zhong, T.Shen, Y.Yang, Z.Lin, and C.Zhang. Joint sub-bands learning with clique structures for wavelet domain super-resolution. In Advances in Neural Information Processing Systems, pages 165-175, 2018.
[46] C.Dong, C.C.Loy, and X.Tang. Accelerating the super-resolution convolutional neural network. In European conference on computer vision, pages 391-407, 2016.
[47] Y.Cheng, J.Yan, and Z.Wang. Enhancement of Weakly Illuminated Images by Deep Fusion Networks. In 2019 IEEE International Conference on Image Processing (ICIP), pages 924-928, 2019..
[48] X.Deng. Enhancing image quality via style transfer for single image super-resolution. IEEE Signal Processing Letters, 25(4): 571-575, 2018.
[49] X.Deng, R.Yang, M.Xu, and P.L.Dragotti. Wavelet Domain Style Transfer for an Effective Perception-distortion Tradeoff in Single Image Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 3076-3085, 2019.
[50] Y.Chen, H.Fang, B.Xu, Z.Yan, Y.Kalantidis, M.Rohrbach, S. Yan, and J.Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. arXiv preprint arXiv:1904.05049, 2019.
[51] Q.Chang, K. Hung, and J.Jiang. Deep learning based image super-resolution for nonlinear lens distortions. Neurocomputing 275: 969-982, 2018.
[52] V.Fakour-Sevom, E.Guldogan, and J.K.Kämäräinen. 360 panorama super-resolution using deep convolutional networks. In Int. Conf. on Computer Vision Theory and Applications (VISAPP) (Vol. 1), 2018.
[53] J. Houghton. Finding The No-Parallax Point. http://www.johnhpanos.com/epcalib.htm, 2013.
[54] R.,Littlefield. Theory of the "No-Parallax" Point in Panorama Photography, 2006.
[55] D.P.Kingma, and J.Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[56] M.Haris, G. Shakhnarovich, and N.Ukita. Deep back-projection networks for super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1664-1673, 2018.
[57] A.Ignatov et al. NTIRE 2019 Challenge on Image Enhancement: Methods and Results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[58] Z.,Wang, A.C.Bovik, H.R. Sheikh, and E.P.Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4): 600-612, 2004.
[59] R. Zhang, P.Isola, A.A.Efros, E.Shechtman, and O.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.
[60] C.Ma, C.Y.Yang, X.Yang, and M.H.Yang. Learning a no-reference quality metric for single-image super-resolution. Computer Vision and Image Understanding, 158:1-16, 2017.
[61] A.Mittal, R.Soundararajan, and A.C.Bovik. Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters, 20(3): 209-212, 2012.