

GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild – Supplementary Materials

Chao Wang¹ Ana Serrano² Xingang Pan^{1,3} Bin Chen¹ Karol Myszkowski¹
Hans-Peter Seidel¹ Christian Theobalt¹ Thomas Leimkühler¹
¹MPI Informatik ²Universidad de Zaragoza ³Nanyang Technological University
{chaowang, binchen, karol, theobalt, thomas.leimkuehler}@mpi-inf.mpg.de xingang.pan@ntu.edu.sg
hpseidel@mpi-sb.mpg.de

Abstract

This document provides more details on the following topics:

- *S1: Details of GAN inversion.*
- *S2: Analysis of the dataset statistics.*
- *S3: Analysis of the failure case.*
- *S4: Comparisons of the generated results and the training dataset.*
- *S5: Image quality and dynamic range of the generated results for other categories dataset.*
- *S6: More generated HDR samples, and comparisons with Vanilla-GAN results. The original HDR images are also provided in the attachment.*
- *S7: The gradual transition between two environment maps in IBL, which can be more clearly observed in the attached video.*
- *S8: GlowGAN-generated HDR images can serve as a valuable source of data augmentation for various downstream tasks, such as inverse tone mapping.*
- *S9: More results on the exposure distribution variance.*
- *S10: More visualization comparisons of inverse tone mapping results. We also provide original HDR files in the attachment.*

The provided HDR-format images can be displayed with HDR viewers like <http://www.hdrilabs.com/picturenaut/>.

1. Details of GAN Inversion

As described in the main paper Sec. 3.2, to obtain a more realistic inversion result, we adopt a two-stage optimization with the following objective:

$$[e^*, \mathbf{w}^*, \theta_S^*] = \operatorname{argmin}_{e, \mathbf{w}, \theta_S} \Phi \left(C(S_{\theta_S}(\mathbf{w})), \hat{\mathbf{I}} \right). \quad (1)$$

In the first stage, only the latent code \mathbf{w} and exposure e are optimized. We run 1000 iterations with a learning rate of 0.1. The learning rate ramps up from 0 to 0.1 linearly in the first 50 steps and ramps down to 0 in the last 250 steps. In the second stage, we only optimize the generator weights θ_S while fixing the remaining parameters following [5]. We run 350 iterations with a learning rate of 3×10^{-5} in this stage. The optimization takes about 15 minutes on one RTX 8000 GPU.

2. Analysis of the Dataset Statistics

We conducted an analysis of image statistics on the *Landscape* dataset. We show a histogram of the mean value per image in Fig. 1. Furthermore, we categorize a pixel as overexposed if its value exceeds 0.95 and as underexposed if it falls below 0.05. We then compute the proportions of overexposed and underexposed pixels across the entire dataset, finding that overexposed pixels account for 2.5%, while underexposed pixels make up 9.2%.

3. Analysis of the Failure Case

The failure case shown in Fig. 10 in main paper can be attributed to the significant presence of high-frequency content in the overexposed LDR image. To illustrate this effect, we conducted an experiment where we add varying amounts of Gaussian noise to an image prior to GAN inversion, as demonstrated in Fig. 2. While GAN inversion can handle reasonable amounts of

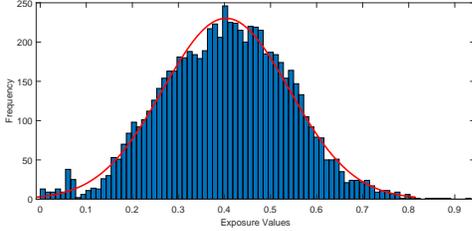


Figure 1: Histogram of mean pixel intensities.

noise as typically encountered in images (Input 1), we see that inversion quality declines with increased noise level (Inputs 2 and 3).

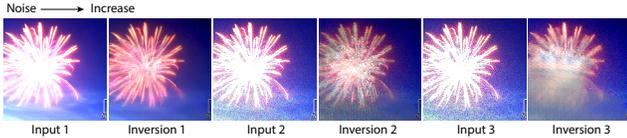


Figure 2: Influence of noise on GAN inversion.

4. Comparisons of the Generated Results and the Training Dataset

In Fig. 3 we show multiple LDR samples of our model next to their nearest neighbors in the dataset, according to the LPIPS distance [6]. We see that our samples are clearly different, strongly indicating that our model synthesizes new images rather than memorization.



Figure 3: Our samples vs. their nearest neighbors in the training dataset.

5. Analysis of the Quality and Dynamic Range of the generated results for other categories.

We provide objective evaluation of the generated results for datasets of other categories ($\sigma_e^2 = 1.0$) in Table 1. Image quality is evaluated using FID and KID, while dynamic range extension is measured using DR50 and DR90. The FID and KID scores exhibit a reasonable alignment, indicating favorable image quality. The

DR50 and DR90 metrics vividly showcase a substantial expansion of the dynamic range.

Categories	FID↓	KID($\times 10^4$)↓	DR50	DR90
Lightning	3.29	4.67	15.5	21.1
Fireplace	6.00	12.4	19.6	23.3
Fireworks	3.53	4.05	18.4	23.8
Windows	5.69	3.55	14.9	18.6

Table 1: Image quality and dynamic range evaluation for the generated results of other datasets ($\sigma_e^2 = 1.0$).

6. Visualization of Generated Samples

We show more HDRI samples generated from our models. To show the superiority of our algorithm, we provide the samples generated from a Vanilla GAN and GlowGAN with different exposure in the Landscapes and Lightning datasets shown in Fig. 4 to Fig. 11. We can see that the Vanilla-GAN samples can only represent a limited dynamic range, while GlowGAN samples have a much wider dynamic range. In addition, a larger variance σ_e^2 can extend the dynamic range to some extent. For fairness, we normalize the HDR images such that their luminance’s integrals are the same, following [2]. This is to ensure that all images are roughly aligned around the same scale. We also show tone-mapped results using [3] in Fig. 12 to Fig. 14. In addition, high resolution (512×512) samples for Landscapes and Lightning are shown in Fig. 16 and Fig. 17.

7. Image Based Lighting

The HDR images that GlowGAN generates serve as a cheap source of environment maps for Image Based Lighting. Additionally, thanks to the semantically meaningful latent space of GlowGAN, we can interpolate between two latent codes to achieve a gradual transition effect between two light sources, *e.g.*, from sunset to daytime, as shown in Fig. 19. The gradual transition effect can be more clearly observed in the attached video.

8. Data augmentation for ITM

GlowGAN’s ability to generate abundant HDR images presents a promising application for data augmentation in downstream tasks such as inverse tone mapping (ITM). We explore this idea using two supervised methods, HDRCNN [1] and ExpandNet [4]. We collected real HDR images to create HDR-LDR training pairs. We also used GlowGAN to generate HDR images and corresponding HDR-LDR pairs. We then trained two

ITM models using three strategies: a) training with 16k real samples, b) training with 8k real and 8k generated samples (Mix), and c) using the same data as in a) plus additional 10k GlowGAN-generated samples (Aug). We evaluated the models’ performance on a test set of 220 images, shown in Table 2. We observe that ITM quality is roughly the same in strategy a) and b), while performance improved consistently using the GlowGAN-based data augmentation strategy c). As can be seen in Fig. 18, the visual quality of ITM outputs is enhanced.

Model	Data	HDR-VDP3 \uparrow	PU21-VSI \uparrow	PU21-PSNR \uparrow
HDRCNN	16k	6.31 \pm 1.46	0.950 \pm 0.037	27.02 \pm 6.08
HDRCNN-Mix	16k	6.37 \pm 1.49	0.949 \pm 0.038	27.03 \pm 6.07
HDRCNN-Aug	26k	6.55 \pm 1.42	0.953 \pm 0.037	27.50 \pm 5.98
ExpandNet	16k	6.34 \pm 1.64	0.955 \pm 0.038	28.89 \pm 5.89
ExpandNet-Mix	16k	6.34 \pm 1.66	0.957 \pm 0.034	28.88 \pm 5.90
ExpandNet-Aug	26k	6.46 \pm 1.65	0.956 \pm 0.036	29.30 \pm 6.16

Table 2: Impact of incorporating GlowGAN-generated HDR images as data augmentation on ITM performance.

9. Effects of the Exposure Distribution Variance

Table 3 shows the effects of the exposure distribution variance σ_e^2 on image quality and dynamic range evaluated on the Lightning dataset. The results are consistent with that on the Landscapes dataset (Table 1 in the main paper). We also study its effect on inverse tone mapping in Table 4, which shows that this application favors $\sigma_e^2 = 1$.

Model	σ_e^2	FID \downarrow	KID($\times 10^4$) \downarrow	DR50	DR90
SG-XL ¹	–	3.15	4.64	11.1	12.3
Ours	1.0	3.29	4.67	15.5	21.1
Ours	3.0	3.34	4.78	15.8	21.3
Ours	5.0	3.60	5.95	16.0	21.2

¹ Refers to a vanilla StyleGAN-XL model.

Table 3: Effects of σ_e^2 on quality and dynamic range for the Lightning dataset.

σ_e^2	PU21-PIQE(\downarrow)
1.0	31.8 \pm 5.1
3.0	32.8 \pm 6.1
5.0	33.3 \pm 6.3

Table 4: Inverse tone mapping quality for different σ_e^2 .

10. Visualization of Inverse Tone Mapping Comparisons

We include all of the ITM results used in the user study, as shown in Fig. 23 to Fig. 43. We observe that our method has a strong advantage in the completion of overexposed regions. To verify the stability of our method, we use different random seeds for optimization on the Landscapes dataset, and count the frequency of PU21-PIQE. Fig. 20 shows the distribution. From 30 different initial seeds, we see that our method performs stably with an average PU21-PIQE of 31.97 and a standard deviation of 0.14. Furthermore, we show the different inversion results in Fig. 21 from the same LDR target to illustrate the diversity of our method. In addition, Fig. 22 shows the robustness of our methods for different overexposed levels, where we can always generate plausible content.

References

- [1] Eilertsen et al. HDR image reconstruction from a single exposure using deep cnns. *ACM TOG*, 36(6):1–15, 2017. 2, 4
- [2] Roland W Fleming, Ron O Dror, and Edward H Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of vision*, 3(5):3–3, 2003. 2
- [3] Rafał Mantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. In *ACM SIGGRAPH 2008 papers*, pages 1–10. 2008. 2, 8, 9, 10
- [4] Demetris Marnierides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Computer Graphics Forum*, volume 37, pages 37–49. Wiley, 2018. 2
- [5] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. *arXiv preprint arXiv:2106.05744*, 2021. 1
- [6] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 2



Figure 4: Landscapes Sample 1. (From left to right, the exposure stops [1] are -8, -5, -2, 0.)

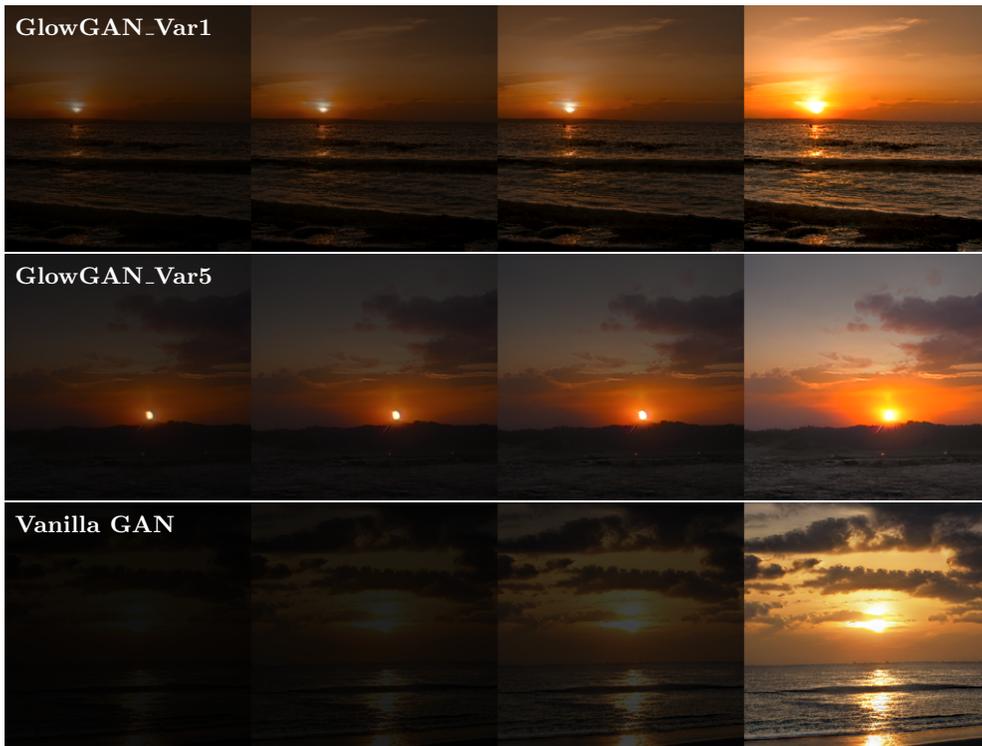


Figure 5: Landscapes Sample 2. (From left to right, the exposure stops are -8, -5, -2, 0.)



Figure 6: Landscapes Sample 3. (From left to right, the exposure stops are -8, -5, -2, 0.)

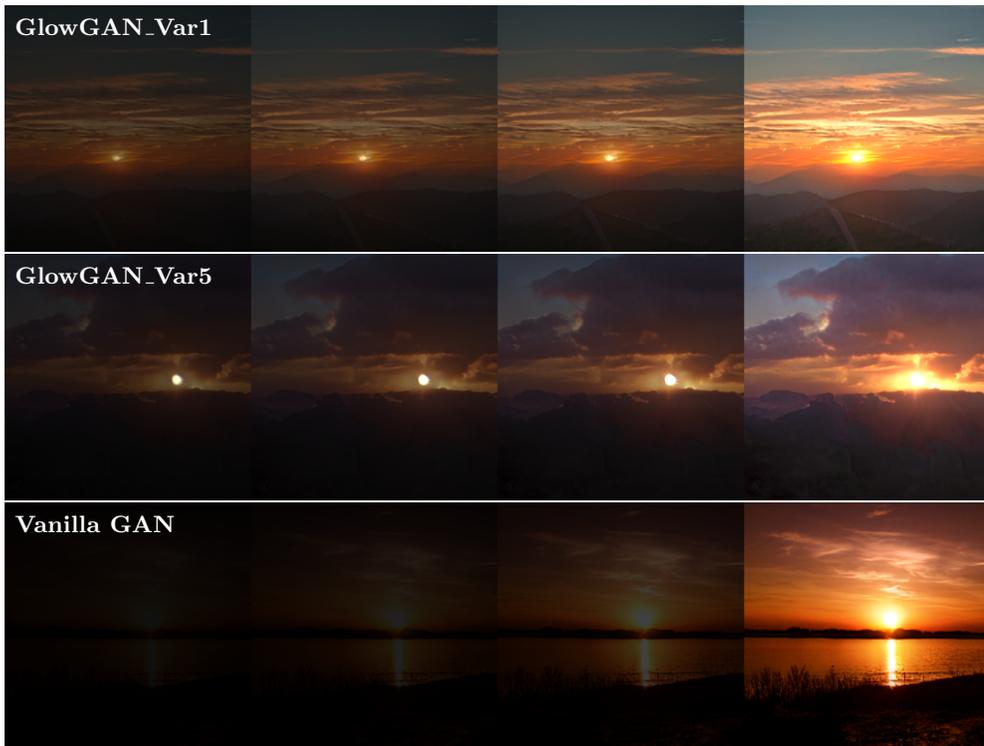


Figure 7: Landscapes Sample 4. (From left to right, the exposure stops are -8, -5, -2, 0.)



Figure 8: Lightning Sample 1. (From left to right, the exposure stops are -8, -5, -2, 0.)

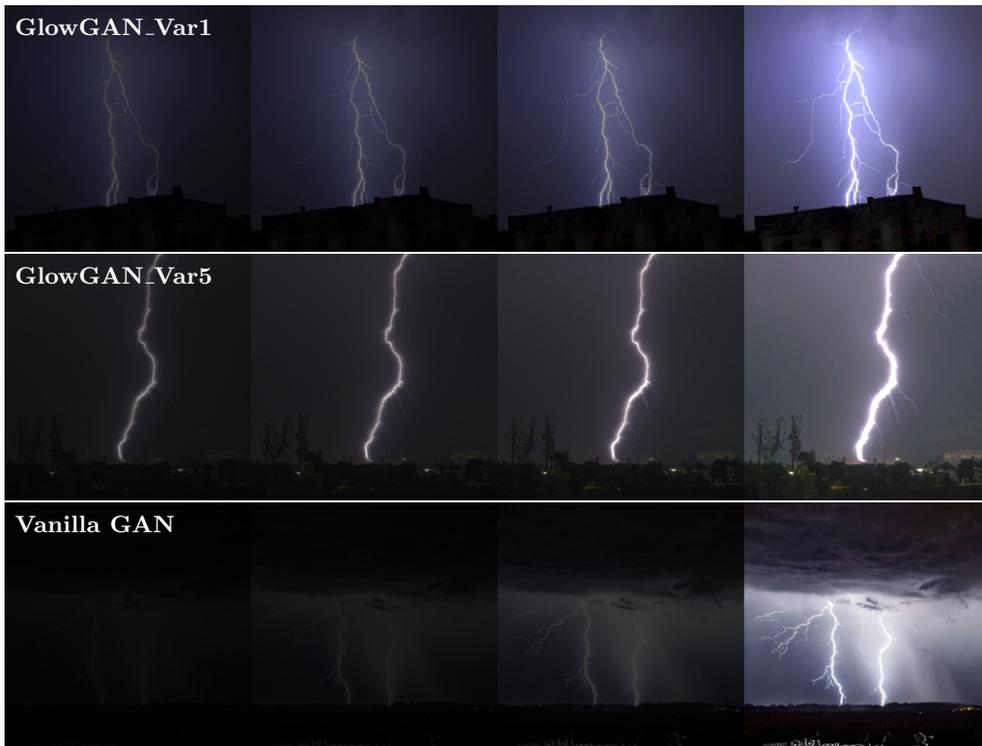


Figure 9: Lightning Sample 2. (From left to right, the exposure stops are -8, -5, -2, 0.)



Figure 10: Lightning Sample 3. (From left to right, the exposure stops are -8, -5, -2, 0.)

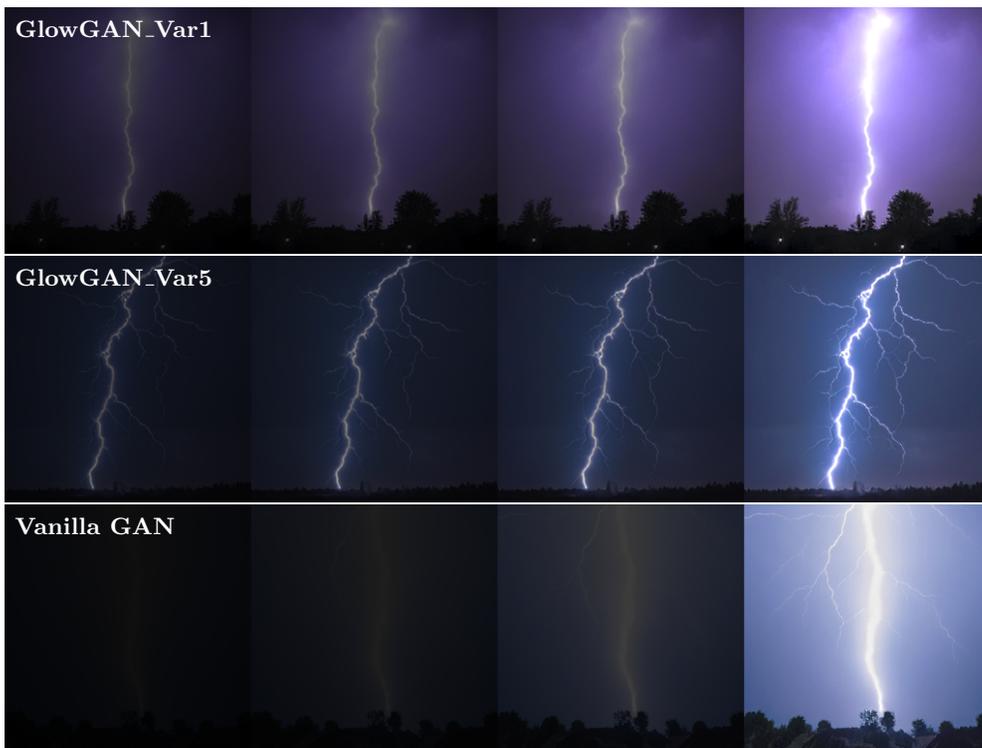


Figure 11: Lightning Sample 4. (From left to right, the exposure stops are -8, -5, -2, 0.)



Figure 12: Windows samples. We use the tone mapper in [3] for visualization.



Figure 13: Fireplaces samples. We use the tone mapper in [3] for visualization.



Figure 14: Fireworks samples. We use the tone mapper in [3] for visualization.



Figure 15: Night city samples. We use the tone mapper in [3] for visualization.

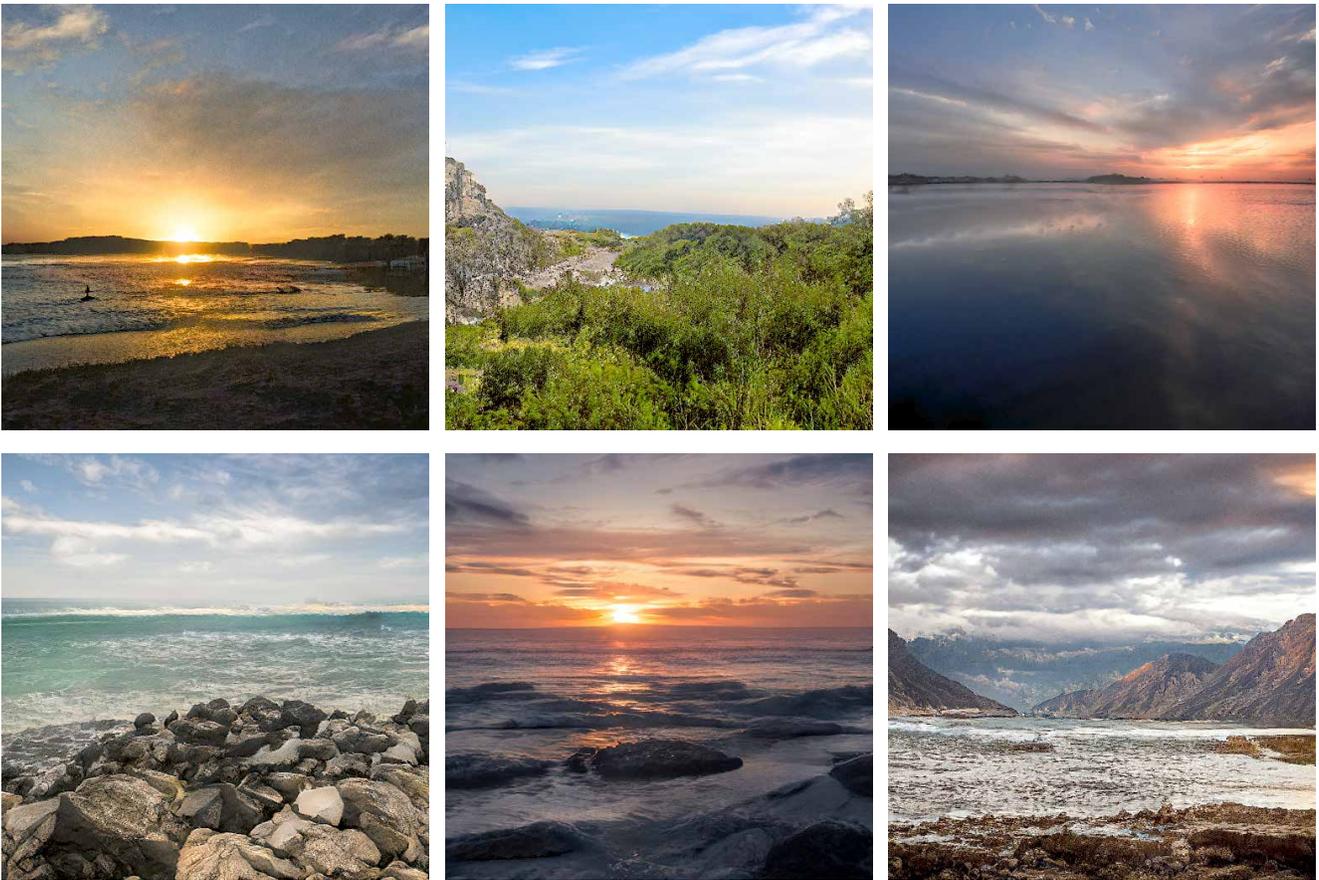


Figure 16: Landscapes samples with resolution 512×512 , we use the tone mapper in [3] for visualization.



Figure 17: Lightning samples with resolution 512×512 , we use the tone mapper in [3] for visualization.

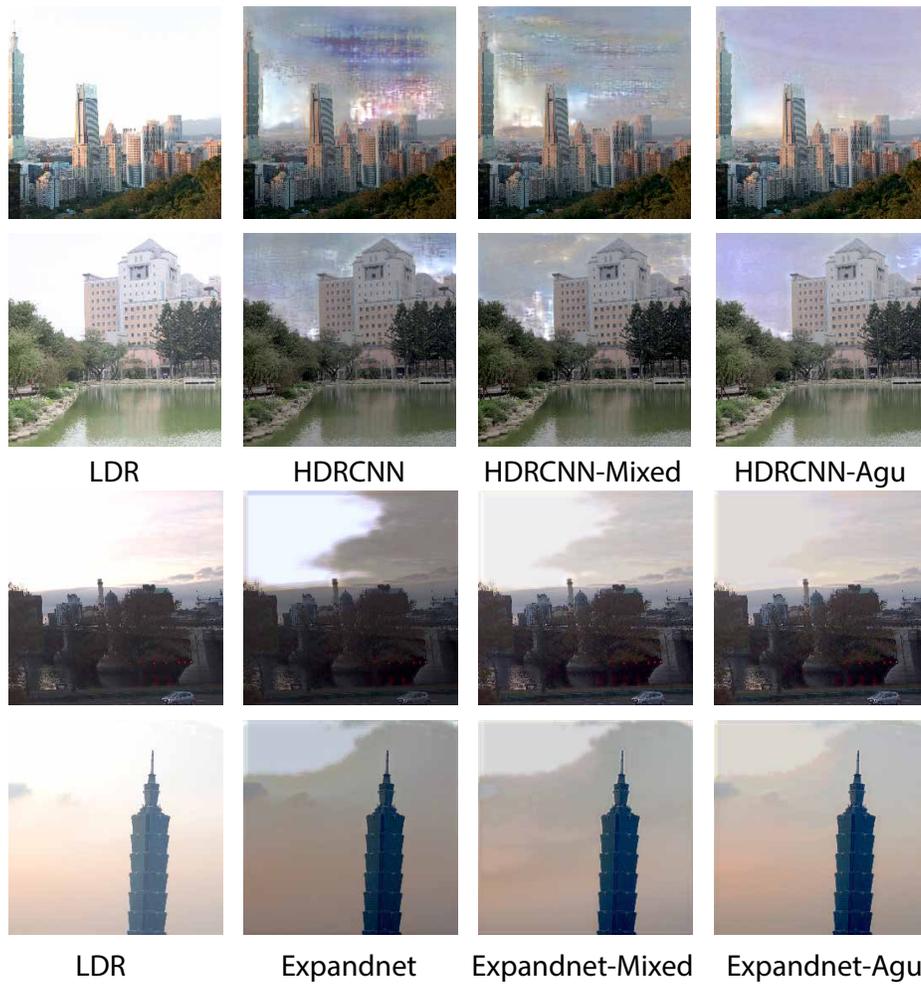


Figure 18: Visual comparison of ITM with and without GlowGAN-generated HDR data augmentation, demonstrating improved visual quality in the augmentation results.



Figure 19: Visualization of the gradual transition effects for IBL.

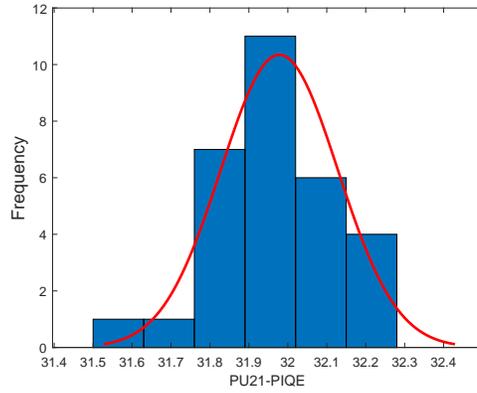


Figure 20: The distribution of PU21-PIQE score from different initial seeds with mean 31.97 and standard deviation 0.14.

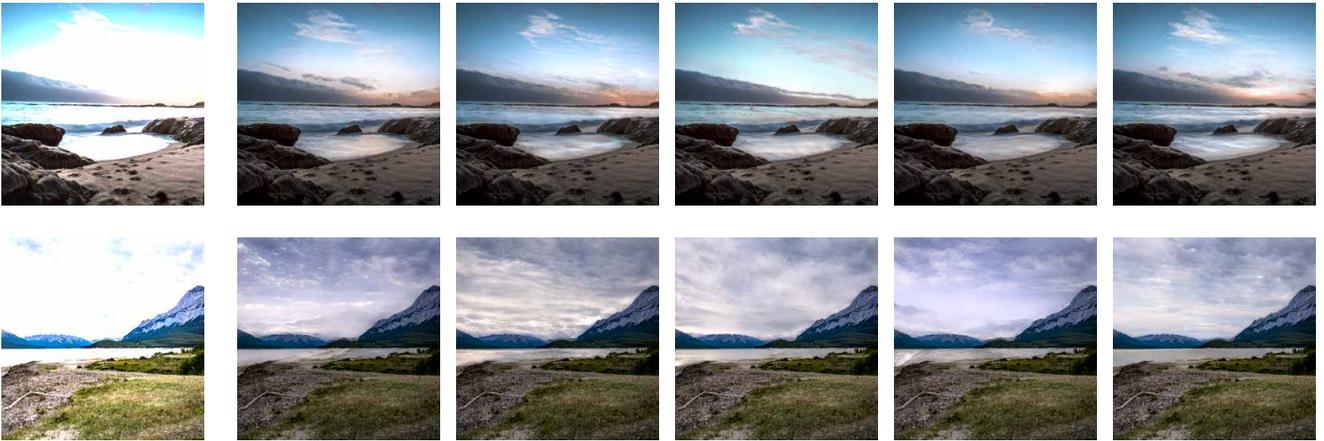


Figure 21: Given an input LDR image, our method can produce diverse ITM results using different random seeds.

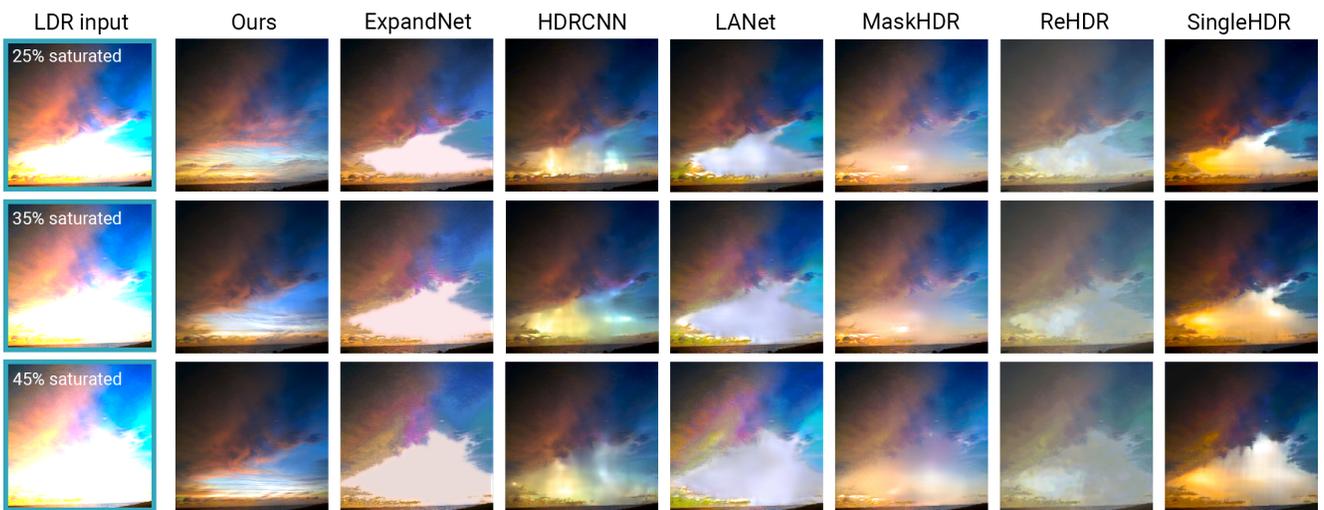


Figure 22: Here we compare the ITM results for different saturated LDRs. We can see our method is more robust and can generate plausible results for the increasing proportion of over-exposed regions.



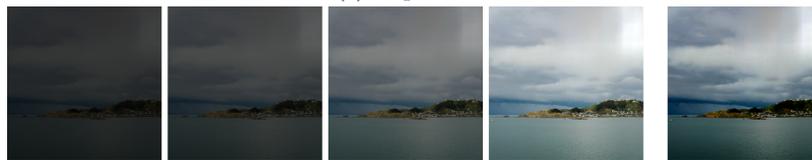
(a) LDR



(b) Ours



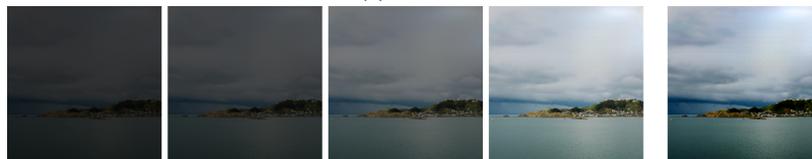
(c) Expandnet



(d) Hdrcnn



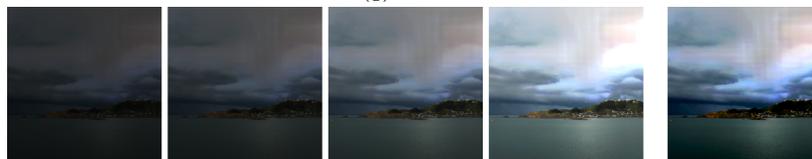
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 23: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 24: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



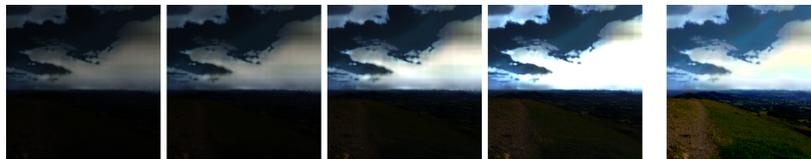
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 25: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 26: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



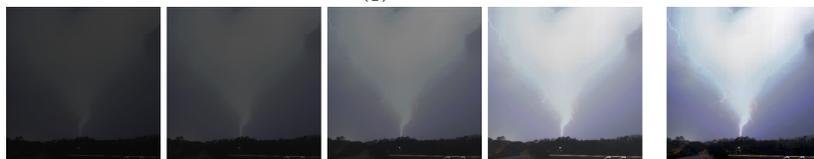
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 27: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5 , -2 , 0 , 3 , and the last is the tone-mapped result.

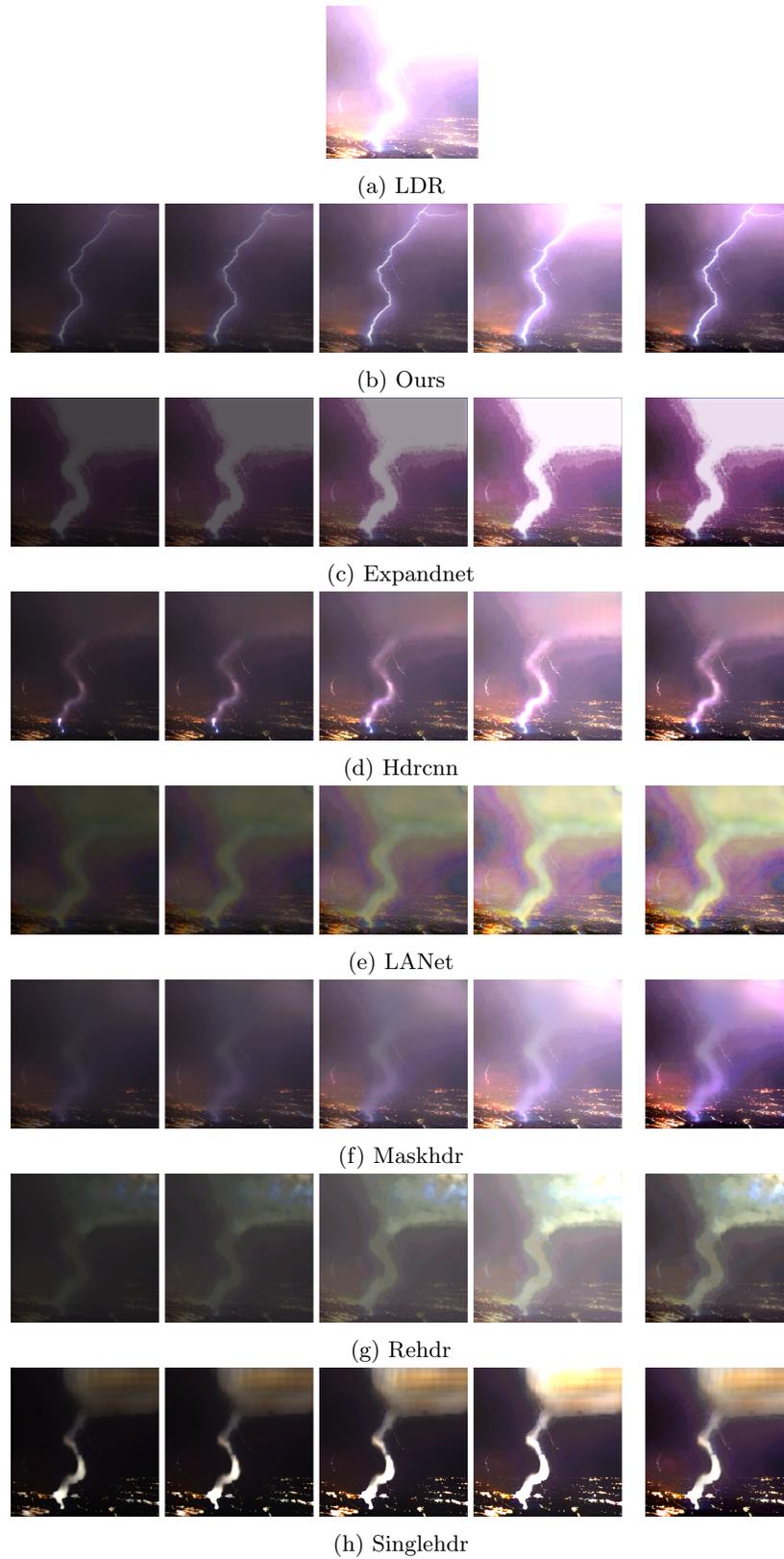


Figure 28: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5 , -2 , 0 , 3 , and the last is the tone-mapped result.

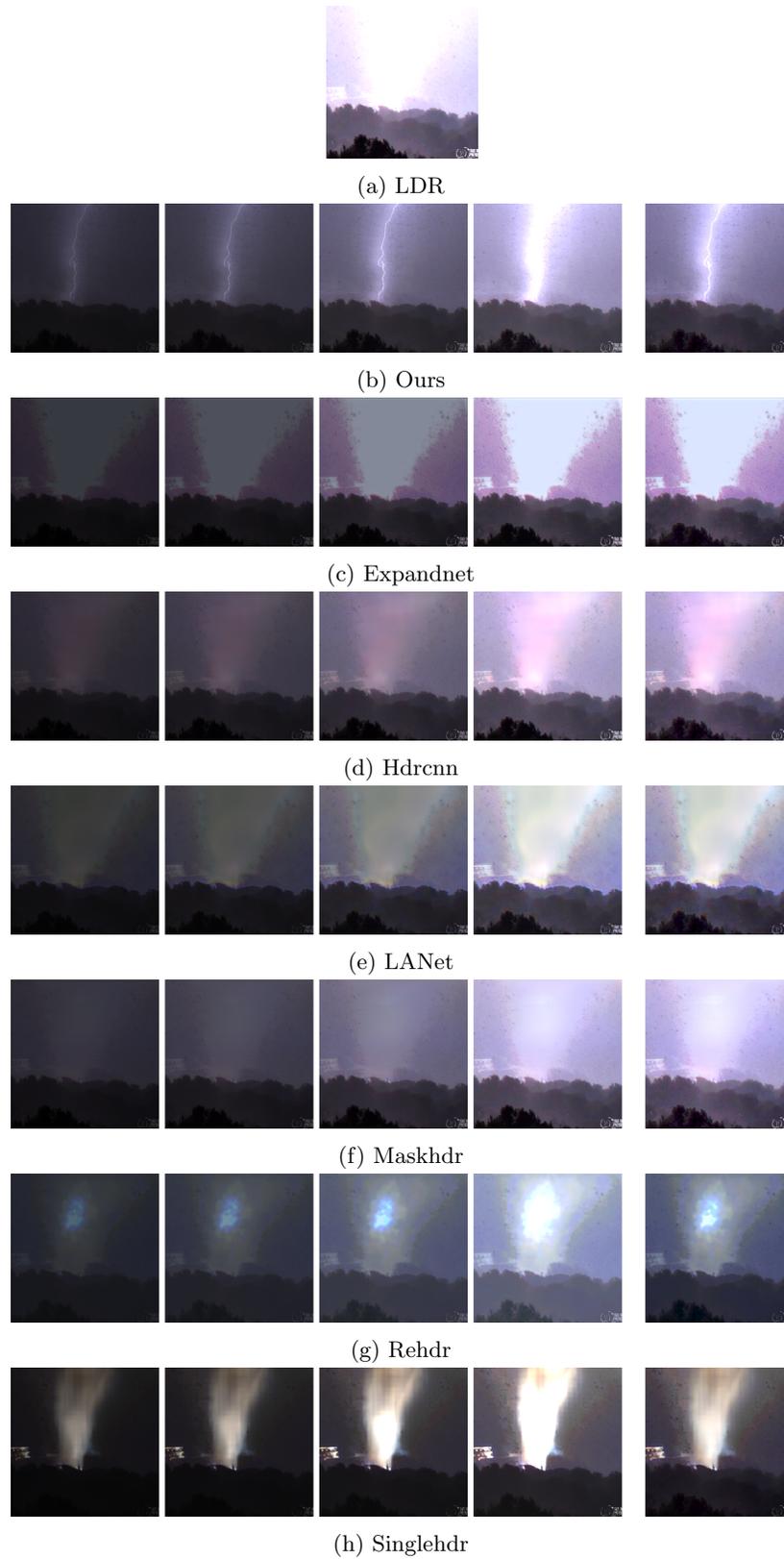


Figure 29: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5 , -2 , 0 , 3 , and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 30: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 31: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 32: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

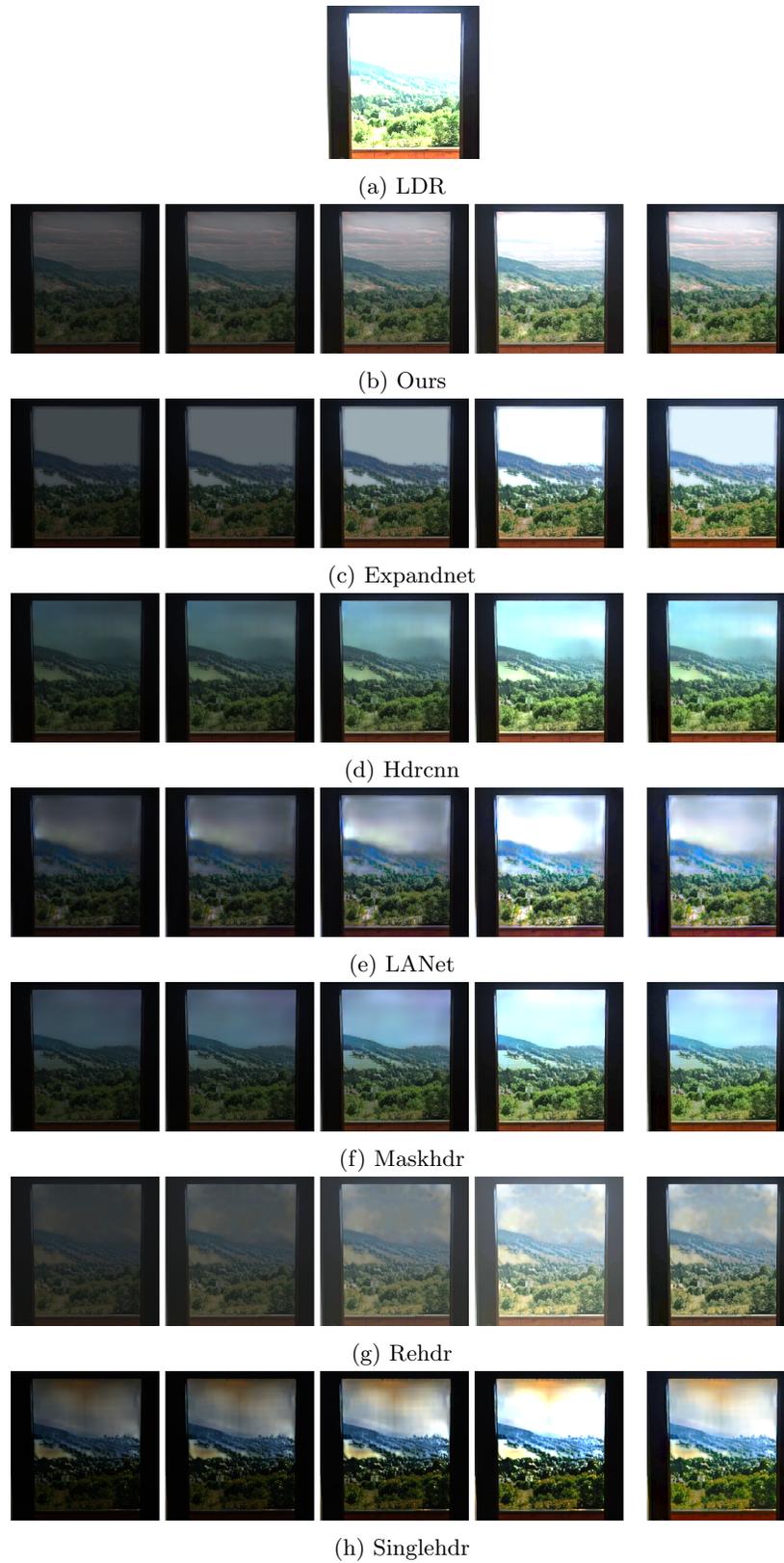
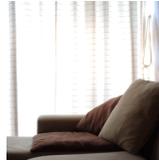


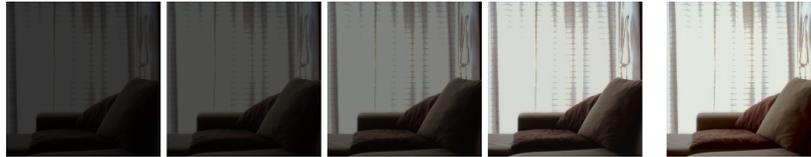
Figure 33: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



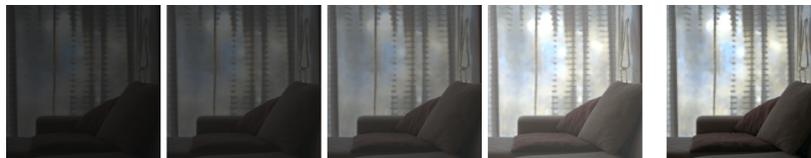
(d) Hdrcnn



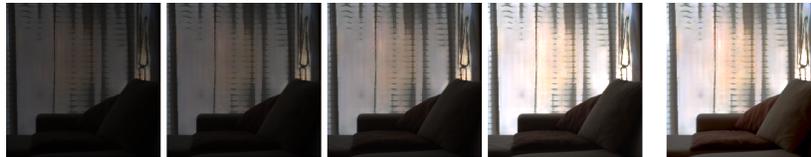
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 34: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

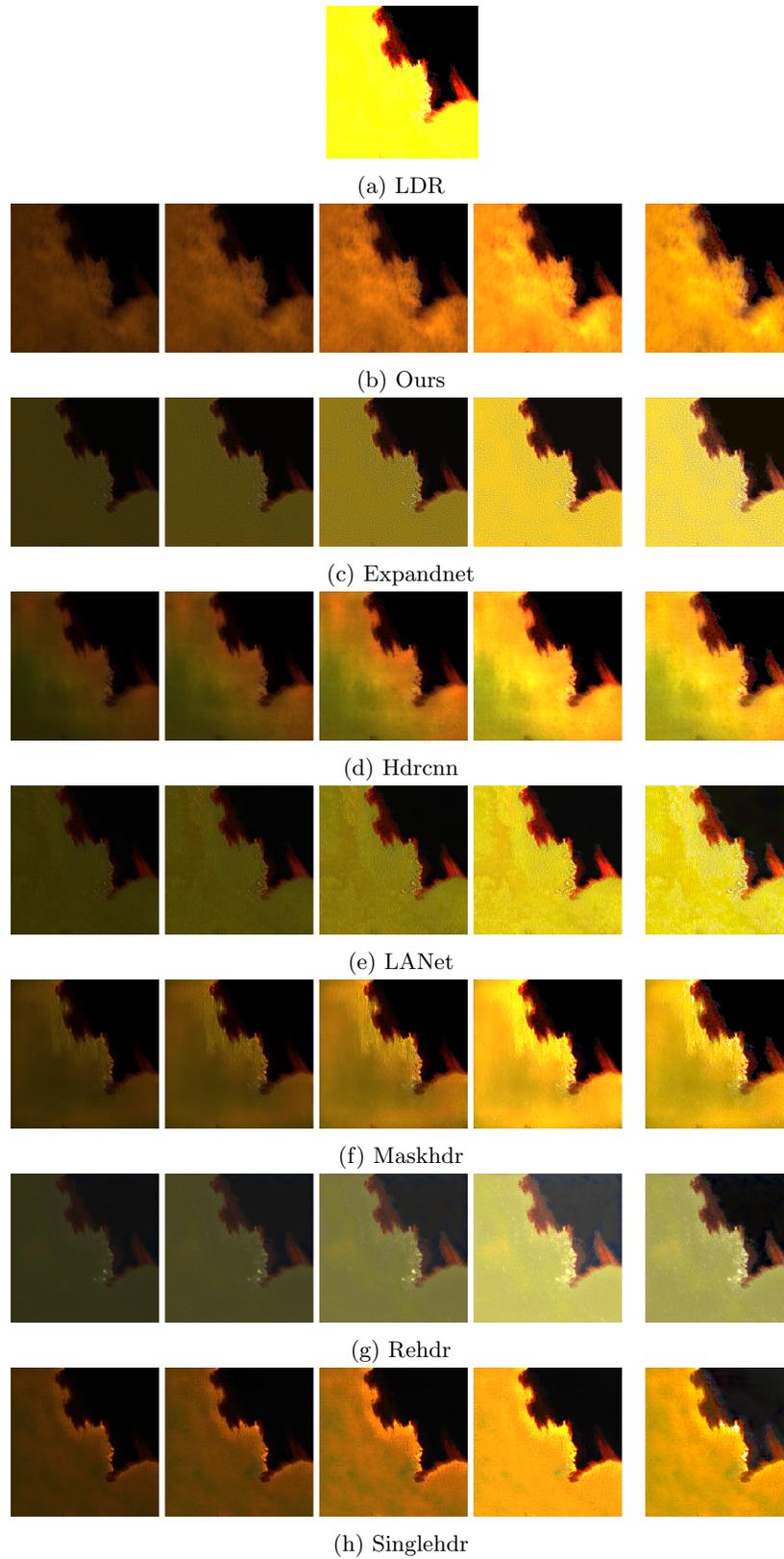


Figure 35: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

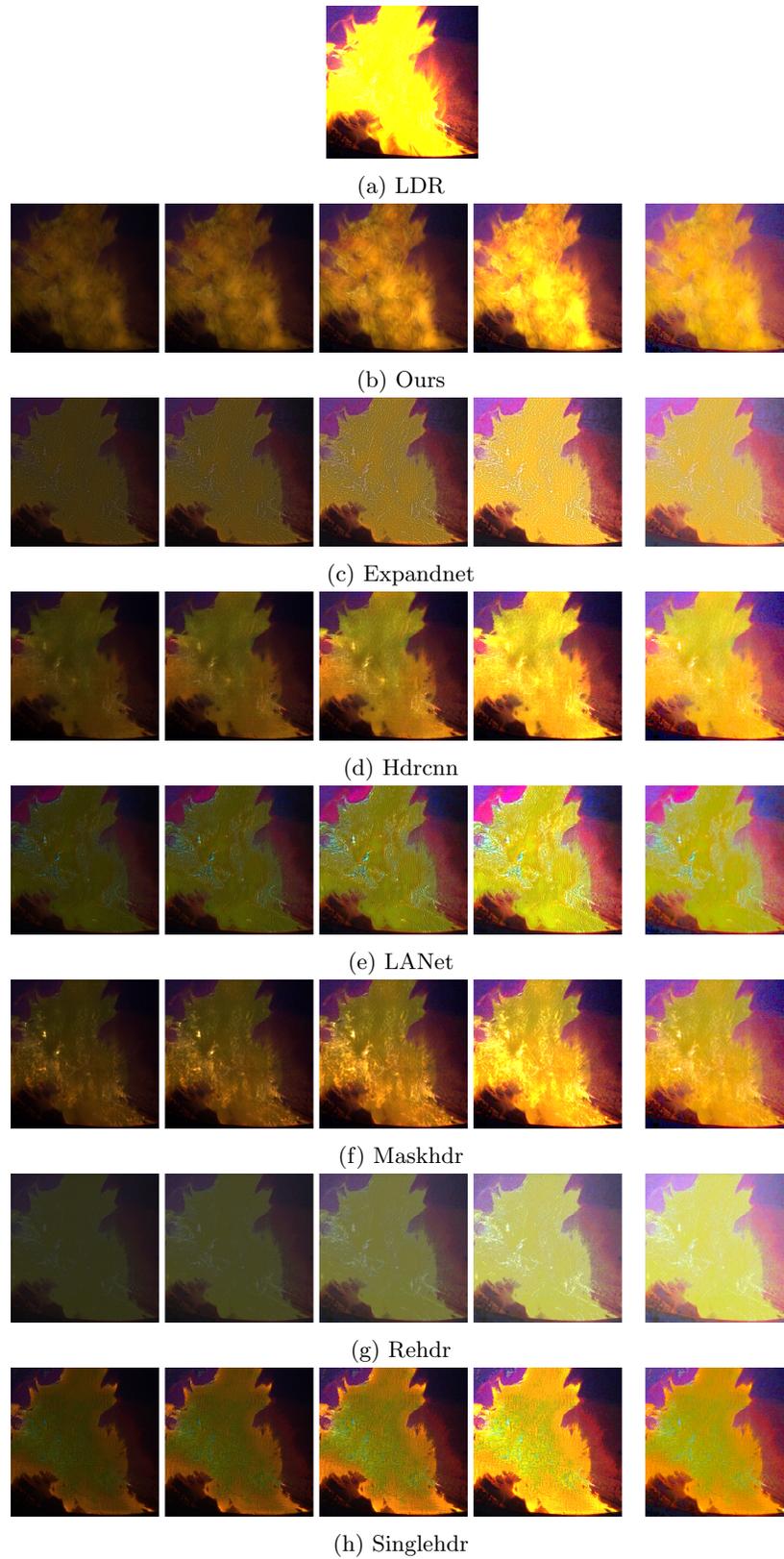


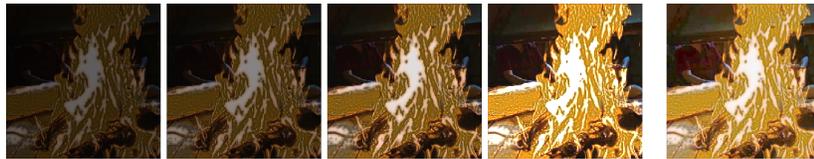
Figure 36: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



(b) Ours



(c) Expandnet



(d) Hdrcnn



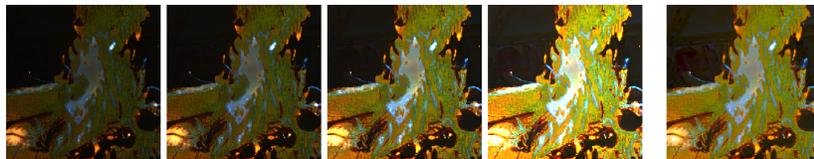
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 37: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.



(a) LDR



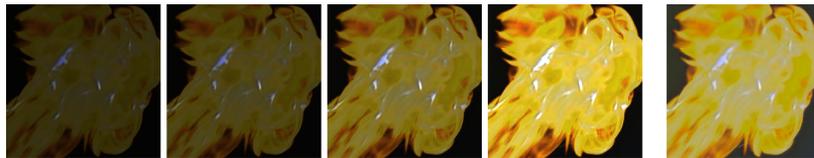
(b) Ours



(c) Expandnet



(d) Hdrcnn



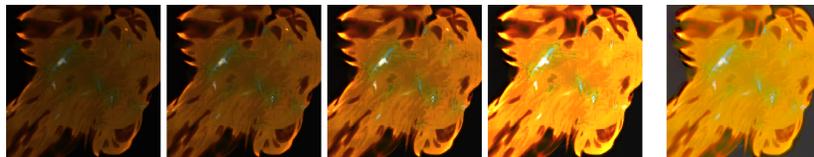
(e) LANet



(f) Maskhdr



(g) Rehdr



(h) Singlehdr

Figure 38: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

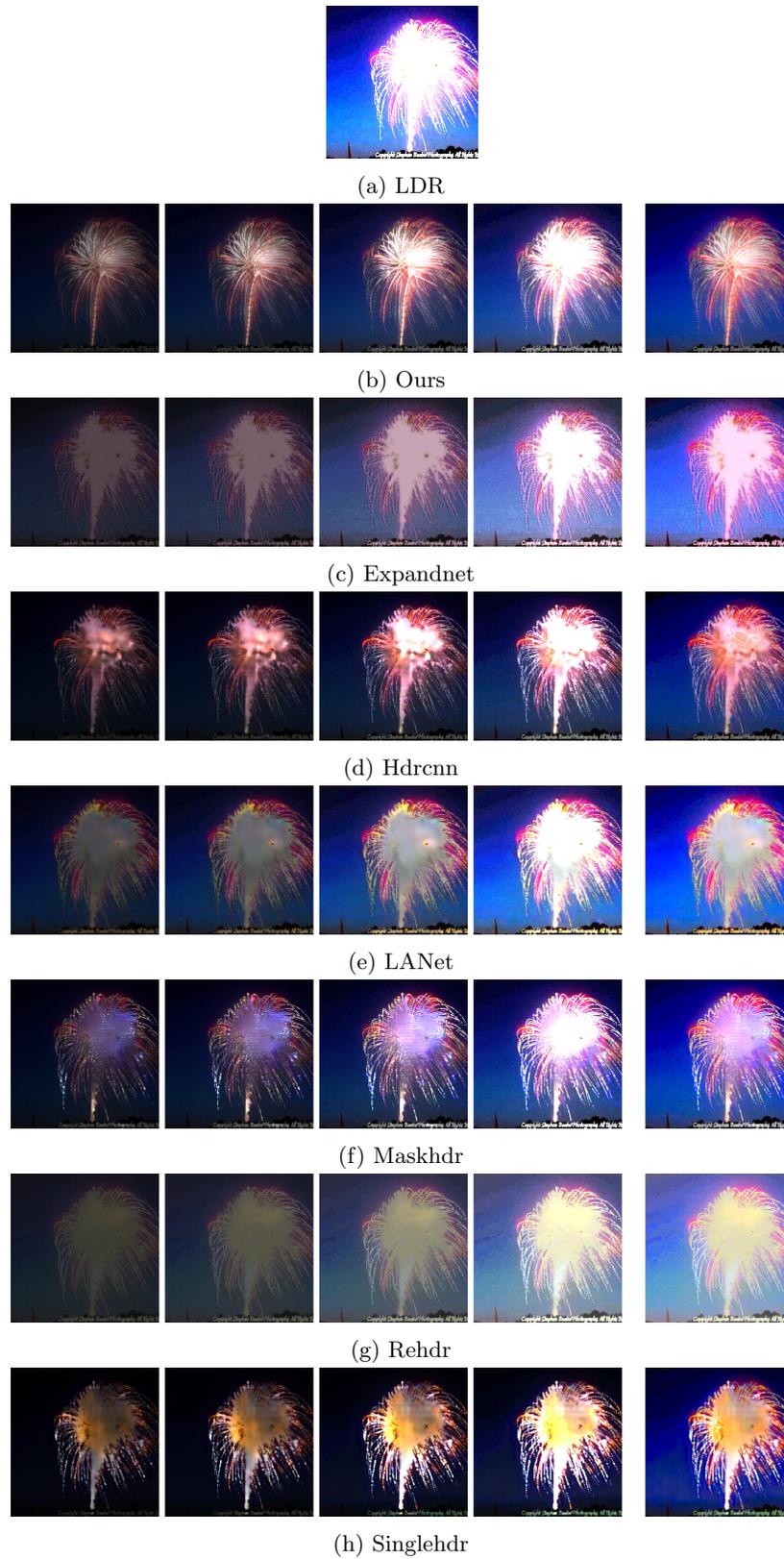


Figure 39: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

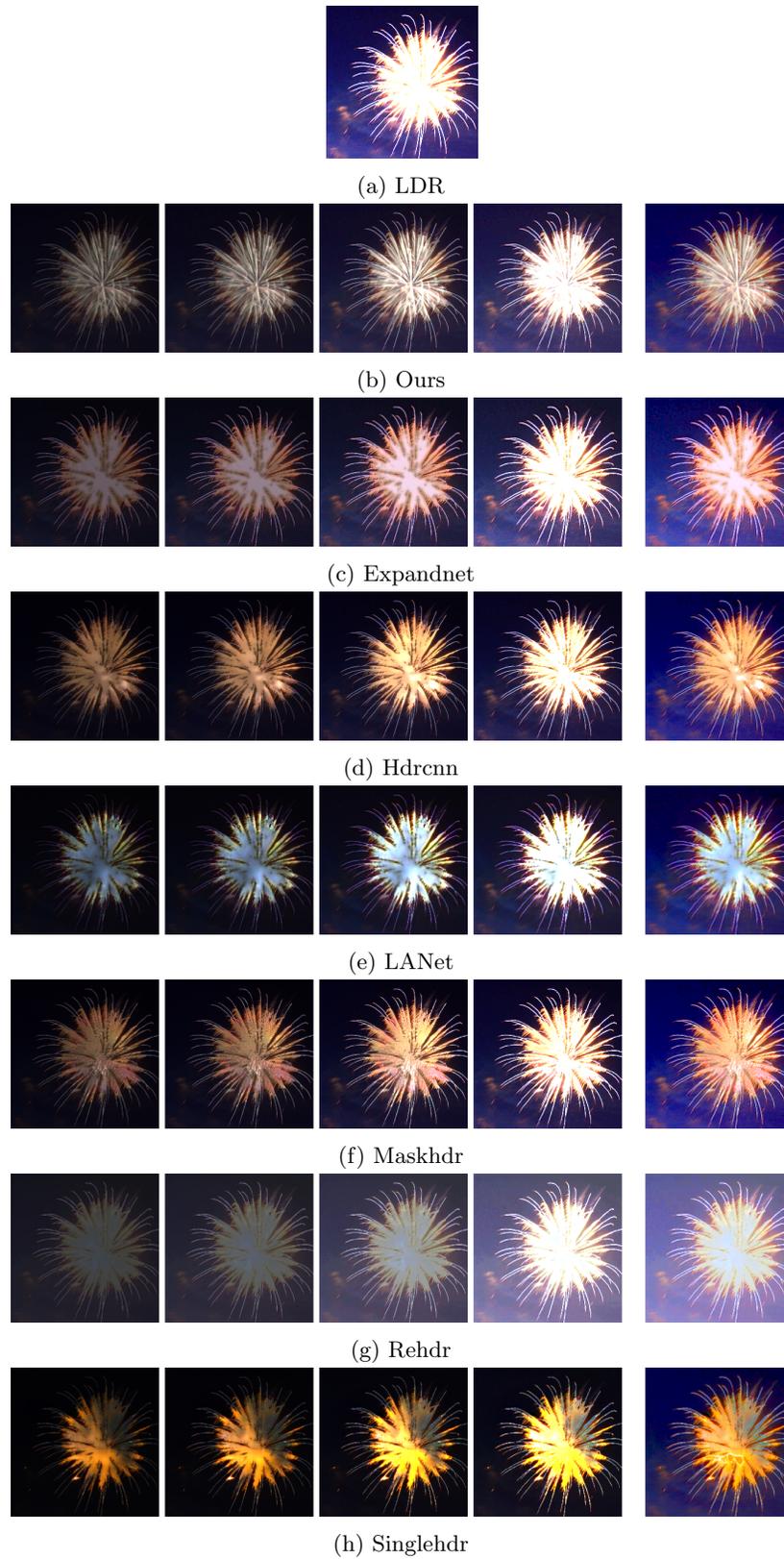


Figure 40: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

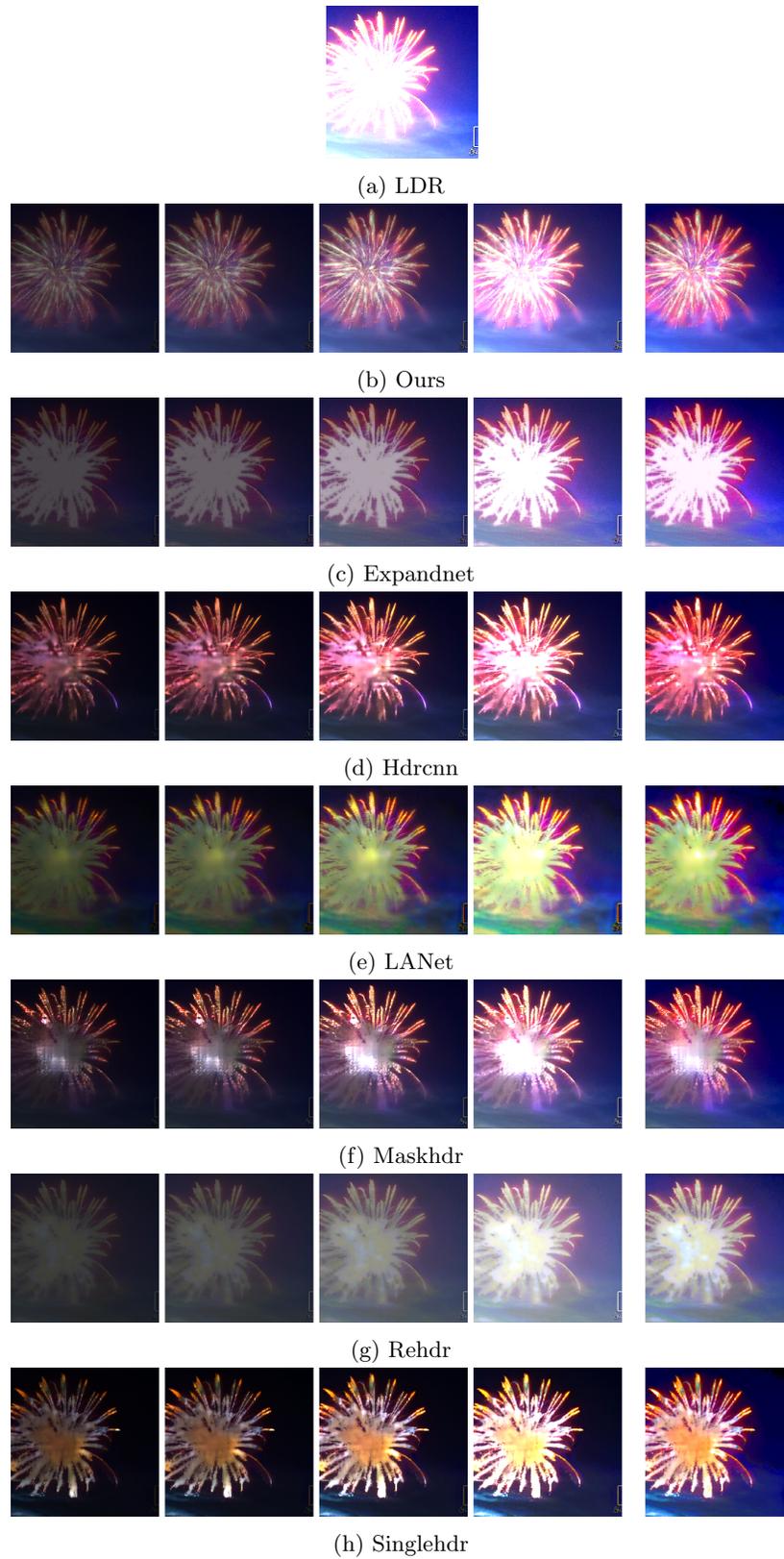


Figure 41: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

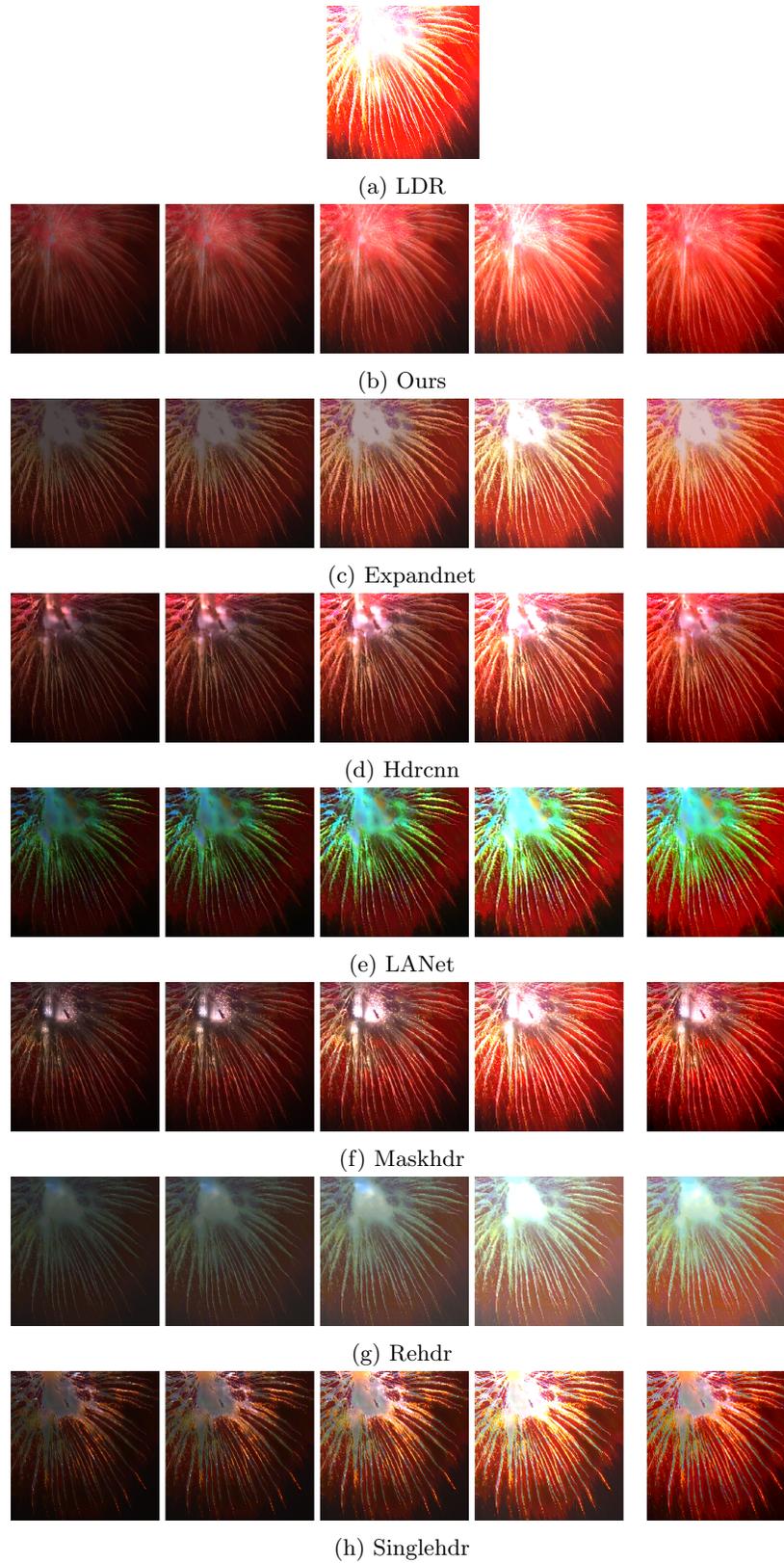


Figure 42: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

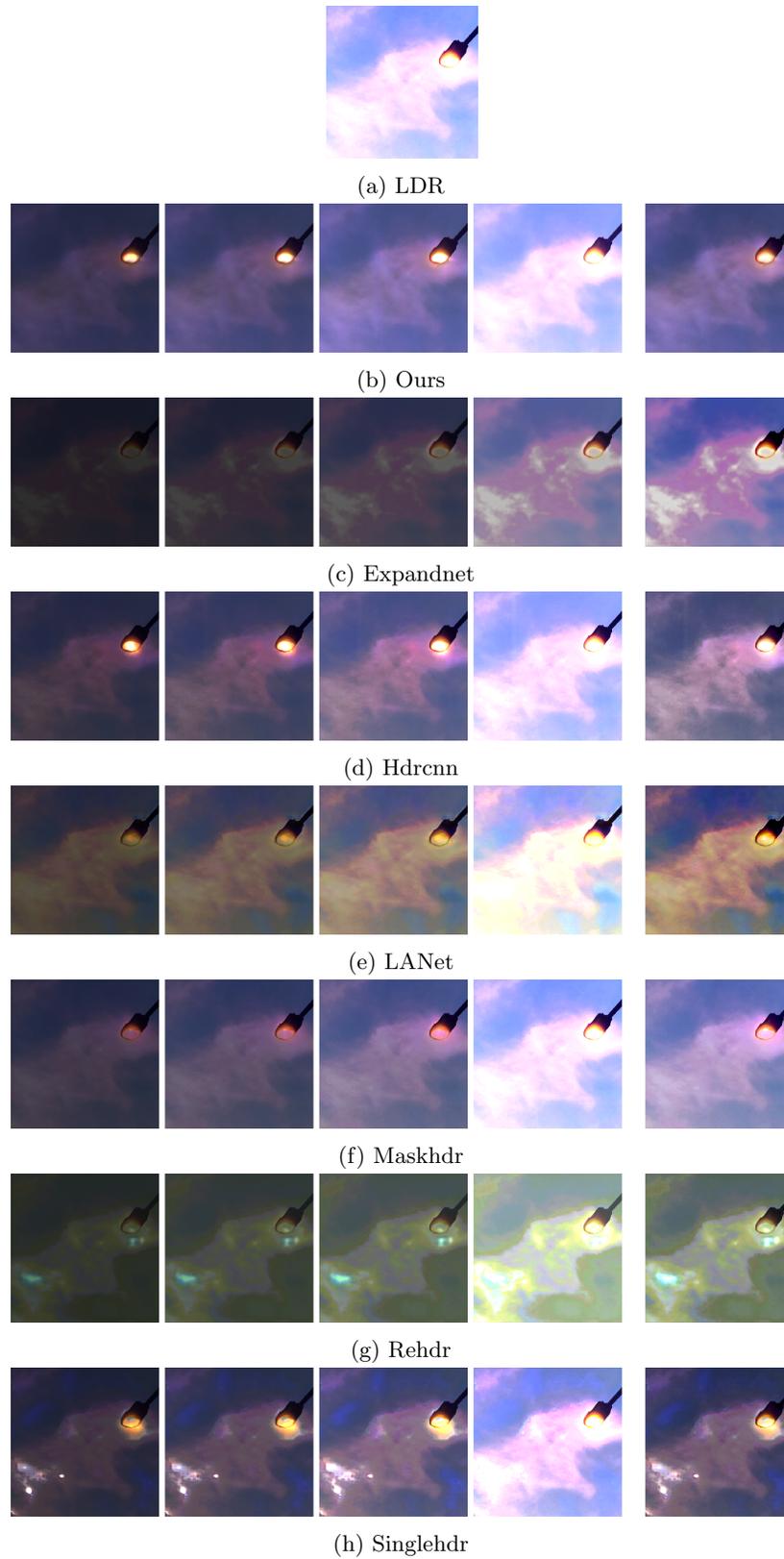


Figure 43: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.

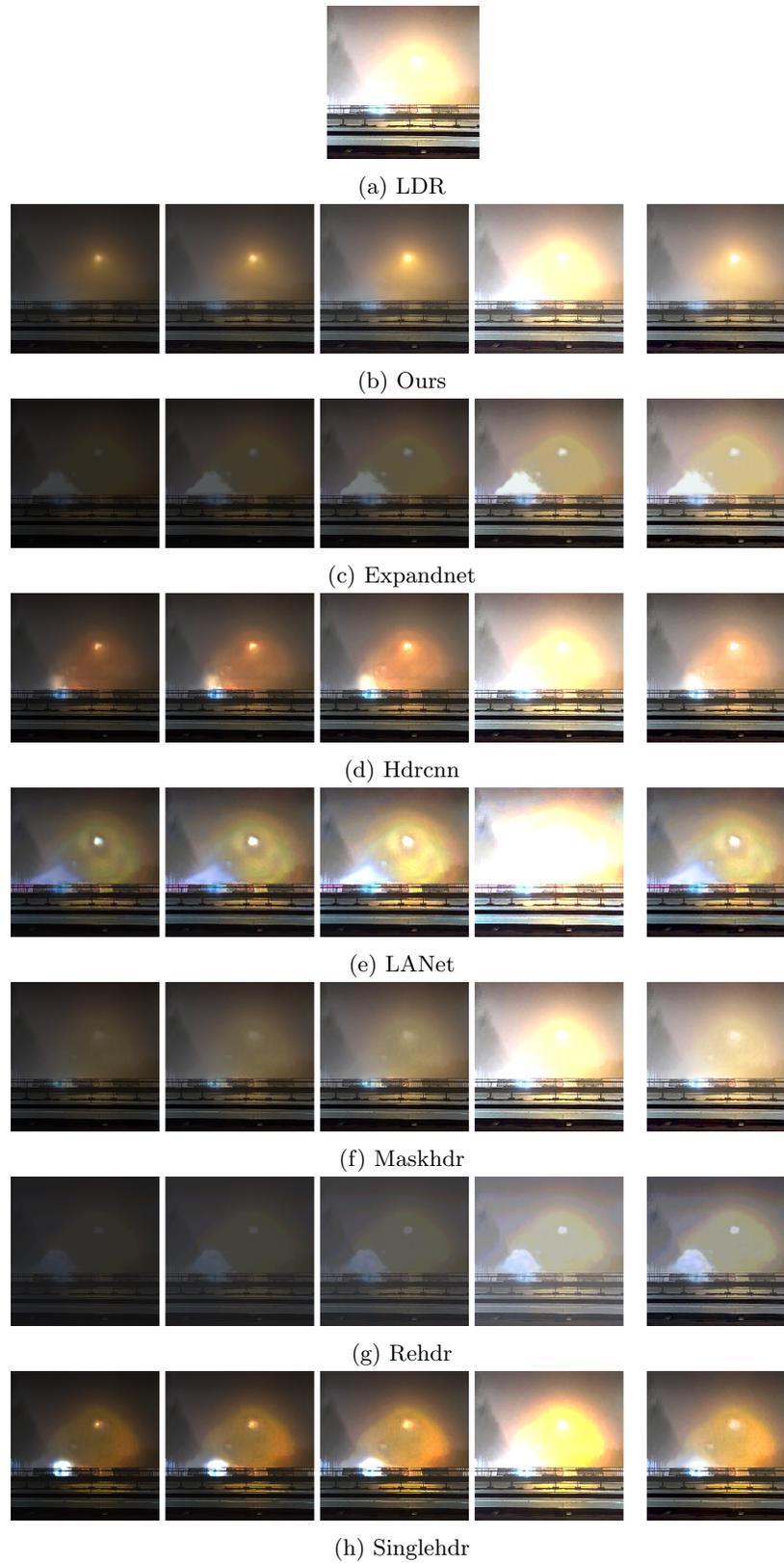


Figure 44: Visualization comparisons with other methods. From left to right, the first four images are different exposure projections of the reconstructed HDR with exposure -5, -2, 0, 3, and the last is the tone-mapped result.