Deep HDR estimation with generative detail reconstruction: Supplementary Materials

Y. Zhang and T. O. Aydin
DisneyResearch|Studios

In this supplementary material, we provide further implementation and network architecture details. Additional visual results are shown for a more thorough ablation study as well as comparison between our results and the current state-of-the-art.

1. Additional Objective evaluations

In Table 1, we report additional objective evaluations that utilize the perceptually uniformity (PU) encoding. Our proposed method also outperforms existing HDR reconstruction methods in terms of PU-PSNR and PU-SSIM.

<table>
<thead>
<tr>
<th>Method</th>
<th>PU-PSNR mean (σ)</th>
<th>PU-MS-SSIM mean (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDR CNN [EKD*17]</td>
<td>44.4923 (5.5717)</td>
<td>0.9343 (0.0973)</td>
</tr>
<tr>
<td>DrTM [EK17]</td>
<td>33.7938 (1.9358)</td>
<td>0.8519 (0.1133)</td>
</tr>
<tr>
<td>ExpandNet [MBRHD18]</td>
<td>35.6328 (5.1561)</td>
<td>0.8660 (0.1161)</td>
</tr>
<tr>
<td>Santos et al. [STKK20]</td>
<td>45.8392 (6.8725)</td>
<td>0.9483 (0.1867)</td>
</tr>
</tbody>
</table>

Baseline

<table>
<thead>
<tr>
<th>Method</th>
<th>PU-PSNR mean (σ)</th>
<th>PU-MS-SSIM mean (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-net only</td>
<td>43.3657 (2.1593)</td>
<td>0.9226 (0.1145)</td>
</tr>
<tr>
<td>D-net only</td>
<td>42.9325 (2.0148)</td>
<td>0.9092 (0.1127)</td>
</tr>
<tr>
<td>R-net only</td>
<td>43.7787 (1.9572)</td>
<td>0.9126 (0.1132)</td>
</tr>
<tr>
<td>(B + D) net</td>
<td>44.8475 (2.4824)</td>
<td>0.9367 (0.1213)</td>
</tr>
<tr>
<td>(B + D) net</td>
<td>45.2573 (2.1467)</td>
<td>0.9432 (0.1144)</td>
</tr>
</tbody>
</table>

Proposed (B + D + R) net

Table 1: PU encoded HDR reconstruction performance of our method compared to various baselines and state-of-the-art methods.

2. Network architectures

We report our implementation details of the network architectures. For simplicity, we denote the layers with K (kernel size), D (dilation), S (stride size) and C (number of channels).

Base layer reconstruction network & Refinement network are adapted from U-net architecture. In the encoder, all convolutional layers followed by max pooling layer. The skip connections from each encoder layer to its corresponding decoder layer have been added. The decoder uses bilinear upsampling (BU) to upsample feature maps. All convolutional layers except the last one are followed by ReLU activation functions.

Encoder: K3S1C64 - K3S1C128 - K3S1C256 - K3S1C512 - K3S1C512

Decoder: K3S1C256 - K3S1C128 - K3S1C64 - K3S1C64 - K3S1C3 - clip

Detail reconstruction network is constructed by two stages stacked together. The first stage is convolutional network with partial convolutional layers. Nearest upsampling (NU) has been employed in the decoder of this network and skip connections have been added from each encoder layer to its corresponding decoder layer. The second stage, a contextual inpainting network contains two parallel encoders: dilated convolutional branch and contextual inpainting branch. The output of the two parallel encoders are concatenated together to feed into the decoder. All convolutional layers except the last one are followed by ReLU activation function.

Stage 1 Encoder: K7S2C64 - K5S2C128 - K5S2C256 - [K3S2C512] ×3.

Stage 1 Decoder: [K3S1C512] ×2 - K3S1C256 - K3S1C128 - K3S1C64 - K3S1C3 - K3S1C3 - sigmoid.

Stage 2 Dilated Convolutional Branch: K5S1D1C32 - K3S2D1C32 - K3S1D1C64 - K3S2D1C64 - K3S1D1C128 - K3S1D2C128 - K3S1D4C128 - K3S1D8C128 - K3S1D16C128 - concat.

Stage 2 Contextual Attention Branch: K5S1D1C32 - K3S2D1C32 - K3S1D1C64 - K3S2D1C64 - K3S1D1C128 - ReLU - contextual attention layer [YLY*18] - K3S1D1C128 - K3S1D1C128 - concat.

Stage 2 Decoder: K3S1D1C128 - K3S1D1C128 - NU(2×) - K3S1D1C64 - K3S1D1C64 - UN(2×) - K3S1D1C32 - K3S1D1C16 - K3S1D1C16 - K3S1D1C3 - clip.

The WGAN-GP loss has been calculated to both masked regions and the whole image respectively.

3. Additional ablation examples

3.1. Detail reconstruction network with & without contextual detail inpainting component

Figure 1 and 2 shows example results of the detail reconstruction network with and without the contextual detail inpainting component.

4. Additional comparisons with other methods

Additional examples of the comparisons between our model and other existing methods can be found in Figure 3 and 4.
Figure 1: Example results of the detail reconstruction network where the network with and without the contextual detail inpainting subnetwork component.
Figure 2: Example results of the detail reconstruction network where the network with and without the contextual detail inpainting subnetwork component.
Figure 3: Comparison of our method’s results with state-of-the-art techniques. Input SDR image (a) and corresponding crops (b) are shown on the right. We present results from DrTMO [EKM17] (c), ExpandNet [MBRHD18] (d), HDRCNN [EKD17] (e) and our model (f). The reference HDR is shown in (g), and HDR-VDP-2 visible difference map is given in (h).
Figure 4: Additional comparison of our method’s results with state-of-the-art techniques. Input SDR image (a) and corresponding crops (b) are shown on the right. We present results from DrTMO [EKM17] (c), ExpandNet [MBRHD18] (d), HDRCNN [EKD17] (e) and our model (f). The reference HDR is shown in (g), and HDR-VDP-2 visible difference map is given in (h).
References


