

Received August 31, 2021, accepted September 27, 2021, date of publication September 29, 2021, date of current version October 14, 2021. *Digital Object Identifier* 10.1109/ACCESS.2021.3116702

# Image Compression-Aware Deep Camera ISP Network

# KWANG-HYUN UHM<sup>®</sup>, KYUYEON CHOI, SEUNG-WON JUNG<sup>®</sup>, (Senior Member, IEEE), AND SUNG-JEA KO<sup>®</sup>, (Fellow, IEEE)

Department of Electrical Engineering, Korea University, Seongbuk-gu, Seoul 02841, South Korea

Corresponding author: Seung-Won Jung (swjung83@korea.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT Future Planning under Grant NRF-2020R1F1A1069009, and in part by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea Government (MSIT) under Grant 2019-0-00268, Development of SW Technology for Recognition, Judgment and Path Control Algorithm Verification Simulation and Dataset Generation.

**ABSTRACT** Several recent studies have attempted to fully replace the conventional camera image signal processing (ISP) pipeline with convolutional neural networks (CNNs). However, the previous CNN-based ISPs, simply referred to as ISP-Nets, have not explicitly considered that images have to be lossy-compressed in most cases, especially by the off-the-shelf JPEG. To address this issue, in this paper, we propose a novel compression-aware deep camera ISP learning framework. At first, we introduce a new use case of compression artifacts simulation network (CAS-Net), which operates in the opposite way of commonly used compression artifacts reduction networks. Then, the CAS-Net is connected with an ISP-Net such that the ISP network can be trained with consideration of image compression. Throughout experimental studies, we show that our compression-aware camera ISP network can produce images with a better tradeoff between bit-rate and image quality compared to its compression-agnostic version when the performance is evaluated after JPEG compression.

**INDEX TERMS** Camera ISP, compression artifacts, convolutional neural network, image compression.

# I. INTRODUCTION

Image signal processing (ISP) pipeline is used in modern digital cameras to convert raw camera sensor data to a highquality human-readable sRGB image. ISP pipeline consists of several operations including image demosaicing, denoising, white balance, color space conversion, gamma correction, tone mapping, and others [1]. Traditionally, each component of ISP pipeline is manually tuned by experts for a given camera, which is time-consuming and may yield the accumulated error in the final reconstructed sRGB image [2].

Recently, deep learning has been extensively applied to various image restoration tasks such as image demosaicing [3]–[5], denoising [6]–[8], super-resolution [9], [10], and tone mapping [11]–[14], and have demonstrated promising results. Since these tasks are required for in-camera ISP pipeline, several research endeavors have been made towards a design of a fully convolutional neural network (CNN)-based camera ISP [15]–[18], referred to as ISP-Net.

However, one of the common shortcomings of the previous ISP-Nets is that they have not explicitly considered that images have to be lossy-compressed in most cases, especially by the off-the-shelf JPEG [19]. Although endto-end learning-based compression has received significant interest [20], the classic JPEG is still in use today for most camera ISPs. If ISP-Nets are trained without considering image compression that is essentially followed, resultant images after compression can be obtained with a sub-optimal tradeoff between bit-rate and image quality.

In this paper, we propose a compression-aware ISP-Net learning framework that incorporates the JPEG compression procedure into the training of ISP-Nets. Since the JPEG compression is non-differentiable, we apply a fully-convolutional compression artifacts simulation network (CAS-Net), which can add JPEG compression artifacts to a given image. The CAS-Net can be simply trained by reversing the input and output required for training compression artifacts reduction networks. The CAS-Net is pretrained and then cascaded with an ISP-Net, and the parameters of the CAS-Net are fixed during the training of the ISP-Net. In this way, the ISP-Net can be trained with consideration of compression artifacts and thus can produce images with a better tradeoff between bit-rate and image quality compared to its compression-agnostic

The associate editor coordinating the review of this manuscript and approving it for publication was Paolo Crippa<sup>(D)</sup>.

# IEEE Access

version. Experimental results demonstrate the effectiveness of our compression-aware camera ISP network.

The rest of the paper is organized as follows. We review the related work on deep learning-based camera ISP network and compression artifacts simulation in Section II. Then we describe the proposed compression-aware camera ISP learning method in Section III. The experimental results and analysis are presented in Section IV, and finally the conclusions are given in Section V.

### **II. RELATED WORK**

#### A. DEEP CAMERA ISP LEARNING

Several studies have been conducted on replacing some subtasks in an ISP pipeline with deep CNNs [3], [9], [21]. Gharbi *et al.* [21] proposed a method to train a CNN model to address demosaicing and denoising jointly and achieved significant improvement in both tasks compared to the previous non-deep learning-based techniques. Liu *et al.* [3] proposed a self-guidance network for demosaicing and denoising based on green-channel guidance and density map guidance to better recover the high-frequency details. JDSR [9] presented a residual-dense squeeze-and-excitation network for joint demosaicing and super-resolution.

On the other hand, several pioneering works have investigated the application of deep CNNs on replacing the entire ISP pipeline [8], [15]–[18], [22], [23]. DeepISP [15] is one of the first attempts that replaces an entire ISP pipeline by CNNs. Specifically, a two-stage network is proposed to extract low-level and high-level features, where the first and second stages apply local operations such as demosaicing and denoising and global operations such as tone adjustment and color correction, respectively. This design principle makes the ISP-Net easy to share information across different tasks. DeepCamera [22] is a light-weight CNN that replaces the ISP pipeline completely. To be more specific, DeepCamera is designed to perform defective pixel correction, denoising, white balancing, exposure correction, demosaicing, color transform, and gamma encoding. Chen et al. [23] develop a deep CNN model to convert a raw low-light sensor data to a long-exposure high-quality sRGB image. To train this model, they collected raw images captured with short-exposure in low-light conditions. PyNET [16] presented a framework that is independent of mobile ISPs by using the raw images from a mobile camera and their corresponding sRGB images from a DSLR camera. Moreover, PyNET inputs a low-resolution image to the bottom layer and applies additional layers with increasing scales to combine global and local features. W-Net [17] improved the standard U-Net [24] by designing a cascaded U-Net model. In addition, the color loss was introduced to make the ISP-Net robust against the misalignment between raw and sRGB image pairs. CycleISP [8] presented a framework for learning both forward and backward pass of ISP using CNNs to synthesize raw images from sRGB images. CameraNet [18] proposed a two-stage network that sequentially applies restoration and enhancement.



**FIGURE 1.** Compression artifacts simulation network: (a) Training procedure and (b) example results (Best viewed on screen).

# B. COMPRESSION ARTIFACTS SIMULATION

The JPEG image compression is widely used in digital cameras for storing sRGB images with reduced bits. The JPEG lossy compression inevitably leads to artifacts such as blocking and blurring artifacts in the compressed images, where the amount of degradation can be adjusted by the JPEG quality factor (QF) parameter. If a neural network is used in conjunction with off-the-shelf image compression methods, *e.g.*, ISP-Net followed by the JPEG compression, compression-aware learning is required to produce high-quality compressed images. However, since the quantization process included in the JPEG compression is inherently non-differentiable, the compression procedure cannot be directly integrated into an end-to-end learning framework.

To overcome the aforementioned problem, several studies have attempted to incorporate image compression into the training of end-to-end neural networks. For example, Ballé *et al.* [20] added uniform noise to the low dimensional features such that the network can take the quantization noise into account during training and applied entropy coding to the low dimensional features for image compression. Recently, towards the design of invertible camera ISP [2], a differentiable JPEG simulator (DJS) was proposed to reconstruct raw data from JPEG images. Since the rounding function used in the quantization step is non-differentiable, a differentiable



FIGURE 2. Comparison between (a) the compression-agnostic learning and (b) the proposed compression-aware learning of ISP-Net.

approximation of the rounding function is proposed using the Fourier series expansion. Other works on differentiable rounding function can also be found in [25]-[28]. For example, Shin and Song [25] approximated the rounding operation using the third-order polynomial such that it has non-zero derivatives almost everywhere. Theis et al. [26] replaced the derivative of the rounding function with the derivative of its approximation, where the identity function was used for the smooth approximation. In other words, the rounding is performed as usual in the forward path while the gradients are simply bypassed in the backward path. Gong et al. [27] proposed a differentiable soft quantization function that approximates discrete rounding using tanh functions. Recently, Son et al. [29] investigated the utility of CNNs in explicitly imitating image degradation caused by image compression. Specifically, by considering the characteristics of the image compression process, they introduced the auxiliary codec network (ACN) that can synthesize compression artifacts such as contouring and ringing artifacts in the output image. The effectiveness of the ACN was demonstrated by training their compact representation network with the ACN for image compression.

We apply the same principle of the aforementioned methods [2], [20], [29] to train the network located in front of image compression, i.e., ISP-Net, in an end-to-end manner. To this end, we use the standard U-Net structure for CAS-Net due to their simplicity and generalizability and train it using pairs of JPEG compressed and original images as shown in Fig. 1(a). Since quality factors (QFs) used in modern camera ISPs are relatively high (80 and 90 were used in our experiment), resultant JPEG compressed images suffer from weak but non-marginal compression artifacts; our CAS-Net successfully mimics such JPEG compression artifacts including blocking and contouring artifacts as shown in Fig. 1(b). The performance comparison and analysis of compression artifacts simulation methods will be provided in Section IV.

# III. COMPRESSION-AWARE CAMERA ISP NETWORK

## A. FRAMEWORK

Given a set of raw images  $\mathcal{X}$  and their corresponding sRGB images  $\mathcal{Y}$ , our goal is to learn an ISP-Net  $f : \mathcal{X} \to \mathcal{Y}$ , such that, for a pair  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ , the compressed version of the reconstructed sRGB image c(f(x)) matches the target compressed image c(y), where  $c(\cdot)$  denotes the JPEG compression. To facilitate our compression-aware ISP-Net optimization process, we replace the non-differentiable JPEG compression procedure with our differentiable CAS-Net  $g: y \mapsto c(y)$ , so that the ISP-Net f is trained to match the compression artifacts simulation result of the reconstructed sRGB image g(f(x)) to that of the target sRGB image g(y). The overall pipeline of the proposed method is illustrated in Fig. 2(b), and for comparison, the compression-agnostic ISP-Net learning is shown in Fig. 2(a). We note that the CAS-Net g is pretrained and then fixed during the training of ISP-Net f, and the CAS-Net g is only used in the training phase and not used in the testing phase, where the actual JPEG compression is applied to the output of the ISP-Net f(x) at test time. We perform simple bilinear interpolation for demosaicing of Bayer raw images, and the demosaiced images are then fed to the ISP-Net f to produce sRGB images.

#### **B. LOSS FUNCTION**

We first train the CAS-Net g with the given sRGB images  $\mathcal{Y}$  and their JPEG compressed counterparts using  $L_{cas}$ , which is expressed as:

$$\mathcal{L}_{cas} = \mathbb{E}_{y} \| g(y) - c(y) \|_{1}, \tag{1}$$

where  $\|\cdot\|_1$  measures the *L*1 loss.

1

Next, we train the ISP-Net f with the given raw images  $\mathcal{X}$  and the corresponding sRGB images  $\mathcal{Y}$  using the sRGB reconstruction loss, denoted as  $L_{isp}$ , which measures the L1-distance between the reconstructed sRGB image f(x) and the ground-truth sRGB image y.  $L_{isp}$  is defined as:

$$\mathcal{L}_{isp} = \mathbb{E}_{x} \| f(x) - y \|_{1}.$$
<sup>(2)</sup>

Note that the training of ISP-Net using only  $L_{isp}$  corresponds to compression-agnostic ISP learning as illustrated in Fig. 2(a).

After the CAS-Net g and ISP-Net f are trained, we finetune the ISP-Net f using the combined loss  $L_{total}$ , which is formulated as follows:

$$\mathcal{L}_{total} = \lambda_{isp} \mathcal{L}_{isp} + \lambda_{comp} \mathcal{L}_{comp}, \qquad (3)$$

where  $\lambda_{isp}$  and  $\lambda_{comp}$  are hyper-parameters that balances the two loss terms, and  $\mathcal{L}_{comp}$  is the compression loss which computes the *L*1-distance between the compression artifacts simulation result of the reconstructed sRGB image, *i.e.*, g(f(x)), and that of the ground-truth sRGB image, *i.e.*, g(y).  $\mathcal{L}_{comp}$  can be expressed as:

$$\mathcal{L}_{comp} = \mathbb{E}_{x} \| g(f(x)) - g(y) \|_{1}.$$
(4)

The use of  $\mathcal{L}_{comp}$  encourages the ISP-Net f to produce sRGB images that are effective for image compression.

# C. NETWORK ARCHITECTURE

We use the U-Net structure with channel attention module [17] for the ISP-Net f, which has been demonstrated to be effective in the raw to RGB mapping task. The model f consists of the encoder, decoder, and skip connections between them. The encoder part is composed of four levels of convolutional blocks, where each block has three  $3 \times 3$  convolutional layers, each of which is followed by a Leaky ReLU. After each block,  $2 \times 2$  max-pooling with stride 2 is applied for down-sampling. The decoder part is also composed of four-level convolutional blocks with  $2\times$ bilinear up-sampling, and the channel attention module is added at the end of each block, where the channel attention vector is obtained using the global average pooling, two fully connected layers, and the sigmoid function. The output of the channel attention module is then channel-wise multiplied with the feature map. The channel dimension of the first level of the encoder and decoder is 32 and then doubled at each block level. The architecture of the CAS-Net g is similar to the ISP-Net f without the channel attention module, which is simply referred to as U-Net in SectionIV.

# **IV. EXPERIMENTAL RESULTS**

### A. DATASET

To train and test our proposed framework, we used the raw images from the MIT-Adobe FiveK dataset [30]. More specifically, we collected 487 raw images captured by the Nikon D700 camera. Likewise with [2], [8], we rendered the sRGB images from the collected raw images using the LibRaw library [31], which provides the most representative in-camera ISP pipeline. Then, we compressed the rendered sRGB images using the JPEG with the two QFs of 80 and 90. We split the dataset into a ratio of 80:5:15 for training, validation, and testing of our model. Note that the model was trained individually for each QF.

# **B. IMPLEMENTATION DETAILS**

We used PyTorch [32] and a single NVIDA Titan Xp GPU. During network training, we randomly cropped the images with the patch size of 448 × 448, and normalized the pixel values in the range of [0, 1]. The common data augmentations including flipping and rotation were applied. We used the Adam optimizer [33] with a learning rate of 0.0001 to train our CAS-Net and ISP-Net, while the learning rate was decayed by a factor of 0.1 when fine-tuning the ISP-Net using  $\mathcal{L}_{total}$ . The CAS-Net was trained using  $\mathcal{L}_{cas}$  for 30 epochs with a batch size of 64, and the ISP-Net was pretrained using  $\mathcal{L}_{isp}$  for 30 epochs with a batch size of 32. Finally, the ISP-Net was fine-tuned using  $\mathcal{L}_{total}$  for 10 epochs with a batch size of 16. We empirically set the hyper-parameters  $\lambda_{isp}$  and  $\lambda_{comp}$ to 0.1 and 1, respectively.



FIGURE 3. Visual comparison of compression artifacts simulation methods. For each result, the error map of the local region is visualized and the PSNR / SSIM values are presented.

#### TABLE 1. Quantitative results (PSNR / SSIM) of different methods for compression artifacts simulation.

| QF | Method         |                |                |  |  |
|----|----------------|----------------|----------------|--|--|
|    | U-Net          | ACN            | DJS            |  |  |
| 80 | 43.84 / 0.9805 | 43.58 / 0.9777 | 42.35 / 0.9703 |  |  |
| 90 | 44.69 / 0.9800 | 44.59 / 0.9796 | 43.02 / 0.9715 |  |  |

# C. COMPRESSION ARTIFACTS SIMULATION RESULTS

We compare the quantitative and qualitative performance of three compression artifacts simulation methods on our test set, namely U-Net, ACN [29], and DJS [2]. Note that U-Net and ACN are trained using sRGB-JPEG image pairs while DJS does not require training. We train U-Net and ACN models for each QF of JPEG compression. To evaluate the quality of the simulated JPEG images produced by the above methods, we measure peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [34] between simulated and real JPEG-compressed images.

Table 1 shows the average PSNR and SSIM values of all methods on the test set. It can be seen that U-Net achieves the highest PSNR and SSIM scores at both QFs. The performance gap between DJS and U-Net (1.49dB at QF = 80) is much more significant than between ACN and U-Net

137828

(0.27dB at QF = 80). There is a clear advantage of using the CNN-based methods over the approximated rounding function-based methods for compression artifacts simulation in terms of PSNR and SSIM. Though the architecture of ACN is specially designed to imitate the JPEG compression process by rearranging the  $8 \times 8$  non-overlapping blocks of the input image into the channel axis, we found that the simple U-Net structure is capable of achieving higher performance. We believe our CAS-Net can be generalized to other codecs such as JPEG2000 [35] and HEVC [36].

Fig. 3 shows qualitative comparisons of different methods for the JPEG simulation. We visualize the error map of the compression simulation for each method. As can be seen, DJS results in a significant difference between the simulated and real compressed images, indicating its limited performance in compression artifacts simulation. Meanwhile, neural network-based methods, *i.e.*, U-Net and ACN, show much fewer differences.

### D. COMPRESSION-AWARE ISP LEARNING RESULTS

We evaluate the effectiveness of our compression-aware ISP learning method on the test set. To this end, we compare the performance of our method to that of the compression-agnostic ISP learning baseline in terms of the

# **IEEE**Access



FIGURE 4. Visual comparison between compression agnostic and compression aware ISP-Nets: (left) GT sRGB, (center) resultant image of the compression-agnostic ISP-Net followed by JPEG compression (QF = 90), and (right) resultant image of the compression-aware ISP-Net followed by JPEG compression (QF = 90).

bit-rate and image quality. We apply the JPEG compression to the sRGB images reconstructed by the ISP-Net from the raw data with the two QFs, and then measure the average bits per pixel (BPP) and PSNR of the compressed images. The PSNR is computed between the compressed version of the reconstructed sRGB images and that of the ground-truth sRGB images. The three methods for the compression artifacts simulation, which are mentioned in Section IV-C, are used as variants of the JPEG simulation module g in the training of our compression-aware ISP-Net. Table 2 summarizes the average PSNR and BPP results. The compression-agnostic baseline corresponds to the ISP-Net that is trained using only  $\mathcal{L}_{isp}$ . It can be clearly seen that the proposed compression-aware ISP-Net outperforms the baseline by a large margin in terms of PSNR at the similar BPP on both QFs. When the CAS-Net is not fixed during the training of the ISP-Net, the performance is not better than its parameter-fixed version because an unnecessary update of the CAS-Net makes the ISP-Net ineffective at the test stage. Note that the DJS is a parameter-free module. Among

| QF | Method               | PSNR / BPP     |               |                      |               |  |
|----|----------------------|----------------|---------------|----------------------|---------------|--|
| 80 | Compression-agnostic | 35.54 / 0.541  |               |                      |               |  |
|    |                      | CAS-Net freeze | U-Net         | ACN                  | DJS           |  |
|    | Compression-aware    | X              | 36.74 / 0.567 | 36.76 / <b>0.545</b> | n.a.          |  |
|    |                      | $\checkmark$   | 36.94 / 0.547 | <b>36.80</b> / 0.548 | 36.90 / 0.547 |  |
|    | Compression-agnostic | 35.88 / 0.900  |               |                      |               |  |
| 90 |                      | CAS-Net freeze | U-Net         | ACN                  | DJS           |  |
|    | Compression-aware    | X              | 37.15 / 0.916 | 37.24 / <b>0.911</b> | n.a.          |  |
|    |                      | √              | 37.22 / 0.904 | <b>37.40</b> / 0.916 | 37.10/0.917   |  |

#### TABLE 2. Quantitative comparison results (PSNR / BPP) of different solutions for ISP-Net learning at JPEG QFs of 80 and 90.



FIGURE 5. Comparison of the average BPP and PSNR for the results of the compression-agnostic and compression-aware ISP-Nets, followed by JPEG compression with different QFs. "Ground-truth" indicates the performance for the compressed ground-truth sRGB images with the different QFs.

TABLE 3. Ablation study on the effect of parameter settings.

| QF | $\lambda_{isp}$ | $\lambda_{comp}$ | Method               |                      |                      |
|----|-----------------|------------------|----------------------|----------------------|----------------------|
|    |                 |                  | U-Net                | ACN                  | DJS                  |
| 80 | 1               | 1                | 36.86 / <b>0.546</b> | <b>36.80</b> / 0.548 | 36.76 / <b>0.544</b> |
|    | 1               | 0.1              | 36.71 / 0.548        | 36.73 / 0.548        | 36.60 / 0.549        |
|    | 0.1             | 1                | <b>36.94</b> / 0.547 | 36.80 / 0.548        | <b>36.90</b> / 0.547 |
| 90 | 1               | 1                | 37.17 / 0.910        | 37.37 / <b>0.915</b> | 37.09 / <b>0.905</b> |
|    | 1               | 0.1              | 37.21 / 0.912        | 37.24 / 0.917        | 37.07 / 0.911        |
|    | 0.1             | 1                | 37.22 / 0.904        | <b>37.40</b> / 0.916 | <b>37.10</b> / 0.917 |

the three CAS-Net variants, the U-Net achieved the best performance on QF = 80, and the ACN achieved the best performance on QF = 90 in terms of PSNR. We highlight that with any CAS-Net variants, a large performance gain over the compression-agnostic baseline can be achieved, which demonstrates the effectiveness of our ISP-Net learning framework.

Fig. 4 shows the qualitative comparison results for the compression-agnostic and compression-aware ISP Nets on QF = 90. We visualize the error map between the resultant image of the ISP-Net followed by the JPEG compression and the ground-truth sRGB image. The PSNR and BPP

of the results are also shown. The results obtained by the compression-agnostic baseline show large errors, which indicates that the model produces sub-optimal results considering that the reconstructed sRGB images need to be passed through the JPEG compression. On the other hand, it can be clearly seen that the proposed compression-aware ISP-Net followed by JPEG compression produces images much less distortion compared to the baseline method. These results show the potential of our approach for the practical application of ISP-Nets.

#### E. ABLATION STUDY

We perform an ablation study on the selection of the hyperparameters  $\lambda_{isp}$  and  $\lambda_{comp}$ . We compare the performance of our proposed model with different parameter settings, where  $\lambda_{isp}$  and  $\lambda_{comp}$  are varied from {0.1, 1}. Table 3 shows the results. We observe that the parameter setting of  $\lambda_{isp} = 0.1$ and  $\lambda_{comp} = 1$  achieves the best performance in overall in terms of PSNR for all CAS-Net variants on both QFs. It can be clearly seen that lowering  $\lambda_{comp}$  negatively affects the performance both in PSNR and BPP, which demonstrates the importance of  $\mathcal{L}_{comp}$  for the training of our compressionaware ISP-Net.

### F. GENERALIZATION TO UNSEEN QFs

Fig. 5 shows the average PSNR and BPP of the test images obtained after applying the JPEG compression with different QFs to the reconstructed sRGB images. Although the CAS-Net was trained using the fixed QF of 90, the proposed compression-aware ISP-Net shows consistent improvements over the baseline for all test QFs from 80 to 96 with the interval of 2, demonstrating its generalizability.

#### **V. CONCLUSION**

In this letter, we proposed a compression-aware camera ISP learning strategy for effective sRGB reconstruction with respect to image compression. To this end, the fully convolutional CAS-Net was used to mimic the non-differentiable JPEG compression procedure and incorporated into the training of the ISP-Net to make the reconstruction process consider image compression. Experimental results demonstrated that the images obtained by our compression-aware ISP-Net have better image quality compared to the compressionagnostic ISP-Net. Several future studies are being considered. First, in this study, we applied the U-Net structure for both CAS-Net and ISP-Net. Advanced architecture design can further boost the performance, which is left for our future study. Second, ISP-Net can be extended to take multiple frames as an input for better image reconstruction. Finally, we plan to develop a fully end-to-end ISP-Net that produces compressed bitstreams from Bayer raw images without relying on JPEG.

# ACKNOWLEDGMENT

(Kwang-Hyun Uhm and Kyuyeon Choi contributed equally to this work.)

#### REFERENCES

- H. C. Karaimer and M. S. Brown, "A software platform for manipulating the camera imaging pipeline," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 429–444.
- [2] Y. Xing, Z. Qian, and Q. Chen, "Invertible image signal processing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 6296–6387.
- [3] L. Liu, X. Jia, J. Liu, and Q. Tian, "Joint demosaicing and denoising with self guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2020, pp. 2240–2249.
- [4] W. Xing and K. Egiazarian, "End-to-end learning for joint image demosaicing, denoising and super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 3507–3516.
- [5] S. M. A. Sharif, R. Ali Naqvi, and M. Biswas, "Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops* (CVPRW), Jun. 2021, pp. 233–242.
- [6] Y. Wang, X. Song, and K. Chen, "Channel and space attention neural network for image denoising," *IEEE Signal Process. Lett.*, vol. 28, pp. 424–428, Feb. 2021.
- [7] D.-W. Kim, J. R. Chung, and S.-W. Jung, "GRDN:Grouped residual dense network for real image denoising and GAN-based real-world noise modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops* (CVPRW), Jun. 2019, pp. 2086–2094.
- [8] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "CycleISP: Real image restoration via improved data synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2696–2705.
- [9] X. Xu, Y. Ye, and X. Li, "Joint demosaicing and super-resolution (JDSR): Network design and perceptual optimization," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 968–980, Jun. 2020.
- [10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
- [11] N. Robidoux, L. E. G. Capel, D.-E. Seo, A. Sharma, F. Ariza, and F. Heide, "End-to-end high dynamic range camera pipeline optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 6297–6307.
- [12] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, "Deep tone mapping operator for high dynamic range images," *IEEE Trans. Image Process.*, vol. 29, pp. 1285–1298, 2020.
- [13] C. Guo and X. Jiang, "Deep tone-mapping operator using image quality assessment inspired semi-supervised learning," *IEEE Access*, vol. 9, pp. 73873–73889, 2021.
- [14] K. Panetta, L. Kezebou, V. Oludare, S. Agaian, and Z. Xia, "TMO-Net: A parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale HDR dataset," *IEEE Access*, vol. 9, pp. 39500–39517, 2021.
- [15] E. Schwartz, R. Giryes, and A. M. Bronstein, "DeepISP: Toward learning an end-to-end image processing pipeline," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 912–923, Feb. 2019.
- [16] A. Ignatov, L. Van Gool, and R. Timofte, "Replacing mobile camera ISP with a single deep learning model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 536–537.
- [17] K.-H. Uhm, S.-W. Kim, S.-W. Ji, S.-J. Cho, J.-P. Hong, and S.-J. Ko, "W-Net: Two-stage U-Net with misaligned data for raw-to-RGB mapping," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3636–3642.

- [18] Z. Liang, J. Cai, Z. Cao, and L. Zhang, "CameraNet: A two-stage framework for effective camera ISP learning," *IEEE Trans. Image Process.*, vol. 30, pp. 2248–2262, 2021.
- [19] W. B. Pennebaker and J. L. Mitchell, JPEG Still Image Data Compression Standard. New York, NY, USA: Van Nostrand Reinhold, 1992.
- [20] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–23.
- [21] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," ACM Trans. Graph., vol. 35, no. 6, pp. 1–12, Nov. 2016.
- [22] S. Ratnasingam, "Deep camera: A fully convolutional neural network for image signal processing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3868–3878.
- [23] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Computer-Assisted Intervent*, 2015, pp. 234–241.
- [25] R. Shin and D. Song, "JPEG-resistant adversarial images," in Proc. Neural Inf. Process. Syst. Workshops, Dec. 2017, pp. 1–6.
- [26] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–19.
- [27] R. Gong, X. Liu, S. Jiang, T. Li, P. Hu, J. Lin, F. Yu, and J. Yan, "Differentiable soft quantization: Bridging full-precision and low-bit neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4852–4861.
- [28] J. Choi and B. Han, "Task-aware quantization network for JPEG image compression," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 309–324.
- [29] H. Son, T. Kim, H. Lee, and S. Lee, "Enhanced standard compatible image compression framework based on auxiliary codec networks," 2020, arXiv:2009.14754. [Online]. Available: http://arxiv.org/abs/2009.14754
- [30] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. CVPR*, Jun. 2011, pp. 97–104.
- [31] LibRaw. (2021). *Raw Image Decoder*. [Online]. Available: https://www.libraw.org/
- [32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. for Learn. Represent., 2015, pp. 1–15.
- [34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [35] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek, "An overview of JPEG-2000," in *Proc. DCC Data Compress. Conf.*, Mar. 2000, pp. 523–541.
- [36] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.



**KWANG-HYUN UHM** received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include image processing and computer vision.



**KYUYEON CHOI** received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2020, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include image processing and computer vision.



**SUNG-JEA KO** (Fellow, IEEE) received the B.S. degree in electronic engineering from Korea University, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from The State University of New York at Buffalo, in 1986 and 1988, respectively.

From 1988 to 1992, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Michigan at Dearborn. In 1992, he joined the Department of

Electronic Engineering, Korea University, where he is currently a Professor. He has published more than 210 international journal articles. He holds more than 60 registered patents in fields, such as video signal processing and computer vision. He is a member of the National Academy of Engineering of Korea. He was a recipient of the Best Paper Award from the IEEE Asia-Pacific Conference on Circuits and Systems in 1996; the Hae-Dong Best Paper Award from the Institute of Electronics and Information Engineers (IEIE) in 1997; the 1999 LG Research Award; the Research Excellence Award from Korea University in 2004; the Technical Achievement Award from the IEEE Consumer Electronics (CE) Society in 2012; the 15-Year Service Award from the TPC of ICCE in 2014; the Chester Sall Award (First Place Transaction Paper Award) from the IEEE CE Society in 2017; and the Science and Technology Achievement Medal from the Korean Government in 2020. He has served as the General Chairperson for ITC-CSCC 2012 and the General Chairperson for IEICE 2013. He was the President of IEIE, in 2013; the Vice President of the IEEE CE Society, from 2013 to 2016; and the Distinguished Lecturer of the IEEE, from 2015 to 2017. He is also an Editorial Board Member of the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS.



**SEUNG-WON JUNG** (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2005 and 2011, respectively. From 2011 to 2012, he was a Research Professor with the Research Institute of Information and Communication Technology, Korea University. From 2012 to 2014, he was the Research Scientist of Samsung Advanced Institute of Technology, Yongin, South Korea. From 2014 to 2020, he was

an Assistant Professor with the Department of Multimedia Engineering, Dongguk University, Seoul. In 2020, he joined the Department of Electrical Engineering, Korea University, where he is currently an Associate Professor. He has published more than 60 peer-reviewed articles in international journals. His current research interests include image processing and computer vision. He received the Hae-Dong Young Scholar Award from the Institute of Electronics and Information Engineers in 2019.

. . .