Learning a Practical SDR-to-HDRTV Up-conversion using New Dataset and Degradation Models, Supplementary Material

Cheng Guo[1,2], Leidong Fan[3,2], Ziyu Xue[4,1] and Xiuhua Jiang[2,1]
[1]State Key Laboratory of Media Convergence and Communication, Communication University of China
[2]Peng Cheng Laboratory [3]Peking University
[4]Academy of Broadcasting Science, National Radio and Television Administration
{guocheng,jiangxiuhua}@cuc.edu.cn, fanleidong@stu.pku.edu.cn, xueziyu@abs.ac.cn

Abstract

This material will provide additional content to accomplish the main paper, and extra experiment result.

1. Additional Content

1.1. Different types of HDR (main paper §2.1)

In main paper Tab.1, we distinguish **our scope** with other HDR-related works from not only application scenario, but also the type of target HDR: 'linear HDR' or 'HDRTV'. Here, we further explain their discrepancy by Fig.1.

At bottom-right Fig.1, we plot the relative luminance at same position (dashed line) to show their discrepancy: As seen, 'linear HDR' dedicate to record the full absolute range of luminance, while that of 'HDRTV' is only slightly more significant than SDR. This confirms the theory in [3] that SI-HDR is intolerable to the absent of highlight energy, while result from SDR-to-HDRTV up-conversion may still contain over-exposure which barely affect viewing experience. That is, our task emphasize more in viewing experience and less in over-exposure hallucination.

Theoretically, 'linear HDR' could be put into 'HDRTV' application after scale&normalization, color space transform (CST), and PQ/HLG non-linearization. However, such solution is usually not practical since: (1) Most 'linear HDR' record relative luminance (*i.e.* 1.0 pixel value means k*nit* luminance, k is unknown), thus it's hard to decide how the relative luminance should be scaled and normalized to $1000nit$. (2) Most 'linear HDR' is not assumed in WCG, when we put it in WCG container, simple CST will not produce any pixel in advanced color volume (outside BT.709), or we have to append extra gamut expansion method ([4–9] *etc*.) to obtain BT.2020 color volume. This is also the reason why SI-HDR (whose output is 'linear HDR') methods are not selected as competitors in main paper's experiment.

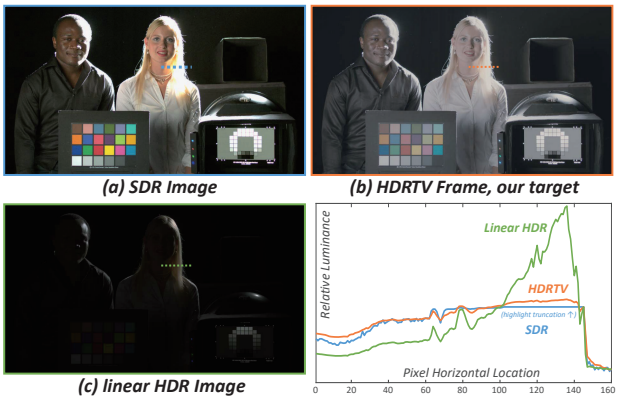

Figure 1. Different version of same scene, **(c)** and **(b)** are from HdM-HDR dataset [1], **(a)** is degraded from **(c)** by NTIRE [2]. **Note that**: (1) As explained in main paper Fig.1, **HDRTV appears duller than SDR in print version**. Here we take a more accessible example: *Suppose HDRTV and SDR record the same real luminance/color, since HDRTV has larger container, pixel representing same luminance/color will have bigger value in SDR* (e.g. *1) than HDRTV* (e.g. *0.7). In this case, due to smaller value, HDRTV will appear dimmer (luminance) and more desaturated (color) if it's unanimously visualized as SDR.* (2) 'linear HDR' also appear dark if directly visualized, since its linear pixel value concentrates in lower part after normalization.

1.2. HDRTV-tailored metrics (main paper §3.2)

Metrics from main paper Tab.4 are crucial since they serve as the theoretical basis of both HDRTV4K dataset and assessment criteria. Therefore, we define them in detail:

FLHP (Fraction of HighLight Pixels). First, each pixel of normalized HDR encoded value $\mathbf{E}' = [R', G', B']^{\mathrm{T}}$ is transferred to linear domain $\mathbf{E} = [R, G, B]^{\mathrm{T}}$ by PQ EOTF:

$$\mathbf{E} = \left(\frac{max[(\mathbf{E}'^{m_2} - c_1), 0]}{c_2 - c_3\mathbf{E}'^{m_2}}\right)^{m_1} \tag{1}$$

where m_1=16384/2610, m_2=524228/2523, c_1=3424/4096,

c_2=77216/4096 and c_3=76544/4096. Then, luminance Y is taken from tri-stimulus $\mathbf{S} = [X, Y, Z]^{\mathrm{T}}$ (*2nd row*):

$$\mathbf{S} = \mathbf{ME},\ \mathbf{M} = \begin{bmatrix} 0.6370 & 0.1446 & 0.1689 \\ 0.2627 & 0.6780 & 0.0593 \\ 0.0000 & 0.0381 & 1.0610 \end{bmatrix} \quad (2)$$

We treat >100nit (SDR's upper bound, correspond to 0.1 in normalized linear PQ1000) as 'highlight' part and count its spatial ratio as **FHLP** = $numel(Y{>}0.1)/(3840{\times}2160)$, where $numel(\cdot)$ means number of elements.

EHL (Extent of HighLight). As claimed in main paper, **EHL** is appended to compensate the shortcoming of **FHLP**, as illustrated in Fig.2. It's defined as the average pixel (i) distance between HDR's origin luminance (Y, from Eq.2 *2nd row*) and its highlight-clipped version:

$$\frac{1}{n}\sum_{i=1}^{n}\sqrt{[Y_i - clamp(Y_i, 0, 0.1)]^2},\ n = 3840 \times 2160 \quad (3)$$

where $clamp(\cdot)$ maps all highlight part in Y to 100nit.

Figure 2. The motive of **EHL**. We provide oscillogram to see how much energy is in highlight part (**EHL**) and heatmap (SDR's luminance range depicted as grayscale) to manifest both the extent (**EHL**) and spatial portion of highlight pixels (**FHLP**). The necessity of **EHL** is proven *e.g.* (b) and (c) share similar **FHLP** but distinct **EHL**. Also, (a)(b) are with same **EHL** but different **FHLP**.

FWGP (Fraction of Wide-Gamut Pixels). Calculated as **FWGP** = $numel(\mathbf{E}_{\mathbf{SDR}_{\mathbf{OOG}}} \notin (\mathbf{0}, \mathbf{1}))/(3840{\times}2160)$ where $\mathbf{E}_{\mathbf{SDR}_{\mathbf{OOG}}}$ is from main paper Eq.4. As explained there, OOG/WCG pixels will fall outside valid range if represented by RGB primaries of a narrower gamut (BT.709).

EWG (Extent of Wide-Gamut). Proposed by [10], the normalized pixel(i)-average distance between original tri-stimulus $\mathbf{S}$ and its gamut-hard-clipped version $\mathbf{S}_{HC}$:

$$\frac{1}{maxD}\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{S}_i - \mathbf{S}_{HCi}\|_2,\ maxD \approx 0.2751 \quad (4)$$

where $\mathbf{S}$ is from Eq.1&2. To get $\mathbf{S}_{HC}$, hard-clip is applied to the BT.709 version of $\mathbf{E}$, then the hard-clipped pixel ($\mathbf{E}_{709}$) is converted back to gamut-invariant tri-stimulus:

$$\mathbf{S}_{HC} = \mathbf{ME}_{709},\ \mathbf{M} = \begin{bmatrix} 0.4214 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \quad (5)$$

where $\mathbf{E}_{709}$ is derived using main paper Eq.4&5 (feed $\mathbf{E}$ as '$\mathbf{E}_{SDR}$' there, and treat '$\mathbf{E}_{SDR_{709}}$' as $\mathbf{E}_{709}$ here in Eq.5). The motivation of **EWG** is similar to **EHL**, examples are shown in Fig.3.

Measurement of HDR/WCG is highly valued in media industry, HDR/WCG area is usually display as *zebra pattern* in some monitor *e.g.* SNOY BVM-X300 to guide the content producer. Similarly, we use 4 metrics above on the *extent of HDR/WCG* to select better label HDR, and assess the quality of method's output HDR. Extremely, if some HDR's **FHLP**, **EHL**, **FWGP** and **EWG** all drop down to 0, it means this HDR is just a SDR content with HDR/WCG container, and HDRTV's advance on HDR/WCG volume will be completely untapped when it's displayed.

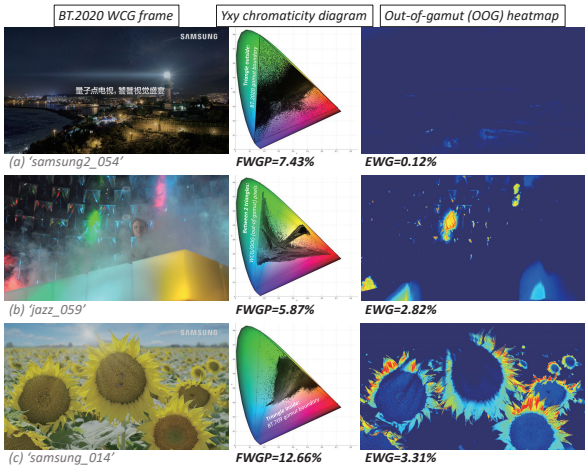

Figure 3. Example of **FWGP** and **EWG**, frames are from current HDRTV1K [11] training set. We provided *Yxy chromaticity diagram* to see which color is in WCG (outside BT.709) volume (**FWGP**), and OOG heatmap to identify WCG pixels' location and extent (**EWG**). As seen, (a)(b) share similar **FWGP** but distinct **EWG**, while (c) reach a **EWG** closed (b) with doubled **FWGP**.

Then, we start to introduce 3 metrics on the *overall-style*, which serve as an important reference in both DMs and assessment criteria. First, single HDR frame's **ALL (Average Luminance Level)** is the pixel-average of luminance channel Y from *2nd row* in Eq.2. **HDRBQ** is defined in [12], and had been proven of higher relevance to subjective score than **ALL**. Higher metrics means brighter view HDRTV frame will deliver when correctly visualized.

ASL (Average Saturation Level) is derived as follow: Start with linear RGB in BT.2020 primaries ($\mathbf{E}$ in Eq.1), we

first transfer it to LMS response by:

$$[L, M, S]^{\mathrm{T}} = \mathbf{ME},\ \mathbf{M} = \begin{bmatrix} 1688 & 2146 & 262 \\ 683 & 2951 & 462 \\ 99 & 309 & 3688 \end{bmatrix} /4096 \quad (6)$$

Then, LMS is converted to PQ-nonlinear using OETF:

$$E' = (\frac{c_1 + c_2 E^{1/m_1}}{1 + c_3 E^{1/m_1}})^{1/m_2},\ E \in \{L, M, S\} \quad (7)$$

Finally, IC_tC_p is derived from $E' = [L', M', S']^{\mathrm{T}}$:

$$[I, C_t, C_p]^{\mathrm{T}} = \mathbf{M}E',\ \mathbf{M} = \begin{bmatrix} 2048 & 2048 & 0 \\ 6610 & -13613 & 7003 \\ 17933 & -17390 & -543 \end{bmatrix} /4096 \quad (8)$$

where luminance component $I \in$ [0,1] and chrominance components $C_tC_p \in$ [-0.5,0.5]. Since I and C_tC_p are designed to be independent, from [13] we know that the length ($\|\mathbf{C}\|$) of chrominance vector $\mathbf{C} = [C_t, C_p]^{\mathrm{T}}$ represent pixel's saturation (angle means hue). Therefore, frame's overall saturation (**ASL**) is calculated as the pixel-average of $\|\mathbf{C}\|$, as formulated in main paper Tab.4. Example of **ASL**'s C_t-C_p chrominance plane is shown in Fig.4.

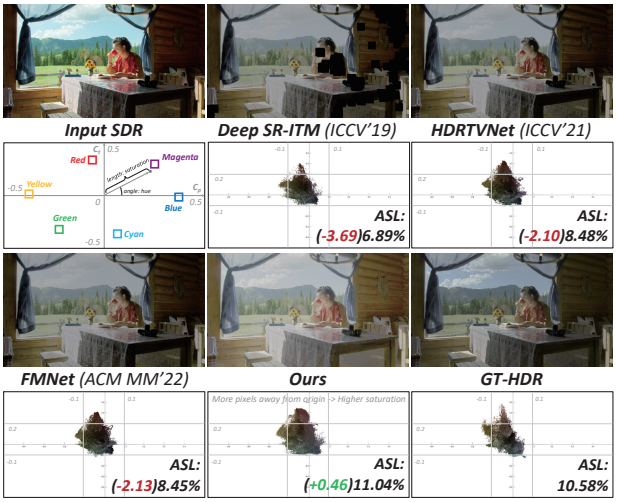

Figure 4. Example of **ASL**'s C'_t-C'_p plane: distance to original point indicate single pixel's saturation, while its angle means hue. We use the same scene as main paper Fig.1, as seen, result with higher **ASL** do 'spread' wider among C'_t-C'_p chrominance plane.

Note that since the scale of HDR's IC_tC_p consists with SDR's $Y'C'_bC'_r$ *i.e.* $C'_bC'_r$ also range in [-0.5,0.5], **ASL** become the only metric which is numerically comparable between HDR&SDR. For SDR's **ASL** (used in main paper Tab.6 *etc.*), we change $\mathbf{C}$ to $[C'_b, C'_r]^{\mathrm{T}}$ [14] which are derived from SDR *gamma*-encoded value $\mathbf{E}'_{SDR} = [R', G', B']^{\mathrm{T}}$:

$$C'_b = \frac{B' - Y'}{1.8556},\ C'_r = \frac{R' - Y'}{1.5748} \quad (9)$$

where $Y' = 0.2627R' + 0.6780G' + 0.0593B'$.

Finally, we append 3 metrics on *intra-frame diversity* which are (1) not HDR-exclusive and (2) not used in assessment criteria. **SI (Spatial Information)** and **CF (ColorFullness)** are respectively defined in Annex 6 of [15] and [16], while single frame's **stdL** is the pixel-average of the variance of luminance Y channel (Eq.2 *2nd row*). These metrics all involve variance, meanwhile Annex 2 of [17] provide some entropy-based metrics. We exclude the latter since they are highly relevant to the former.

1.3. Illustration of different degradation models (DMs) (main paper Tab.6)

In main paper Tab.6, §2.3 & §3.3, we mentioned that current DMs fail for undue style change and inadequate degradation capability. This phenomenon will be illustrated by different degraded SDR in Fig.5 and DMs' corresponding LUTs (look-up tables) in Fig.6.

Figure 5. SDR images generated by different degradation models (DMs), from same HDR. As seen, tone-mapping operators (TMOs) like ***2446a*** and ***Reinhard*** dedicate to preserve as much information from HDR, so their SDRs are of less **FOEP** (fraction of over-exposed pixels). Also, from **ALL** (average luminance level) and **ASL** (average saturation level) we now that current DMs in first row tend to exaggeratedly alter the style from HDR to SDR, so network will learn a vise-versa SDR-HDR style tendency which do not follow the technical and artistic intend of origin SDR.

2. Detailed Experiment

2.1. Why conventional metrics fail (main paper §4.2)

It's mentioned in the footnote at main paper §4.2 that conventional distance-based metrics *e.g.* PSNR, SSIM, ΔE [18] even HDR-VDP-3 [19] could not meet our assessment criteria. We prove this phenomenon by:

(1) Proof by contradiction. The 1st(our)/2nd(*DaVinci*) methods in subjective experiment only score 7th/8th PSNR and 7th/6th ΔE, coincidentally PSNR and ΔE are purely distance-based. This indicate that higher conventional metrics only stand for closer value with GT. However, in this 'perceptual-motivated' SDR-to-HDRTV up-conversion, result is allowed to be better than GT. So there occurs case *e.g.* both brighter (ours) and dimer (Deep SR-ITM) score

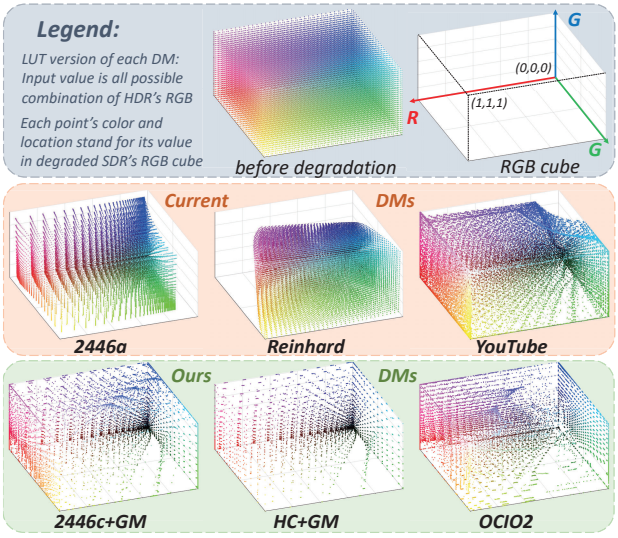

Figure 6. Since those DMs are all global (pixel-independent) operations, we visualize them as LUT (look-up table). Points in SDR($\mathbf{x}'$)' RGB cube are the corresponding output value of all possible input HDR($\mathbf{y}'$) value. More points on R/G/B=1 planes and sparser points between 0-1 mean more clipping occurs, and more points near R=G=B (achromatic axis) stand for degraded SDR's less saturation. As seen, more clipping occurs in ours DMs. Also, from the shape we know that their characteristics are closer.

the same low, since their pixel value are both far form GT.

(2) Dataset shift (drifting), first found by [3, 20]. As claimed in main paper, the reason why '***YouTube***-DM' methods score higher conventional metrics, is that the 'LQ-GT relationship' of 3 of 12 test clips has been 'seen' by these methods. That is, methods only score higher on test set with similar distribution/characteristics to their training set. In our task, such characteristics is partly determined by label HDR (GT), but more by DM (GT-to-LQ). We further verify this by testing 3 methods (trained with different representative DM(s)) on 6 types of input SDR (degraded from same GT by 6 different DMs as listed in Tab.1&2&3). For better reproducibility, we use current benchmark—117 HDR frames from HDRTV1K [11] test set.

Training set DM	Test set DM	G.	PSNR↑	SSIM↑	ΔE↓	VDP3↑
YouTube	*2446a*	✓	14.99	0.589	139.1	5.968
	Reinhard	✓	24.33	0.833	56.04	8.215
	YouTube	×	**37.13**	**0.970**	**9.135**	**8.595**
	2446c+GM	✓	21.28	0.907	67.14	7.559
	HC+GM	✓	15.22	0.533	127.6	6.255
	OCIO2	✓	23.12	0.807	54.04	8.152

Table 1. Metrics of **HDRTVNet** [11] (trained with ***YouTube*** DM), on different types of SDR. '**G.**' stands for 'generalization experiment' *i.e.* if test set DM is different with training set.

As seen, method's conventional metrics are better only when testing on 'familiar' LQ-GT relationship. This phenomenon is similar to, for example, an image retouching network trained on *Expert C* of MIT-Adobe FiveK dataset will score lower when testing on *Expert A*. The differences lies in: (1) The 'multiple LQ-GT relationship' in MIT-Adobe FiveK is caused by various GT (*Expert A-E*), while that in our task is by diversified LQ (*i.e.* DMs). (2) All 'LQ-GT relationship' in MIT-Adobe FiveK dataset will help the network learn an enhancement, while some 'LQ-GT relationship' in our task lead to deterioration.

Fig.7 can better explain **(1)** and **(2)**:

Training set DM	Test set DM	G.	PSNR↑	SSIM↑	ΔE↓	VDP3↑
Reinhard	*2446a*	✓	13.55	0.521	160.2	5.065
	Reinhard	×	**32.32**	**0.929**	**19.65**	**8.406**
	YouTube	✓	25.14	0.858	42.10	6.962
	2446c+GM	✓	17.83	0.835	95.08	6.264
	HC+GM	✓	14.18	0.484	143.6	5.366
	OCIO2	✓	18.68	0.765	84.23	5.701

Table 2. Result of **SR-ITM-GAN** [21] (trained by ***Reinhard*** DM).

Training set DM	Test set DM	G.	PSNR↑	SSIM↑	ΔE↓	VDP3↑
2446c+GM + HC+GM + OCIO2	*2446a*	✓	24.71	0.917	48.20	8.011
	Reinhard	✓	16.34	0.736	117.7	7.803
	YouTube	✓	18.13	0.859	93.69	8.144
	2446c+GM	×	24.59	**0.923**	**41.33**	**8.905**
	HC+GM	×	20.39	0.731	66.18	7.797
	OCIO2	×	**25.84**	0.868	42.57	7.843

Table 3. **Our method**'s performance on 6 types of input SDR.

2.2. More visuals (main paper §4.2)

After clarifying the deficiency of conventional metrics, we stick to new assessment criteria and provide more visual comparison based on it. Results of different competitors are provided in Fig.8&9&10, while more illustration on ablation studies are provided in Fig.11.

2.3. Correct display of HDRTV (main paper §4.3)

As explained earlier, HDR will appear dim in print version. That is, HDRTV frames in Fig.8&9&10 are only comparable between each other, *i.e.* A dimer than B in print version will also be dimer when correctly visualized. However, since the motive of our task is to promote viewing experience [22], we need to check if result HDR is better than SDR. A glance on how this is achieved has been shown in main paper Fig.7, here we provide more in Fig.12&13&14

2.4. Sole comparison of model capability

In main paper, all learning-based competitors are not retrained with our data, on the purpose of assessing if they can be practically used in media industry with their original intention. This experiment is part of the project 'quality assessment of UHD video enhancement', and it assess not only the network capability but more training strategy.

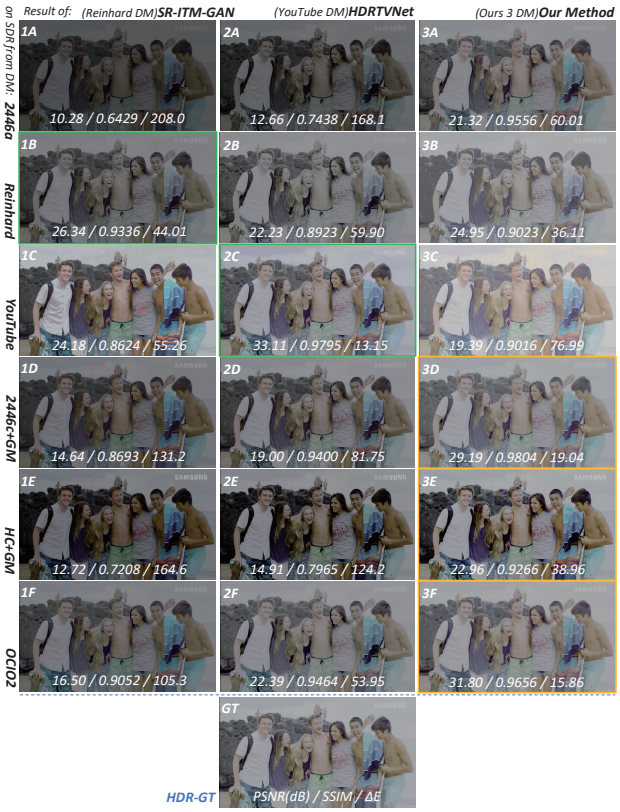

Figure 7. Feeding 3 representative methods (trained with distinct DMs) with various SDR (degraded from same GT using different DM). We use frame *'078'* from HDRTV1K [11] test set, and illustrate each result's **PSNR/SSIM/ΔE**. We notice that: **(1)** result both worse/dimer (***2D***) and more vivid (***3C***) score the same low, and **(2)** method scores higher when comes to SDR degraded by same DM as training (highlighted with green and yellow boxes), since such 'LQ-GT relationship' has been seen. Such demo should suffice the theory **why conventional distance-based metrics fail**.

Therefore, we append a 'sole' comparison of network capability. Here, with our training settings, each method is trained and tested using old benchmark HDRTV1K dataset [11] (we also used their SDR so the DM is still ***YouTube***). Results are provided in Tab.4.

References

[1] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays," in *Digital photography X*, vol. 9023, pp. 279–288, 2014. 1

[2] E. Pérez-Pellitero *et al.*, "Ntire 2022 challenge on high dynamic range imaging: Methods and results," in *Proc. CVPR*, pp. 1009–1023, 2022. 1

[3] G. Eilertsen, S. Hajisharif, P. Hanji, A. Tsirikoglou, R. K. Mantiuk, and J. Unger, "How to cheat with metrics in single-image hdr reconstruction," in *Proc. ICCV*, pp. 3998–4007, 2021. 1, 4

[4] J. Preiss, M. D. Fairchild, J. A. Ferwerda, and P. Urban, "Gamut mapping in a high-dynamic-range color space," in *Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications*, vol. 9015, pp. 79–85, 2014. 1

[5] M. Takeuchi, S. Saika, Y. Sakamoto, Y. Matsuo, and J. Katto, "A study on color-space conversion method considering color information restoration," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–2, 2018. 1

[6] H. Le, M. Afifi, and M. S. Brown, "Improving color space conversion for camera-captured images via wide-gamut metadata," in *Color and Imaging Conference*, vol. 2020, pp. 193–198, 2020. 1

[7] H. Anderson, E. Garcia, and M. Gupta, "Gamut expansion for video and image sets," in *14th International Conference of Image Analysis and Processing-Workshops (ICIAPW 2007)*, pp. 188–191, 2007. 1

[8] M. Takeuchi, Y. Sakamoto, R. Yokoyama, S. Heming, Y. Matsuo, and J. Katto, "A gamut-extension method considering color information restoration using convolutional neural networks," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 774–778, 2019. 1

[9] H. Le, T. Jeong, A. Abdelhamed, H. J. Shin, and M. S. Brown, "Gamutnet: Restoring wide-gamut colors for camera-captured images," in *Color and Imaging Conference*, vol. 2021, pp. 7–12, 2021. 1

[10] L. Bai, Y. Yang, and G. Fu, "Analysis of high dynamic range and wide color gamut of uhdtv," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 1750–1755, 2021. 2

[11] X. Chen, Z. Zhang, J. S. Ren, L. Tian, Y. Qiao, and C. Dong, "A new journey from sdrtv to hdrtv," in *Proc. ICCV*, pp. 4500–4509, 2021. 2, 4, 5, 7

[12] S. Ploumis, R. Boitard, and P. Nasiopoulos, "Image brightness quantification for hdr," in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 640–644, 2021. 2

[13] Dolby, "What is ictcp? - introduction." https://professional.dolby.com/siteassets/pdfs/ictcp_dolbywhitepaper_v071.pdf. 3

[14] ITU, Geneva, Switzerland, *Recommendation ITU-R BT.709-6: Parameter values for the HDTV standards for production and international programme exchange*, 6 ed., 6 2015. 3

[15] ITU, Geneva, Switzerland, *Recommendation ITU-R BT.500-14: Methodologies for the subjective assessment of the quality of television images*, 14 ed., 10 2019. 3

[16] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Human vision and electronic imaging VIII*, vol. 5007, pp. 87–95, 2003. 3

[17] ITU, Geneva, Switzerland, *Report ITU-R BT.2245-10: HDTV and UHDTV including HDR-TV test materials for assessment of picture quality*, 10 ed., 9 2022. 3

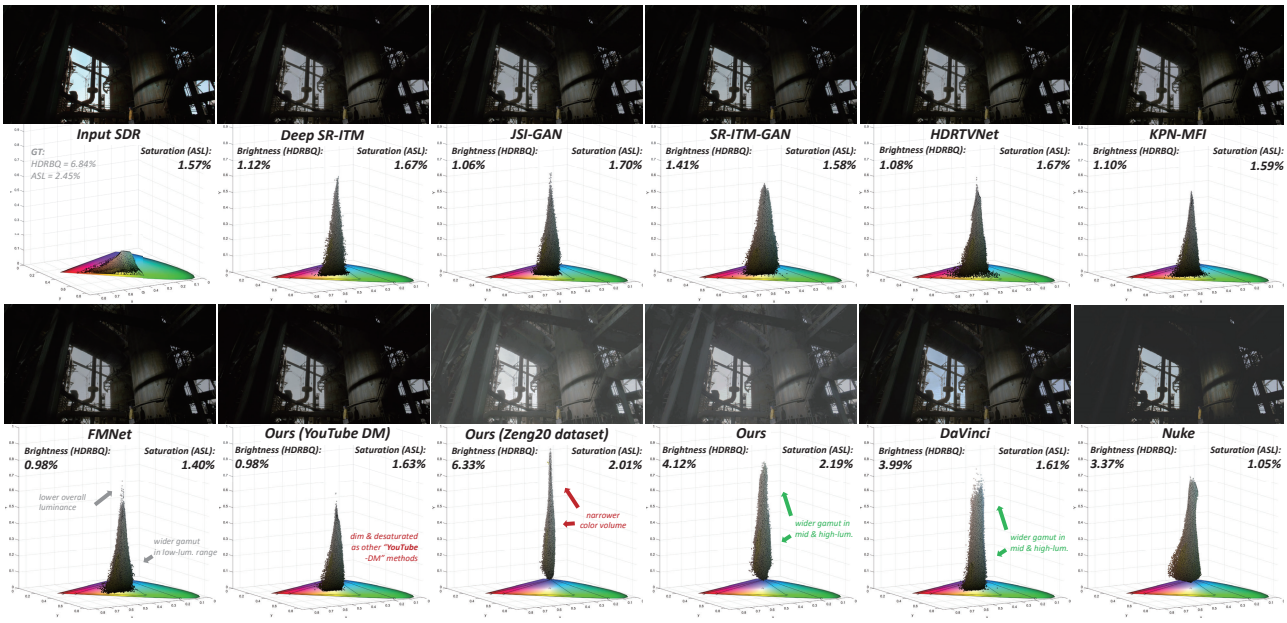

Figure 8. Testing on a scene with high latitude/contrast. Current methods make no enhancement on both saturation and brightness, meanwhile recover little information from dark area. Our method avoids above deficiencies. Also, when our label HDR is changed to Zeng20 which is of least *extent of HDR/WCG*, our network still learns information recovery and saturation, but produce way narrower color volume. When we change DM to ***YouTube*** while keeping label HDR from HDRTV4K dataset unchanged, our network learns similar dim and desaturated result as other methods which are also trained with ***YouTube*** DM.

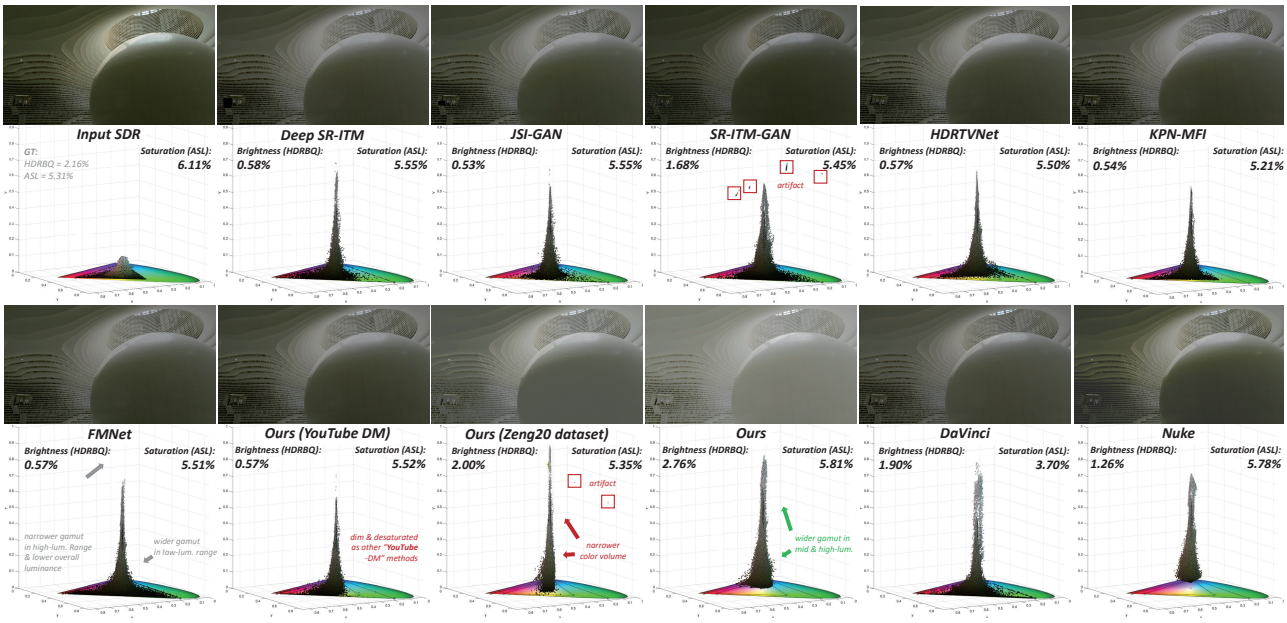

Figure 9. Testing on a scene with less colorfulness. Conclusion similar to Fig.8 can be drawn. What's special is that: our method and other commercial methods (*DaVinci* and *Nuke*) ia able to keep a 'plump' color volume in higher luminance range even for less saturated scene, while other learning-based methods' *3D Yxy chromaticity diagrams* are 'sharp' at high-lumniance range.

[18] ITU, Geneva, Switzerland, *Recommendation ITU-R BT.2124-0: Objective metric for the assessment of the potential visibility of colour differences in television*, 0 ed., 1 2019. 3

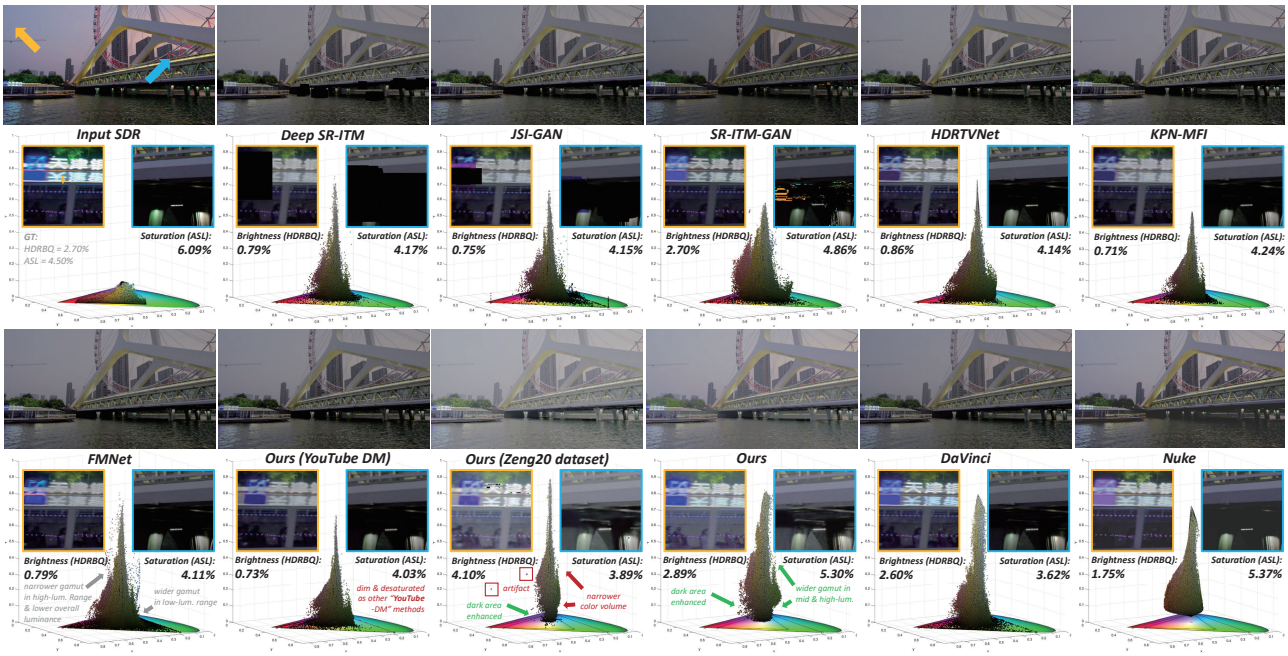

Figure 10. Performance on low-luminance condition. Conclusion similar to Fig.8&9 can be drawn. Also, Deep SR-ITM-GAN, JSI-GAN and SR-ITM-GAN generate artifact at dark area, while that of ours (Zeng20 dataset) is on highlight part. Our method again recovers more in dark area (blue box) while maintaining good look at bright area (yellow box).

Network	#param↓	runtime↓	PSNR(dB)↑	SSIM↑	ΔE↓	VDP3↑
Deep SR-ITM	2.87M	1.21s	36.09	0.961	11.351	7.940
JSI-GAN	1.06M	0.87s	35.99	0.951	10.854	8.102
SR-ITM-GAN	515k	0.61s	36.83	0.966	9.177	8.465
HDRTVNet(full)	37.20M	1.20s	37.39	0.949	**9.070**	**8.701**
KPN-MFI	3.37M	1.69s	37.33	0.957	9.667	8.560
FMNet	1.24M	0.70s	**37.51**	**0.978**	9.270	8.586
LSN (ours)	**325k**	**0.53s**	37.13	0.970	9.135	8.595

Table 4. Only network capability is compared, when all methods are trained with same training settings, and tested on same HDRTV1K [11] test set (with 117 frames). We highlight the 1st/2nd score with bold/underline. Note that, since some methods (Deep SR-ITM, JSI-GAN and KPN-MFI) are not applicable to UHD (3840×2160) resolution due to GRAM OOM (out-of-memory) under our 12 GB GPU, the runtime is all counted when feeding HD (1920×1080) SDR frame(s). As seen, our method scores the 4th/2nd/2nd/2nd the PSNR/SSIM/ΔE/VDP3 with the minimum number of parameters (#param) and runtime.

[19] K. Wolski, D. Giunchi, *et al.*, "Dataset and metrics for predicting local visible differences," *ACM Trans. Graph.*, vol. 37, no. 5, pp. 1–14, 2018. 3

[20] P. Hanji, R. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, "Comparison of single image hdr reconstruction methods—the caveats of quality assessment," in *Proc. SIGGRAPH*, pp. 1–8, 2022. 4

[21] H. Zeng, X. Zhang, Z. Yu, and Y. Wang, "Sr-itm-gan: Learning 4k uhd hdr with a generative adversarial network," *IEEE Access*, vol. 8, pp. 182815–182827, 2020. 4

[22] ITU, Geneva, Switzerland, *Report ITU-R BT.2381-0: Requirements for high dynamic range television (HDR-TV) systems*, 0 ed., 7 2015. 4

Figure 11. Ablation studies, conclusion similar to main paper can be drawn. That is, when we keep label HDR from new HDRTV4K dataset unchanged, but use other DMs *e.g*. ***YouTube*** and ***Reinhard***, network (LSN) will learn a similar dark (yellow boxes and arrows) and desaturated *style* like other network trained on these DMs. Moreover, ***Reinhard*** TMO produce little degradation in SDR, so LSN didn't learn to recover highlight information in HDR. Also, when we keep 3 DMs (***OCIO2***, ***2446c+GM*** and ***HC+GM***) un altered, and use label HDR from other datasets, LSN's learned *style* tendency remain similar. Yet, when it comes to Zeng20 dataset which contains less HDR/WCG volume, our LSN will not 'recognize' then anf produce artifact in to-be-recovered highlight area. In the case of slightly-inferior HDRTV1K dataset, the difference is relatively less significant, but still noticeable from the expanded luminance value indicated with yellow arrow.

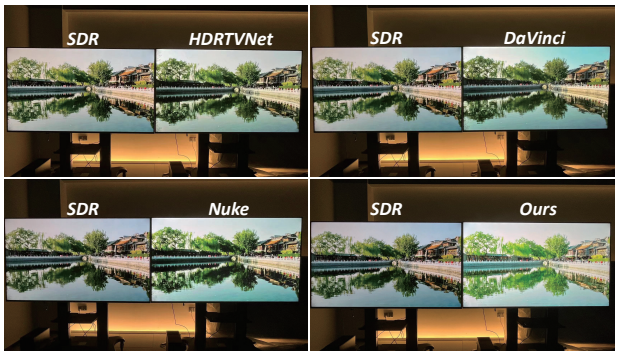

Figure 12. How the participant of subjective experiment feel, when simultaneously watching **SDR** (left) **and HDRTV** (right) which are **both correctly visualized**. We illustrate only 1 learning-based method, HDRTVNet, since ***YouTube***-DM methods all look similar on HDRTV. Photos here are shot DLSR with same setting. As seen, our result is more vivid, since HDRTV's advance on HDR/WCG volume is utilized to greater extent. Please zoom in.

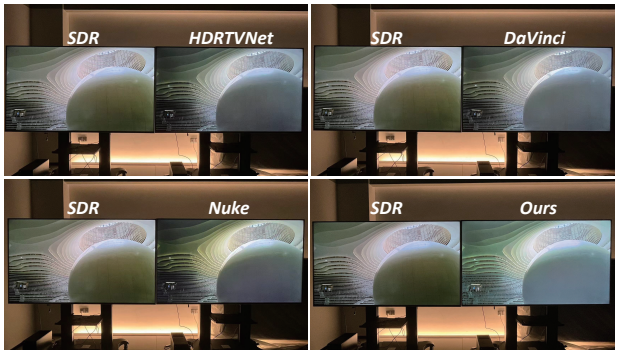

Figure 13. Conclusion similar to Fig.12 can be drawn.

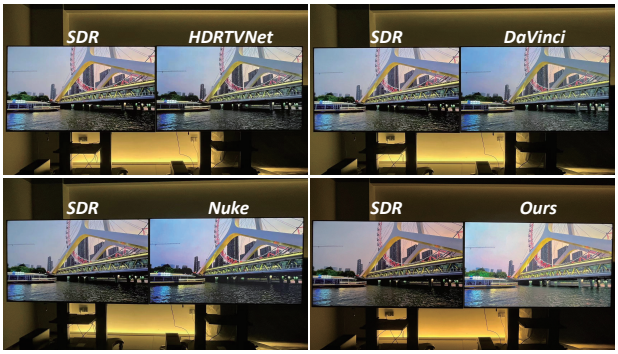

Figure 14. Conclusion similar to Fig.12&13 can be drawn.