OoDHDR-Codec: Out-of-Distribution Generalization for HDR Image Compression

Linfeng Cao[1], Aofan Jiang[1], Wei Li[2], Huaying Wu[1], Nanyang Ye[1]*

[1] Shanghai Jiao Tong University
[2] Huawei Noah's Ark Lab
linfengcao1996@gmail.com, stillunnamed@sjtu.edu.cn, liwei472@hisilicon.com, 987941090@qq.com, ynylincoln@sjtu.edu.cn

Abstract

Recently, deep learning has been proven to be a promising approach in standard dynamic range (SDR) image compression. However, due to the wide luminance distribution of high dynamic range (HDR) images and the lack of large standard datasets, developing a deep model for HDR image compression is much more challenging. To tackle this issue, we view HDR data as distributional shifts of SDR data and the HDR image compression can be modeled as an out-of-distribution generalization (OoD) problem. Herein, we propose a novel out-of-distribution (OoD) HDR image compression framework (OoDHDR-codec). It learns the general representation across HDR and SDR environments, and allows the model to be trained effectively using a large set of SDR datasets supplemented with much fewer HDR samples. Specifically, OoDHDR-codec consists of two branches to process the data from two environments. The SDR branch is a standard black-box network. For the HDR branch, we develop a hybrid system that models luminance masking and tone mapping with white-box modules and performs content compression with black-box neural networks. To improve the generalization from SDR training data on HDR data, we introduce an invariance regularization term to learn the common representation for both SDR and HDR compression. Extensive experimental results show that the OoDHDR codec achieves strong competitive in-distribution performance and state-of-the-art OoD performance. To the best of our knowledge, our proposed approach is the first work to model HDR compression as OoD generalization problems and our OoD generalization algorithmic framework can be applied to any deep compression model in addition to the network architectural choice demonstrated in the paper. *Code available at* https://github.com/caolinfeng/OoDHDR-codec.

Introduction

High dynamic range (HDR) technology, originated in the 1970s (Mann and Ali 2016), is leading a visual revolution at present. Compared with standard dynamic range (SDR) images, HDR technology has a larger dynamic range ratio between the maximum and minimum luminance, which greatly broadens the boundaries of the displayable color gamut and allows to preserve more information of the real scenes. As an example shown in Figure 1(a), the SDR image (left column) losses most of the details at high luminance region, while the tone-mapped HDR image (right column) presents the scene containing extreme sunlight and shade, providing a more lifelike visual experience. Undoubtedly, the maturity and popularity of HDR technology can greatly enrich the interactive experience of traditional multimedia.

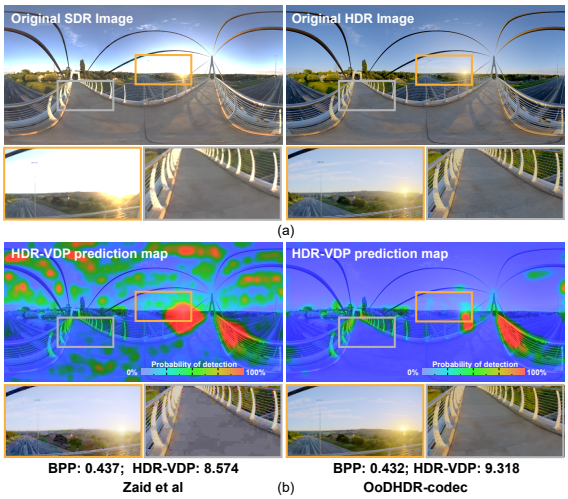
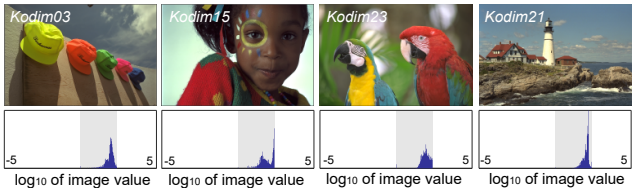


Figure 1: (a) Comparison between SDR and HDR images of "pedestrian overpass" from *HDRI HAVEN*[1]. (b) Visualization of results with approximately 0.43 bpp. Visual Difference Predictor for HDR images (HDR-VDP) (Mantiuk et al. 2011) predicts the distortions are visible to the human observer or not, where the higher the value, the better the image quality (*Better viewed on screen*).

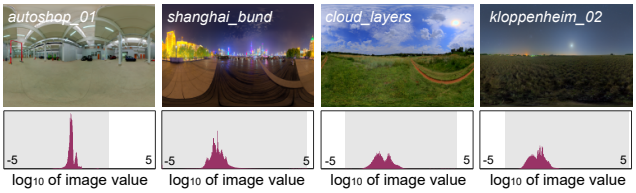
Although several HDR encoding formats including Radiance RGBE (.hdr) (Ward 1991), OpenEXR (.exr) (Kainz, Bogart, and Stanczyk 2009), LogLuv TIFF (.tiff) (Larson 1998), etc. have been proposed to facilitate the application of HDR technology, it still faces many challenges for prevalence. In addition to the hardware cost, significantly higher storage and transmission bandwidth costs also prevent HDR from being widely used in streaming media (e.g., YouTube) and social media (e.g., TikTok, Twitter) platforms. Thus, it is

*Nanyang Ye is the Corresponding author.

[1]https://hdrihaven.com/hdris/



(a) Data distribution of SDR examples



(b) Data distribution of HDR examples

Figure 2: Distribution shifts among SDR and HDR data. Gray area indicates the upper and lower limits of data distribution (log_{10} of value)

urgent and essential to develop efficient solutions for HDR image compression.

The deep neural network (DNN) is thought to be a promising method to implement the compression, of which the parameters are globally optimized instead of being designed in a hand-crafted way. However, the implementation of learning-based HDR compression especially in an end-to-end manner is rather challenging. Different from SDR with massive public resources available, HDR technology has gradually become a hot topic in recent years with the breakthrough development of hardware (shooting, display, etc.), and the public data source is still limited. In addition, the luminance information reflected by HDR varies across camera parameters, sensor characteristics and shooting conditions, so the pixel value is impossible to have a standard and unified distribution. Therefore, how to design an end-to-end HDR deep compression model and train it with good performance in practical use has become an open question.

On the other hand, there are numerous standard SDR datasets publicly available for training (e.g., DIV2K (Agustsson and Timofte 2017), Flickr2K (Lim et al. 2017), etc.) and evaluation (like Kodak), which greatly facilitate the development of deep SDR compression.

So can SDR data be used to make up for the insufficient training samples in deep HDR compression? Theoretically, if the HDR data share the same features as those of the SDR in coding space, SDR data can be leveraged for HDR compression learning. Fig. 2 presents some distribution examples of SDR and HDR, where the gray area indicates the upper and lower limits of data distribution (log_{10} of value). We observe that compared with SDR of which the data is concentrated in a fixed interval, HDR data is more widely distributed and varies in scope. Nevertheless, there is no essential distinction between HDR and SDR in the form of data distribution within the grey area, but more like the data distribution shift in one dimension. Inspired by it, we view the HDR values as the distribution shift of SDR images and model the HDR compression as an out-of-distribution (OoD) problem. Specifically, we propose a novel OoD image compression framework (OoDHDR-codec) which utilizes the joint data of SDR and HDR environments, augmented with training regularization, so as to learn the latent representations that are strongly correlated with compression performance, and thus allows the model not have to be trained on a large HDR collection.

In addition to data shift, too wide distribution of HDR is another factor that increases the difficulty of model design and training. To address this, we design a hybrid model incorporating white-box modules for differentiable perceptual encoding and tone-mapping (also their inverse) operations to initially compress the distribution interval of HDR, with a black-box neural network to implement the content compression. The linearly segmented luminance masking and tone-mapping curves guarantee the differentiability and reversibility of our framework.

We deploy our OoDHDR-code framework on a deep compression model, and showcase the compression performance on SDR and HDR. It achieves strong competitive in-distribution performance compared to other SDR compression approaches (comparable performance with the learning-based methods and outperforms the latest conventional standard VVC (Ohm and Sullivan 2018)), and also shows superior compression performance on HDR (OoD) data over other HDR compression approaches. The ablation study presents significantly higher generalization ability on HDR (OoD) data compared to the standard model, and also demonstrates the effectiveness of perceptually uniform (PU) encoding and tone-mapping operation (TMO).

Related Work

HDR Compression

With the development of HDR technology, there have been numerous HDR image and video compression researches. The HDR file formats mainly includes Radiance RGBE (.hdr) (Ward 1991), OpenEXR (.exr) (Kainz, Bogart, and Stanczyk 2009), LogLuv TIFF (.tiff) (Larson 1998). For further HDR compression, the algorithms can be roughly divided into two categories: backward and non-backward compatible methods (Mukherjee et al. 2019). The first category generates two or more layers containing an 8-bit SDR version (Zaid and Houimli 2017; Mai et al. 2011), which allows content can be displayed on legacy SDR devices. The non-backward compatible methods encode the content into higher bit-depth (eg., 10-bit or 12-bit) streams that are supported by modern video/image codecs (Mantiuk et al. 2004; Garbas and Thoma 2011; Miller, Nezamabadi, and Daly 2013; Mukherjee et al. 2019). For backward solution, Li et al. first adopt forward and inverse tone-mapping to optimize the HDR compression (Li, Sharan, and Adelson 2005). JPEG HDR is also a typical compression algorithm. It converts HDR content into an SDR version by a global TMO and the information for HDR reconstruction is encoded with JPEG legacy compliant codestream (Ward and Simmons 2006; Richter 2013; Zaid and Houimli 2017). Apart from the global TMO, Pendu et al. use local TMO with a template-

based inter-layer prediction (ILP) for HDR reconstruction (Pendu, Guillemot, and Thoreau 2015).

However, these methods utilize the conventional standards for image content compression, which rely on handcrafted design of each module and are not expected to be the general and optimal solutions for all image contents.

Learning-based Compression

Recently, deep learning for SDR image compression has attracted considerable attention. Google Inc propose three variants of RNN based en/decoder to compress progressive images and residuals (Toderici et al. 2017, 2016). Li et al. design fully convolution networks with residual block as en/decoder (Li et al. 2018). Different from using the round function in the conventional codecs, deep compression employs some differentiable operations (e.g., K-means algorithm (Mentzer et al. 2018), mapping function with Bernoulli distribution noise (Toderici et al. 2016)) for soft-quantization. To further reduce the redundancy of quantization code, the entropy model for entropy coding (e.g., Huffman coding) is proposed and has been widely investigated for coding navigation (Toderici et al. 2017; Minnen, Ballé, and Toderici 2018; Mentzer et al. 2018).

However, the implementation of learning-based HDR compression especially in an end-to-end manner is rather difficult. The main challenges are as follows:

1. Distribution differences between HDR and SDR domains lead to SDR datasets not being suitable for HDR model training. Also, the standard deep SDR compression models designed for smaller luminance range are not compatible with HDR data. This is because the wide luminance distribution of HDR data is harder to be represented by fewer bits of information.

2. Large distribution differences within HDR domain leads to the potential shifts between training and testing data. The luminance distribution of test dataset might deviate from training dataset, which is different from SDR image compression where pixel values are constrained in a smaller range. Therefore, the traditional training paradigm (empirical risk minimization) for independent and identically distributed (IID) data is not optimal for HDR tasks.

Methodology

To deal with the aforementioned problems, we argue that it is critical to bridge the distribution gap between HDR and SDR domains, and enhance the robustness of network. By investigating the data distributions in two domains (SDR and HDR), we find that although there is luminance distribution between these two domains, they share similar semantic features for image compression. For example, to compress an image containing a cat, it is essential to preserve the shape information of this image both for its SDR and HDR version. This inspires us to learn the invariant representations from SDR and HDR for more efficient data representations. To this end, we model the HDR compression as an OoD generalization problem and propose an algorithmic framework to learn the invariant representations in HDR and SDR domains (as Fig. 3) for image compression. The details of implementation are as follows:

To eliminate the impact of too wide HDR data distribution on training, we design a white-box pre-processing module that contains a PU transform and a global TMO, which converts the HDR data to a more concentrated-distribution space. The content compression is achieved by a DNN. It firstly transforms the input data into the coding space with a deep encoder, then quantizes the latent representation and generates the bit-stream using entropy model and lossless coding (e.g., arithmetic coding (Rissanen and Langdon 1981)), and finally transform the information back to the image space with entropy model and deep decoder. To find a more general coding space and thus effectively extract the latent representation of two domains (SDR & HDR), we perform regularization of the losses and fusion them together during training, allowing the DNN to generalize to different distributions.

Preprocess Module

Perceptual Unit Transform Due to the wide illumination range distribution of HDR scenes, we employ a perceptual transfer function to map the linear physical luminance (cd/m^2) to a perceptual uniform space derived from the contrast sensitivity function of the visual system, which helps to avoid the distortion of TMO caused by too large dynamic range (Mantiuk et al. 2011). Specifically, the transfer function we adopt is known as Perceptually Uniform (PU) encoding, which is defined as the reciprocal integral of detection threshold:

$$P(L) = \int_{L_{min}}^{L} \frac{1}{T(l)} dl \tag{1}$$

where L_{min} represents the minimum encoded luminance. L is the absolute luminance. $T(L)$ represents the detection thresholds defined as below:

$$T(L) = S \cdot \left(\left(\frac{C_1}{L} \right)^{C_2} + 1 \right)^{C_3} \tag{2}$$

Where S refers to the absolute sensitivity constant. C_1, C_2, C_3 are parameters obtained by contrast sensitivity fitting, in which we use the parameters from (Mantiuk et al. 2011). For forward/inverse perceptual transform and backward propagation in networks, we use piece-wise (five segments) linear functions to fit the PU encoding curves, which keeps the problem analytically tractable during forward and inverse calculations. The piece-wise PU curve is shown in Fig. 4(a). The forward and inverse modules are denoted as "PU" and "IPU" in Fig. 3.

Global Tone-Mapping Operator To correct the data biases caused by too wide data distribution, we apply a TMO operator to further map the perceptual values into a more concentrated-distribution space (same distribution boundary with SDR). There are plenty of HDR tone-mapping implementations, mainly based on the "S-shaped" curve or power function curve with exponent (gamma) less than 1. However, for most of the tone-mapping curves, it is very hard to find its exact inverse function expression, which inevitably increases the calculation difficulty and introduces the reconstruction error (especially for large HDR luminance values).

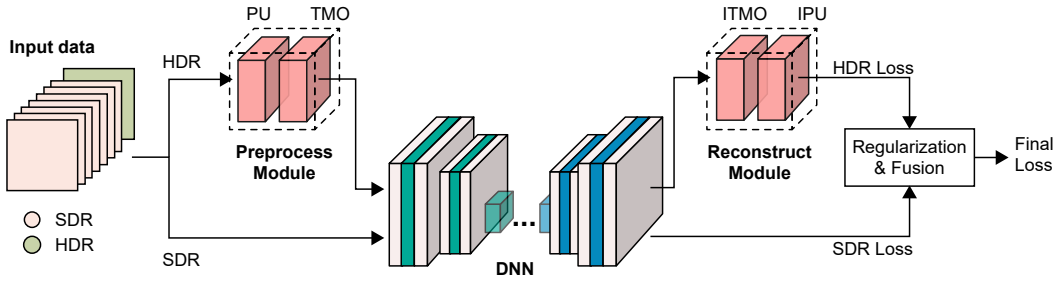


Figure 3: Overview of the proposed OoDHDR-codec framework. The input data consists of SDR and HDR images in a certain proportion (7:1 in figure). The framework consists of two branches to process the corresponding data streams. After compression by DNN, the losses of two environments are regularized and fusion together to get the final loss.

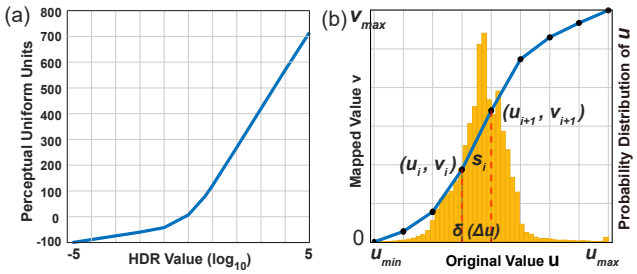


Figure 4: (a) Piece-wise PU transform function, converts absolute luminance into approximately perceptual uniform values. (b) Piece-wise tone-mapping curve, background histogram represents the data distribution in original space.

To address this, we introduce the mapping approach in (Mai et al. 2011), which models the mapping curve uniquely with a series of nodes (u_i, v_i) and linear functions to construct our TMO (shown in Figure 4 b). For each segment, the interval is set as a constant value δ. The forward mapping function is defined as:

$$v(u) = (u - u_i)s_i + v_i, \quad \text{if} \quad u_i < u < u_{i+1} \tag{3}$$

where u represents the original PU value, v is the mapped value and s_i is the slope of the ith linear function. similarly, the inverse mapping function can be defined as:

$$\tilde{u}(v) = \begin{cases} \frac{v - v_i}{s_i} + u_i, & \text{if} \quad v_i < v < v_{i+1} \quad \text{and} \quad s_i > 0 \\ \frac{(u_i + u_{i+1})}{2}, & \text{if} \quad s_i = 0 \end{cases} \tag{4}$$

where $\tilde{u}$ is the inverse mapped value. For the segment with a slope of zero, the $\tilde{u}$ is assigned with the mean value of its lower and upper bounds.

To keep the simplicity of calculation, we utilize the closed-form solution in (Mai et al. 2011) with the assumption that the original u and distorted $\tilde{u}$ has the same slope due to the local linearity of the mapping curve, and the slope s_i is given by:

$$s_i = \frac{v_{\max} {p_i}^{1/3}}{\delta \sum_{i=1}^{N} {p_i}^{1/3}}, \quad p_i = \sum_{u=u_i}^{u_{i+1}} p(u) \tag{5}$$

where $v_{\max}$ is the maximum value, $p(u)$ is the probability of the original PU value, δ denotes the interval of the piecewise function.

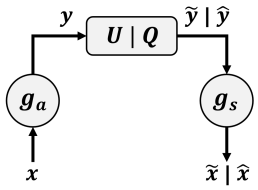


Figure 5: Operational diagram of the compression model.

Compression Model Generalization

Learning-based Compression Formulation The image compression based on transform coding approach (Goyal 2001) (shown in Fig. 5) can be expressed as:

$$y = g_a(x; \phi_1); \hat{y} = Q(y); \hat{x} = g_s(\hat{y}; \phi_2) \tag{6}$$

where input data x is firstly transformed into a coding space with a parametric analysis transform $g_a(x; \phi_1)$ to extract the latent representation y for quantization and entropy coding $(U|Q)$, and thus realize the data lossy compression. For image recovery, the quantized latent representation $\hat{y}$ is sent to a synthesis transform g_s to reconstruct image $\hat{x}$.

The learning-based image compression can be regarded as a rate-distortion optimization (Ballé et al. 2018), and the loss function can be expressed as

$$L = Rate(\hat{y}) + \lambda Dist(x, \hat{x}) \tag{7}$$

where $Rate$ and $Dist$ represent the entropy coding rate and distortion measures respectively. $\lambda \in \mathbb{R}^+$ is a hyperparameter for rate and quality trade-off.

Invariance Regularization To make up for the insufficient HDR training samples, we utilize a large SDR dataset augmented with a small HDR dataset as union training environment $\varepsilon = \{e_{sdr}, e_{hdr}\}$. Herein, our goal is to use these multiple datasets to learn generalizable encoders and decoders (g_a, g_s), which performs well across e_{sdr}, e_{hdr} and also other environments with different distributions that do not appear in the training set.

We also need to note that although e_{hdr} data is augmented with e_{sdr}, naively training models on the augmented dataset

will lead to degenerated performance as the model will tend to overfit on the SDR dataset. While a simple method is to propose a loss function that is the weighted sum of losses on HDR and SDR to avoid overfitting, this may hurt the SDR compression performance if the weight for HDR loss is large. To improve the OoD generalization performance of the proposed algorithmic scheme, in addition to the traditional compression loss, we add invariance regularization terms that enforce the model to achieve optimalty simultaneously at two environments.

To minimize the expected loss overall environments ε, inspired by (Arjovsky et al. 2019), we introduce invariance regularization as a penalty term to learn data representations eliciting invariant features across multiple environments. Different from (Arjovsky et al. 2019) that is designed for classification tasks, we calculate the gradients of losses on the reconstructed image. The loss terms of two environments are as follows:

$$\begin{aligned} L_{sdr} =& 1 - ms\text{-}ssim(x_{sdr}, \hat{x}_{sdr}) \\ & - \lambda_4 \left\| \nabla_{w=1} ms\text{-}ssim(x_{sdr}, w^T \hat{x}_{sdr}) \right\|_2^2 \end{aligned} \tag{8}$$

$$L_{hdr} = l_1(x_{hdr}, \hat{x}_{hdr}) + \lambda_4 \left\| \nabla_{w=1} l_1(x_{hdr}, w^T \hat{x}_{hdr}) \right\|_2^2 \tag{9}$$

where the first term in Eq. (8), (9) measures the quality distortion while the second is the corresponding gradient norm penalty term for invariance regularization which enforce the model to achieve optimality simultaneously for SDR and HDR. Two losses share the same hyper-parameter $\lambda_4 \in \mathbb{R}^+$ for balancing the numerical scale of losses.

Based on this, the final loss function is can be formulated as follows:

$$L_{final} = \lambda_1(\lambda_2 L_{sdr} + (1-\lambda_2) L_{hdr}) + \lambda_3 \left\| w_{dnn} \right\|_2^2 + Rate(\hat{y}) \tag{10}$$

where the hyper-parameter $\lambda_1 \in \mathbb{R}^+$ is used for rate and quality trade-off, $\lambda_2 \in \mathbb{R}^+$ balances losses from two environments. w_{dnn} denotes the model weights, the corresponding regularization term is adjusted by hyper-parameter $\lambda_3 \in \mathbb{R}^+$. For distortion measurement, MS-SSIM metrics (Wang, Simoncelli, and Bovik 2003) is employed in e_{sdr} and the pixel-wise L_1 loss (MAE) is employed in e_{hdr}.

Experiments

In this section, we deploy the OoDHDR-codec algorithmic framework on a widely used deep image compression model structure, augmented with simplified attention model (Cheng et al. 2020) and hyperprior (Minnen, Ballé, and Toderici 2018). It is worth noting that our proposed method is generic and can be applied to any deep model. The channel number of the latent representation y is set to 192 (More network details can be found in the supplemental material). For evaluation, we conduct the compression experiments on both SDR (in-distribution) and HDR (OoD) datasets.

Experimental Setup

Datasets The environment e_{sdr} is constructed by SDR images from DIV2K (Agustsson and Timofte 2017) and Flickr2K (Lim et al. 2017) datasets with around 3500 samples. To construct e_{hdr}, we collect HDR images from (Funt and Shi 2010; Kalantari and Ramamoorthi 2017; Yeganeh and Wang 2013; Narwaria et al. 2013; Debevec and Malik 2008; Ward et al. 2006), *pfstools* resources [2], *HDRI HEVEN*, with 480 samples in total and all of them are in Radiance RGBE (.hdr) (Ward 1991) format that contains the absolute luminance value (cd/m^2). For in-distribution evaluation, a widely used Kodak lossless image benchmark containing 24 SDR (768 × 512) images is tested. For OoD evaluation, we conduct the tests on 80 HDR images (2048 × 1024) from *HDRI HAVEN* (not be selected deliberately for a fair comparison). The detail information of HDR datasets is listed in supplemental material.

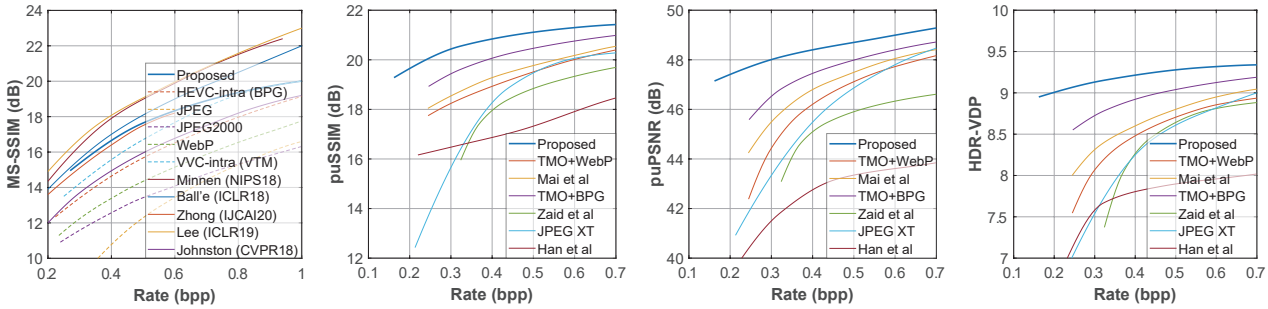
Training Details Our framework is implemented with PyTorch 1.6.0, CUDA v11.4 on NVIDIA 2080Ti GPU. The deployment of deep compression model is realized based on the CompressAI library (Bégaint et al. 2020). During training, the model is optimized using Adam (Kingma and Ba 2014) with a batch size of 32 (7:1 for the ratio of SDR to HDR). Each image is randomly cropped with patch size of 256×256, and a regular data augmentation including random rotation and flipping are conducted. The model is trained up to 250 epochs with initial learning rate of $1e^{-3}$, which decreases to $1e^{-4}$ at 200th epoch. To obtain the models with different rate control, the hyper-parameter λ_1 is chosen from [12, 40, 150, 300], and the channel number of latent representation is fixed at 192. For each model, the hyper-parameter of λ_2, λ_3 and λ_4 is set to 0.95, $1e^{-5}$ and 1, the segment number of TMO is set to 10.

Evaluation Metrics To assess the performance, we use several perceptual and structural quality image metrics for comparison between the original images and the degraded ones. For in-distribution (SDR) evaluation, the quality is measured by MS-SSIM, and the rate is measured by bits per pixel (bpp). For OoD (HDR) evaluation, the quality is measured with HDR-VDP 3.0.6 (Mantiuk et al. 2011), puPSNR and puSSIM (Aydın, Mantiuk, and Seidel 2008) as these metrics deliver the assessment that is highly correlated with subjective evaluation for HDR images (Mukherjee et al. 2016).

In-Distribution Performance Evaluation

Comparison Approaches For in-distribution evaluation, we compare our compression model with widely-used compression standards and recently proposed methods, including JPEG, JPEG2000, WebP (Banerjee and Arora 2011), BPG(444) (Bellard 2015) and VTM(444), and recent deep compression work by (Ballé et al. 2018), (Minnen, Ballé, and Toderici 2018), (Lee, Cho, and Beack 2019), (Zhong, Akutsu, and Aizawa 2020) and (Johnston et al. 2018). The implementation of JPEG, JPEG2000, WebP and BPG are realized with CompressAI library, and the results of VTM(444) are traced from the CompressAI paper (Bégaint et al. 2020). The results of the learning-based approach are traced from their corresponding paper.

[2] http://pfstools.sourceforge.net/hdr_gallery.html



(a) In-distribution performance (b) OoD performance.1 (c) OoD performance.2 (d) OoD performance.3

Figure 6: (a) In-distribution performance evaluation on Kodak with MS-SSIM metric. (b)(c)(d) OoD performance evaluation on HDR testset with puSSIM, puPSNR, HDR-VDP metrics respectively (The higher the better).

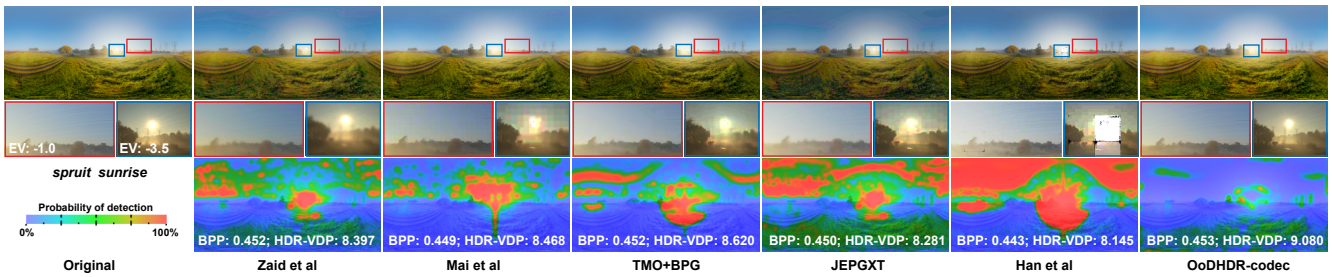


Figure 7: Visual quality comparison of example images from HDR test set with approximately 0.44 bpp. The images are evaluated on HDR-VDP (Mantiuk et al. 2011) metric (the higher the HDR-VDP value, the better the image quality) and the prediction maps are presented to indicate the probability of distortion detection by the human visual system at the corresponding image region. For a more intuitive comparison of details, the cropped and zoomed-in images are presented under suitable exposure values (EVs) (*Better viewed on screen*).

Rate-distortion Performance The rate-distortion results on Kodak are shown in Fig. 6 (a). Where the quality metric is converted to (-10log$_{10}$(MS-SSIM)) form for a clearer presentation. The results of conventional codecs are presented with dotted lines while the learning-based methods are presented with solid lines. These results indicate that our method achieves a competitive or even better result compared to several learning-based methods (Zhong, Akutsu, and Aizawa 2020; Johnston et al. 2018) which are purely trained and tested in-domain. Our method also outperforms the widely used image-standard BPG, and yields better performance in MS-SSIM compared to VTM, the intra-frame codec of the next-generation compression standard Versatile Video Coding (VVC) (Ohm and Sullivan 2018).

Out-of-Distribution Performance Evaluation

Comparison Approaches To evaluate the generalization of our method in HDR (OoD) images, we compare our method with the representative HDR compression approaches (schemes):

1. (Mai et al. 2011): Under framework of combining traditional SDR codecs (H.264/AVC in their work) with reversible TMO, which is the most widely used framework for HDR backward-compatible compression. (Same framework embedded with WebP and BPG have also been implemented, denoted as "TMO+WebP" and "TMO+BPG").
2. (Zaid and Houimli 2017): Under the framework of combining conventional SDR compression standard (JPEG in their work) with a ratio image to store the information used for HDR reconstruction, which is a another typical framework for backward-compatible HDR compression.
3. JPEGXT (Artusi et al. 2019): An extension of JPEG towards high-dynamic range photography.
4. (Han et al. 2020): Deep HDR compression method. DNN for residual compression and post-processing.

Rate-distortion Performance Figure 6(b)(c)(d) presents the average ratio-distortion (RD) curves on the HDR test set. In three metrics, our method outperforms all other HDR compression approaches, both conventional and learning-based compression schemes. Especially in low bit rates, the gap between our method and other approaches is even more pronounced. These results indicate that our approach also yields a strong OoD compression performance.

Visual Results Fig. 7 presents the visual quality evaluation from HDR test set. It can be seen that our method presents more clear and textural details (including the power lines in the sky, and the transmission tower under sunlight), while obvious artifacts and noises occur in other approaches. The smaller detection region and lower probability in HDR-VDP prediction maps also reveal the better overall perfor-

Model	HDR (OoD)* puPSNR (dB)	puSSIM	HDR-VDP	Rate (bpp)	SDR (In-Distribution)† PSNR (dB)	MS-SSIM	Rate (bpp)
Standard$_{sdr}$	27.875	0.971	4.67	0.401	**26.799**	**0.969**	0.268
Standard$_{hdr}$	35.846	0.970	7.490	0.269	25.746	0.956	0.273
OoDHDR	**37.899**	**0.972**	**7.92**	0.223	26.572	0.968	0.269
Standard$_{sdr}$+TMO	45.399	0.987	8.536	0.202			
Standard$_{hdr}$+TMO	45.228	0.982	8.756	0.206	* Evaluation on HDR test set		
OoDHDR+TMO	**46.192**	**0.987**	**8.840**	0.202	†Evaluation on Kodak (SDR set)		

Table 1: Ablation studies of OoD training framework and TMO module

Segment (N)	puPSNR (dB)	HDR-VDP	Rate (bpp)
N = 2	44.701	8.198	0.100
N = 5	46.546	8.547	0.129
N = 10	46.837	8.823	0.148
N = 20	**47.145**	**8.950**	0.163

Table 2: Testing results of OoDHDR-codec with different segment number (N) of TMO

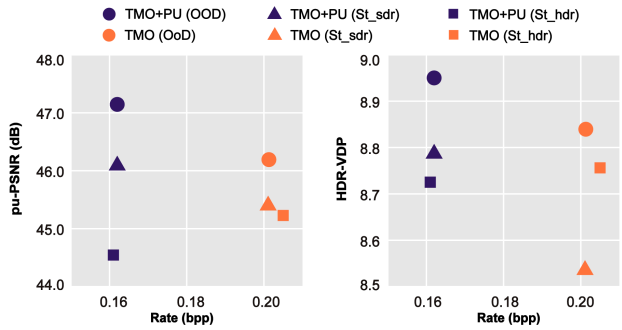


Figure 8: Ablation study of PU encoding. "OoD" denotes OoDHDR model, "Std_*" denotes the Standard model with its corresponding training set (The higher the better).

mance of our method on the example images. More visualization results in supplementary.

Ablation Studies

Generalization of Model To further investigate the generalization ability and measure the contribution of our proposed framework, we compare the following models:
(1) Standard$_{sdr}$: Backbone model of the framework, trained with SDR datasets. (2) Standard$_{hdr}$: Backbone model of the framework, trained with HDR datasets. (3) OoDHDR: Proposed method. (4) Standard$_{sdr}$+TMO: Standard$_{sdr}$ model augmented with TMO module during testing. (5) Standard$_{hdr}$+TMO: Standard$_{hdr}$ model augmented with TMO module during testing. (6) OoDHDR+TMO: OoDHDR model augmented with TMO module during testing.
These models are optimized with λ_1 = 12 and trained up to 250 epochs. For each model with HDR data during training, the TMO module is adopted for stable training.

The corresponding results are depicted in Table. 1. For HDR (OoD) compression, our method notably outperforms the baseline models in all metrics, which reveals the strong OoD generalization of our proposed method, and indicates the effectiveness of the proposed invariance regularization within two domains in the framework for learning the latent representations that are shared by two domains (SDR and HDR). Besides, it can be seen that the overall performance of all models improved significantly by applying the TMO module, which indicates that the TMO module in our method plays an essential role in performance enhancement. This is because TMO initially converts the data into a more concentrated-distribution space, which helps to reduce the generalization gap for the models. We also test the performance with different TMO segmentation numbers. The results of OoDHDR-codec with PU and TMO module are shown in Table. 2. As expected, with the increase of the TMO segment number, the performance of OoDHDR-codec can be improved, confirming the effectiveness of the TMO module in our method. For SDR (in-domain) compression, it can be observed that the baseline model trained with SDR has a slight advantage compared to our method. This is reasonable because our method tries to find a more general representation transform across all domains while avoiding overfitting in an individual domain.

Effectiveness of Perceptually Uniform (PU) Encoding We test the effects of the PU model, which is used for HDR image coding to avoid the distortion of TMO caused by too large dynamic range. The results are depicted in Fig. 8. We can observe that with the adoption of PU module, the coding bitrate of all models reduced significantly, and two models present an enhanced performance. This might be due to the PU transform paying more attention to the coding of high luminance areas, while choosing to further compress the low luminance value into a narrower interval. This helps the transformed image containing richer details at high luminance, and saves the coding costs at extremely low luminance area, thus realizing the coding gain.

Conclusion

In this paper, we propose a novel OoD HDR image compression framework (OoDHDR-codec), which improves the compression model's OoD generalization ability for HDR image compression. Extensive experimental results indicate that our OoDHDR codec shows strong generalization capabilities across multiple datasets. To the best of our knowledge, our proposed approach is the first work to model HDR compression as an OoD generalization problem to achieve state-of-the-art HDR image compression performance.

Acknowledgements

Nanyang Ye was supported by National Natural Science Foundation of China under Grant 62106139, in part by National Key RD Program of China 2018AAA0101200, in part by National Natural Science Foundation of China under Grant (No. 61829201, 61832013, 61960206002, 62061146002, 42050105, 62032020), in part by the Science and Technology Innovation Program of Shanghai (Grant 18XD1401800), and in part by Shanghai Key Laboratory of Scalable Computing and Systems, and in part by BIREN Tech.

References

Agustsson, E.; and Timofte, R. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 126–135.

Arjovsky, M.; Bottou, L.; Gulrajani, I.; and Lopez-Paz, D. 2019. Invariant Risk Minimization. arXiv:1907.02893.

Artusi, A.; Mantiuk, R. K.; Richter, T.; Hanhart, P.; Korshunov, P.; Agostinelli, M.; Ten, A.; and Ebrahimi, T. 2019. Overview and evaluation of the JPEG XT HDR image compression standard. *Journal of Real-Time Image Processing*, 16(7): 413–428.

Aydın, T. O.; Mantiuk, R.; and Seidel, H.-P. 2008. Extending Quality Metrics to Full Luminance Range Images. In *Proceedings of the Human Vision and Electronic Imaging*, 109–118.

Ballé, J.; Minnen, D.; Singh, S.; Hwang, S. J.; and Johnston, N. 2018. Variational image compression with a scale hyperprior. arXiv:1802.01436.

Banerjee, S.; and Arora, V. 2011. WebP Compression Study. https://developers.google.com/speed/webp/docs/webp_study. Accessed: 2021-09-09.

Bégaint, J.; Racapé, F.; Feltman, S.; and Pushparaja, A. 2020. CompressAI: a PyTorch library and evaluation platform for end-to-end compression research. arXiv:2011.03029.

Bellard, F. 2015. BPG image format. https://bellard.org/bpg. Accessed: 2021-09-09.

Cheng, Z.; Sun, H.; Takeuchi, M.; and Katto, J. 2020. Learned Image Compression With Discretized Gaussian Mixture Likelihoods and Attention Modules. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7936–7945.

Debevec, P. E.; and Malik, J. 2008. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the ACM Special Interest Group for Computer GRAPHICS*, 1–10.

Funt, B.; and Shi, L. 2010. The rehabilitation of MaxRGB. In *Proceedings of the Color and Imaging Conference*, 256–259.

Garbas, J.-U.; and Thoma, H. 2011. Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 829–832.

Goyal, V. 2001. Theoretical foundations of transform coding. *IEEE Signal Processing Magazine*, 18(5): 9–21.

Han, F.; Wang, J.; Xiong, R.; Zhu, Q.; and Yin, B. 2020. HDR Image Compression with Convolutional Autoencoder. In *Proceedings of the IEEE International Conference on Visual Communications and Image Processing*, 25–28.

Johnston, N.; Vincent, D.; Minnen, D.; Covell, M.; Singh, S.; Chinen, T.; Jin Hwang, S.; Shor, J.; and Toderici, G. 2018. Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4385–4393.

Kainz, F.; Bogart, R.; and Stanczyk, P. 2009. Technical introduction to OpenEXR. *Industrial light and magic*, 21.

Kalantari, N. K.; and Ramamoorthi, R. 2017. Deep High Dynamic Range Imaging of Dynamic Scenes. *ACM Transactions on Graphics*, 36(4): 1–12.

Kingma, D.; and Ba, J. 2014. Adam: A Method for Stochastic Optimization. arXiv:1412.6980.

Larson, G. W. 1998. LogLuv Encoding for Full-Gamut, High-Dynamic Range Images. *Journal of Graphics Tools*, 3(1): 15–31.

Lee, J.; Cho, S.; and Beack, S.-K. 2019. Context-adaptive Entropy Model for End-to-end Optimized Image Compression. arXiv:1809.10452.

Li, M.; Zuo, W.; Gu, S.; and Zhao, D. 2018. Learning Convolutional Networks for Content-weighted Image Compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3214–3223.

Li, Y.; Sharan, L.; and Adelson, E. H. 2005. Compressing and Companding High Dynamic Range Images with Subband Architectures. *ACM Transactions on Graphics*, 24(3): 836–844.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 136–144.

Mai, Z.; Mansour, H.; Mantiuk, R.; Nasiopoulos, P.; Ward, R.; and Heidrich, W. 2011. Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression. *IEEE Transactions on Image Processing*, 20(6): 1558–1571.

Mann, S.; and Ali, M. 2016. Chapter 1 - The Fundamental Basis of HDR: Comparametric Equations. In *High Dynamic Range Video*, 1–59. Academic Press.

Mantiuk, R.; Kim, K. J.; Rempel, A. G.; and Heidrich, W. 2011. HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Transactions on Graphics*, 30(4): 1–14.

Mantiuk, R.; Krawczyk, G.; Myszkowski, K.; Seidel, H.-P.; and Informatik, M. 2004. Perception-motivated High Dynamic Range Video Encoding. *ACM Transactions on Graphics*, 23: 733–741.

Mentzer, F.; Agustsson, E.; Tschannen, M.; Timofte, R.; and Van Gool, L. 2018. Conditional Probability Models for Deep

Image Compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4394–4402.

Miller, S.; Nezamabadi, M.; and Daly, S. 2013. Perceptual Signal Coding for More Efficient Usage of Bit Codes. *Society of Motion Picture and Television Engineers Motion Imaging Journal*, 122(4): 52–59.

Minnen, D.; Ballé, J.; and Toderici, G. 2018. Joint Autoregressive and Hierarchical Priors for Learned Image Compression. arXiv:1809.02736.

Mukherjee, R.; Debattista, K.; Bashford-Rogers, T.; Vangorp, P.; Mantiuk, R.; Bessa, M.; Waterfield, B.; and Chalmers, A. 2016. Objective and subjective evaluation of High Dynamic Range video compression. *Signal Processing: Image Communication*, 47: 426–437.

Mukherjee, R.; Debattista, K.; Rogers, T.-B.; Bessa, M.; and Chalmers, A. 2019. Uniform Color Space-Based High Dynamic Range Video Compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(7): 2055–2066.

Narwaria, M.; Da Silva, M. P.; Le Callet, P.; and Pepion, R. 2013. Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality. *Optical Engineering*, 52(10): 1 – 16.

Ohm, J.-R.; and Sullivan, G. J. 2018. Versatile video coding–towards the next generation of video compression. In *Proceedings of the Picture Coding Symposium*, volume 2018.

Pendu, M. L.; Guillemot, C.; and Thoreau, D. 2015. Local inverse tone curve learning for high dynamic range image scalable compression. *IEEE Transactions on Image Processing*, 24(12): 5753–5763.

Richter, T. 2013. Backwards Compatible Coding of High Dynamic Range Images with JPEG. In *Proceedings of the Data Compression Conference*, 153–160.

Rissanen, J.; and Langdon, G. 1981. Universal modeling and coding. *IEEE Transactions on Information Theory*, 27(1): 12–23.

Toderici, G.; O'Malley, S. M.; Hwang, S. J.; Vincent, D.; Minnen, D.; Baluja, S.; Covell, M.; and Sukthankar, R. 2016. Variable Rate Image Compression with Recurrent Neural Networks. arXiv:1511.06085.

Toderici, G.; Vincent, D.; Johnston, N.; Hwang, S.; Minnen, D.; Shor, J.; and Covell, M. 2017. Full Resolution Image Compression with Recurrent Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5306–5314.

Wang, Z.; Simoncelli, E.; and Bovik, A. 2003. Multiscale structural similarity for image quality assessment. In *Proceedings of the Asilomar Conference on Signals, Systems Computers*, volume 2, 1398–1402.

Ward, G. 1991. II.5 - REAL PIXELS. In *Graphics Gems II*, 80–83. Morgan Kaufmann.

Ward, G.; and Simmons, M. 2006. JPEG-HDR: a backwards-compatible, high dynamic range extension to JPEG. In *Proceedings of the ACM Special Interest Group for Computer GRAPHICS*, 3–es.

Ward, G.; et al. 2006. High dynamic range image encodings.

Yeganeh, H.; and Wang, Z. 2013. Objective Quality Assessment of Tone-Mapped Images. *IEEE Transactions on Image Processing*, 22(2): 657–667.

Zaid, A. O.; and Houimli, A. 2017. HDR image compression with optimized JPEG coding. In *Proceedings of the European Signal Processing Conference*, 1539–1543.

Zhong, Z.; Akutsu, H.; and Aizawa, K. 2020. Channel-Level Variable Quantization Network for Deep Image Compression. arXiv:2007.12619.