IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 21, 2020, accepted October 28, 2020, date of publication November 2, 2020,
date of current version November 11, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3035086*

# Detail Restoration and Tone Mapping Networks for X-Ray Security Inspection

**HYO-YOUNG KIM[1], (Member, IEEE), SEUNG PARK[2], YONG-GOO SHIN[3], SEUNG-WON JUNG[1], (Senior Member, IEEE), AND SUNG-JEA KO[1], (Fellow, IEEE)**

[1]Department of Electrical Engineering, Korea University, Seoul 02841, South Korea
[2]Medical AI Research Center, Samsung Medical Center, Seoul 06351, South Korea
[3]Division of Smart Interdisciplinary Engineering, Hannam University, Daejeon 34430, South Korea

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

This work was supported by the Institute for Information and communications Technology Promotion (IITP) grant funded by the Korean Government (MSIT), Development of SW Technology for Recognition, Judgment and Path Control Algorithm Verification Simulation and Dataset Generation, under Grant 2019-0-00268.

**ABSTRACT** X-ray imaging is one of the most widely used security measures for maintaining airport and transportation security. Conventional X-ray imaging systems typically apply tone-mapping (TM) algorithms to visualize high-dynamic-range (HDR) X-ray images on a standard 8-bit display device. However, X-ray images obtained through traditional TM algorithms often suffer from halo artifacts or detail loss in inter-object overlapping regions, which makes it difficult for an inspector to detect unsafe or hazardous objects. To alleviate these problems, this article proposes a deep learning-based TM method for X-ray inspection. The proposed method consists of two networks called detail-recovery network (DR-Net) and TM network (TM-Net). The goal of DR-Net is to restore the details in the input HDR image, whereas TM-Net aims to compress the dynamic range while preserving the restored details and preventing halo artifacts. Since there are no standard ground-truth images available for the TM of X-ray images, we propose a novel loss function for unsupervised learning of TM-Net. We also introduce a dataset synthesis technique using the Beer-Lambert law for supervised learning of DR-Net. Extensive experiments comparing the performance of our proposed method with state-of-the-art TM methods demonstrate that the proposed method not only achieves visually compelling results but also improves the quantitative performance measures such as FSITM and HDR-VDP-2.2.

**INDEX TERMS** Convolutional neural network, high dynamic range, tone mapping, unsupervised learning, X-ray imaging.

## I. INTRODUCTION

To detect unsafe or hazardous objects quickly and in a non-invasive manner, X-ray inspection systems have been extensively used in many security applications [1]. Unlike conventional imaging systems, which measure the light reflected by an object, X-ray inspection systems capture high-dynamic-range (HDR) images by measuring the amount of photons passing through objects [2]. To visualize HDR X-ray images that have 12- to 16-bit precision on a standard 8-bit display device, X-ray inspection systems often apply tone mapping (TM), which shrinks the intensities of the HDR image to the target display range [1], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan.

In general, TM methods can be classified into global and local TM methods. Global TM methods apply the same mapping function to all the pixels in the HDR image [4]–[11]. Ward [5] used a simple linear function to compress image contrast instead of the absolute luminance. Ferwerda *et al.* [6] proposed a visual adaptation model based on psychophysical experiments incorporating threshold visibility, color appearance, visual acuity, and sensitivity over time. Larson *et al.* [4] applied a histogram adjustment technique to preserve the histogram distribution of the original HDR image. Drago *et al.* [7] employed adaptive logarithmic mapping to preserve details and contrast. Reinhard and Devlin [9] proposed a global operator based on the photoreceptor response of the cones in the human eye. Khan *et al.* [10] used a luminance histogram to construct a

lookup table for TM. Khan *et al.* [11] applied a histogram-based TM after perceptual quantization to enhance the dark regions and compress the bright ones. Although global TM is simple, fast, and can preserve the intensity orders of the original scenes, it does not sufficiently enhance local image contrast and often results in loss of detail, which are fatal drawbacks in security inspection applications using X-ray images.

By contrast, local TM methods process each pixel differently according to its neighboring pixel values [12]–[26]. Fattal *et al.* [22] designed a novel local TM operator based on gradient attenuation. They compressed the drastic irradiance changes by reducing large gradients under a multi-scale framework. Reinhard *et al.* [23] proposed a local TM algorithm based on an automated dodging and burning algorithm. Li *et al.* [24] used a symmetrical analysis-synthesis filter bank for local gain control and luminance compression. Ma *et al.* [20] presented a tone mapped image quality index (TMQI) and performed dynamic range compression by optimizing this index. Laparra *et al.* [21] proposed a perceptually optimized image rendering method that minimizes the loss of contrast between the input HDR and its tone-mapped images. However, this method requires a high computational complexity due to its iterative optimization process. In contrast to global TM, local TM is generally capable of improving contrast ratios while preserving local details. However, local TM often yields annoying artifacts called halo artifacts around the high-frequency edges, as well as an imbalance in the global scene brightness.

Recently, deep learning-based TM operators have been proposed, which are mainly based on the generative adversarial network (GAN) [27]–[29]. GAN-based TM methods convert the style of input HDR images into that of predefined low-dynamic-range (LDR) target images. However, there is no learning-based TM method specialized for X-ray images to the best of our knowledge.

Indeed, applying existing TM algorithms developed for natural scenery images to X-ray inspection systems is not an ideal solution. X-ray inspection systems aim to detect suspicious objects by scanning suitcases and luggage, which often contain multiple objects stacked on top of each other or overlapping [30]. In other words, when visualizing an X-ray image on a display for a human inspector, the local details should be preserved to allow the inspector to detect unsafe or hazardous objects. However, as mentioned above, the conventional global TM methods are prone to the loss of local details, which makes visual inspection difficult. The traditional and learning-based local TM methods can preserve the local details, but they suffer from halo artifacts which distract the inspector.

This article presents a new deep learning-based TM method for X-ray inspection systems. The proposed method consists of two different networks called the detail-recovery network (DR-Net) and the TM network (TM-Net). The goal of DR-Net is to restore the details of the input HDR image. The DR-Net, which is based on a convolutional neural network (CNN), produces the HDR image with fine details by enhancing the detail layer of the input HDR image. To train DR-Net in a supervised learning manner, we propose a data synthesis technique. Our synthesis technique is motivated by the previous technique called threat image projection (TIP) [31], which generates synthesized X-ray images based on the Beer-Lambert law [32]. In addition to the X-ray image synthesis, we also generate ground-truth (GT) detail layers for DR-Net training.

After restoring the detail layer of the input HDR image by DR-Net, the output LDR image is generated using TM-Net, which focuses on compressing the dynamic range while preserving the restored details. However, it is also hard to train TM-Net since there are no standard GT LDR images for X-ray TM. If the LDR images obtained using traditional TM methods are used as GT for TM-Net training, TM-Net cannot behave very differently from traditional TM methods. To overcome this problem, we introduce an unsupervised learning framework. Specifically, we design a loss function that simultaneously optimizes the detail preservation of the input HDR image and the prevention of halo artifacts. The results of extensive experiments comparing the performance of the proposed method with state-of-the-art TM methods demonstrate that the proposed method achieves visually compelling results by enhancing local details in heavily overlapped areas while also preventing halo artifacts. In addition, the proposed method improves the quantitative scores, including the feature similarity index for tone-mapped images (FSITM) [33] and HDR-visible difference predictor (HDR-VDP)-2.2 [34].

In summary, this article presents three major contributions. (*i*) We propose a novel TM framework specialized for X-ray inspection systems that achieves superior performance compared to conventional TM methods. (*ii*) We introduce a data synthesis technique to train DR-Net that restores the detail layer of the HDR X-ray image in a supervised manner. (*iii*) We present a novel loss function which guides TM-Net to learn TM without requiring GT LDR images.

## II. RELATED WORK

A detailed review of the traditional and recent TM methods can be found in [35], [36]. In this section, we briefly review the TM methods closely related to ours, i.e., image decomposition-based and learning-based methods. The conventional X-ray image synthesis technique [31] that we adopted and extended for our image synthesis is also explained.

### A. IMAGE DECOMPOSITION-BASED TM

One of the most common choices for local TM is image decomposition [35]. In image decomposition-based TM methods, a smoothing filter is applied to the input image. The filtered image is then used to extract the detail and base layers. The detail layer is further refined to enhance local image details, while the base layer is tone-mapped for dynamic

range compression. The LDR image is finally obtained by recombining the refined base and detail layers.

Durand and Dorsey [15] proposed a method using a bilateral filter to decompose the HDR image into the base and detail layers. The method preserves image details but still suffers from halo artifacts. Farbman *et al.* [16] proposed an edge-preserving decomposition method based on the weighted least squares optimization framework. This global optimization-based method tends to produce images with higher quality than the bilateral filter-based method [15], but it has high computational costs. He *et al.* [17] proposed a guided filter which achieves good edge-preserving smoothing properties by using a guidance image for filtering. Moreover, the guided filter can be computed efficiently regardless of the kernel size and intensity range. Liang *et al.* [37] compressed the dynamic range in the gradient domain using a hybrid $l_1 - l_0$ decomposition model. Bae *et al.* [25] proposed a TM method that estimates an optimal detail layer and tone-mapped image without iterative process. Miao *et al.* [26] used multi-layer decomposition and reconstruction to model the properties of brightness, structure, and detail for HDR images. Different strategies were then adopted for each layer to reduce the overall brightness contrast and retain as much scene information.

### B. LEARNING-BASED TM

Recent studies have proposed CNN-based TM operators, which are based on GANs along with reference LDR image datasets [27]–[29]. Zhang *et al.* [27] adopted the improved Wasserstein GAN with gradient penalty, and presented a dataset consisting of tone-mapped images that were manually adjusted by expert photographers. Montulet *et al.* [28] introduced an end-to-end TM approach based on deep convolutional GANs using a dataset consisting of the HDR images and their color corrected versions obtained by top-ranked experts. Rana *et al.* [29] used a multi-scale conditional GAN to solve the problems of conventional GAN-based TM methods [27], [28], such as tiling patterns, local blurring, and saturation. However, Zhang *et al.*'s and Montulet *et al.*'s methods cannot be easily extended to other applications, such as X-ray image inspection, due to their reliance upon experts in dataset generation. Rana *et al.*'s method uses traditional TM methods for generating target LDR images, and thus their method cannot behave very differently from traditional TM methods [29]. To train CNNs without requiring laborious expert retouching or traditional TM methods, this article presents a dataset synthesis technique based on the Beer-Lambert law [32] as well as a novel loss function that can be used to train our networks without paired HDR-LDR images.

### C. X-RAY IMAGE SYNTHESIS

Many studies assume that X-ray image formation obeys the Beer-Lambert law. Based on this assumption, at image location $(x, y)$, the pixel intensity of the X-ray image $I(x, y)$ is defined as

$$I(x, y) = I_0 \exp\left(-\int \mu(x, y, z)dz\right), \tag{1}$$

where $I_0$ is the beam intensity, $z$ represents the depth coordinate, and $\mu$ is the effective attenuation coefficient of the objects in the scene [32].

Based on this image formation model, Rogers *et al.* [31] introduced a data synthesis technique called TIP which generates synthesized threat images that have no significant differences compared to real threat images. More specifically, they synthesize images by multiplying the foreground mask $F(x, y)$ and background mask $B(x, y)$ as follows:

$$I(x, y) = I_0 F(x, y) B(x, y), \tag{2}$$

where

$$\begin{aligned} F(x, y) &= \exp\left(-\int \mu_F(x, y, z)dz\right), \\ B(x, y) &= \exp\left(-\int \mu_B(x, y, z)dz\right). \end{aligned} \tag{3}$$

$\mu_F$ and $\mu_B$ represent the effective attenuation coefficients of the foreground and background masks, respectively. It is worth noting that when $N$ foreground masks are overlapped in the image, $F(x, y)$ in (2) can be replaced with $\prod_{i=1}^{N} F^i(x, y)$, where $F^i$ indicates the $i$-th foreground mask.

Although the aforementioned TIP technique [31] can generate various X-ray images containing multiple overlapped objects, the details of the image are often lost during the image projection process. To explicitly show this problem, we extract the detail layer from the input HDR image using image decomposition with the guided filter [17]. The decomposed layers are obtained as

$$\begin{aligned} I_b(x, y) &= G(\log(I(x, y))), \\ I_d(x, y) &= \log(I(x, y)) - I_b(x, y), \end{aligned} \tag{4}$$

where $I_b$ and $I_d$ are the base and detail layers of the input image, respectively, and $G$ denotes the guided filter. Fig. 1 illustrates the detail loss problem of the TIP technique.[1] As depicted in Fig. 1(b), the detail layer lacks sufficient details where the two objects overlapped. To alleviate this problem, we propose a detail layer synthesis technique that effectively preserves details as shown in Fig. 1(c).

## III. PROPOSED METHOD

As illustrated in Fig. 2, the proposed method is based on local TM using image decomposition. Specifically, we propose two networks: DR-Net and TM-Net. DR-Net first decomposes the input HDR X-ray image into a base layer and a detail layer using the guided filter. Then, the detail layer is passed through a CNN to restore the image details. The base layer and the restored detail layer are recombined into a single reconstructed image. Finally, the tone-mapped image

[1] Unless otherwise mentioned, HDR images are linearly mapped to LDR for visualization.

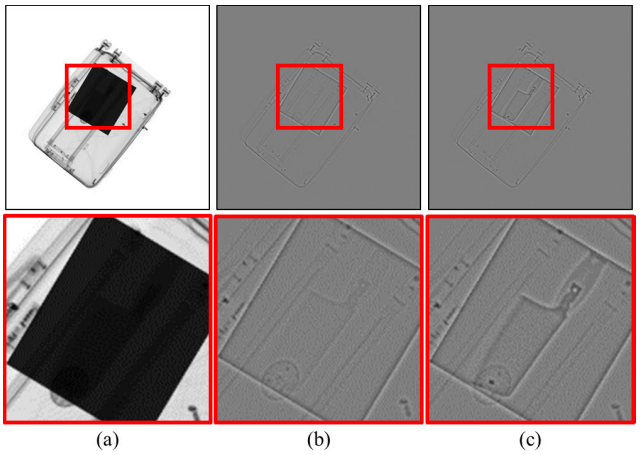



**FIGURE 1. Example of detail loss during the image projection process compared with the proposed detail layer generation technique: (a) The synthesized X-ray image, (b) detail layer obtained by the TIP method, and (c) detail layer obtained by the proposed method.**

is obtained by applying TM-Net. Unlike most previous TM methods [15], [16], [37], TM-Net applies TM not to the base layer but to the restored HDR image to maintain the details restored by DR-Net during the TM process.

U-Net [38] has achieved great success in solving pixel-wise classification problems, including biomedical image segmentation [39], remote sensing image segmentation [40], and image restoration [41]. We adopted U-Net as the baseline network architecture of DR-Net and TM-Net. We modified the convolutional layers to use reflection padding, and changed the deconvolutional layers to bilinear upsampling layers to mitigate the occurrence of checkerboard artifacts.

### A. DR-NET

As explained previously, the detail layer $I_d$ obtained using (4) can suffer from detail loss due to the overlapping of multiple objects, which also happens in general X-ray imaging. Thus, lost details cannot be recovered by directly applying TM to the input HDR X-ray image. We therefore introduce DR-Net to preprocess the input HDR image such that the final tone-mapped LDR image can contain sharper details. To this end, we first propose a method that synthesizes GT detail layers. Specifically, the detail layers of the foreground and background masks are obtained as follows:

$$\begin{aligned} F_d(x,y) &= \log(F(x,y)) - G(\log(F(x,y))), \\ B_d(x,y) &= \log(B(x,y)) - G(\log(B(x,y))), \end{aligned} \tag{5}$$

where $F_d$ and $B_d$ represent the detail layers of the foreground mask $F$ and the background mask $B$, respectively. We then define a GT detail layer $I_{d,gt}$ as follows:

$$I_{d,gt}(x,y) = F_d(x,y) + B_d(x,y). \tag{6}$$

Note the difference between the detail layer obtainable directly from the TIP technique as (4) and our definition in (6). Because the detail layers are first individually extracted from the foreground and background masks and then combined, the proposed detail layer can preserve the details and can thus be treated as GT for DR-Net training. When $N$ foreground masks are overlapped with each other, $F_d(x,y)$ can be defined as $\sum_{i=1}^{N} F_d^i(x,y)$, where $F_d^i$ indicates the detail layer of the $i$-th foreground mask. As shown in Fig. 1(c), the proposed method synthesizes the detail layer with distinct object boundaries, even if there are several instances of inter-object overlap.

During the training stage, DR-Net takes paired detail layers, i.e., $I_d$ from (4) and $I_{d,gt}$ from (6). The loss function $L_{dr}$ is defined to restore the detail layer of the X-ray image as follows:

$$L_{dr} = ||I_{d,r} - I_{d,gt}||_1, \tag{7}$$

where $||\cdot||$ measures the L1 loss between the output of DR-Net, denoted as $I_{d,r}$, and the GT detail layer $I_{d,gt}$. After restoring the detail layer, we obtain the restored HDR image $I_r(x,y)$ as follows:

$$I_r(x,y) = I_{d,r}(x,y) + I_b(x,y). \tag{8}$$

Fig. 3 shows the resultant images of DR-Net. The image regions inside the red boxes shown in Fig. 3 demonstrate that the details of the heavily overlapped regions have been successfully restored.

### B. TM-NET

Because the output of DR-Net is still in HDR, TM is needed to convert HDR to LDR. Without a loss of generality, our TM-Net maps an input HDR image with 16-bit precision to an output LDR image with 8-bit precision.

Since TM is content-dependent, environment-dependent, and application-dependent, it is difficult to define GT LDR images to train TM-Net in a supervised manner. We instead define three different loss terms that can be measured without the need for GT LDR images: structural similarity loss $L_{ss}$, detail preservation loss $L_{dp}$, and relative thickness loss $L_{rt}$. The total loss function $L_{total}$ is defined as a weighted sum of the three loss terms:

$$L_{total} = \lambda_{ss}L_{ss} + \lambda_{dp}L_{dp} + \lambda_{rt}L_{rt}, \tag{9}$$

where $\lambda_{ss}$, $\lambda_{dp}$, and $\lambda_{rt}$ are the weight parameters that control the importance of the three loss terms.

#### 1) STRUCTURAL SIMILARITY LOSS

To preserve the overall structure of the input image, we measure the structural similarity [42] between the input and output of TM-Net. The use of the structural similarity loss can drive TM-Net to preserve the image structures of the HDR image while reducing the dynamic range. $L_{ss}$ is defined as follows:

$$L_{ss} = 1 - \mathrm{SSIM}(I_r, I_o), \tag{10}$$

where SSIM measures the structural similarity [42] between the restored HDR image $I_r$ and the output of TM-Net, denoted

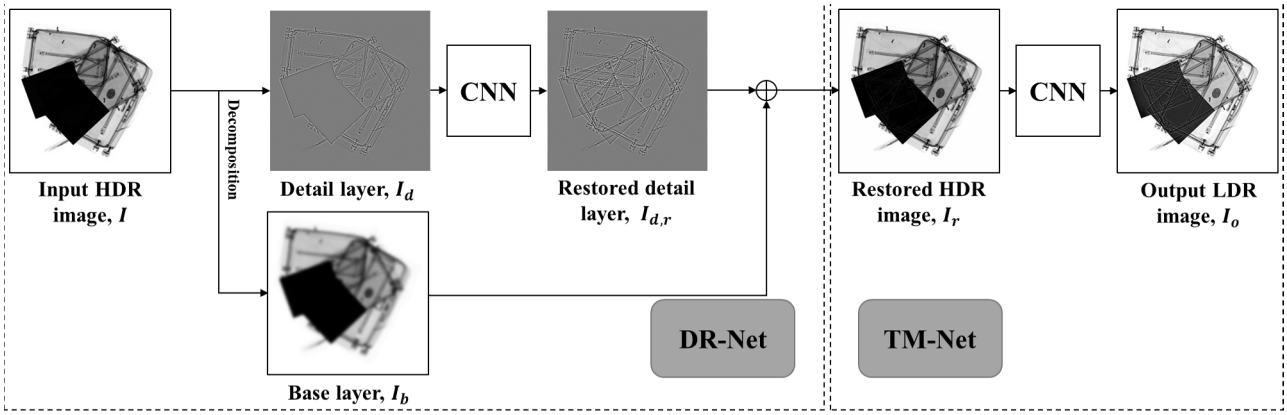



**FIGURE 2.** Overall block diagram of the proposed method.

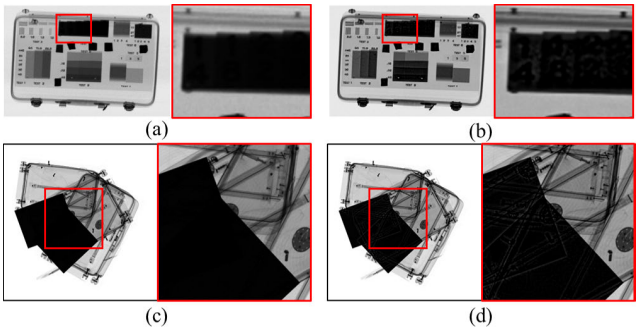



**FIGURE 3.** Example of DR-Net results: (a), (c) The input HDR images and (b), (d) their corresponding resultant images obtained by DR-Net.

as $I_o$. Note that in our implementation, the pixel values of $I_r$ and $I_o$ are normalized to be in the range of [0, 1] so that $L_{ss}$ can be directly measured between the two images. This loss term has also been used in other unsupervised learning tasks [43]–[45].

#### 2) DETAIL PRESERVATION LOSS

Detail loss is inevitable when performing dynamic range compression. To preserve details in the output LDR image, we define the detail preservation loss as follows:

$$L_{dp} = -\sum_{x,y}(G_x(I_r)G_x(I_o) + G_y(I_r)G_y(I_o)), \tag{11}$$

where

$$\begin{aligned} G_x(I_r) &= I_r(x+1,y) - I_r(x,y), \\ G_y(I_r) &= I_r(x,y+1) - I_r(x,y), \\ G_x(I_o) &= I_o(x+1,y) - I_o(x,y), \\ G_y(I_o) &= I_o(x,y+1) - I_o(x,y). \end{aligned} \tag{12}$$

This detail preservation loss is designed to strengthen the details of the resultant LDR images in a spatially adaptive manner. Specifically, since the gradient of the restored HDR image is used as a weighting factor, the resultant LDR images can have pixels with high gradients where their corresponding HDR pixels have high gradients.

#### 3) RELATIVE THICKNESS LOSS

Excessive contrast enhancement can cause a gradient reversal problem, which can lead to halo artifacts. The relative thickness loss is thus presented to maintain the sign of the image gradients before and after TM operation. This loss plays an important role in avoiding over-enhancement, which is a crucial factor for the inspector to estimate the relative thickness of the objects [1]. $L_{rt}$ is defined as follows:

$$\begin{aligned} L_{rt} = \sum_{x,y} &\tanh(\lambda_s G_x(I_r)) - \tanh(\lambda_s G_x(I_o)) \\ &+ \tanh\left(\lambda_s G_y(I_r)\right) - \tanh\left(\lambda_s G_y(I_o)\right). \end{aligned} \tag{13}$$

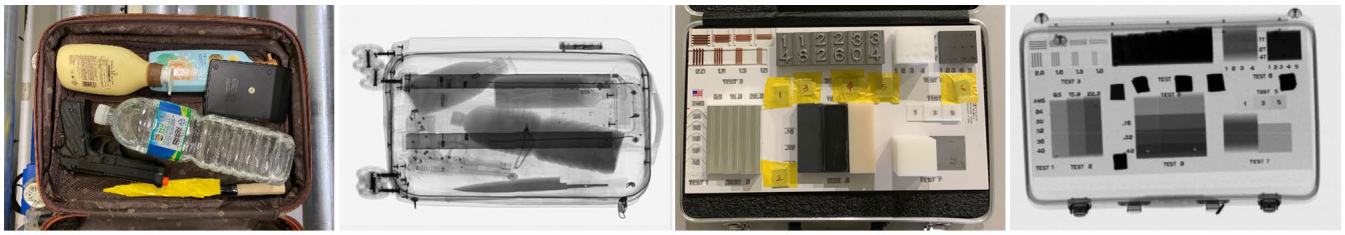
As the sign function is non-differentiable, we use the tanh function instead [46]. The parameter $\lambda_s$ is used to make the tanh function steeper, which makes the tanh function similar to the sign function.

## IV. EXPERIMENTAL RESULTS

In this section, we present qualitative and quantitative performance comparison results to demonstrate the superiority of the proposed method compared to the state-of-the-art TM methods [10], [21], [24], [29], [37].

### A. DATASET AND IMPLEMENTATION DETAILS

There are several X-ray security imaging datasets available [30], [47], [48], but these datasets were developed for computer vision tasks including object classification, detection, segmentation, and unsupervised anomaly detection. To the best of our knowledge, there is no publicly available dataset for TM of X-ray images. We thus first constructed a synthetic dataset. To simulate X-ray images for our target applications, we used knives and firearms as representative threatening objects. Other objects were also included and cluttered inside suitcases to make visual inspection more difficult. Using each object inside the suitcases as an individual foreground mask, we synthesized 10,000 X-ray images and corresponding GT detail layers to train DR-Net as explained in Section III-A. Another 100 synthesized images were used for quantitative performance evaluation.



(a) (b) (c) (d)

**FIGURE 4.** Examples of real X-ray images used for testing TM methods: (a) Suitcase with dangerous objects and (b) its X-ray image, (c) the ASTM F792-08 kit and (d) its X-ray image.

For the test on real image samples, we captured 16-bit HDR X-ray images.[2] Fig. 4(a) shows a test sample of a typical suitcase containing threatening objects. Fig. 4(c) shows another test sample of the ASTM F792-08 kit, which is widely used for evaluating the performance of X-ray inspection systems.[3]

For the performance comparison, we used the widely used global [10] and local [21], [24], [37] TM methods with the author-provided source codes. For the test of the learning-based method [29], we followed the authors' procedure for generating training image pairs. Specifically, we used the same 10,000 synthesized HDR X-ray images as input and obtained multiple tone-mapped LDR images for each HDR image using the other existing TM methods [10], [21], [24], [37]. Among multiple tone-mapped LDR images for each HDR image, we chose the image with the highest TMQI [49] as a target LDR image, resulting paired HDR and HDR images. The author-provided code with the default settings was then used for training their TM network [29].

The proposed DR-Net and TM-Net were trained from scratch for 100k iterations with a learning rate of 1e-5, followed by another 100k iterations with the learning rate linearly decayed to 0. We used the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the batch size was set to 1, which are commonly used settings in various deep learning networks. The weight parameters were empirically chosen as $(\lambda_{ss}, \lambda_{dp}, \lambda_{rt}) = (1, 300, 100)$ to control the relative importance of each loss term, and $\lambda_s$ was set to 0.1 to make the relative thickness loss trainable. Our whole training process was conducted using a single Titan X GPU. More results, datasets, and code can be found on our project webpage.[4]

### B. QUANTITATIVE PERFORMANCE EVALUATION

For the objective performance assessment, we adopted widely used measures, including the measure of enhancement by entropy (EME) [50], PixDist [51], TMQI [49], FSITM [33], and HDR-VDP-2.2 [34]. EME approximates the average contrast in an image by dividing it into blocks, computing a score based on the minimum and maximum gray-levels in each block, and then averaging the scores. PixDist is used as a criterion to measure the level of contrast enhancement, and a high PixDist represents an image with a widely spread histogram. TMQI is a widely accepted metric for TM performance evaluation, which evaluates the structural fidelity and naturalness of tone mapped images. FSITM measures the similarity of the original HDR and converted LDR images using local phase information. The FSITM score is high for visually pleasing images with vivid appearances of the real-world scene. HDR-VDP-2.2 can predict whether or not the difference between two images is visible to a human observer. HDR-VDP-2.2 takes into account several aspects such as the diagonal display size, display resolution, and viewing distance.

**TABLE 1.** Quantitative performance comparison of the conventional and proposed TM methods.

| Method | EME | PixDist | TMQI | FSITM | HDR-VDP-2.2 |
|---|---|---|---|---|---|
| Laparra *et al.* | **43.7535** | 34.4379 | 0.7776 | 0.8014 | 51.5406 |
| Khan *et al.* | 3.9158 | <u>45.4804</u> | **0.7825** | <u>0.8853</u> | 52.4322 |
| Li *et al.* | 14.8975 | 24.7131 | 0.7631 | 0.8677 | 53.4005 |
| Liang *et al.* | 8.2632 | 23.5373 | 0.7633 | 0.8327 | <u>53.6889</u> |
| Rana *et al.* | 4.1968 | **46.6118** | <u>0.7795</u> | 0.8697 | 52.7200 |
| Proposed | <u>16.5676</u> | 39.785 | 0.7715 | **0.9306** | **55.1932** |

Table 1 lists the quality scores obtained for 100 synthetic test images, where the best and second-best scores are bold-faced and underlined, respectively. Laparra *et al.*'s method [21] achieved the highest EME score, and Khan *et al.*'s method [10] showed outstanding performance in terms of PixDist and TMQI due to the global contrast enhancement. Li *et al.*'s [24] and Liang *et al.*'s [37] methods exhibited moderate performance in all performance measures. Rana *et al.*'s method [29] showed a similar performance to that of Khan *et al.*'s because the their training images were mostly selected from Khan *et al.*'s method, which showed the highest TMQI as listed in Table 1. The proposed method acquired the highest FSITM and HDR-VDP-2.2 scores along with the second highest EME score, indicating that the proposed method produces images with higher quality in different aspects.

### C. ABLATION STUDY

Because the proposed method includes DR-Net and TM-Net, we first evaluated the effectiveness of the individual

[2]http://www.sens-tech.com/index.php/security/carry-on-baggage-screening

[3]https://www.astm.org/DATABASE.CART/HISTORICAL/F792-08.htm

[4]https://github.com/hykim0/DRnTM-Net

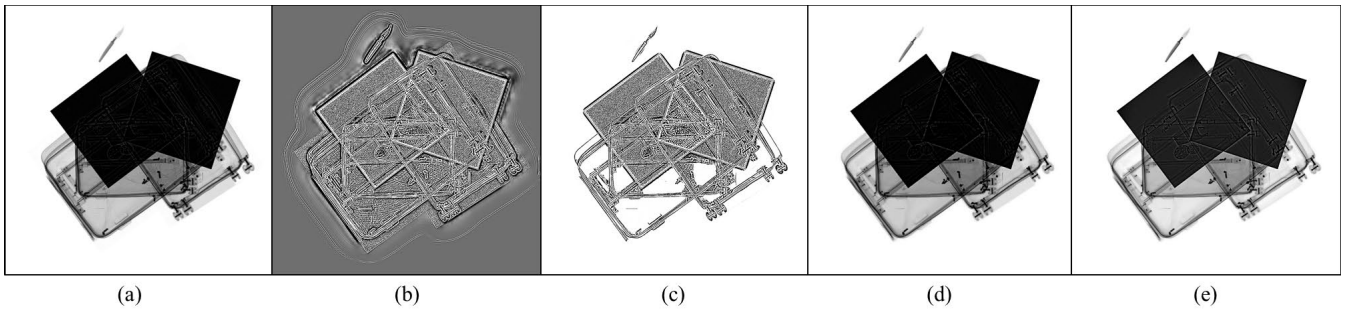



**FIGURE 5.** Visual comparison of TM-Nets trained with different loss combinations: (a) Input detail-restored HDR image, and the resultant images obtained by TM-Nets trained using (b) $L_{dp}$, (c) $L_{dp}$ and $L_{ss}$, (d) $L_{dp}$ and $L_{rt}$, and (e) $L_{dp}$, $L_{ss}$, and $L_{rt}$.

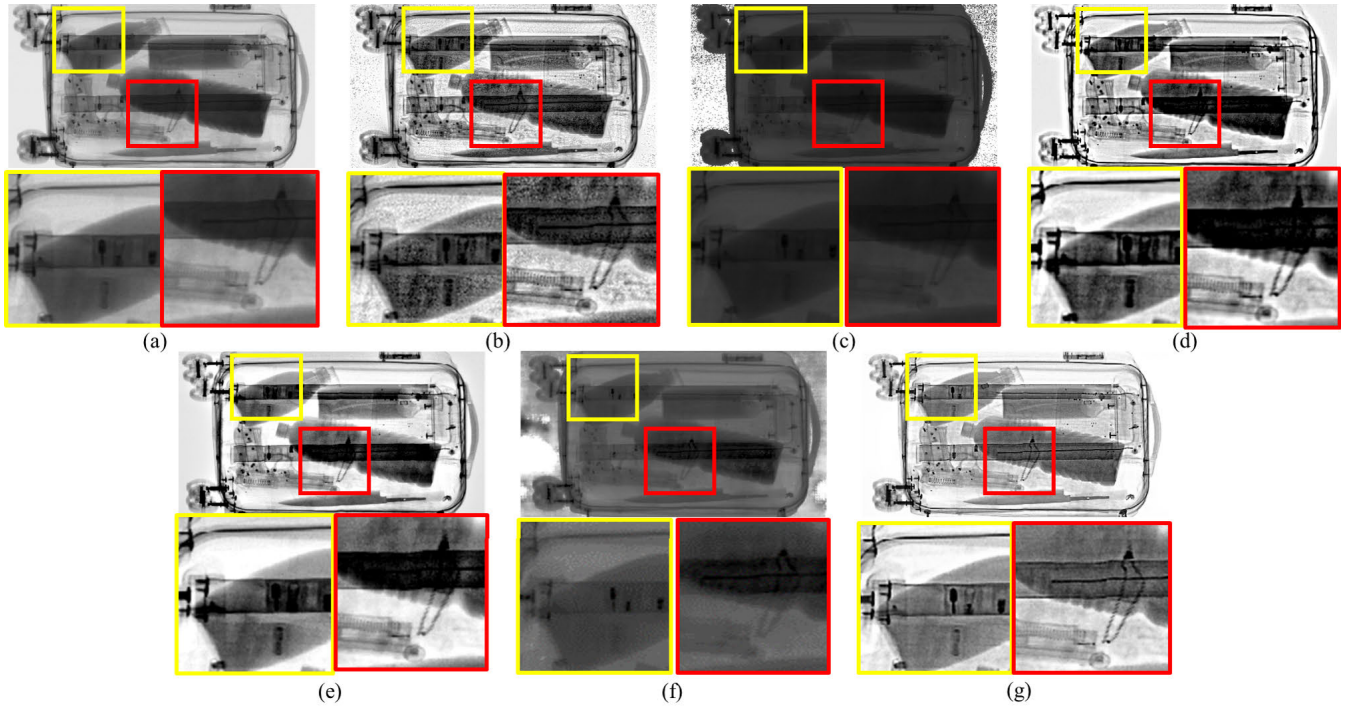



**FIGURE 6.** Comparison results with different TM methods for the suitcase X-ray image. (a) Input HDR image and the results of (b) Laparra *et al.*'s, (c) Khan *et al.*'s (d) Li *et al.*'s, (e) Liang *et al.*'s, (f) Rana *et al.*'s, and (g) the proposed method.

**TABLE 2.** Objective performance comparison of TM methods for the original and detail-restored HDR images.

| Input | Method | EME | PixDist | TMQI | FSITM | HDR-VDP-2.2 |
|---|---|---|---|---|---|---|
| Original HDR | Li *et al.* | 14.8975 | 24.7131 | 0.7631 | 0.8677 | 53.4005 |
| | Liang *et al.* | 8.2632 | 23.5373 | 0.7633 | 0.8327 | 53.6889 |
| | Proposed | 10.3366 | **40.4007** | 0.7638 | **0.9427** | 54.5499 |
| Detail restored HDR | Li *et al.* | **33.7016** | 16.1134 | 0.7468 | 0.8259 | 49.9713 |
| | Liang *et al.* | 20.8025 | 15.5551 | 0.7548 | 0.8401 | 49.7713 |
| | Proposed | 16.5676 | 39.785 | **0.7715** | 0.9306 | **55.1932** |

networks. Li *et al.*'s and Liang *et al.*'s TM methods were used for this performance comparison because their methods produce images with high local contrast without severe over-enhancement. First, the same original HDR images were used for the inputs of TM-Net and the two compared TM methods [24], [37]. As presented in Table 2, the proposed TM-Net yielded the highest PixDist and FSITM scores, as well as the second highest TMQI and HDR-VDP-2.2 scores. Second, the same detail-restored HDR images obtained by DR-Net were used for the inputs of TM-Net and the other two methods. Table 2 shows that TM-Net improved the scores especially when used with DR-Net. The EME score obtained using both DR-Net and TM-Net was higher than that obtained from TM-Net alone, demonstrating the effectiveness of DR-Net for local detail improvement. The use of the detail-restored HDR images also contributed to local detail improvement for Li *et al.*'s and Liang *et al.*'s TM methods as can be noticed from the increase of the EME scores. However, the PixDist scores were considerably decreased and the FSITM and HDR-VDP-2.2 scores were also decreased, indicating the over-enhancement by DR-Net when used with the other two TM methods.

Next, we evaluated the effectiveness of the loss terms in (10)-(13) which were used in TM-Net training. To this end, we applied both DR-Net and TM-Net, but the TM-Nets

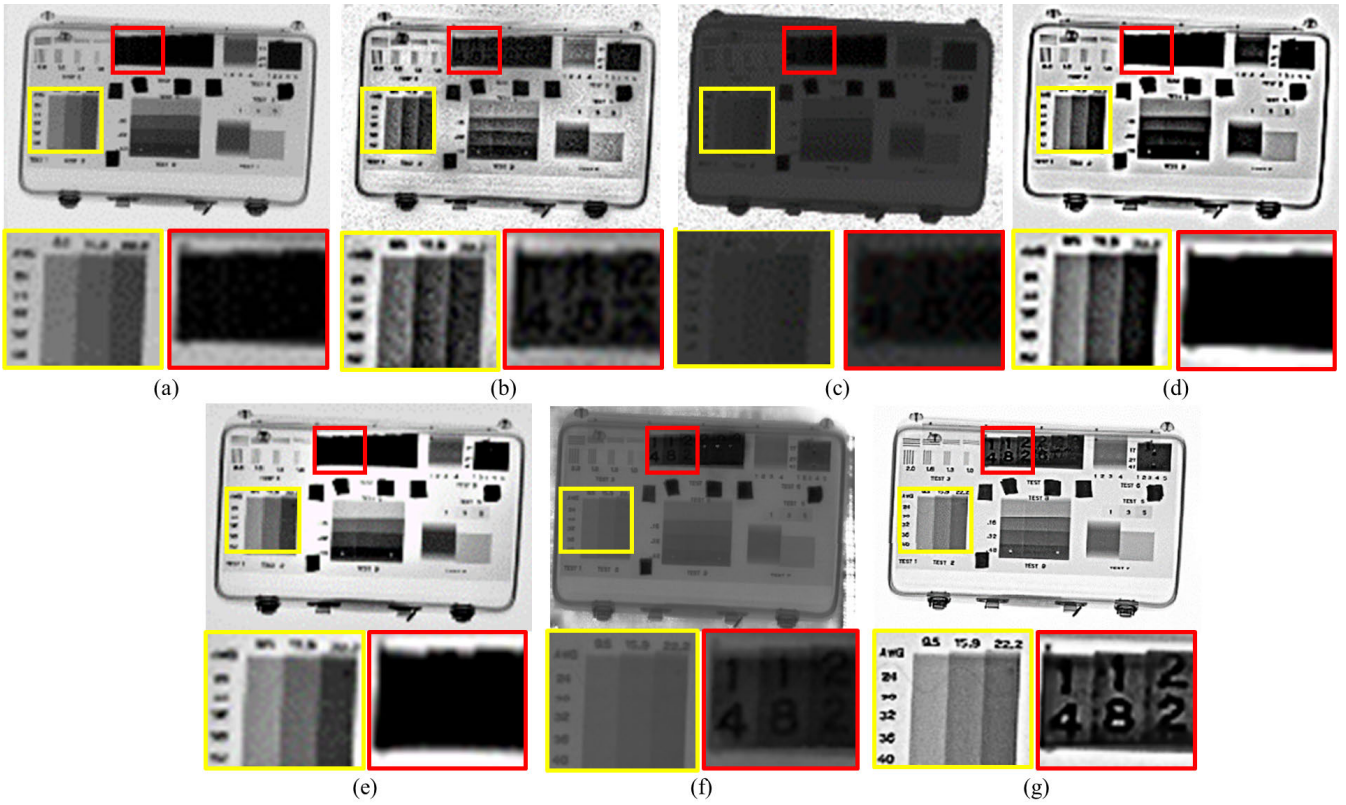



**FIGURE 7.** Comparison results with different TM methods for the ASTM F792-08 kit X-ray image. (a) Input HDR image and the results of (b) Laparra *et al.*'s (c) Khan *et al.*'s (d) Li *et al.*'s, (e) Liang *et al.*'s, (f) Rana *et al.*'s, and (g) the proposed method.

were trained using only one or two loss terms. As demonstrated in Fig. 5, the three loss terms were all found to be necessary to prevent over-enhancement and preserve overall structure. Note that $L_{dp}$ was always included in the loss function because TM-Net maintained the input image as the output when trained without using $L_{dp}$.

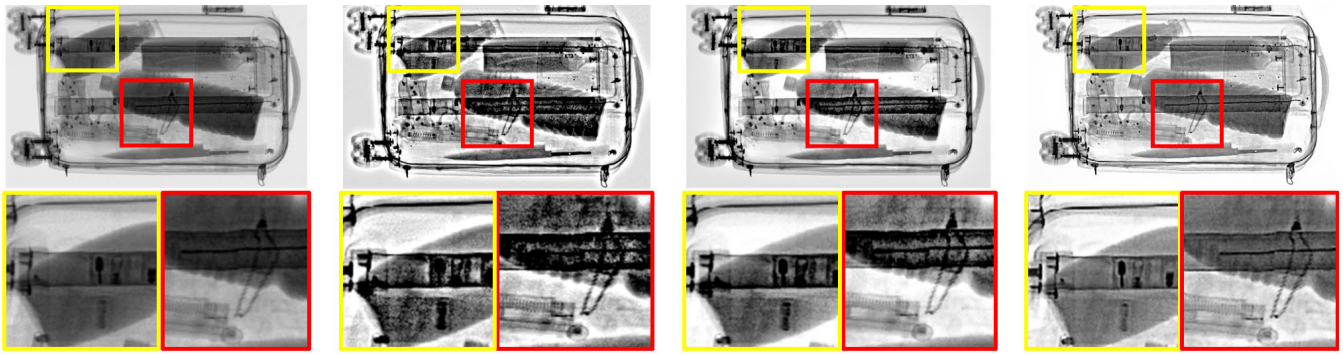
Finally, we changed the baseline network architecture from U-Net to the context aggregation network (CAN) [52]. Here, we maintained the entire framework and only changed the network architectures of DR-Net and TM-Net. The results of the proposed method with CAN were 17.8264, 41.6288, 0.7684, 0.8938, and 54.8046 for EME, PixDist, TMQI, FSITM, and HDR-VDP-2.2, respectively, which are comparable to the scores obtained with U-Net. Based on this result, we can conclude that the proposed method does not strongly depend on the baseline network architectures used.

### D. RESULTS ON REAL IMAGES

We performed qualitative performance comparison of different TM methods on the real samples as shown in Fig. 4. Fig. 6 compares the TM results of the real suitcase X-ray image obtained by the proposed and conventional methods [10], [21], [24], [29], [37]. Khan *et al.*'s global TM method [10] was not able to enhance the local contrast. Laparra *et al.*'s [21] and Li *et al.*'s [24] local TM methods improved the local contrast, but noticeable halo artifacts were present. Liang *et al.*'s decomposition-based TM method [37] produced images with high local contrast with less noticeable halo artifacts, but the details in the inter-object overlapped regions were not clearly rendered. Rana *et al.*'s learning-based TM method [29] could enhance the global contrast, but local contrast was not effectively improved because of the biased distribution of the target LDR images toward globally enhanced images. As shown in Fig. 6(g), the proposed method could significantly improve the visibility of the inter-object overlapped regions.

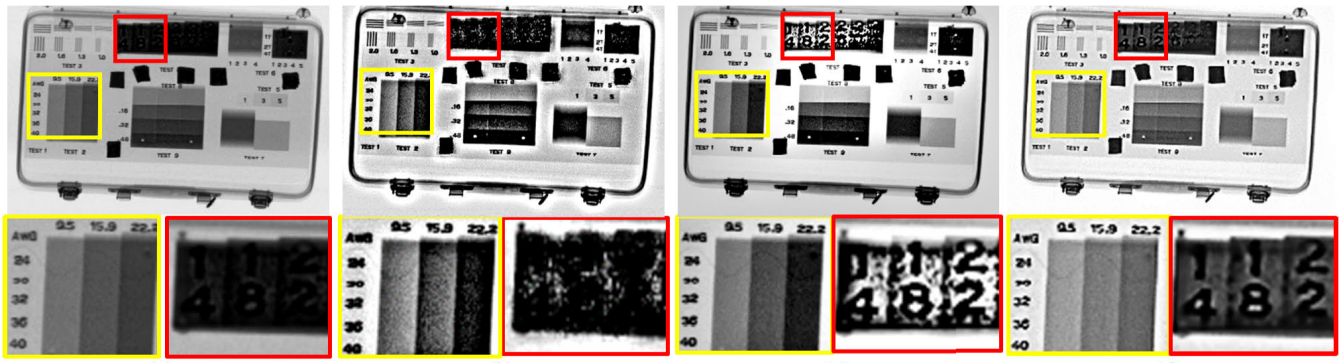
Fig. 7 shows another results on the kit test sample. Similarly, Khan *et al.*'s [10] and Rana *et al.*'s [29] methods suffered from low local contrast. The other local TM methods produced images with high local contrast, but Laparra *et al.*'s [21] method produced noisy images while Liang *et al.*'s [37] and Li *et al.*'s [24] methods suffered from halo artifacts. As shown in Fig. 7(g), the proposed method improved image details with fewer halo artifacts.

We also applied the same detail-restored HDR images as input for Li *et al.*'s and Liang *et al.*'s TM methods as well as our TM-Net. When comparing Figs. 6(d) and (e) with Figs. 8(b) and (c), or Figs. 7(d) and (e) with Figs. 9(b) and (c), it can be found that the proposed DR-Net can further improve the performance of Li *et al.*'s and Liang *et al.*'s TM methods in terms of the detail visibility and sharpness of images. Li *et al.*'s method improved the overall details, but it is still difficult to recognize the numbers in the red box of Fig. 9(b) due to severe noise amplification. Halo artifacts were more severe for Liang *et al.*'s method when used with DR-Net



(a) (b) (c) (d)

**FIGURE 8.** Visual comparison of LDR results of different TM methods on the detail-restored suitcase image obtained by the proposed DR-Net: (a) Input detail-restored HDR image and the results of (b) Li *et al.*'s, (c) Liang *et al.*'s, and (d) the proposed TM-Net.



(a) (b) (c) (d)

**FIGURE 9.** Visual comparison of LDR results of different TM methods on the detail-restored kit image obtained by the proposed DR-Net: (a) Input detail-restored HDR image and the results of (b) Li *et al.*'s, (c) Liang *et al.*'s, and (d) the proposed TM-Net.

as shown in Figs. 8(c) and 9(c). These side effects can also be noticed from the decrease of FSITM and HDR-VDP-2.2 scores when DR-Net was used with Li *et al.*'s and Liang *et al.*'s methods as shown in Table 2.

## V. CONCLUSION

A novel learning-based TM method was proposed to effectively visualize HDR X-ray images on existing display devices for X-ray inspection systems. The proposed method is composed of two networks, where DR-Net restores details in the inter-object overlapping regions of the image and TM-Net converts the detail-restored HDR images to LDR images with detail and structure preservation capability. We also proposed a GT detail layer synthesis technique for supervised DR-Net training and introduced loss terms that can be used to train TM-Net in an unsupervised manner. The experimental results show that the proposed method achieves visually compelling images with improved details and fewer halo artifacts, and is superior to the state-of-the-art TM methods in terms of quantitative performance measures.

## REFERENCES


[1] J. L. Glover, L. T. Hudson, and N. G. Paulter, "Improved threat identification using tonemapping of high-dynamic-range X-ray images," *J. Test. Eval.*, vol. 46, no. 4, pp. 1462–1467, May 2018.

[2] T. Madmad and C. De Vleeschouwer, "Bilateral histogram equalization for X-ray image tone mapping," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2019, pp. 3507–3511.

[3] S. H. Park and E. D. Montag, "Evaluating tone mapping algorithms for rendering non-pictorial (scientific) high-dynamic-range images," *J. Vis. Commun. Image Represent.*, vol. 18, no. 5, pp. 415–428, Oct. 2007.

[4] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graphics*, vol. 3, no. 4, pp. 291–306, Oct./Dec. 1997.

[5] G. Ward, "A contrast-based scalefactor for luminance display," *Graph. Gems*, vol. 4, pp. 415–421, Aug. 1994.

[6] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 249–258.

[7] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, Nov. 2003.

[8] G. Qiu, J. Duan, and G. D. Finlayson, "Learning to display high dynamic range images," *Pattern Recognit.*, vol. 40, no. 10, pp. 2641–2655, Oct. 2007.

[9] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graphics*, vol. 11, no. 1, pp. 13–24, Jan. 2005.

[10] I. R. Khan, S. Rahardja, M. M. Khan, M. M. Movania, and F. Abed, "A tone-mapping technique based on histogram using a sensitivity model of the human visual system," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3469–3479, Apr. 2018.

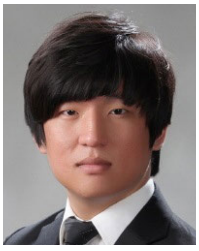
[11] I. R. Khan, W. Aziz, and S.-O. Shim, "Tone-mapping using perceptual-quantizer and image histogram," *IEEE Access*, vol. 8, pp. 31350–31358, Feb. 2020.

[12] E. H. Land, "The retinex," *Amer. Sci.*, vol. 52, no. 2, pp. 247–264, Jun. 1964.

[13] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[14] L. Meylan and S. Susstrunk, "High dynamic range image rendering with a retinex-based adaptive filter," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2820–2830, Sep. 2006.

[15] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. ACM SIGGRAPH*, 2002, pp. 257–266.
[16] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.
[17] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[18] H. Yeganeh, S. Wang, K. Zeng, M. Eisapour, and Z. Wang, "Objective quality assessment of tone-mapped videos," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 899–903.
[19] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.
[20] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
[21] V. Laparra, A. Berardino, J. Ballé, and E. P. Simoncelli, "Perceptually optimized image rendering," 2017, *arXiv:1701.06641*. [Online]. Available: http://arxiv.org/abs/1701.06641
[22] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Techn.*, 2002, pp. 249–256.
[23] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Techn.*, 2002, pp. 267–276.
[24] H. Li, X. Jia, and L. Zhang, "Clustering based content and color adaptive tone mapping," *Comput. Vis. Image Understand.*, vol. 168, pp. 37–49, Mar. 2018.
[25] G. Bae, C. Y. Jang, S. I. Cho, and Y. H. Kim, "Non-iterative tone mapping with high efficiency and robustness," *IEEE Access*, vol. 6, pp. 35720–35733, Jun. 2018.
[26] D. Miao, Z. Zhu, Y. Bai, G. Jiang, and Z. Duan, "Novel tone mapping method via macro-micro modeling of human visual system," *IEEE Access*, vol. 7, pp. 118359–118369, Aug. 2019.
[27] N. Zhang, C. Wang, Y. Zhao, and R. Wang, "Deep tone mapping network in HSV color space," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2019, pp. 1–4.
[28] R. Montulet, A. Briassouli, and N. Maastricht, "Deep learning for robust end-to-end tone mapping," in *Proc. Brit. Mach. Vis. Conf.*, 2019, p. 194.
[29] 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, Sep. 2020.
[30] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao, and Q. Ye, "SIXray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2119–2128.
[31] T. W. Rogers, N. Jaccard, E. D. Protonotarios, J. Ollier, E. J. Morton, and L. D. Griffin, "Threat image projection (TIP) into X-ray images of cargo containers for training humans and machines," in *Proc. IEEE Int. Carnahan Conf. Secur. Technol.*, Oct. 2016, pp. 1–7.
[32] H. H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. New York, NY, USA: Academic, 1996.
[33] H. Ziaei Nafchi, A. Shahkolaei, R. Farrahi Moghaddam, and M. Cheriet, "FSITM: A feature similarity index for tone-mapped images," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1026–1029, Aug. 2015.
[34] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imag.*, vol. 24, no. 1, Jan. 2015, Art. no. 010501.
[35] G. Eilertsen, R. K. Mantiuk, and J. Unger, "A comparative review of tone-mapping algorithms for high dynamic range video," *Comput. Graph. Forum*, vol. 36, no. 2, pp. 565–592, May 2017.
[36] Y. Salih, W. B. Md-Esa, A. S. Malik, and N. Saad, "Tone mapping of HDR images: A review," in *Proc. 4th Int. Conf. Intell. Adv. Syst.*, Jun. 2012, pp. 368–373.
[37] Z. Liang, J. Xu, D. Zhang, Z. Cao, and L. Zhang, "A hybrid l1-l0 layer decomposition model for tone mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4758–4766.
[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[39] R. Bermúdez-Chacón, P. Márquez-Neila, M. Salzmann, and P. Fua, "A domain-adaptive two-stream U-Net for electron microscopy image segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 400–404.
[40] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
[41] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When image denoising meets high-level vision tasks: A deep learning approach," 2017, *arXiv:1706.04284*. [Online]. Available: http://arxiv.org/abs/1706.04284
[42] 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.
[43] Y.-G. Shin, S. Park, Y.-J. Yeo, M.-J. Yoo, and S.-J. Ko, "Unsupervised deep contrast enhancement with power constraint for OLED displays," *IEEE Trans. Image Process.*, vol. 29, pp. 2834–2844, Nov. 2020.
[44] R. Mahjourian, M. Wicke, and A. Angelova, "Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5667–5675.
[45] Z. Yin and J. Shi, "GeoNet: Unsupervised learning of dense depth, optical flow and camera pose," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1983–1992.
[46] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach," *IEEE Access*, vol. 7, pp. 93326–93338, Jul. 2019.
[47] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.
[48] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "GDXray: The database of X-ray images for nondestructive testing," *J. Nondestruct. Eval.*, vol. 34, no. 4, p. 42, Nov. 2015.
[49] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
[50] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 741–758, Mar. 2007.
[51] Z. Chen, B. R. Abidi, D. L. Page, and M. A. Abidi, "Gray-level grouping (GLG): An automatic method for optimized image contrast enhancement—Part I: The basic method," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2290–2302, Aug. 2006.
[52] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2497–2506.



**HYO-YOUNG KIM** (Member, IEEE) received the B.S. degree in electrical engineering from Korea University, in 2013, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, computer vision, and artificial intelligence.

**SEUNG PARK** received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2013 and 2020, respectively. He was a Research Professor with Korea University, in 2020. He currently joined the Samsung Medical Center as an AI Researcher. His current research interests include X-ray imaging and medical image analysis.

**YONG-GOO SHIN** received the B.S. and Ph.D. degrees in electronics engineering from Korea University, Seoul, South Korea, in 2014 and 2020, respectively. He is currently an Assistant Professor with the Division of Smart Interdisciplinary Engineering, Hannam University. His research interests include digital image processing, computer vision, and deep learning.

**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. He was a Research Professor with the Research Institute of Information and Communication Technology, Korea University, from 2011 to 2012. He was a Research Scientist with the Samsung Advanced Institute of Technology, Yongin, South Korea, from 2012 to 2014. He was an Assistant Professor with the Department of Multimedia Engineering, Dongguk University, Seoul, from 2014 to 2020. In 2020, he joined the Department of Electrical Engineering, Korea University, where he is currently an Associate Professor. He has published over 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.

**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–Dearborn. In 1992, he joined the Department of Electronic Engineering, Korea University, where he is currently a Professor. He has published over 210 international journal articles. He also holds over 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 1999 LG Research Award. He received 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 Research Excellence Award from Korea University, in 2004, and the Technical Achievement Award from the IEEE Consumer Electronics (CE) Society, in 2012. He received the 15-Year Service Award from the TPC of ICCE in 2014 and the Chester Sall Award (First Place Transaction Paper Award) from the IEEE CE Society in 2017. He was honored with the Science and Technology Achievement Medal from the Korean Government, in 2020. He has served as the General Chairman for ITC-CSCC 2012 and IEICE 2013. He was the President of the 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 a member of the Editorial Board of the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS.