Supplementary Material for:

Unpaired Learning for High Dynamic Range Image Tone Mapping

Yael Vinker    Inbar Huberman-Spiegelglas    Raanan Fattal

Content:

1. Appendix: pixFID ........................................ 2

2. Ablation Study - Additional Tests and Images . . . . . . . . . . . . . . 3
   a. Compression Level Estimation ......................... 3
   b. Augmenting the Skip-Connection ................. 4

3. Discriminators at Different Image Scales ........... 5

4. Use Study .......................................................... 6

5. Results by Low Light Enhancement Methods ...... 7

6. Additional visual comparisons ...................... 8
   a. Images from HDR+ dataset [2]. ................... 9

7. Examples training images and pipeline .......... 26
   a. LDR images from the DIV2K dataset [1] . .... 27
   b. Tone-mapping pipeline process ..................... 28
   c. Additional tone-mapped images ................. 31

8. Fail Cases ........................................................... 32

9. References .......................................................... 33
Appendix

The FID measures the distance between the distribution of two image classes by modeling their neural activations using multivariate Gaussians and measuring their Fréchet distance. More specifically, the 2048-dimensional activation vectors produced at the last layer of the Inception-v3 classification network are used. Fitting a Gaussian stably is typically done with using 10,000 samples or more. Since each image produces one sample vector, this is the number of images needed from each class in order to estimate the FID.

Datasets of HDR images do not offer these numbers of samples and do not permit the use of the FID. To overcome this limitation we consider the pooled activations in the last Inception layer of the Inception Module B. This layer consists of a spatial array of $8 \times 8$ pixels, each containing a 768-dimensional activation vector. Thus, compared to the original FID where every image provides a single sample vector, switching to this layer results in 64 samples per image.

We term this modified distance pixFID and compare it to the original FID in Figure 1. We used only 1,000 images to evaluate the pixFID in this comparison, next to 50,000 images used to evaluate the FID. The comparison shows the response of these distances to three types of disturbances that were applied to natural images. In case of the Gaussian blur and the additive noise, the pixFID followed the trend of the FID and increased as the disturbance magnitude increased, implying its ability to identify such low-level discrepancies.

In case of a geometric swirl distortion, which does not change the color and shape distribution of small image patches, the pixFID showed an inferior sensitivity to the transformation magnitude. This behaviour is consistent with the fact that activations from a lower level in the network correspond to smaller receptive field and hence to a visual sensitivity to smaller scale features in the image.

Nevertheless, HDR tone-mapping procedures typically does not suffer from geometric or semantic distortions, but from low-level effects such as loss of local contrast, exaggerated details, and edge-related halos.

---

Figure 1. The response of the FID, red plot, and the pixFID, blue plot, to three types of image disturbances are shown. The FID is computed over 50,000 images whereas the pixFID over 1,000 images.
Ablation Study - Additional Tests and Images

Compression Level Estimation

Below we compare the results produced by our per-image compression factor estimation, next to using a fixed value = 1000 as done in “Deep High Dynamic Range Imaging of Dynamic Scenes” SIGGRAPH 2017. This fixed value (center column) led to a slight over compression in the luminance map of the Motorcycle and a clear under compression in the Belgium House image below. In both these examples, our image dependent factor (right column) brought most of the pixels to a visible range and maintained a consistent level of image contrasts.
As expected, also the final images produced by training and inferring by applying a fixed level of compression show an inconsistent appearance due to the different dynamic range they had at the input.

**Augmenting the Skip-Connection with Square-Root Transformation**

The following table shows the images resulting with and without augmenting the skip connections with a smooth transformation. The lack of one (left column) shows a coarse action over the image with noticeable over- and undershooting effects. The use of a square-root (right column) almost completely eliminates these effects.
**Effect of the Discriminators at Different Image Scales**

The following table shows the results obtained by training our network using discriminator at different scales. From this comparison it is clear that when omitting discriminators at certain scales (coarse scale or two in the right columns of the table), the network does not reproduce the missing (by the prior TRC step) contrasts at these scales.

<table>
<thead>
<tr>
<th>One scale</th>
<th>Two scales</th>
<th>Three scales (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
</tbody>
</table>
User Study

Graph reports the results of the user study we conducted

In order to further assess the visual quality of our results, we conducted a user-study that follows the protocol used in Liang[10]. Specifically, it consists of eight participants who were asked to rate between 1 to 8 images produced by the methods in Liang[10], Paris[7], Rana[12], Gu[8], Farbman[9] as well as ours. The images were displayed next to each other (at a random order). The participants were naive as to the purpose of the test and were asked to score the images based on how realistic and unprocessed they appear.

The graph shows that our method obtained the highest mean score 6.43 with a tolerate std of 1.6. The mean opinion score (in bold) and std of each method are shown in the graph.
Results by Low Light Enhancement Methods

While Low light methods also manipulate the image brightness, they are not adapted to operate on images with very high dynamic range.

Jiang et al. [16]

Guo and Li et al. [15]
Additional Visual Comparisons

from the HDR+ dataset [2]
Additional Visual Comparisons
from the HDR Survey dataset [14]


Rana [12]  Ours
Zhang [11]  

Ours
Zhang [11]  

Ours
Examples Training Images
Examples of LDR images from the DIV2K dataset [1]
Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))
Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))
Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))
Network’s action: $Y_c$ (shown in color) and the network’s output on it are shown.
Fail Cases

Over contrast due to the adversarial loss in the coarsest scale, which encourage higher contrasts

By omitting this term we can reduce this exaggerated appearance
References


