Article

Deep HDR Hallucination for Inverse Tone Mapping

Demetris Marnerides [1], Thomas Bashford-Rogers [2] and Kurt Debattista [1],*

[1] WMG, University of Warwick, Coventry CV4 7AL, UK; Demetris.Marnerides@warwick.ac.uk
[2] Department of Computer Science and Creative Technologies, University of the West of England, Bristol BS16 1GY, UK; Tom.Bashford-Rogers@uwe.ac.uk
* Correspondence: K.Debattista@warwick.ac.uk

Abstract: Inverse Tone Mapping (ITM) methods attempt to reconstruct High Dynamic Range (HDR) information from Low Dynamic Range (LDR) image content. The dynamic range of well-exposed areas must be expanded and any missing information due to over/under-exposure must be recovered (hallucinated). The majority of methods focus on the former and are relatively successful, while most attempts on the latter are not of sufficient quality, even ones based on Convolutional Neural Networks (CNNs). A major factor for the reduced inpainting quality in some works is the choice of loss function. Work based on Generative Adversarial Networks (GANs) shows promising results for image synthesis and LDR inpainting, suggesting that GAN losses can improve inverse tone mapping results. This work presents a GAN-based method that hallucinates missing information from badly exposed areas in LDR images and compares its efficacy with alternative variations. The proposed method is quantitatively competitive with state-of-the-art inverse tone mapping methods, providing good dynamic range expansion for well-exposed areas and plausible hallucinations for saturated and under-exposed areas. A density-based normalisation method, targeted for HDR content, is also proposed, as well as an HDR data augmentation method targeted for HDR hallucination.

Keywords: high dynamic range; inverse tone mapping; deep learning

1. Introduction

High Dynamic Range (HDR) imaging [1] permits the manipulation of content with a high dynamic range of luminance, unlike traditional imaging typically called standard or Low Dynamic Range (LDR). While HDR has recently become more common, it is not the exclusive mode of imaging yet, and a significant amount of content exists which is still LDR. Inverse tone mapping or dynamic range expansion [2] is the process of converting LDR images to HDR to permit legacy LDR content to be viewed in HDR.

The problem of inverse tone mapping is essentially one of information recovery, with arguably its hardest part being the inpainting/hallucination of over-exposed and under-exposed regions. In these regions, there is not sufficient information in the surrounding pixels of the LDR input for interpolation, compared to regions that are, for example, LDR only due to quantisation and thus preserve some colour and structure information. In addition, inpainting is content and context specific, while other aspects of dynamic range expansion are not very dependent on semantic aspects of the scene.

Deep neural networks are powerful function approximators and CNNs have been shown to be good candidates to target the inpainting problem [3,4]. However, in this case, it is not only the network architecture that needs to be carefully selected, but the optimisation loss is of paramount importance as well. For example, in other inverse tone mapping works [5–8] the CNNs presented are trained with standard regression losses (L_1, L_2 and/or cosine similarity) and yield good results when considering well exposed areas, but they either do not reconstruct any missing information of saturated areas, or provide some reconstruction in these areas but of not sufficient quality.

In this work, a GAN-based solution to the Inverse Tone Mapping problem is proposed, providing simultaneous dynamic range expansion and hallucination. GANs can provide a

solution to the hallucination problem through the introduction of a learnable loss function, i.e., the discriminator network, which being a CNN itself is powerful enough to capture multiple modalities. The method is end-to-end, producing a linear HDR image from a single LDR input. To handle the extreme nature of the exponentially distributed linear HDR pixel values of natural images a novel HDR pixel distribution transform is proposed as well, mapping the HDR pixels to a more manageable Gaussian distribution.

In summary, the primary contributions of this work are:

- The use of GANs for the reconstruction of missing information in over-exposed and under-exposed areas of LDR images.
- An HDR pixel distribution adaptation method, for normalising HDR content, which is exponentially skewed towards zero.
- A data augmentation method for the use of more readily available LDR datasets to train CNNs for HDR hallucination and inverse tone mapping.
- Comparisons with variants of the proposed GAN-based method using different network architectures and modules.

2. Background and Related Work

Inverse Tone Mapping Operators (ITMOs), attempt to generate HDR from LDR content. They can generally be expressed as:

$$\tilde{I}_{\text{HDR}} = f_{\text{ITM}}(I_{\text{LDR}}), \text{ where } f_{\text{ITM}} : [0, 255] \rightarrow \mathbb{R}^{+} \tag{1}$$

where the predicted HDR image is denoted as $\tilde{I}_{\text{HDR}}$ and the ITMO as f_{ITM}. If we consider a badly exposed part of an image, the missing content from the original scene is completely lost and can take multiple reasonably valid shapes and forms, e.g., the over-exposed part of a road may have contained parts of a sidewalk or alternatively a pedestrian. We can therefore say that the ITM problem is an ill-posed one.

The main procedure followed by ITMOs that are non-learning based is composed of the following steps [1]:

1. Linearisation: Apply inverse CRF and remove gamma.
2. Expansion: Well-exposed areas are expanded in dynamic range. The operation can be local or global in luminance space.
3. Hallucination: Reconstruct missing information in badly-exposed areas. Not all methods have this step.
4. Artefact removal: Remove or reduce quantisation or compression artefacts.
5. Colour correction: Correct colours that have been altered due to saturation of only a single or two channels from the RGB image.

The main distinction between ITMOs is whether the expansion is local or global. Global methods apply the same luminance expansion function per pixel, that has the same form independently of pixel location. The method by Landis [9], which was one of the earlier ones, expands the range using power functions, ultimately to have the content displayed on an HDR monitor. Akyüz et al. [10] expand single LDR exposures using linear transformations and gamma correction, while Masia et al. [11,12] propose a multi-linear model for the estimation of the gamma value.

Local methods use analytical functions to expand the range, which unlike global methods, depend on local image neighbourhoods. Banterle et al. [2], expand the luminance range by applying the inverse of the any invertible Tone Mapping Operator (TMO) and use a smooth low frequency expand map to interpolate between the expanded and the LDR image luminance. Rempel et al. [13] also use an expand map, termed brightness enhancement function (BEF). However, this is computed through the use of a Gaussian filter in conjunction with an edge-stopping function to maintain contrast. Kovaleski and Oliviera [14] extend the work of Rempel et al. [13] by changing the estimation of the BEF. Subsequently, Huo et al. [15] further extend these methods by removing the thresholding used by Kovaleski and Oliviera.

A method that includes inpainting of missing content is proposed by Wang et al. [16]. It is partially user-based, where the user adds the missing information on the expanded LDR image. Kuo et al. [17] attempt to automatically produce inpainted content, using textures that are similar to the surroundings of the over-exposed areas. It is, however, limited in the type of content, as the neighbouring textures must be matching and must be of only one single type.

Recently, a variety of deep learning-based ITMOs have been introduced. Zhang and Lalonde [18] predict HDR environment maps from captured LDR panoramas, while Eilertsen et al. [7], predict content for small over-saturated areas of the LDR input using a CNN, and use an inverse CRF to linearise the rest of the input. Endo et al. [8] predict multiple exposures from a single LDR input using a CNN and fuse them using a standard merging algorithm. Marnerides et al. [5] use a multi-branch architecture for predicting HDR from LDR content in an end-to-end fashion. The spectrally consistent UNet architecture, termed GUNet [6], has also been used to for end-to-end HDR image reconstruction. Lee et al. [19] use a dilated CNN to infer multiple exposures from a single LDR image sequentially, using a chain-like structure, which are then fused. Han et al. [20] propose using a neuromorphic camera in conjuction with a CNN to help guide HDR reconstruction from a single LDR capture, while Sun et al. [21] use a Diffractive Optical Element in conjunction with a CNN and Liu et al. [22] invert the capturing pipeline and use a CNN in conjunction for the same purposes. Sharif et al. [23] use a two stage network for reconstructing HDR images focusing on de-quantisation, while Chen et al. [24] combine inverse tone mapping with denoising. Concurrently to this work, Santos et al. [25] follow a feature-masking methodology along with a perceptual loss to simultaneously expand well-exposed areas and hallucinate badly exposed areas. In contrast to this work, the author's do not follow a GAN-based approach and the CNN architectures are more complicated, implementing masked features at various layers.

3. Method

The proposed Deep HDR Hallucination (DHH) method attempts to expand the range of well-exposed areas and simultaneously fill in missing information in under/over-exposed areas in an end-to-end fashion. First, an overview of the method is given, followed by the presentation of the GAN-based loss function and the proposed HDR pixel distribution transformation for HDR content normalisation.

3.1. Network Architecture

The proposed network is based on the UNet architecture [26], which has been successfully used previously for image translation tasks using GANs [27,28]. The encoder and decoder, downsample and upsample eight times, forming a total of 9 levels of feature resolutions with feature sizes 3-64-128-256-512-512-512-512-1024, similarly to Isola et al. [27]. At each level, the encoder uses a residual block as proposed by He et el [29] formed of two 4×4 convolutional layers along with Instance Normalisation and the SELU [30] activation, while the decoder fuses the encoder features with the upsampled features using a single convolution. Downsampling and upsampling is performed using residual bilinear modules. A diagram of the architecture is shown in Figure 1.

The discriminator network is similar to the PatchGAN discriminator used by Isola et al. [27] for image translation It consists of five layers with 64-128-256-512-1 features, respectively, as shown in Figure 2. The first three layers reduce the spatial dimensionality using bilinear downsampling with a factor of 2. The final layer output consists of a single channel of size 32×32 (for an input of 256×256). Each pixel of the output is a discriminatory measure of an input patch, signalling whether the input patch is from the data distribution or if it's a generator sample. The discriminator uses the Leaky ReLU activation [31], with $\lambda = 0.2$.
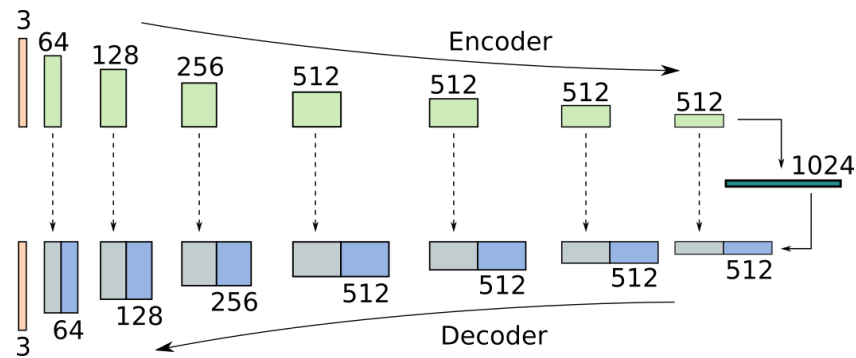
Figure 1. Network architecture based on the UNet architecture [26] with skip connections.
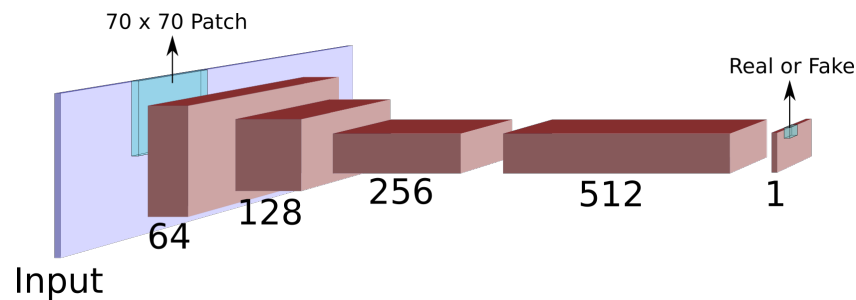

Figure 2. Diagram of the patch-based discriminator network.

3.2. GAN Objective

The original optimisation objective for GANs is often found to be unstable [32–34] and many variations of the loss have been proposed attempting to provide stable alternatives. The loss function DHH adopts is the Geometric GAN method [35] which utilises the hinge loss, commonly used for binary classification:

$$L_{\text{hinge}}(x, y) = \max(0, 1 - xy) \tag{2}$$

where y is the target label (0 or 1) and x is the classifier prediction (as a probability). Applying it in the context of GAN, given a generator network, G, and a discriminator network, D, the GAN losses are given by:

$$L_{D,\text{GAN}} = \mathbb{E}_{I_{\text{HDR}}}[\max(0, 1 - D(I_{\text{HDR}}))] + \mathbb{E}_{I_{\text{LDR}}}[\max(0, 1 + D(G(I_{\text{LDR}})))] \tag{3}$$

$$L_{G,\text{GAN}} = -\mathbb{E}_{I_{\text{LDR}}}[D(G(I_{\text{LDR}}))] \tag{4}$$

where I_{HDR} are target HDR images, corresponding to LDR inputs, I_{LDR} from the dataset. The final predictions, $\tilde{I}_{\text{HDR}}$ are the generator network outputs:

$$\tilde{I}_{\text{HDR}} = G(I_{\text{LDR}}) \tag{5}$$

The hinge loss objective was found to provide stable training for GANs in imaging problems [36] including the large scale "BigGAN" study [37] for image generation. In addition, the generator is optimising two additional losses which help stabilise the training and improve the quality of the results, one being the perceptual loss [38] and the other being the feature matching loss [28], which is similar to the perceptual loss. The optimisation loss for the generator network is given by:

$$L_{G,\text{total}} = L_{G,\text{GAN}} + L_{\text{perc}}(\tilde{I}_{\text{HDR}}, I_{\text{HDR}}) + L_{\text{fm}}(\tilde{I}_{\text{HDR}}, I_{\text{HDR}}). \tag{6}$$

The perceptual loss, L_{perc}, is the same as the one used in [38]. It is based on a VGG architecture [39], trained on the ImageNet dataset [40]. Perceptual losses can recover improved textures since they take into account pixel inter-correlations locally since they

use early layers of the pre-trained network. The feature matching loss, L_{fm}, is calculated as in [28] in the same way that the perceptual loss is calculated on the VGG features, but on the PatchGAN discriminator instead. As pointed out in [28] it can help generate more consistent structures, which in our case are the the under/over exposed areas. The discriminator loss is the same as in Equation (3).

3.3. Data Augmentation

Generative imaging problems require large datasets, of the order of tens of thousands to millions of samples, along with large models for improved results, as demonstrated by Brock et al. [37]. Readily available HDR datasets are small in size, for example, in the works by Eilertsen et al. [7] and Marnerides et al. [5] the datasets used are of the order of 1000 HDR images. For this reason, a data augmentation method is proposed, which uses a pre-trained GUNet model [6] for dynamic range expansion. A large readily available LDR dataset is used which is then expanded using the pre-trained model in order to simulate an HDR dataset. To create input-output pairs the following procedure is followed:

1. Sample an image, I_d from the LDR dataset.
2. Expand the image range using a trained GUNet to predict the HDR image I_{HDR}.
3. Crop and resize the image to 256×256 using the approach from Marnerides et al. [5].
4. Use the Culling operator (clipping of top an bottom 10% of values).

The culling operator on the last step discards information from the expanded HDR target image which the hallucination network attempts to reconstruct. The dataset is sourced from Flickr [41] and consists of 45,166 training images from 17 different categories using the following tags: Cloud, fields, lake, landscape, milkyway, mountain, nature, panorama, plains, scenery, skyline, sky, sunrise, sunset, sun, tree, water.

3.4. HDR Pixel Transform

The HDR reconstruction loss used in recent CNN-based ITM methods [5], comprise of an L_1 loss along with a cosine similarity term. This specific loss, particularly the cosine similarity term is used to accommodate the unique distribution of HDR image pixel values, shown in red in Figure 3. The cosine term, being magnitude independent, helps improve colour consistency in the darker areas, where deviations between the RGB channels do not influence the main regression loss (L_1) greatly thus producing deviations in chrominance.

However, having normally distributed data is beneficial for training CNNs, hence a variety of methods attempt to induce normalised outputs from activations [30,42]. Thus, instead of adjusting the training loss function to fit the distribution, the distribution can be normalised. The proposed method applies such a transformation on the original HDR pixels for inputting into both the generator and the discriminator. The blue histograms in Figure 3 show the HDR pixel distributions after applying a transformation function, such that the resulting distribution is approximately Gaussian.

To derive the transformation function it is first noted that the HDR pixels x approximately follow an exponential distribution of the form:

$$x \sim \lambda e^{-\lambda x} \tag{7}$$

with the parameter lambda estimated from the HDR dataset to be $\lambda \approx 1$. To transform an exponentially distributed variable X to become a Gaussian variable Y the following equation is used:

$$Y = \Phi^{-1}(F_Y(Y)) \tag{8}$$

where Φ is the Cumulative Distribution Function (CDF) of the Gaussian distribution and F_Y is the CDF of the exponential distribution:

$$\Phi(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \tag{9}$$
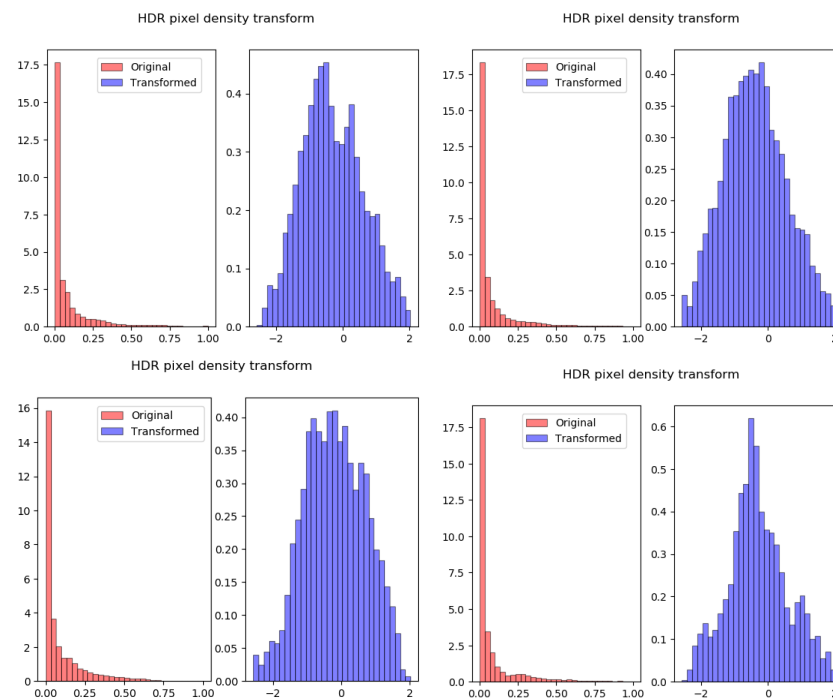
Figure 3. Histograms for HDR pixel distributions before and after the density transform. The four subfigures represent four different samples of 256×256 HDR image mini-batches of size 32. There is some variation observed in the resulting Gaussian distributions, due to the variation in HDR content and the approximation that HDR pixels follow an exponential distribution.

4. Results and Discussion

This section presents both qualitative and quantitative results. For both, DHH is trained using four Nvidia P100 GPUs for approximately 700,000 iterations using a mini-batch of size 36. The generator and discriminator updates are performed with the same frequency (alternating every one step). The learning rates used were $1e^{-4}$ for the generator and $4e^{-4}$ for the discriminator, following the Two Timescale Update Rule (TTUR) [43] using the Adam optimiser with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Both networks make use of spectral normalisation [44] to help with training stability by controlling their Lipschitz constant via limiting the highest singular value (spectral norm) of the linear transformation weight matrices.

4.1. Qualitative

Figure 4 presents sample predictions during training. The first column is the input image, while the second and third columns are the tone-mapped target and prediction images, respectively. The fourth column is the masks of the badly-exposed areas in the input, highlighting the pixels that are completely black or white. The last two columns highlight the badly exposed areas in the target and the prediction respectively.

It is observed that the well-exposed areas of the image are reconstructed well, matching the corresponding areas in the target image. In the badly exposed areas, textures are reconstructed to match the surroundings. The hallucinations blend well with the rest of the image and also exhibit texture variations. For example in the first row, the sky is reconstructed and varies from blue to light red towards the horizon. The second and fourth rows introduce variations in the clouds. However, semantic detail is not reconstructed fully, for example potentially the shape of the sun in the third row.

Figure 5 shows a scan-line plot of a prediction from the HDR test set, using the DHH method, along with the LDR input and HDR target predictions. The LDR input is saturated and constant for most of the line for all three channels (white), while the target and predictions vary and have strong correlations. The target and prediction values are

scaled to be close to the LDR values for visualisation purposes. It is observed that the prediction of the blue channel is lower than the target relative to the prediction of the other two. This is in agreement with the image, since the prediction is more yellow along the scan-line whereas the target image is blue. The tone is reproduced adequately and blends well with the surroundings.
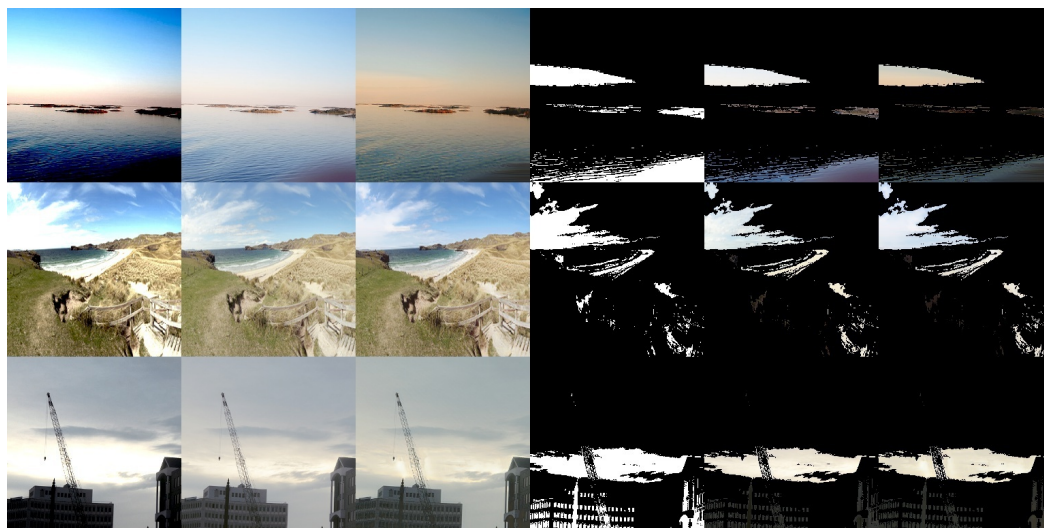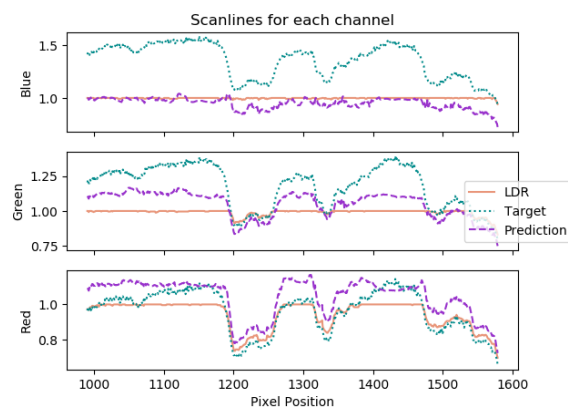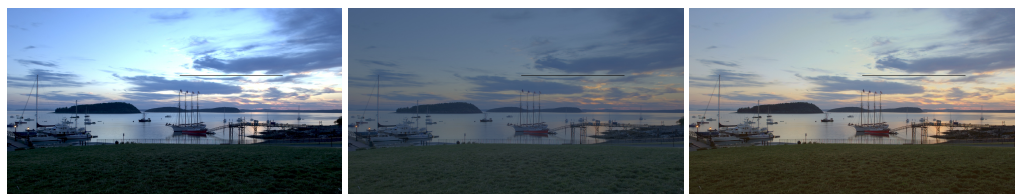

Figure 4. Samples for the DHH method (training). Columns 1–3 are the input LDR image and corresponding tone mapped HDR target and prediction, respectively. Column 4 is a mask showing the fully saturated ($[255, 255, 255]$) and fully under-exposed ($[0, 0, 0]$) pixels of the input. Columns 5 and 6 show the masked target and masked prediction, respectively.

(**a**) Scan-line plot

(**b**) Input LDR   (**c**) Target HDR   (**d**) Prediction HDR

Figure 5. Scan-line plot for a test-set image prediction using DHH. The saturated areas in the LDR are reconstructed well in the prediction and correlate with the target HDR.

Figure 6 shows test-set predictions using DHH at various resolutions. The first column is the LDR input (culling) while columns 2–4 are (tone mapped) predictions from inputs of resolutions 256×256, 512×512 and 1024×1024, respectively. It is observed that the dynamic range expansion is consistent at different resolutions; however, the hallucinations

change structure depending on the input size. In hard cases that are not common in the dataset, for example the side of the building in Figure 6a, the inpainting is not of good quality and the textures do not match the surroundings. The clouds that are washed out in the input LDR are reconstructed well. Likewise, the part of the sky around the sun which is saturated in Figure 6b, along with the grass in the bottom left which is under-exposed, is also reconstructed well. Same for the washed out sky in Figure 6c.

(**a**) BenJerrys


(**b**) BandonSunset(1)
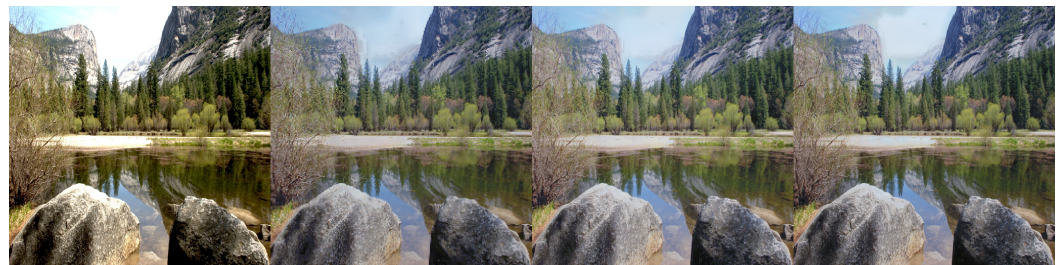
(**c**) MirrorLake

Figure 6. Test-set predictions for inputs at varying resolutions. First column is the input LDR, followed by predictions at resolutions of 256×256, 512×512 and 1024×1024, respectively.

4.2. Quantitative

While DHH is focused more on qualitative results and is not expected to outperform other methods quantitatively as metrics do not account for hallucinated content, the results serve to demonstrate if any further quality in the rest of the image is being lost compared to the standard methods. In addition, quantitative results serve as a validation for the use of the HDR pixel transform proposed in Section 3.4 and the hdr data augmentation regime from Section 3.3.

Figures 7a and 8b show quantitative results of the DHH method for the test-set compared with the CNN methods presented in ExpandNet [5], GUNet [6] and the method by Eilertsen et al. [7], using the perceptually uniform (PU) [45] encoded PSNR, SSIM [46], MS-SSIM [47] and HDR-VDP-2 [48].
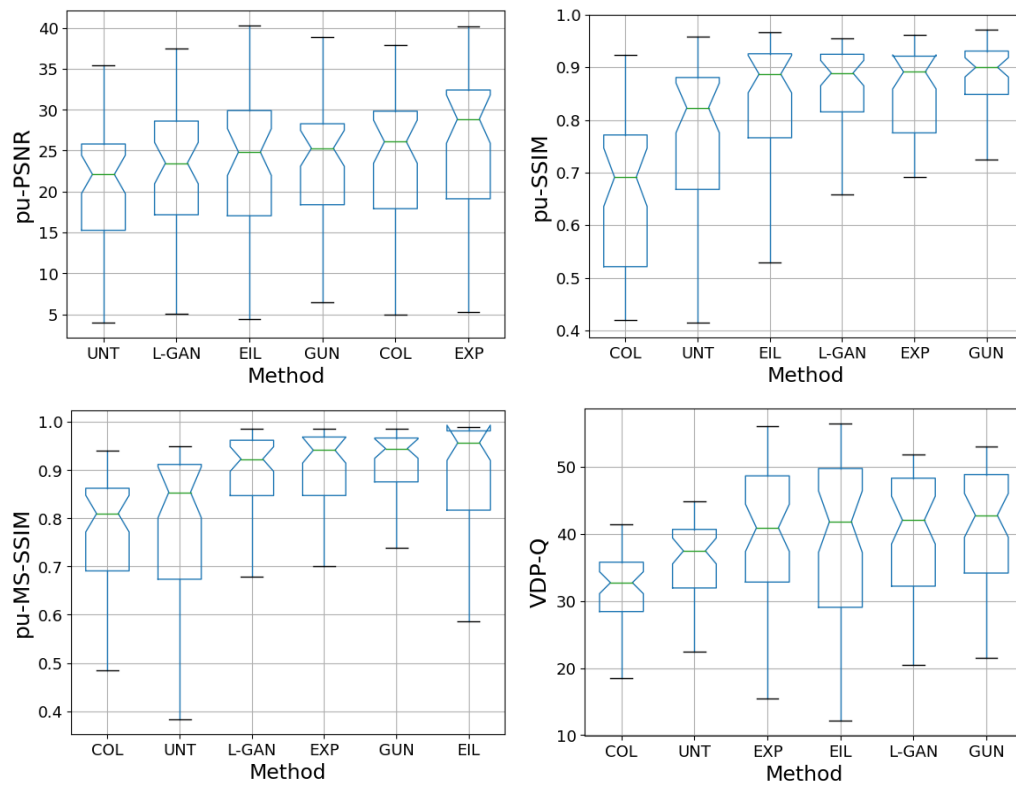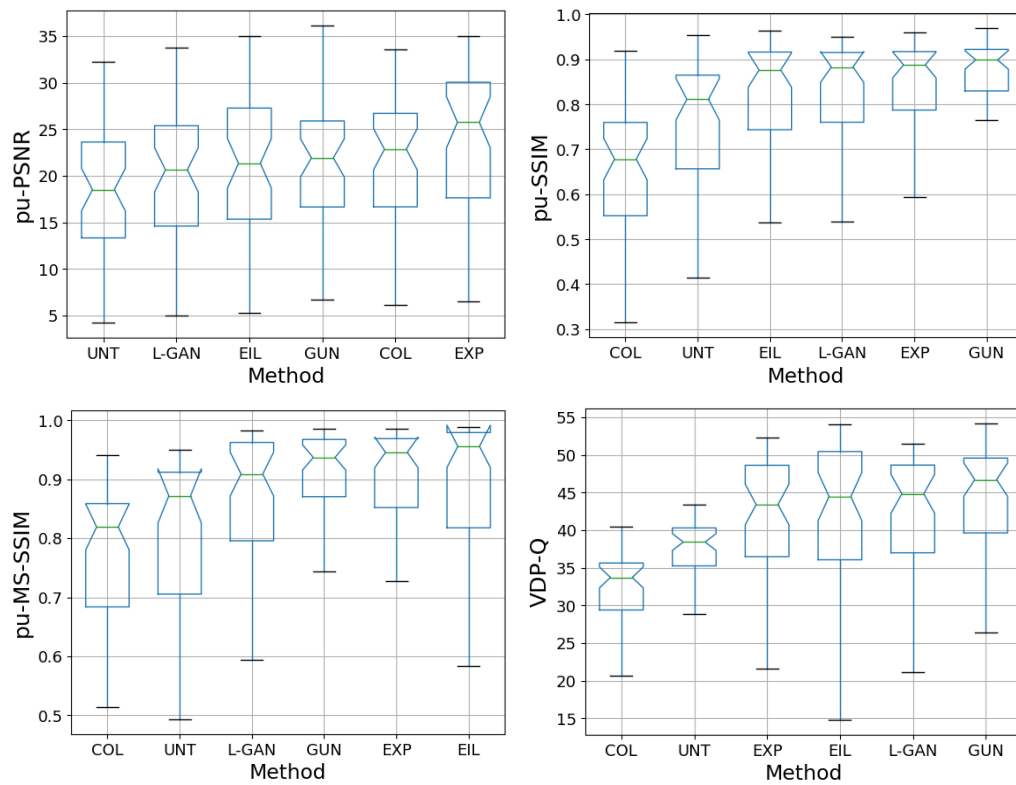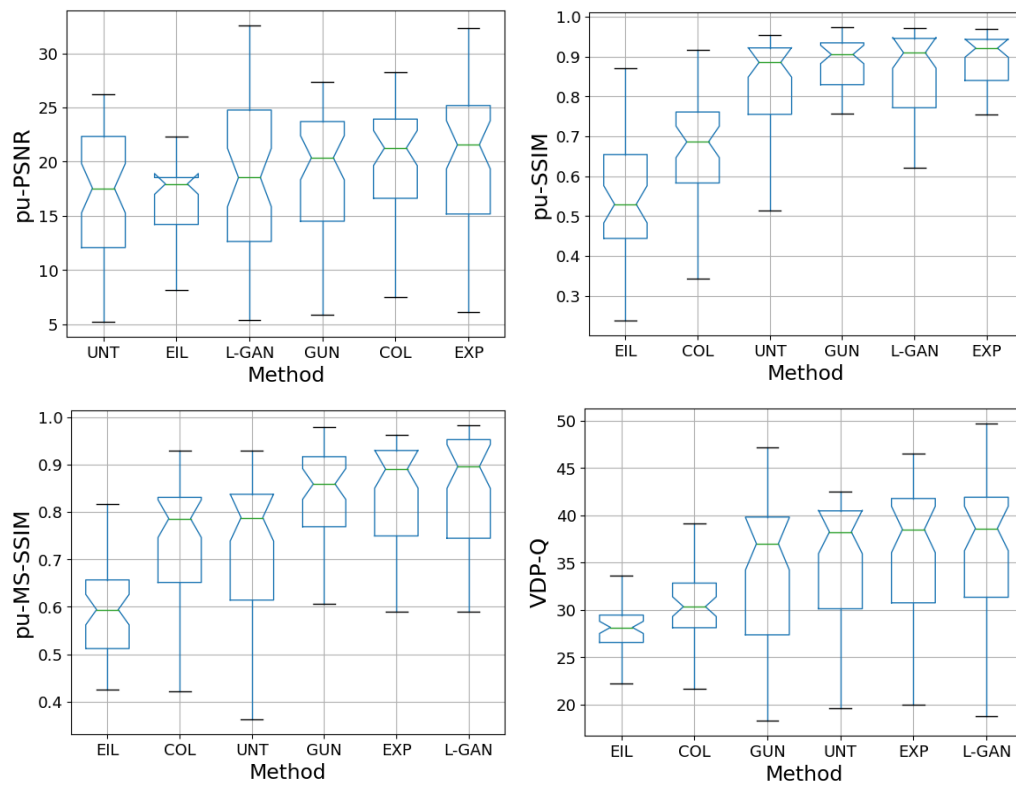
(**a**) optimal—scene-referred
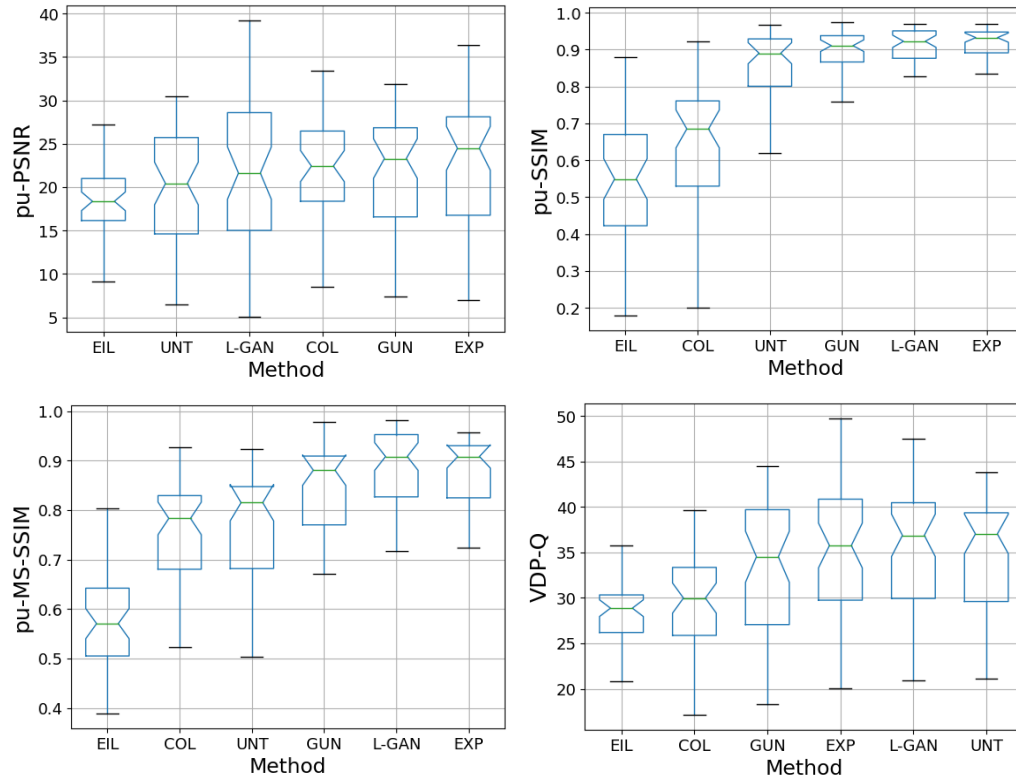
(**b**) culling— scene-referred

Figure 7. Box plots of all metrics for scene-referred HDR obtained from LDR culling and optimal exposures. The DHH method is referred to as L-GAN in the plots.

(**a**) optimal—display-referred

(**b**) culling—display-referred

Figure 8. Box plots of all metrics for display-referred HDR obtained from LDR culling and optimal exposures. The DHH method is referred to as L-GAN in the plots.

The results for the DHH method are comparable with those of ExpandNet and GUNet and are not significantly different in the culling case. While DHH does not dramatically improve on the quantitative performance of the other two methods, especially in the optimal exposure case where most of the content is well-exposed, these results are significant in multiple ways. First, they indicate that using a pre-trained network for data augmentation can produce good outcomes. The DHH network was trained using data that was fundamentally LDR unlike the HDR data that the other networks were trained with. The LDR data was expanded using the GUNet method and then tone mapped using exposure, to remove content and use it as the network inputs. In addition, the loss used was based solely on GAN and perceptual losses that are CNN based, without using standard regression losses, which shows the effectiveness of these losses in comparison with more traditional ones. These results also demonstrate that even though hallucination is achieved to some degree, it is not at the cost of the dynamic range expansion of well-exposed areas.

4.3. Ablation Studies

Ablation studies were conducted to complement the results obtained using the DHH method. Alternative single-model end-to-end architectures are considered as well as the effects of the normalisation layer. A GUNet [6] architecture is considered, which was shown to perform well for dynamic range expansion of well-exposed areas, while keeping the performance of UNet architectures. An additional Autoencoder network, sometimes termed a context encoder, which was previously used for LDR image inpainting [4] is trained and compared. A UNet architecture is also considered, similar to DHH, but with Batch normalisation [42], instead of the Instance Normalisation layers used in DHH. All the networks of the ablation studies similar in size. The main encoder/decoder parts use features with sizes 64-128-256-512. The bottlenecks are of size 512 and 8 layers deep, with layers 2–7 using dilated convolutions of dilation size 2-4-8-8-8-8-8, respectively. All networks use residual modules, the SELU activation [30] and 4×4 kernels for all layers except the Pre-Skip and Post-Fuse modules which use 1×1 kernels. All UNet networks use bilinear downsampling and upsampling which was shown [6] to have the least distortions in the spectrum compared to other non-guided UNet configurations. The Guided filter parameters for the GUNet architectures are the same as in the original [6]. An additional version of the DHH method is trained for consistency, using the above model configuration, which was chosen to be smaller to save computational resources, similarly to the other models in the ablation studies.

Figure 9 also presents samples for the GUNet and Autoencoder methods. The GUNet architecture successfully expands the well exposed areas and also attempts to hallucinate fully saturated and under-exposed areas. However the hallucinations are predominantly flat without any texture variations. This is an expected outcome due to the nature of the GUNet which has a strong guidance from the input features. The autoencoder network provides hallucinations of varying textures in the badly exposed areas; however, it distorts the well exposed areas, adding defects of repeated patterns. The lack of faithful reproduction of the well exposed areas can be attributed to the lack of skip connections.

Figure 10 compares the effects of using batch normalisation and instance normalisation. Batch normalisation tends to "average out" the hallucination predictions, creating outputs close to the LDR input in tone. For example, the saturated part of the sky is more grey/white in the BNH prediction rather than the INH where it's blue. The overall tone of the INH prediction is closer to the target HDR image, compared to that of BNH. This suggests that using instance normalisation provides improved hallucination predictions.
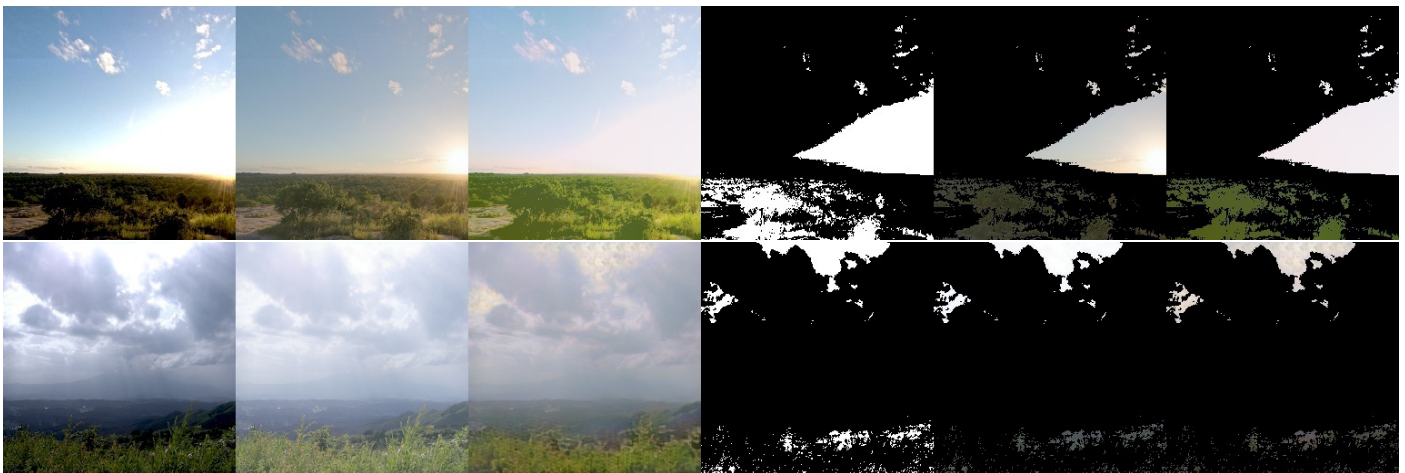
Figure 9. Training set samples for the GUNet and Autoencoder ablation studies. First column is the input LDR, followed by the target, prediction, mask, masked-target and masked prediction. Targets and predictions are tone-mapped. The masks show the fully over/under exposed regions.
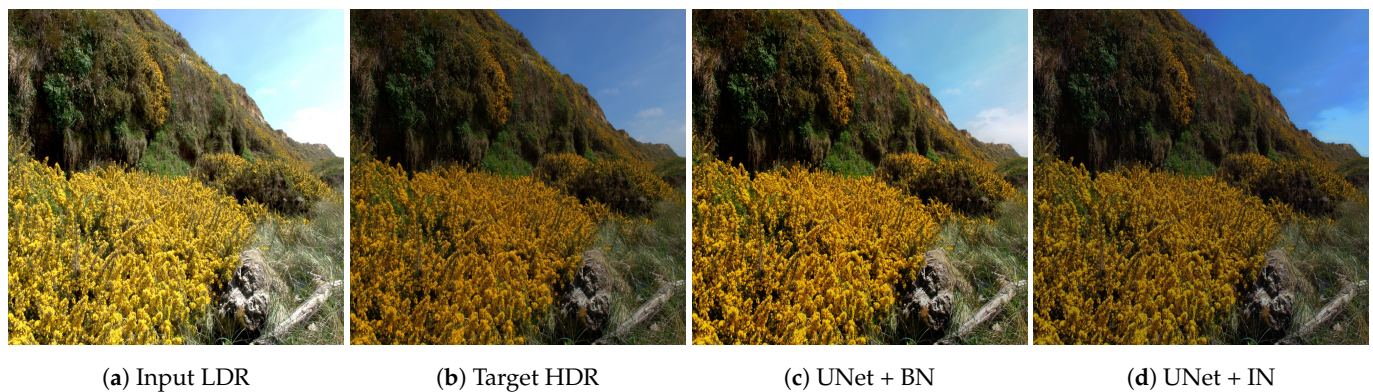

(**a**) Input LDR   (**b**) Target HDR   (**c**) UNet + BN   (**d**) UNet + IN

Figure 10. Effect of batch normalisation and instance normalisation. Instance normalisation predictions are closer to the target HDR in tone.

5. Conclusions

This work presented DHH, a GAN-based method, that expands the dynamic range of LDR images and fills in missing information in over-exposed and under-exposed areas. The method is end-to-end using a single CNN, performing both tasks simultaneously. Due to the scarcity of HDR data, a data augmentation regime is proposed for simulating LDR-HDR pairs for HDR hallucination, using a pre-trained CNN for dynamic range expansion. An additional HDR pixel distribution transformation method is proposed, for the normalisation of HDR content for use in deep learning methods. Methods involving masking that combine an expansion and a hallucination network each focusing on different aspects of the problem were inspected. This approach produced visible boundary artefacts and tone mismatch for both grey and RGB linear masks. End-to-end methods using a single network provided improved results. Networks with no skip connections, similar to context encoders previously used for other inpainting tasks produced noticeably more artefacts for well exposed areas and are may not be ideal for a complete end-to-end solution. Normalisation layers have an effect on the global contrast of the prediction, giving more consistent predictions for larger inputs. Batch Normalisation can cause results to be more desaturated and closer to the LDR input, while Instance Normalisation that gives predictions closer to the target HDR. Overall, the results compare well with state-of-the-art methods. Improvements can be directed towards the application of models on larger sized images with larger areas to hallucinate. In addition, the models can be adjusted such that they are able to better reproduce semantic details and not be limited to matching textures.

Future work will look into the application of the method to HDR video, ensuring temporal consistency across frames. Other applications that may be considered include colourisation, de-noising and single image superresolution along with inverse tone-mapping.

References

1. Banterle, F.; Artusi, A.; Debattista, K.; Chalmers, A. *Advanced High Dynamic Range Imaging*, 2nd ed.; AK Peters (CRC Press): Natick, MA, USA, 2017.
2. Banterle, F.; Ledda, P.; Debattista, K.; Chalmers, A. Inverse tone mapping. In Proceedings of the GRAPHITE'06, Kuala Lumpur, Malaysia, 29 November–2 December 2006; p. 349. [CrossRef]
3. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* **2017**, *36*, 1–14. [CrossRef]
4. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
5. Marnerides, D.; Bashford-Rogers, T.; Hatchett, J.; Debattista, K. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. *Comput. Graph. Forum* **2018**, *37*, 37–49. [CrossRef]
6. Marnerides, D.; Bashford-Rogers, T.; Debattista, K. Spectrally Consistent UNet for High Fidelity Image Transformations. *arXiv* **2020**, arXiv:2004.10696.
7. Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R.K.; Unger, J. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **2017**, *36*, 178. [CrossRef]
8. Endo, Y.; Kanamori, Y.; Mitani, J. Deep reverse tone mapping. *ACM Trans. Graph.* **2017**, *36*, 171–177. [CrossRef]
9. Landis, H. Production-Ready Global Illumination. Available online: https://www.semanticscholar.org/paper/Production-Ready-Global-Illumination-Landis/4a9de79235445fdf346b274603dfa5447321aab6 (accessed on 2 June 2021).
10. Akyüz, A.O.; Fleming, R.; Riecke, B.E.; Reinhard, E.; Bülthoff, H.H. Do HDR displays support LDR content? A psychophysical evaluation. *ACM Trans. Graph.* **2007**, *26*, 38. [CrossRef]
11. Masia, B.; Agustin, S.; Fleming, R.W.; Sorkine, O.; Gutierrez, D. Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions. *ACM Trans. Graph.* **2009**, *28*, 1–8. [CrossRef]
12. Masia, B.; Serrano, A.; Gutierrez, D. Dynamic range expansion based on image statistics. *Multimed. Tools Appl.* **2017**, *76*, 631–648. [CrossRef]
13. Rempel, A.G.; Trentacoste, M.; Seetzen, H.; Young, H.D.; Heidrich, W.; Whitehead, L.; Ward, G. LDR2HDR: On-the-Fly Reverse Tone Mapping of Legacy Video and Photographs. *ACM Trans. Graph.* **2007**, *26*, 39. [CrossRef]
14. Kovaleski, R.P.; Oliveira, M.M. High-Quality Reverse Tone Mapping for a Wide Range of Exposures. In Proceedings of the 27th SIBGRAPI Conference on Graphics, Patterns and Images, Rio de Janeiro, Brazil, 26–30 August 2014; pp. 49–56.
15. Huo, Y.; Yang, F.; Dong, L.; Brost, V. Physiological inverse tone mapping based on retina response. *Vis. Comput.* **2014**, *30*, 507–517. [CrossRef]
16. Wang, L.; Wei, L.Y.; Zhou, K.; Guo, B.; Shum, H.Y. High Dynamic Range Image Hallucination. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.446.7833&rep=rep1&type=pdf (accessed on 2 June 2021).
17. Kuo, P.H.; Liang, H.J.; Tang, C.S.; Chien, S.Y. Automatic high dynamic range hallucination in inverse tone mapping. In Proceedings of the 2014 IEEE 16th International Workshop on Multimedia Signal Processing (MMSP), Jakarta, Indonesia, 22–24 September 2014; pp. 1–6.
18. Zhang, J.; Lalonde, J.F. Learning high dynamic range from outdoor panoramas. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4519–4528.
19. Lee, S.; An, G.H.; Kang, S. Deep Chain HDRI: Reconstructing a High Dynamic Range Image from a Single Low Dynamic Range Image. *IEEE Access* **2018**, *6*, 49913–49924. [CrossRef]
20. Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; Shi, B. Neuromorphic camera guided high dynamic range imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1730–1739.

21. Sun, Q.; Tseng, E.; Fu, Q.; Heidrich, W.; Heide, F. Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.

22. Liu, Y.L.; Lai, W.S.; Chen, Y.S.; Kao, Y.L.; Yang, M.H.; Chuang, Y.Y.; Huang, J.B. 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), Seattle, WA, USA, 14–19 June 2020.

23. Sharif, S.; Naqvi, R.A.; Biswas, M.; Sungjun, K. A Two-stage Deep Network for High Dynamic Range Image Reconstruction. Available online: https://www.semanticscholar.org/paper/A-Two-stage-Deep-Network-for-High-Dynamic-Range-Sharif-Naqvi/64236160dcc06a1370f2358c3e44b44d9054e796 (accessed on 2 June 2021).

24. Chen, X.; Liu, Y.; Zhang, Z.; Qiao, Y.; Dong, C. HDRUNet: Single Image HDR Reconstruction with Denoising and Dequantization. Available online: https://arxiv.org/abs/2105.13084 (accessed on 2 June 2021).

25. Santos, M.S.; Tsang, R.; Khademi Kalantari, N. Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. *ACM Trans. Graph.* **2020**, *39*, 1–10. [CrossRef]

26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Available online: https://arxiv.org/abs/1505.04597 (accessed on 2 June 2021).

27. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2016**, arXiv:1611.07004.

28. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.

29. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.

30. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. Available online: https://arxiv.org/abs/1706.02515 (accessed on 2 June 2021).

31. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on Machine Learning (ICML), Atlanta, GA, USA, 16–21 June 2013; p. 3.

32. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv* **2017**, arXiv:1710.10196.

33. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved Techniques for Training GANs. *arXiv* **2016**, arXiv:1606.03498.

34. Kodali, N.; Abernethy, J.; Hays, J.; Kira, Z. On Convergence and Stability of GANs. *arXiv* **2017**, arXiv:1705.07215.

35. Lim, J.H.; Ye, J.C. Geometric GAN. *arXiv* **2017**, arXiv:1705.02894.

36. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. *arXiv* **2018**, arXiv:1805.08318.

37. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv* **2018**, arXiv:1809.11096.

38. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.

39. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

40. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

41. Flickr. 2018. Available online: https://www.flickr.com/ (accessed on 2 June 2021).

42. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.

43. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *arXiv* **2017**, arXiv:1706.08500.

44. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.

45. Aydın, T.O.; Mantiuk, R.; Seidel, H.P. Extending quality metrics to full luminance range images. In *Human Vision and Electronic Imaging XIII*; International Society for Optics and Photonics: London, UK, 2008; Volume 6806, p. 68060B.

46. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]

47. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.

48. Narwaria, M.; Mantiuk, R.; Da Silva, M.P.; Le Callet, P. HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images. *J. Electron. Imaging* **2015**, *24*, 010501. [CrossRef]