# DCDR-UNet: Deformable Convolution Based Detail Restoration via U-shape Network for Single Image HDR Reconstruction

Joonsoo Kim, Zhe Zhu, Tien Bau, Chenguang Liu
DMS Lab, Samsung Research America, USA
{joonsoo.k, zhe.zhu, t.bau, cheng.liu1}@samsung.com


Figure 1. HDR reconstruction from a single LDR image. By learning the receptive field of each overexposed pixel location so that the receptive field can include non-overexposed information for restoring a overexposed object properly, our method restores the overexposed objects such as power lines, a car and sky accurately. These differences are better visible in pdf version with zoom-in.

## Abstract

*Single image based HDR reconstruction methods using deep neural network have been proposed to mainly restore the lost details in the overexposed region. However, they cannot restore the details well if the overexposed region becomes large because the receptive fields of their networks are not large enough to cover the region. Also, they cannot restore the partially overexposed small object well if the non-overexposed portions of the object are sparse. In this paper, we propose new deep neural network, namely DCDR-UNet (Deformable Convolution Based Detail restoration via U-shape network), for single image HDR reconstruction. By introducing a new block called Deformable Convolution Residual Block (DCRB) and our loss function, we show how deformable convolution can be well utilized to solve the problems of the existing methods in single image HDR reconstruction. Our experimental results show that our method achieves much better results than all the existing methods quantitatively and qualitatively.*

## 1. Introduction

While modern displays can render HDR (High dynamic range) content, SDR content, which includes only low dynamic range (LDR) of the HDR scenes, is still dominant in the market. Therefore, there have been demands to generate an HDR image from an LDR image. To satisfy the demands, single-image based HDR reconstruction methods using deep neural network have been proposed [1, 2, 6, 16, 20–22, 28]. These methods enable the creation of details in an HDR image from a single LDR image, without requiring additional exposures or specialized hardware. To do this, they use the multiple pairs of LDR and HDR images to train their networks, so their networks learn how to restore the details and the tones in very bright region (overexposed) in the LDR images. Through their experiments they show that the details and tones in the bright region in an LDR image can be restored to match an HDR image.

Generally, there are two types of overexposed objects (or regions) to be restored in an LDR image. The first one is a partially overexposed object. Most of existing methods

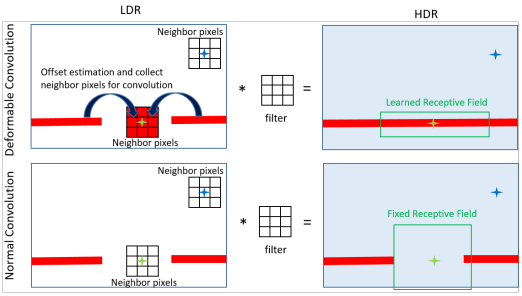


Figure 2. The examples of using deformable convolution and normal convolution for detail restoration

[1, 6, 16, 21, 22, 28] focus on restoring this type of overexposed object. When they restore this type of an object, the non-overexposed region of the object is mainly used to restore the overexposed region of it. For example, if there are clouds in the sky and small portion of them are overexposed, the existing methods use the non-overexposed region of cloud to restore the overexposed region. However, their methods show poor restoration results when the overexposed portion of the object is much larger than the non-overexposed portion of it. In the 2nd and 3rd images of Figure 1, the large portions of the power lines and the car are overexposed and the small portions of them are only visible. In these cases, one of the existing methods, HDRUNET [1], cannot restore the object, which makes them to be seen as unnaturally. That is because the fixed receptive field of HDRUNET cannot handle the HDR image restoration from small non-overexposed portion of the object in the 2nd image. In the 3rd image, different non-overexposed pixels at different overexposed pixels (blue and red point) are used for restoration because of the limited and fixed receptive field of HDRUNET.

The second type is an entirely overexposed object. This is more challenging case than the partially overexposed object. Especially, when the entirely overexposed region is large and the size of a HDR reconstruction network is not large, the reconstructed HDR image looks poor. The first LDR image in Figure 1 shows that the large sky region is entirely overexposed. Different from partially overexposed objects, there is no hint to infer the entirely overexposed region as sky. To infer this region as sky, the HDR reconstruction networks are trained from the multiple pairs of LDR and HDR images that includes a sky region next to buildings. However, if a network has a fixed receptive field (e.g. HDRUNET) and it is smaller than the overexposed sky, the reconstructed sky shows unnatural textures. Note that the network with a fixed large receptive field can have similar problems when the overexposed region becomes larger or the image resolution becomes larger.

In this paper, we propose new deep neural network, namely DCDR-UNet (Deformable Convolution Based Detail restoration via U-shape network), for single image HDR reconstruction. In our method, we utilize a deformable convolution [4] to solve the problems of the existing methods. The use of the deformable convolution enables our network to learn the receptive field on each pixel location from training images. Note that the purpose of using deformable convolution is not simply increasing the receptive field of the network but determining the receptive field adaptive to each pixel. Figure 2 shows a good example of it. In this example, an entire red line becomes disconnected and sky becomes white in an LDR image because of overexposure. To restore them using deformable convolution, the neighbor pixels of the overexposed line pixel (green point) and sky pixel (blue point) are collected from the line pixels (non-overexposed) and the sky (overexposed) accordingly. Then, the same convolution filter, which is trained to convert white pixel to blur and keep the red pixel as red, is applied to the LDR image to reconstruct the HDR image. To utilize deformable convolution effectively, we introduce a new residual block called Deformable Convolution Residual Block (DCRB). This block combines offset estimation, deformable convolutions and SFT (spatial feature transform) layers [29] in residual block fashion. For more accurate offset estimation, which is very important for learned receptive field on each pixel, the offset estimation in placed on the input path of DCRB block while the deformable convolution is placed in the residual path of it. To utilize our DCRB correctly, we use the loss function that combines a pixel loss and a perceptual loss. Our experiments show more clearly that the combination of the DCRB and our loss function can generate great synergy, which can greatly improve the reconstructed image quality. The main contributions of our work can be summarized as follows:

- As far as we know, we are the first one that utilizes deformable convolution in single image HDR reconstruction.
- We show how to utilize deformable convolution effectively for single image HDR reconstruction by introducing both our DCRB and our loss function.
- We show how the overexposed object or region in an LDR image is well reconstructed through our experimental results and analysis.

## 2. Related Works

### 2.1. Multiple Exposure HDR Reconstruction

While modern displays are capable of rendering HDR content, SDR content is still prevalent. The common way to obtain an HDR image is to fuse multiple SDR images with different exposure, i.e. multi-exposure fusing (MEF) [5]. Those methods have achieved promising results, but they also suffer from artefacts such as ghosting and tearing, especially when there is motion in the scene. Thus earlier works [10, 12, 26, 27] focused on mitigating these kinds of artefacts. With deep learning leading the way in many research areas, researchers began using neural networks

in HDR reconstruction. With the very first deep learning based HDR reconstruction methods [15] performing image alignment and merging by a convolutional neural network (CNN), later improvements [25] were made by replacing the conventional optical flow in the alignment step with CNN. More recently end-to-end HDR reconstruction [32–34] or reconstruction through generative adversarial network [24] have also been studied. However, those approaches won't work for large amount of existing legacy content. Thus, in this paper we focus on single image HDR reconstruction.

### 2.2. Single Image HDR Reconstruction

With the remarkable performance of neural networks in various image reconstruction tasks, a natural idea is to use deep learning methods to reconstruct multiple exposure SDR images from a single SDR input, and then synthesize the HDR image from the reconstructed ones. This strategy is also branded as "reverse tone mapping" [8]. The reconstruction of the multiple exposure images can be achieved via chained sub-networks [19], and the synthesized image quality can be further improved by using GAN style training [20, 21].

Another way in single image HDR reconstruction is to directly predicting the lost details in the overexposed regions, mostly use neural networks. In [6] the overexposed region is first detected using pixel intensity, then the missing details in the overexposed region is restored using U-net structure. This method does not consider the tone mapping of normal exposed region between LDR and HDR images. In [22] the reverse process of camera pipelines is learned through multiple different networks. First, quantization error is restored through the first network. Then, inverse tone mapping of normal exposed region is done. Last, the same U-net structure of [6] is used to restore the details of the overexposed region. In [16] the idea of recurrent neural network is adopted in convolutional neural network (CNN). An LDR image is run through CNN multiple times to generate an HDR image. Through multiple iterations, this network can increase the receptive field of its network with small number of its parameters. However, the multiple iteration also increases the inference time a lot. In [1] a condition network with SFT layer is adopted in U-net for the detail reconstruction in HDR image. Due to the SFT layer, the details in the overexposed region can be adaptively restored for each input. In [2], three different steps sequentially combined to convert a SDR image to an HDR image. First, adaptive global color mapping is performed using base network and condition network. Then, ResNet structure is adopted to perform local enhancement. Last, Unet structure is used for highlight detail generation. Our work is also related to image exposure correction. A complete survey of this topic is beyond the scope of this paper and the reader is referred to the literature[9, 13, 14, 30, 36] for further details.

## 3. Proposed Method

To utilize deformable convolution for single image HDR reconstruction, we choose the network architecture of HDRUNET as a baseline and modify it for our purpose. The reason why we choose HDRUNET as a baseline is that this architecture has separate sub-networks for each of overexposed regions and normal regions, which can help to train its network efficiently on many HDR reconstruction datasets that have different tones for a overexposed region and a normal exposed region. Also, it has shown good performance in many previous works [1, 28] while its size is relatively small compared to other methods.

Similar to HDRUNET, our DCDR-UNet consists of three modules such as restoration net, condition net and tone net. The restoration net mainly focuses on restoring the lost details of overexposed objects or regions. This network is a U-shape structure and uses an LDR image input with 3 condition maps generated from condition net to restore the details. It mainly consists of multiple special blocks, namely DCRB (Deformable Convolution Residual Block), which combines offset estimation, deformable convolution and SFT layer in residual block. Through the offset estimation, which predicts the relative 2D locations of neighbor pixels to be convolved with a deformable convolution filter, in DCRB, our network learns the proper receptive fields on each pixel location. The deformable convolution filter then determines how the neighbor pixels within the receptive fields should be used for the best restoration quality. The DCRB is the main difference between our network and HDRUNET. Due to the restoration capability of DCRB, we can improve the reconstructed HDR quality much more than HDRUNET even though we use the less number of DCRBs (8) than the number of the residual blocks (12) in HDRUNET. It eventually makes our network smaller than HDRUNET. To utilize the DCRB correctly, the loss function is very important. In this paper, we combine pixel loss and perceptual loss for HDR images which are encoded with hyper tangent. Note that the perceptual loss is also used in [2, 16, 22], but no one uses the loss with any deformable convolution based module like DCRB. It is well described in 3.4 and 4.3. The condition network generates different condition maps for each scale ($\times$ 1, $\times$ 0.5, $\times$ 0.25), and the condition maps are used as inputs for SFT layer and DCRB in the restoration net. Through condition maps, our restoration network can generate the lost details in the overexposed region more adaptively. Different from HDRUNET, our condition network downsamples the image first and apply multiple convolutions followed by upsampling, which makes our condition map be smooth and be generated with larger neighbor regions. For LDR and HDR images, there are generally tone differences in not only an overexposed region but also a normal exposed region. To force the restoration net to be utilized in restoring the lost

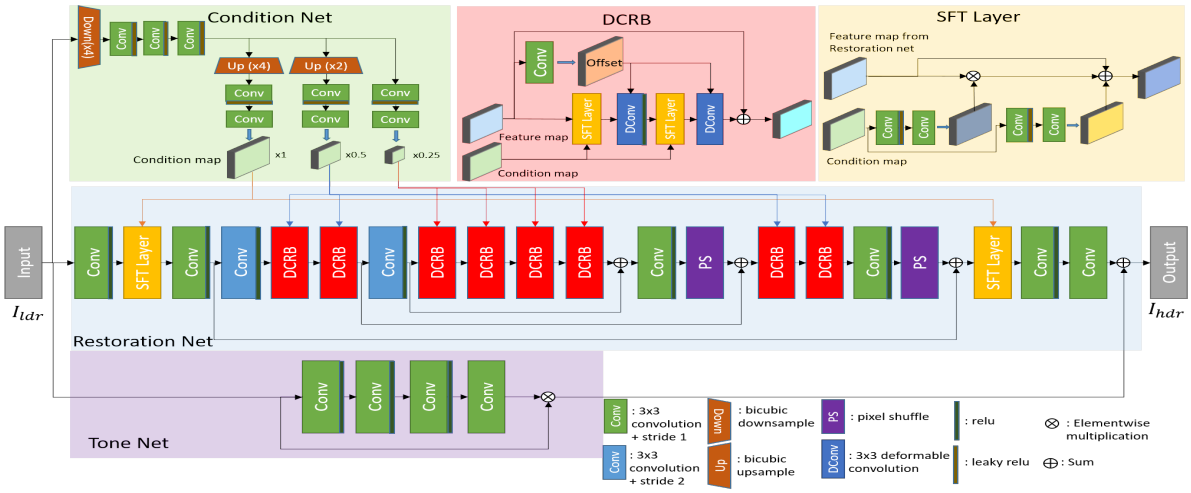


Figure 3. Overall block diagram of DCDR-UNet

details and tones in the overexposed region, we also need a separate network that matches tones between the normal exposed regions of LDR and HDR images. For this purpose, we need a tone network. The entire network is well shown in Figure 3.

Let $Net_R()$, $Net_C()$, and $Net_T()$ be the restoration net, condition net, and tone net respectively. Then, our DCDR-UNet, $Net_{DCDR}()$, can generate an HDR image ($I_{hdr}$), which restores the lost details in the overexposed region in an LDR image ($I_{ldr}$), by adding the two outputs of the restoration net and the tone net:

$$I_{hdr} = Net_{DCDR}(I_{ldr}) = Net_R(I_{ldr}, Net_C(I_{ldr})) + Net_T(I_{ldr}) \quad (1)$$

Note that $Net_C(I_{ldr})$ generate three scale condition maps and $Net_R$ accepts multiple inputs including the condition maps as well as $I_{ldr}$.

### 3.1. Restoration Net

The main role of restoration net is to restore details in the overexposed region. In this network, not only an input image but also multiple condition maps generated from the condition net are utilized to restore the lost details in the overexposed region. As shown in Figure 3, our restoration net has a U-shape network with three scales. In the first scale ($\times$ 1), several convolution layers with two SFT layers are used to encode low level features and reconstruct the lost details from the encoded image features. In the second and third scale ($\times$ 0.5, $\times$ 0.25), multiple DCRBs are used to encode and decode the images features to restore the missing details in the overexposed regions.

#### 3.1.1 Spatial Feature Transform (SFT) Layer

During training, our network learns how to restore the lost details in the overexposed regions from multiple similar training images. However, the similar training images have some variations. For example, there would be multiple training images that include sky but the sky have many variations such as different colors, textures or clouds. To make our network learn these variations better, we adopt SFT layer [29].

As shown in Figure 3, we utilize the SFT layer in all the scales of the restoration net. At the first scale, we use the two SFT layers: one in the encoder and the other one in the decoder side. At the second and third scale, we utilize the SFT layer in DCRB. In each DCRB, two SFT layers used with two deformable convolutions. The SFT layers help the DCRB to be able to generate the different details for different image contents better. Let $x$ and $y$ be inputs for our SFT layer: $x$ is a feature map from previous layer in the restoration net and $y$ is a condition map at a certain scale from the condition net. Then, our SFT layer is defined as:

$$SFT(x, y) = x + cv^2(y) \odot x + cv^2(y) \quad (2)$$

where $cv^2(y) = cv \circ cv(y)$ is two sequentially connected convolutional layers which have a $3 \times 3$ filter size and one leaky ReLU between them and $\odot$ is the element-wise multiplication. With a residual style of the SFT layer, our training is more stable.

#### 3.1.2 Deformable Convolution Residual Block (DCRB)

Our DCRB has mainly three components such as offset estimation, SFT layer and deformable convolution. First, the offset estimation predicts the relative 2D locations of $k \times k$ neighbor pixels, which will be convolved with a deformable convolution filter, for each pixel. It will generate an offset feature map with $2 \times k \times k$ channels. This feature map is then used in a deformable convolution that has the $k \times k$ filter size. For example, our deformable convolution has the $3 \times 3$ filter size, then the offset estimation generates an offset feature map that has 18 channels. Estimating the correct

offset is very important in DCRB because choosing the best neighbor pixels for deformable convolution is the key point of learning the receptive field on each pixel. Let $z$ be an input feature map for DCRB. The offset in the DCRB is then estimated by:

$$o(z) = cv(z) \tag{3}$$

where $o(z)$ is the offset feature map with 18 channels and $cv(z)$ is a convolution operation with $3 \times 3$ filter size on $z$.

When the offset is estimated, the input feature map, $z$, is skipped to the output of the block. Also, the same input passes through the first SFT layer followed by the first deformable convolutional layer that uses the estimated offset as another input. Then, the output of the first deformable convolutional layer goes through the second SFT layer followed by the second deformable convolutional layer. The output of the second deformable convolutional layer would be added to the $z$. Note that different from original deformable convolutional layer [4] that estimates the offsets using the direct input of the deformable convolution layer, we estimate the offset from the input of DCRB and use the same offset for the first and second deformable convolution layers. That is because the direct inputs of the deformable convolutional layers have only residual information and they are not enough for estimating accurate offsets while the input of DCRB has entire information of restored details in previous DCRB. Let $z$ and $y$ be the inputs for DCRB: $z$ (an input feature map of DCRB), $y$ (an input condition map from the condition net). The DCRB is then defined as:

$$DCRB(z, y) = z + dcv(o(z), SFT(dcv(o(z), SFT(z, y)), y)) \tag{4}$$

where $dcv(in1, in2)$ is the deformable convolution using two inputs: $in1$ and $in2$ represent an offset feature map and a feature map generated from the previous SFT layer.

### 3.2. Condition Net

In this network, we generate 3 scales condition maps, which will be used in the restoration network. By providing the condition maps, our restoration network can better reconstruct the details for different image contents. First, the down-sampled input passes through multiple convolution layers to generate feature maps and the feature maps are up-sampled to generate different scale feature maps. Then the feature map at each scale passes through another multiple convolutional layers to generate final condition map at each scale. Given an input LDR image $I_{ldr}$, the condition map at each scale is defined as:

$$y_s = Net_C(I_{ldr}) = cv^2(up_{1/s}(cv^3(d_4(I_{ldr})))) \tag{5}$$

where $y_s$ is a condition map the $s \in 1, 0.5, 0.25$ scale, $d_4()$ is a downsample operation by 4, and $up_{1/s}()$ an upscale operation by $1/s$.

### 3.3. Tone Net

In general, there is an image tone difference between LDR and HDR images. To force our restoration net to focus on restoring the lost details in an overexposed region, we have a tone net. The tone net mainly matches the images tones of the non-overexposed region between LDR and HDR. Given an input LDR image, $I_{ldr}$, the tone net is defined as:

$$Net_T(I_{ldr}) = I_{ldr} \odot T_G(I_{ldr}) \tag{6}$$

where $T_G(I_{ldr}) = cv^4(I_{ldr})$ is the local tone gain map and $cv^4$ is four sequentially connected convolutional layers which have a $3 \times 3$ filter size. Note that the last convolutional layer does not have ReLU after convolution. To show that the tone net mainly works for restoring the tones of the non-overexposed region of LDR, we visualize the outputs of the tone net and restoration net in the supplementary material.

### 3.4. Loss Function

In [1], L1 distance between the predicted HDR and ground truth HDR, which are encapsulated with hyper tangents, (called $Tanh\_L1$ distance) shows better restoration results than simple L1 or L2 distance. We also adopt this loss in our loss function. However, we find that using this loss function only is not sufficient to utilize our DCRB well. If we use this loss function, which is one type of pixel loss, only to train our network, our network mainly focuses to restore a dominant part such as sky in the large overexposed region but not partially overexposed small object like thin power lines. For example, an LDR image in the second row of Figure 4 shows that the over exposed region is mostly sky, but the thin power lines are partially overexposed. For the image, our network with $Tanh\_L1$ loss restores the sky regions well but does not restore the partially overexposed power line correctly. This is because $Tanh\_L1$ loss is not enough to emphasize small or thin objects compared to a large sky region. For this problem, we adopt additional VGG loss used in [18] and combine it with the $Tanh\_L1$ loss. Since the VGG loss is a perceptual loss that helps to restore the object in perceptually correct way and it shows the effectiveness in super resolution task in [18], we use this loss as an additional loss. The VGG loss,$L_{VGG}$, is defined as:

$$L_{VGG}(I^{pred}_{hdr}, I^{gt}_{hdr}) = \sum_{(i,j)\in S} \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{pred}_{hdr})_{x,y} - \phi_{i,j}(I^{gt}_{hdr})_{x,y})^2 \tag{7}$$

where $I^{pred}_{hdr}$ is a predicted HDR image from our network, $I^{gt}_{hdr}$ is a ground truth HDR image, $\phi_{i,j}$ is the feature map obtained by the $j^{th}$ convolution before the $i^{th}$ maxpooling layer in the VGG19 network, $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the $\phi_{i,j}$, and $S = \{(1,2), (2,4), (3,8), (4,12), (5,16)\}$ is the set of the feature map indexes to be selected. Then, our final loss

function,$L_{final}$, is defined as:

$$L_{final}(I_{hdr}^{pred}, I_{hdr}^{gt}) = |Tanh(I_{hdr}^{pred}) - Tanh(I_{hdr}^{gt})| + w_{vgg} \cdot L_{VGG}(I_{hdr}^{pred}, I_{hdr}^{gt}) \quad (8)$$

where $w_{vgg} = 0.1$ is the weight that makes a balance between the $Tanh\_L1$ loss and the VGG loss. By minimizing this loss, our network is trained to achieve a detail restoration on not only the large exposed region but also the partially overexposed small object within the region.

# 4. Experimental Results

## 4.1. Experimental Setup

### 4.1.1 Dataset

To validate the effectiveness of our network, we use the public dataset from [3]. According to [7], the unknown maximum peak luminance causes us hard to use the quantitative metrics such as PSNR and SSIM. The maximum peak luminance of the HDR image in this dataset is within 16 bit depth (many of LDR/HDR datasets use "hdr" file extension for HDR images, where the maximum peak luminance is unknown). To evaluate our method more quantitatively, we mainly use this dataset here. This dataset includes 2975 and 1525 pairs of LDR (8 bits) and HDR (16 bits) images for training and test respectively. This dataset includes the pairs of LDR and HDR images, which of them are geometrically aligned, for many HDR scenes. Also, the size of overexposed regions in the images in the dataset varies from small to large, which is effective to validate our network. Note that the HDR images have 16 bits linear color depth, so the normal exposed regions in the HDR images look very dark. For training, we use all the 2975 pairs of LDR and HDR images from training set. First, we collect multiple pairs of 512×512 patches from the pair of LDR and HDR image. To prevent the case that the patches are not selected from overexposed region, we split the entire image into the non-overlapped sub-regions so that the size of each sub-region becomes 512×512. For test, we generate two sets from all the 1525 test images. For the first set (Test set1), we use all the 1525 images to evaluate our network. However, we realize that there are many LDR images that have small overexposed regions in this test set. To evaluate how much our network can restore the lost details from large overexposed regions as well, we collect the LDR images with large overexposed regions and the corresponding HDR images. To select the LDR images with large overexposed regions, we first convert a RGB LDR image to a grayscale LDR image. Then, we count the number of the grayscale pixels that have larger intensity than threshold, $T_{bright}$. If the ratio of the number of the bright pixels and the total number of the pixels in the image is larger than $\lambda$, we select the LDR image with the corresponding HDR image. $T_{bright} = 240$ and $\lambda = 0.029$ are empirically chosen to find enough number of images. The total 424 pairs of LDR and HDR images are chosen for this set. We call this set as Test set2.

### 4.1.2 Evaluation metrics

To evaluate our method with the existing methods, we use 5 metrics: PSNR-L [1], PSNR-$\mu$ [1], SSIM [31], LPIPS [35], and HDR-VDP2 [23], which are widely used in many existing methods [1, 2, 6, 16, 22, 28], for quantitative evaluation. For PSNR-$\mu$, we use $\mu = 10$ while $\mu = 5000$ is used in [1] because higher $\mu$ can remove the restored details in the overexposed region again. For HDR-VDP2, the linear rgb normalized within [0,1] is used with the option "rgb-native" and the "pixels per degree" is set to 24.

### 4.1.3 Implementation Details

To train our network, the LDR (8 bit) and HDR (16 bit) patches are normalized within [0,1]. During training, the batch size set to 16 and the number of training iteration is set to $2 \times 10^5$. All the network parameters are randomly initialized using Kaiming initialization [11] and optimized using Adam optimizer [17].

## 4.2. Ablation Study

We performed an ablation study on three main components for our DCDR-Net in Table 2: DCRB, tone net and condition net. By comparing the entire model with the model without each of three components, we show how much each component contributes to final HDR image reconstruction. Note that "no DCRB" means that we use a normal residual block, which is defined in [1]: the offset estimation is removed and the deformable convolutions are replaced with normal convolutions in DCRB. For "no condition net" case, since condition net is removed, every SFT layer in the restoration net is also replaced with a normal convolutional layer. Table 2 shows that the all the components contribute to improve the reconstructed HDR image quality, but the DCRB improves the most compared to the other two components. It proves that how important DCRB is in our network.

## 4.3. Effectiveness of DCRB and VGG Loss

We performed additional experiments to see how important the combination of DCRB and an additional VGG loss is for single image HDR reconstruction. The results are shown in Table 3 and Figure 4. Table 3 shows that the improvement is more significant between (b) and (a). That is because the large overexposed objects such as overexposed sky or overexposed building are well restored using DCRB with simple $Tanh\_L1$. However, if there are sparse non-overexposed pixels of the partially overexposed small object in the LDR image, DCRB with simple $Tanh\_L1$ does not restore it well as shown in Figure 4 (b). As shown in Figure 4 (c), the additional VGG loss helps the small object be restored better. However, when the DCRB is used with the additional

| | Test Set1 | | | | | Test Set2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | PSNR-L↑ | PSNR-$\mu$↑ | SSIM↑ | LPIPS↓ | HDR-VDP2↑ | PSNR-L↑ | PSNR-$\mu$↑ | SSIM↑ | LPIPS↓ | HDR-VDP2↑ | No.Param↓ |
| HDRCNN [6] | 47.69 | 48.15 | 0.9973 | 0.0057 | 62.73 | 40.64 | 43.40 | 0.9930 | 0.0121 | 59.40 | 27.8M |
| SingleHDR [22] | 39.97 | 39.57 | 0.9937 | 0.0077 | 64.80 | 34.63 | 37.36 | 0.9890 | 0.0146 | 61.10 | 61.0M |
| FHDR [16] | 46.54 | 46.84 | 0.9963 | 0.0046 | 66.58 | 37.96 | 40.57 | 0.9905 | 0.0113 | 61.45 | 0.6M |
| HDRTV [2] | 49.36 | 50.77 | 0.9975 | 0.0037 | 67.75 | 39.74 | 43.16 | 0.9935 | 0.0097 | 62.71 | 37.2M |
| KUNet [28] | 48.03 | 49.43 | 0.9975 | 0.0043 | 64.82 | 38.87 | 42.11 | 0.9935 | 0.0106 | 59.98 | 1.1M |
| HDRUNET [1] | 50.41 | 52.00 | 0.9980 | 0.0036 | 65.47 | 40.78 | 44.29 | 0.9945 | 0.0095 | 60.87 | 1.7M |
| **Proposed Method** | 52.47 | 54.71 | 0.9985 | 0.0026 | 68.68 | 43.25 | 47.09 | 0.9959 | 0.0072 | 63.43 | 1.5M |

Table 1. Quantitative performance comparison (Red : the best, Blue : the second, Green : the third)

| | Test Set1 | | | | Test Set2 | | | |
|---|---|---|---|---|---|---|---|---|
| DCRB | × | ✓ | ✓ | ✓ | × | ✓ | ✓ | ✓ |
| Tone Net | ✓ | × | ✓ | ✓ | ✓ | × | ✓ | ✓ |
| Condition Net | ✓ | ✓ | × | ✓ | ✓ | ✓ | × | ✓ |
| PSNR-L | 51.08 | 52.01 | 52.29 | 52.47 | 41.01 | 42.96 | 42.76 | 43.25 |
| PSNR-$\mu$ | 52.66 | 53.69 | 54.39 | 54.71 | 44.54 | 46.56 | 46.48 | 47.09 |

Table 2. Ablation study on main three components.

| | Test Set1 | | | | Test Set2 | | | |
|---|---|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| DCRB | × | ✓ | × | ✓ | × | ✓ | × | ✓ |
| VGG loss | × | × | ✓ | ✓ | × | × | ✓ | ✓ |
| PSNR-L | 50.46 | 51.24 | 51.08 | 52.47 | 40.70 | 42.26 | 41.01 | 43.25 |
| PSNR-$\mu$ | 52.21 | 53.57 | 52.66 | 54.71 | 44.20 | 45.98 | 44.54 | 47.09 |

Table 3. Effectiveness of DCRB and VGG Loss. The improvement is significant when both DCRB and VGG loss are used (d) compared to using either DCRB (b) or VGG loss (c) solely.

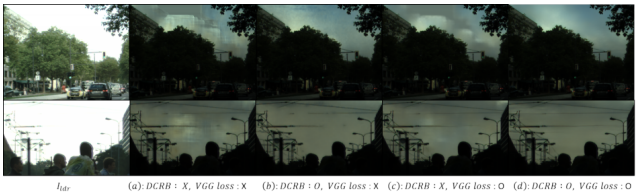


Figure 4. Visual analysis of DCRB and VGG Loss

VGG loss (d), the quality of the reconstructed HDR image is much improved. It proves that our DCRB should be used with the additional VGG loss.

### 4.4. Performance Comparison

#### 4.4.1 Implementation of existing methods

We compare our method against 6 existing methods [1, 2, 6, 16, 22, 28]. Note that the existing methods that provide publicly available training codes are chosen here for fair comparison. Therefore, we retrain all of them using their official implementation on our training dataset. The same $512 \times 512$ patches are used as well. For [6], which does not consider the different tones of normal exposed region between LDR and HDR in default setting, we modify its setting so that its network can be used for entirely region, not only for overexposed regions.

#### 4.4.2 Comparison with existing methods

The Quantitative results are shown in Table 1. According to Table 1, our method achieves much better performances than all the existing methods over all the metrics. Especially, our method achieves better scores than HDRUNET and HDRTV, which are the state of the art in single image based HDR reconstruction, by 2.06 and 2.71 and by 3.11 and 3.94 in test set1 on PSNR-L and PSNR-$\mu$. Also, it achieves better scores than them by 2.47 and 2.80 and by 3.51 and 3.93 in test set2 on PSNR-L and PSNR-$\mu$. The number of training parameters of our network is 1.5M, which is smaller than that of HDRUNET and HDRTV. Even though two existing methods of FHDR and KUNet have less number of parameters than our method, their performances are much lower than our method.

Figure 5 shows the visual quality results of all the methods. Note that we do not apply any tone mapping algorithm to all the images here because a tone mapping sometimes faints the artifacts and details in the restored overexposed region. Because of this, normal exposed regions in all the images look very dark. Instead, we include all the tone mapped results of the same images in the supplemental materials. The images on the $1^{st}$, $5^{th}$ and $7^{th}$ column show that our method can restore the partially overexposed small objects such as the tower, power line and leaves as well as the large overexposed sky better than all the existing methods. For the images on the $2^{nd}$, $4^{th}$ and $6^{th}$ column, the overexposed objects such as cars and buildings are restored very close to the ground truth in our method compared to the existing methods. More qualitative comparisons can be found in the supplementary material.

## 5. Conclusion

In this paper, we propose a DCDR-UNet that firstly utilized deformable convolution in single image HDR reconstruction. Thanks to the combination of DCRB and our loss function, our network can learn the receptive field for each overexposed pixel effectively from training images, which helps our network to restore the lost details in the overexposed region regardless of the size of the region. Our experimental results show that our DCDR-UNet, which has less parameters than the most existing methods, can restore the details of both entirely and partially overexposed objects/regions even when they are very large or small. Also, our network achieves the best quantitative results against all the existing methods over all the 5 metrics.


Figure 5. Qualitative Comparison. Our results show much better visual quality against all the existing methods. Note that the lower left region of each image is zoom-in region of the red bounding box of each image. From the 1st, 5th and 7th column images, we can see that our method can restore very thin objects as well as the overexposed sky naturally while all the existing methods restore the thin object partially or unnatural sky. Especially, for the 5th column images, our method can reconnects the disconnected power lines partially overexposed in LDR while all the existing methods cannot do. For the 2nd and 6th column images, our method can restore the color and texture of the overexposed cars much closer to GT against all the existing methods. For the 3rd image, our method restores sky with partially overexposed object naturally while many of existing methods produce halo artifacts around the partially overexposed object. Note that we do not apply any tone mapping algorithm to all the images here because tone mapping sometimes faint the artifacts and details in the restored overexposed region. Therefore, normal exposed regions in all the images look very dark. Instead, we include all the tone mapped results of the same images in the supplemental materials. These differences are better visible in pdf version with zoom-in.

## References


[1] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 354–363, 2021. 1, 2, 3, 5, 6, 7

[2] X. Chen, Z. Zhang, J. S. Ren, L. Tian, Y. Qiao, and C. Dong. A new journey from SDRTV to HDRTV. *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 4500–4509, 2021. 1, 3, 6, 7

[3] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 6

[4] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017. 2, 5

[5] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, page 369–378, USA, 1997. 2

[6] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.*, 36(6), 2017. 1, 2, 3, 6, 7

[7] G. Eilertsen, S. Hajisharif, P. Hanji, A. Tsirikoglou, R. Mantiuk, and J. Unger. How to cheat with metrics in single-image HDR reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 3998–4007, 2021. 6

[8] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6), 2017. 3

[9] F Eyiokur, Dogucan Yaman, Hazim Kemal Ekenel, and Alexander Waibel. Exposure correction model to enhance image quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 676–686, 2022. 3

[10] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.*, 35(6), 2016. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015. 6

[12] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. HDR deghosting: How to deal with saturation? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 1163–1170, 2013. 2

[13] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6043–6052, 2022. 3

[14] J. Huang, F. Zhao, M. Zhou, J. Xiao, N. Zheng, K. Zheng, and Z. Xiong. Learning sample relationship for exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9904–9913, Los Alamitos, CA, USA, 2023. 3

[15] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4), 2017. 3

[16] Zeeshan Khan, Mukul Khanna, and Shanmuganathan Raman. Fhdr: Hdr image reconstruction from a single ldr image using feedback network. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5, 2019. 1, 2, 3, 6, 7

[17] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *The International Conference on Learning Representations (ICLR)*, 2015. 6

[18] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 5

[19] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 6:49913–49924, 2018. 3

[20] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 596–611, 2018. 1, 3

[21] Siyeong Lee, So Yeon Jo, Gwon Hwan An, and Suk-Ju Kang. Learning to generate multi-exposure stacks with cycle consistency for high dynamic range imaging. *IEEE Transactions on Multimedia*, 23:2561–2574, 2021. 2, 3

[22] Y. Liu, W. Lai, Y. Chen, Y. Kao, M. Yang, Y. Chuang, and J. Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 3, 6, 7

[23] R. Mantiuk, K. Kim, A. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4), 2011. 6

[24] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE Transactions on Image Processing*, 30:3885–3896, 2021. 3

[25] K. Ram Prabhakar, Susmit Agrawal, Durgesh Kumar Singh, Balraj Ashwath, and R. Venkatesh Babu. Towards practical and efficient high-resolution hdr deghosting with cnn. In *Proceedings of the European Conference on Computer Vision (ECCV)*, page 497–513, Berlin, Heidelberg, 2020. Springer-Verlag. 3

[26] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6), 2012. 2

[27] Ana Serrano, Felix Heide, Diego Gutierrez, Gordon Wetzstein, and Belen Masia. Convolutional sparse coding for high dynamic range imaging. In *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics*, page 153–163, Goslar, DEU, 2016. Eurographics Association. 2

[28] Hu Wang, Mao Ye, Xiatian Zhu, Shuai Li, Ce Zhu, and Xue Li. Kunet: Imaging knowledge-inspired single hdr image reconstruction. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1408–1414. International Joint Conferences on Artificial Intelligence Organization, 2022. Main Track. 1, 2, 3, 6, 7

[29] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 4

[30] Yang Wang, Long Peng, Liang Li, Yang Cao, and Zheng-Jun Zha. Decoupling-and-aggregating for image exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18115–18124, 2023. 3

[31] 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. 6

[32] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *The European Conference on Computer Vision (ECCV)*, 2018. 3

[33] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020.

[34] Qingsen Yan, Dong Gong, Javen Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Dual-attention-guided network for ghost-free high dynamic range imaging. *International Journal of Computer Vision*, pages 1–19, 2021. 3

[35] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 6

[36] Yijie Zhou, Chao Li, Jin Liang, Tianyi Xu, Xin Liu, and Jun Xu. 4k-resolution photo exposure correction at 125 fps with 8k parameters. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024. 3