LumaFlux: Structure-Aware SDR-to-HDR Conversion with Monotone Luminance Flows

A frozen flow transformer that reads measurements of the SDR frame in place of a text prompt

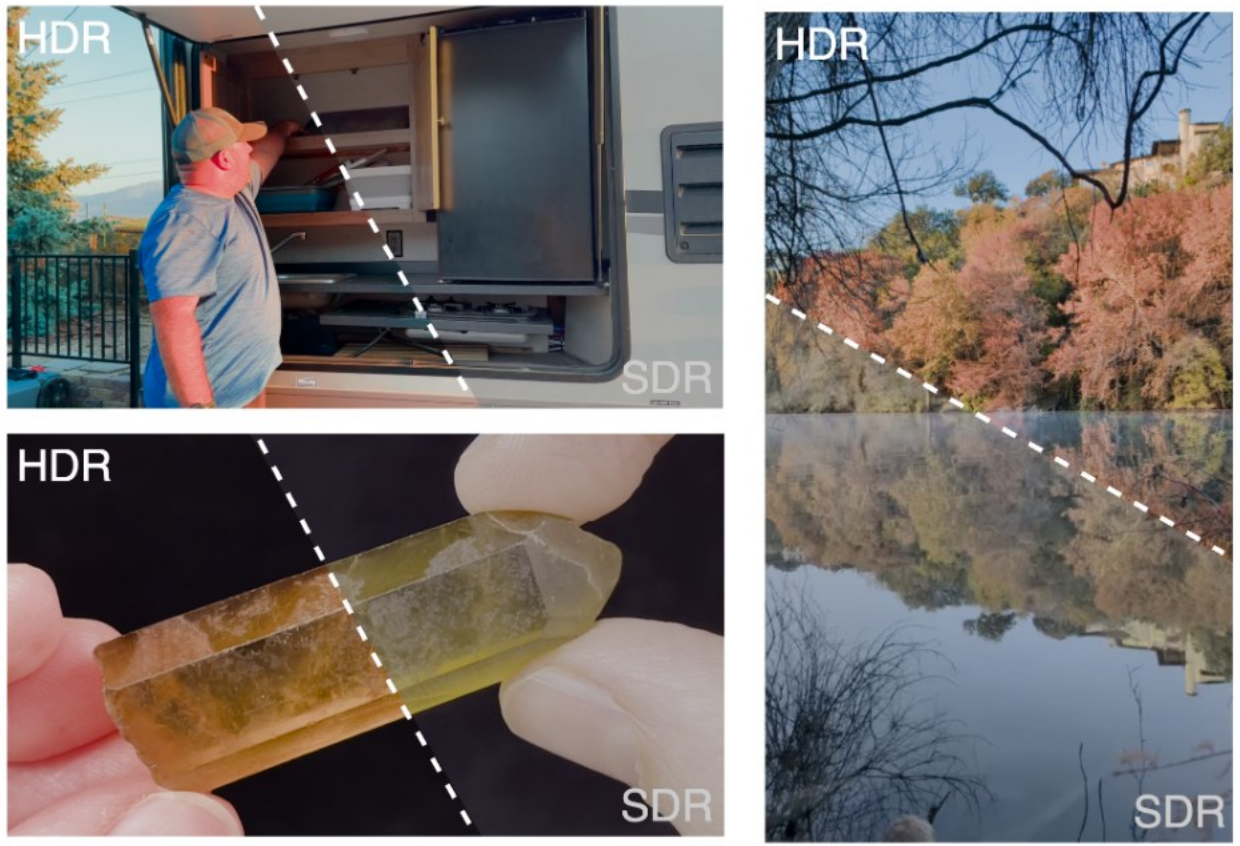
Shreshth Saini, Hakan Gedik, Neil Birkbeck, Yilin Wang, Balu Adsumilli, Alan C. Bovik

The University of Texas at Austin · Google / YouTube

Paper: arXiv:2604.02787

November 2025

What the SDR pipeline throws away



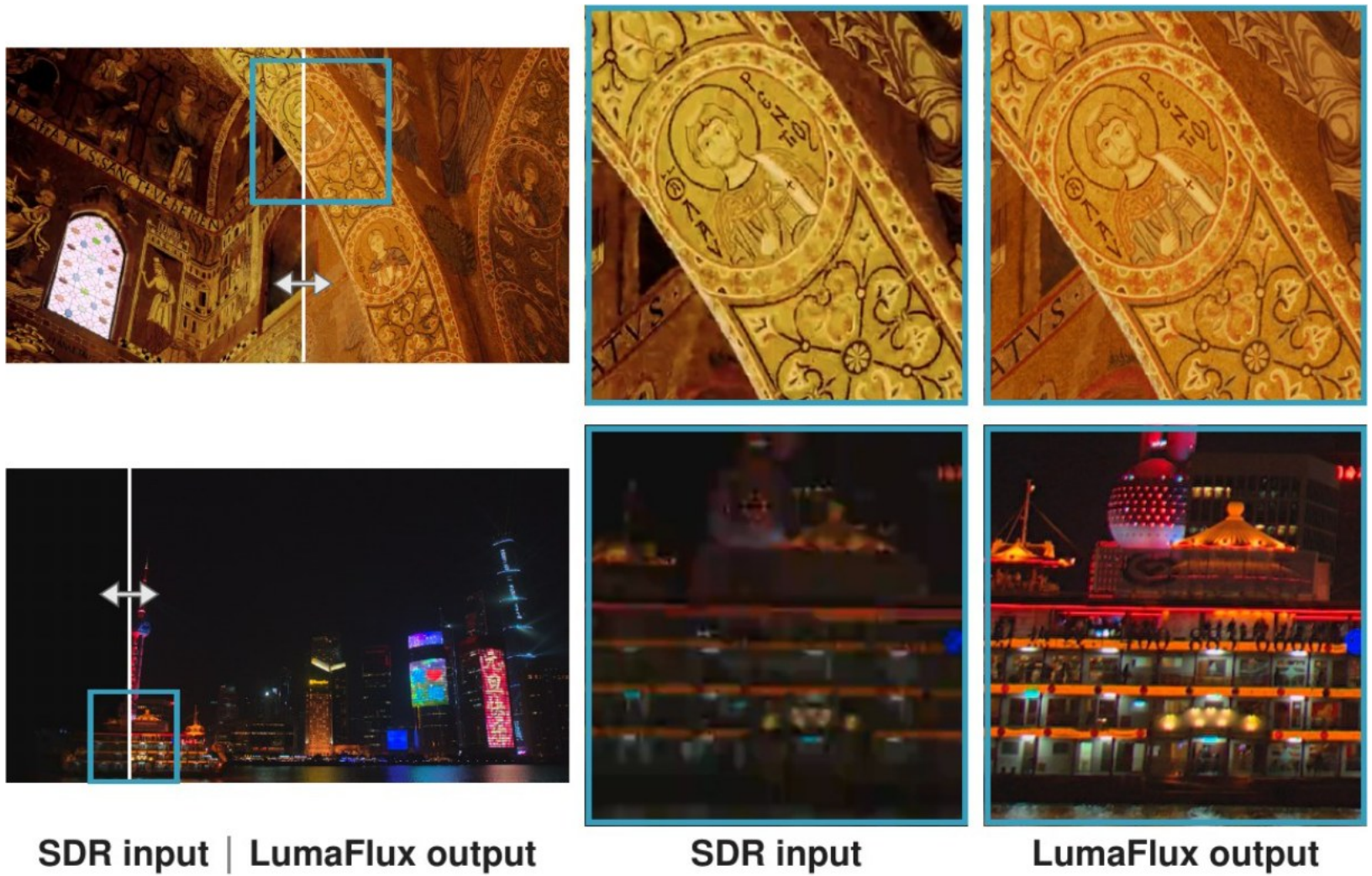
Each frame is split, HDR upper left and SDR lower right. Shadow detail, sky color and specular highlights go missing in SDR.

- Above 100 nit, light clips to white.
- BT.2020 color squeezes into BT.709, and 10 bits become 8.
- Many HDR frames share one SDR frame, so the inverse has to invent.

$$\mathbf{x}_{\text{sdr}} = \Gamma^{709}{}_{\text{OETF}}\left(\mathbf{M}_{709} \cdot \Gamma^{\text{PQ}}{}_{\text{EOTF}}(\mathbf{x}_{\text{hdr}}) / \mathbf{L}_{\text{max}} \right) + \varepsilon$$

The forward map decodes PQ (the perceptual quantizer curve), applies the BT.2020 to BT.709 gamut matrix M_{709}, encodes with the 709 curve, then quantizes and compresses (ε).

SDR input next to LumaFlux output



SDR input | LumaFlux output SDR input LumaFlux output

Left of each white line is the SDR input, right is LumaFlux; the boxed regions are enlarged at right. HDR shown tone-mapped.

Where earlier methods fall short

Analytic operators

Reinhard, BT.2446

One curve for every scene. Clipped highlights stay clipped.

Learned converters

HDRTVNet, FMNet, KUNet

Excellent in domain. They drift when the tone mapping changes.

Generative priors

LEDiff, X2HDR

A real structure prior, but color and tone drift.

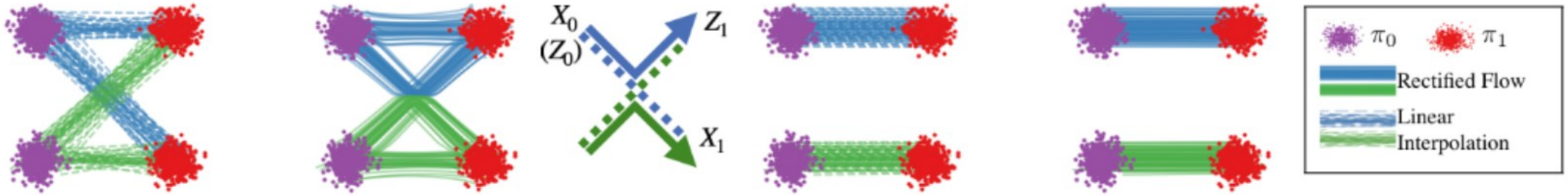
We wanted that prior, pinned to what the SDR frame measures.

Part 1 of 2

How LumaFlux works

What has to change in a frozen SDR image model before it can output calibrated HDR?

LumaFlux starts the flow at the SDR frame



(a) Linear interpolation $X_t = tX_1 + (1-t)X_0$ (b) Rectified flow Z_t induced by (X_0, X_1)

(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$ (d) Rectified flow Z'_t induced by (Z_0, Z_1)

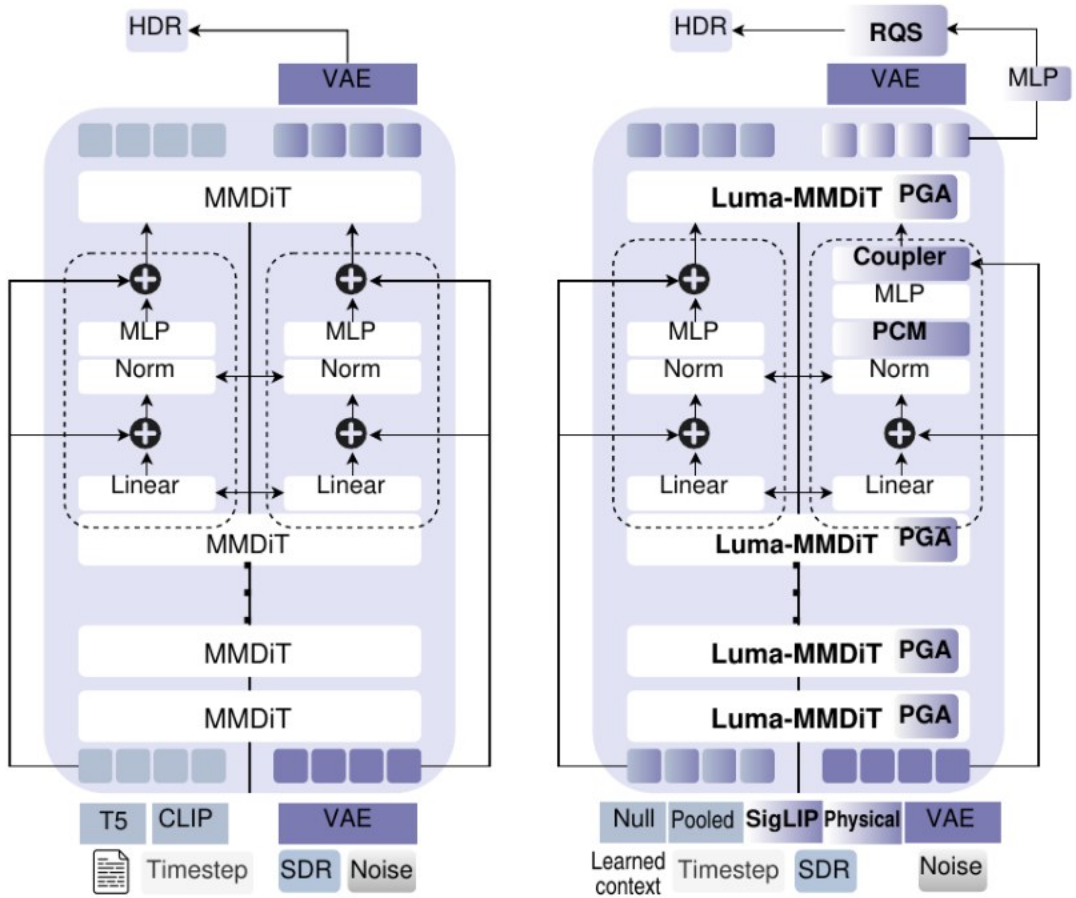
Rectified flow straightens the paths between paired samples, so a few Euler steps follow them. Figure from Liu et al., Flow Straight and Fast.

$$z_1 = z_{SDR} + 0.05\ \varepsilon$$

The encoded SDR frame plus a small perturbation, ε ~ N(0, I).

- Train on the straight path from the HDR latent to z_1.
- Sample in 8 Euler steps, with no strength knob.
- 40 steps cost 4.7× the time and calibrate no better.

LumaFlux in one picture



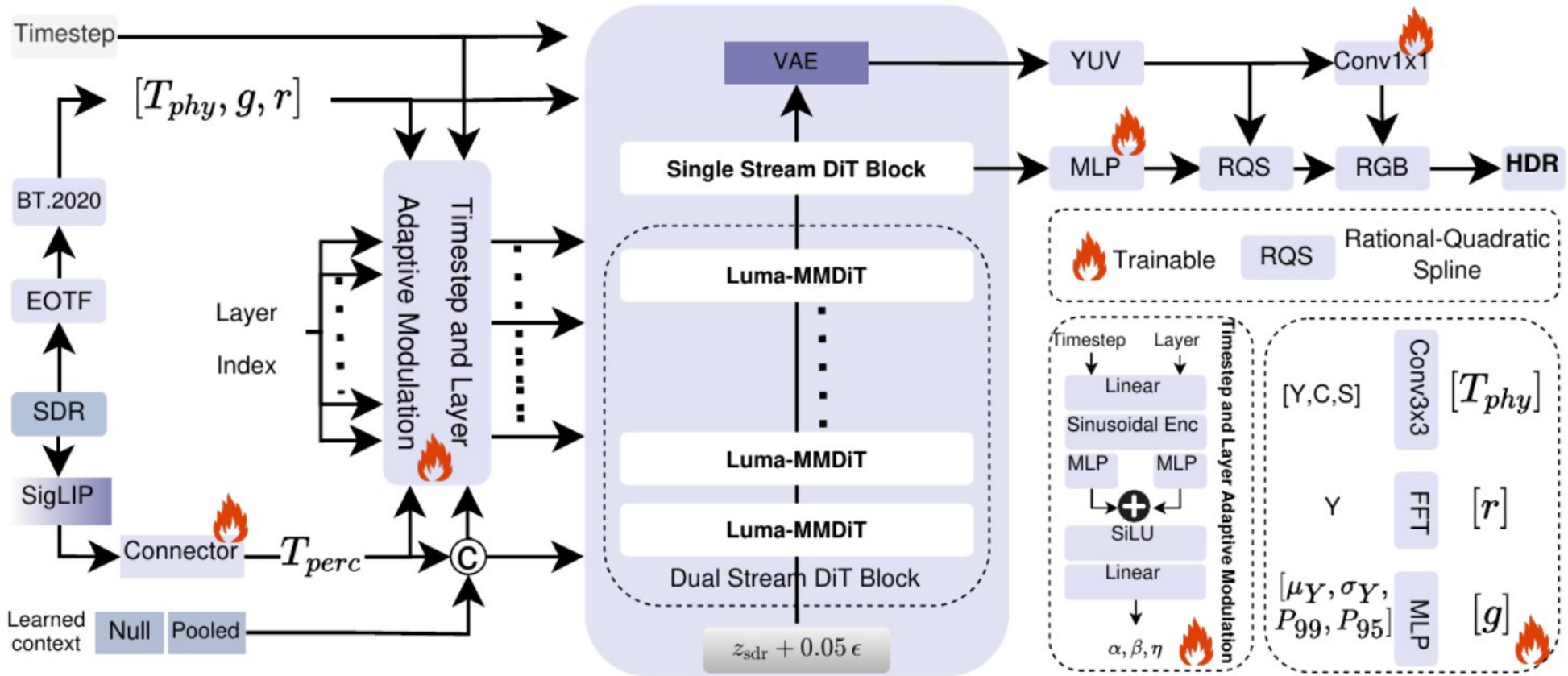
Left: adapting the MM-DiT directly, driven by a text prompt. Right: LumaFlux, with the MM-DiT and VAE frozen.

- **Frozen** FLUX.1-dev MM-DiT, VAE and SigLIP. No text encoder.
- **PGA**, Physically-Guided Adaptation, a LoRA gated by measurements.
- **PCM**, Perceptual Cross-Modulation, and a residual coupler add scene context.
- **RQS**, a monotone rational-quadratic spline, calibrates luma.

0.57%

of 12.48B weights trained, 71.3M parameters

Measurements of the SDR frame steer every frozen block



All 57 MM-DiT blocks are wrapped (19 dual-stream, 38 single-stream, d = 3072, 24 heads). Flames mark the trainable parts.

Per token
luminance, gradient, saturation

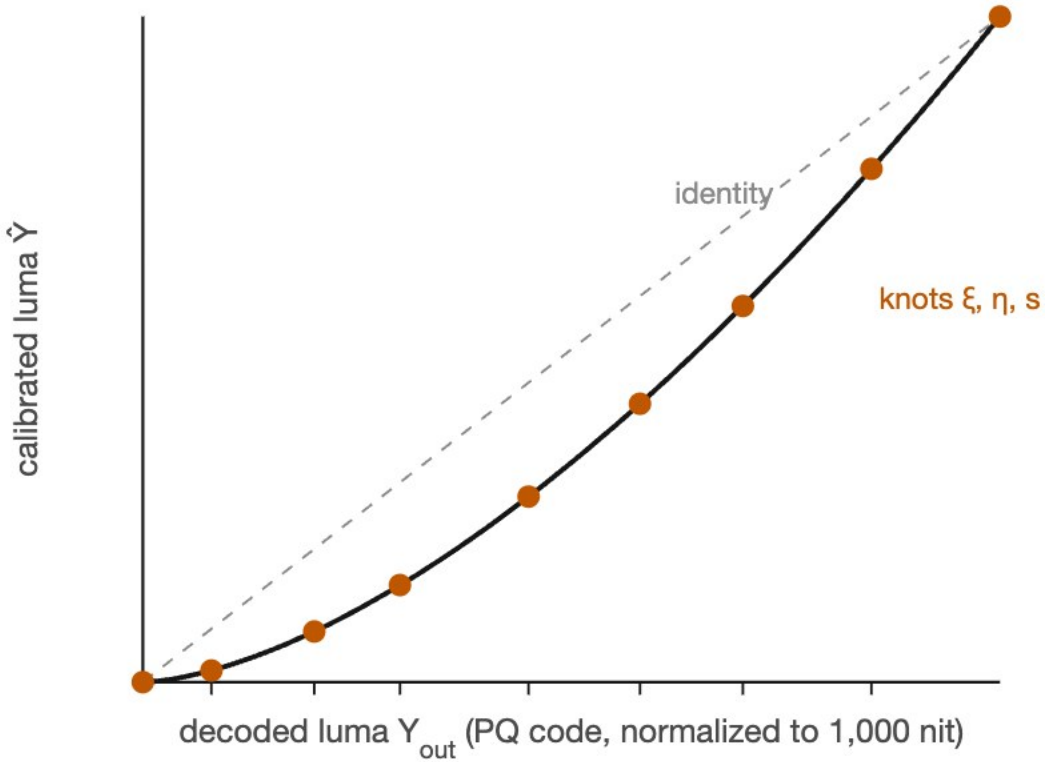
Per frame
luminance statistics, 8 FFT bands

Per block and step
six gains from the conditioner Ψ(t, l)

A monotone spline turns the VAE output into HDR luma

$$\hat{Y} = \mathrm{RQS}(Y_{out};\, \xi,\, \eta,\, s)$$

- A rational-quadratic spline (RQS) with 8 bins, **25 parameters per frame**.
- Widths, heights and slopes are positive, so luma keeps its ordering.
- Removing it costs 9.69 dB. A linear spline costs 0.45 dB.



8 widths ξ, 8 heights η and 9 knot slopes s, from a 21.9K-parameter head that starts at the identity (dashed).

Part 2 of 2

Data, benchmark, results

Does it hold up when every method gets the same inputs and the same scoring code?

A training corpus of user-generated and professional HDR

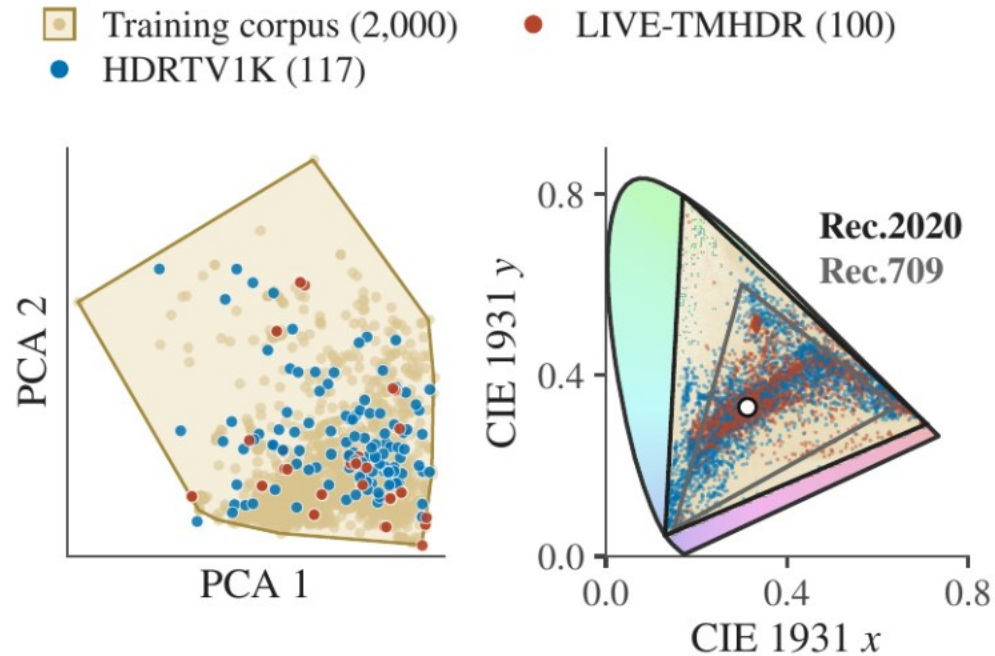
314,396
SDR-HDR pairs

2,092
source videos

171,384 / 143,012
UGC / PGC examples

HDR source	What it adds
CHUG (Saini et al., ICIP 2025)	856 user-generated references, 846 for training and 10 reserved
LIVE-TMHDR	40 references with their 120 expert-graded SDR clips
HDRTV1K	1,235 training pairs
Sol Levante	HDR10 master, Netflix Open Content (CC BY)

- Missing SDR sides are synthesized with eight tone-mapping operators at three compression levels.
- A manifest regenerates everything in about four CPU hours.



(a) PCA of ten frame statistics. The corpus hull holds all 100 LIVE-TMHDR frames and 116 of the 117 HDRTV1K test frames. (b) CIE xy footprint.

LumaEval scores every method the same way

Track	Pairs	What is scored	Checkpoint
Synthetic (primary)	100	20 held-out CHUG and LIVE-TMHDR contents, eight tone-mapping variants, 78 video-variant groups	mixed-corpus
Expert	100	LIVE-TMHDR pairs whose SDR is an expert grade	mixed-corpus
Native	117	HDRTV1K test pairs	mixed-corpus
Legacy HDRTV1K	117	same pairs, five-metric suite, published baseline values	HDRTV1K-specific
Fixed-QP	468	intra-only HEVC on the SDR input at QP 27, 32, 37, 42	both
Temporal	5 clips	Bonfire and four CHUG clips, PU21 flicker energy	mixed-corpus
Text-to-HDR	6 clips	LTX-Video clips, 121 frames each, intrinsic metrics only	mixed-corpus

Outputs are PNG16 PQ BT.2020 at 1,000 nit. Every row is scored with identical inputs, outputs and metric code.

Degradations

Eight tone-mapping variants, from BT.2390 to YouTube LogC.

Metrics

PU21-PSNR and SSIM, ΔE_{ITP} color error, HDR-LPIPS, HDR-VDP-3, FR-HIDROVQA.

FR-HIDROVQA is a full-reference distance on HIDRO-VQA features (Saini et al., WACV 2024).

Baselines

Two analytic operators, four CNNs, PromptIR, LEDiff, X2HDR, and a VAE + RQS control.

Best in every column of the synthetic track

Method	PU21-PSNR ↑	PU21-SSIM ↑	ΔE_{ITP} ↓	HDR-LPIPS ↓	HDR-VDP-3 ↑	FR-HIDRO ↓
BT.2446c	<u>23.74</u>	<u>0.8291</u>	<u>61.60</u>	0.307	5.611	0.753
Reinhard operator	23.62	0.8286	61.76	<u>0.295</u>	<u>5.726</u>	0.753
HDRTVNet++	23.03	0.8192	66.67	0.328	5.417	<u>0.703</u>
FMNet	22.86	0.8176	67.92	0.330	5.412	0.724
ITM-LUT	22.92	0.8142	67.41	0.315	5.496	0.761
VAE + RQS (control)	18.54	0.7647	111.67	0.381	3.673	0.924
LumaFlux (ours)	**24.23**	**0.8294**	**56.54**	**0.269**	**5.812**	**0.631**

Macro-averages over the same 100 frame IDs and 78 video-variant groups. PU21-PSNR-Y is in the backup copy. Bold best, underline second best. FR-HIDRO is the HIDRO-VQA feature distance (Saini et al., WACV 2024).

24.23 dB
PU21-PSNR, 0.49 dB above the strongest baseline

56.54
ΔE_{ITP}, 5.06 lower than the strongest baseline

0.631
FR-HIDROVQA, from 0.703 for HDRTVNet++

One checkpoint, no per-track tuning. Without diffusion transport the control collapses to 18.54 dB.

On human-graded SDR, LumaFlux trails only on ΔE_{ITP} and HDR-VDP-3

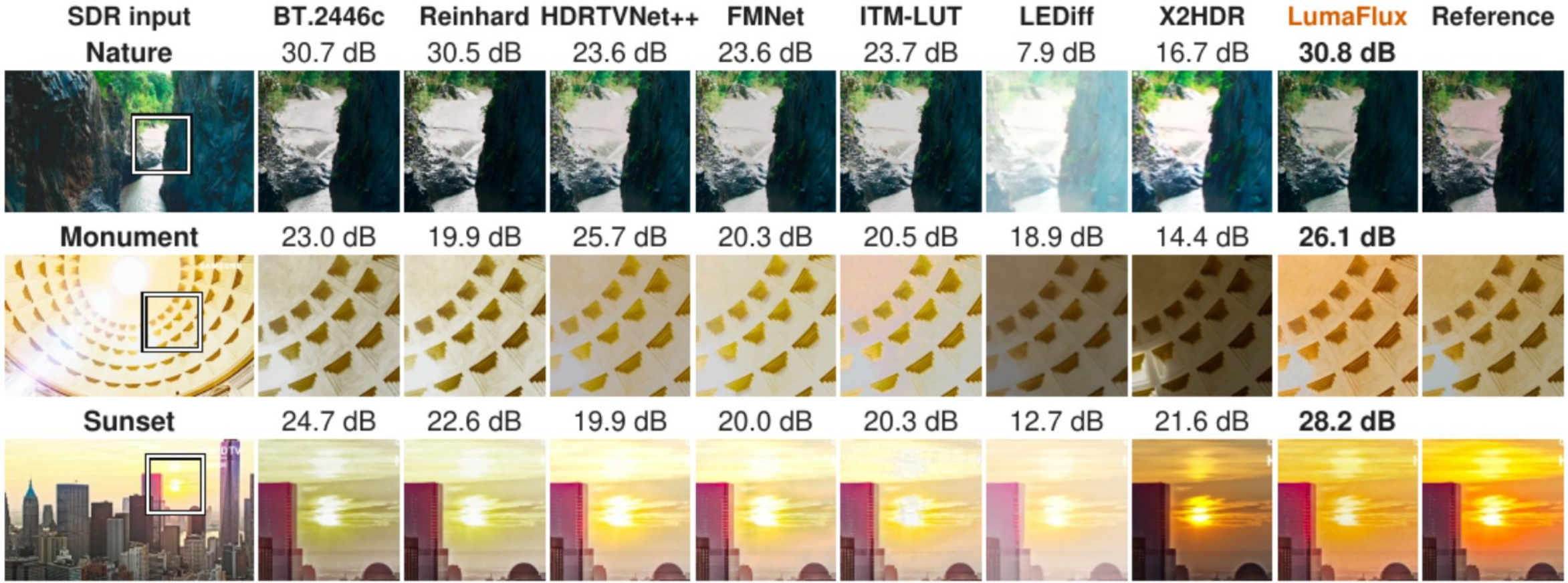
Method	PU21-PSNR ↑	PU21-SSIM ↑	ΔE_{ITP} ↓	HDR-LPIPS ↓	HDR-VDP-3 ↑	FR-HIDRO ↓
BT.2446c	<u>33.16</u>	<u>0.8442</u>	**37.83**	0.148	**7.397**	0.660
Reinhard operator	31.48	0.8270	43.59	0.155	<u>7.173</u>	0.662
HDRTVNet++	31.53	0.8125	44.92	0.147	7.144	0.671
ITM-LUT	31.46	0.8085	45.49	<u>0.142</u>	7.090	<u>0.639</u>
LEDiff	18.30	0.6480	103.15	0.496	4.723	0.898
X2HDR	23.78	0.7578	51.28	0.366	5.532	0.758
LumaFlux (ours)	**33.31**	**0.8480**	<u>40.19</u>	**0.138**	6.919	**0.544**

The 100 LIVE-TMHDR pairs whose SDR side is an expert grade, with the preregistered per-content exposure calibration. Bold best, underline second best. The full table, with every baseline and PU21-PSNR-Y, is in the backup.

FR-HIDROVQA 0.544, 14.9% below the best baseline. BT.2446c keeps the lower color error, 37.83 against 40.19, and the higher HDR-VDP-3.

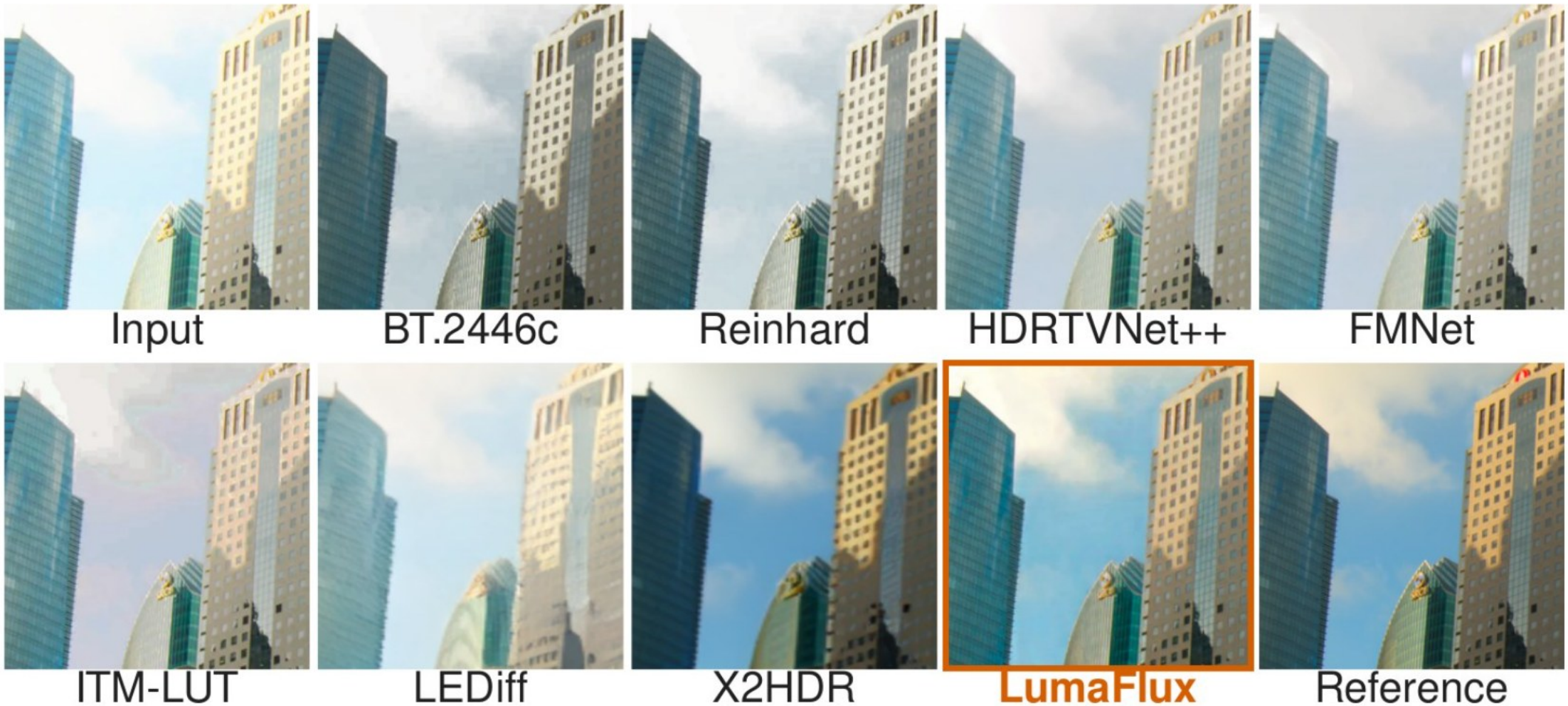
Three scenes, every method

The widest gap is on Sunset, 28.22 dB against 24.67 for BT.2446c.



Full-frame PU21-PSNR above each crop, best in bold. LumaFlux 30.83 dB on Nature (BT.2446c 30.70) and 26.13 on Monument (HDRTVNet++ 25.68). LEDiff (7.90, Nature) and X2HDR (14.36, Monument) are lowest.

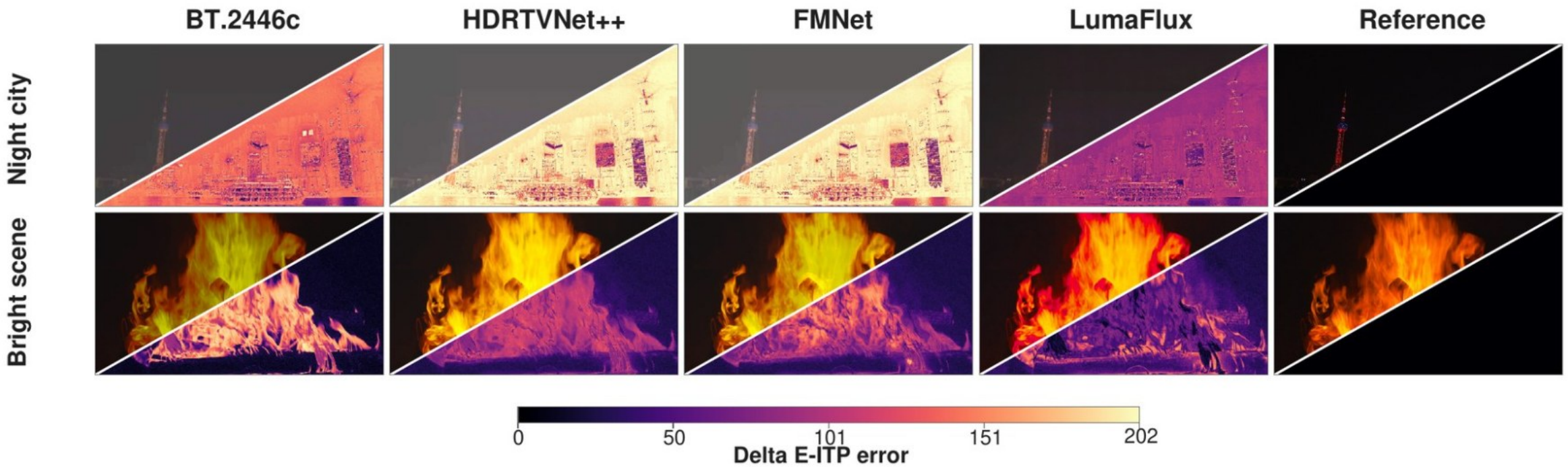
One region of one frame, every method



Tower skyline, BT.2390 tone mapping at CRF 23. Same crop for every method, LumaFlux outlined. HDR shown tone-mapped.

LumaFlux keeps the facade structure and brings back the sky the curves and CNNs wash out.

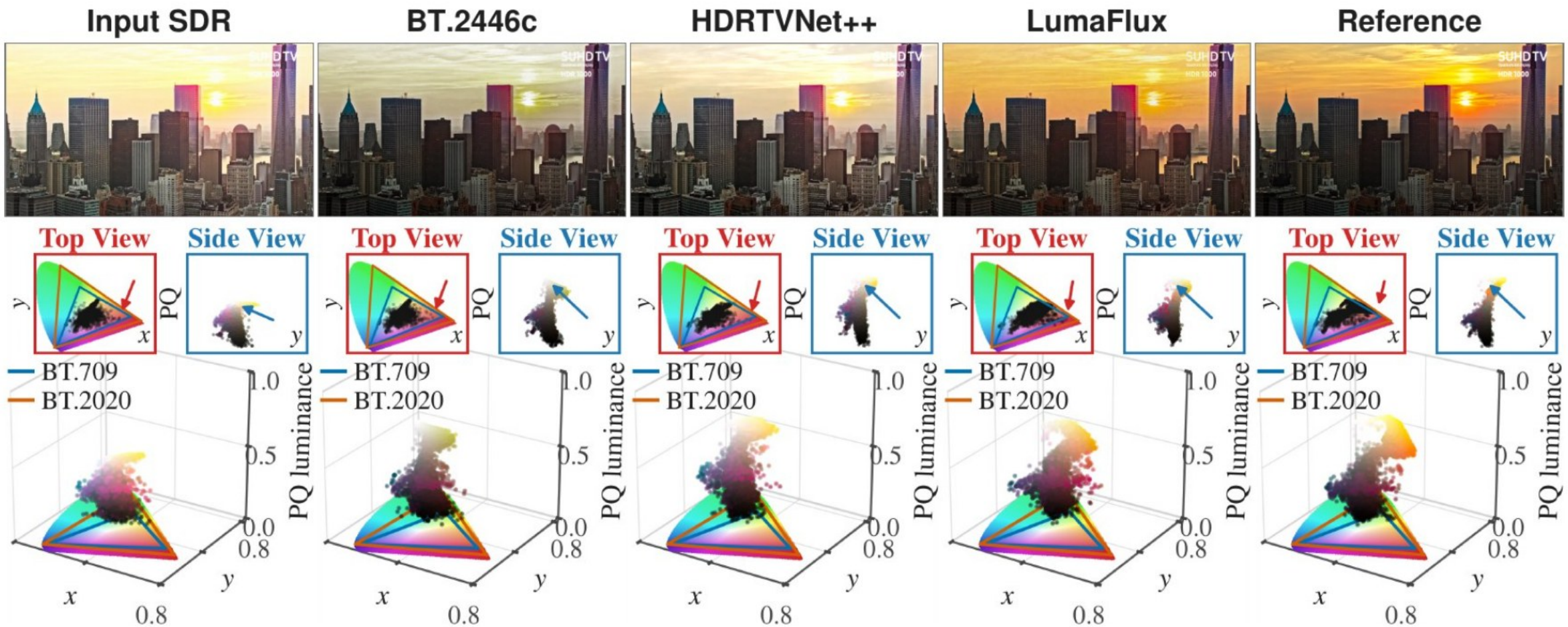
Where the color error sits



Each panel splits into the preview, upper left, and per-pixel ΔE_{ITP}, lower right, on one shared scale. Darker is lower error.

- Night city: the other three light up across sky and towers; LumaFlux stays dark.
- Flame: LumaFlux still overshoots chroma in the core.

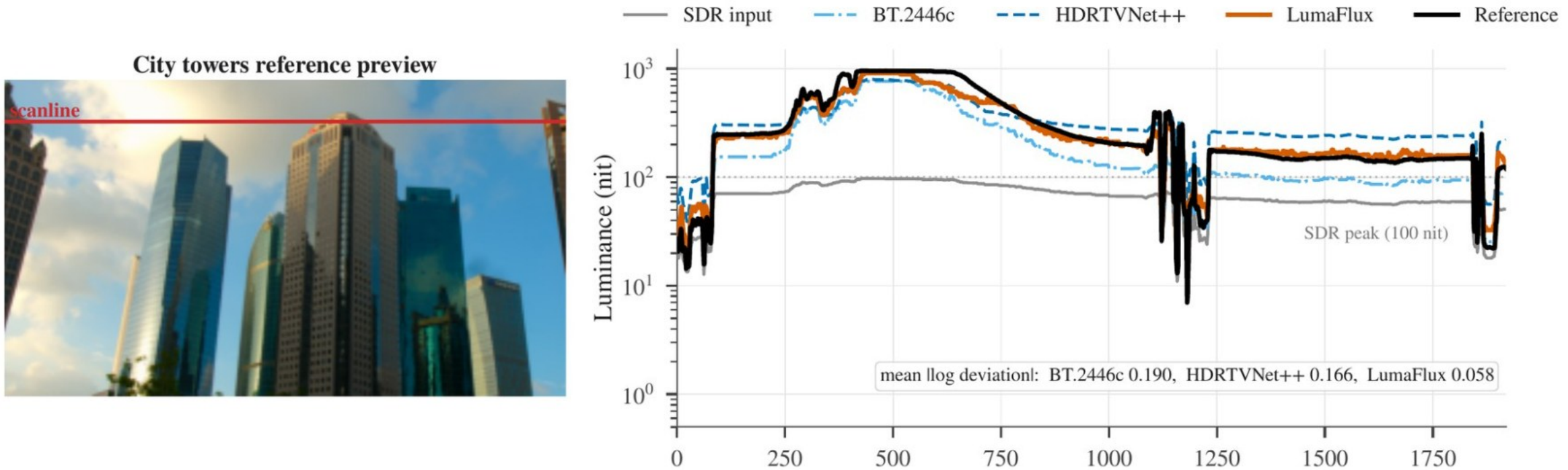
Color volume on one city frame



Frames above, CIE xyY point clouds below. Height is PQ-coded luminance; triangles are BT.709 and BT.2020. HDR shown tone-mapped.

BT.2446c and HDRTVNet++ under-fill the volume; LumaFlux recovers most of what the reference fills.

Luminance along one row tracks the reference



City towers. The red line marks the sampled row; the plot shows luminance along it on a log nit axis.

0.058
LumaFlux, mean absolute log deviation

0.166
HDRTVNet++

0.190
BT.2446c

Removing the spline costs 9.69 dB

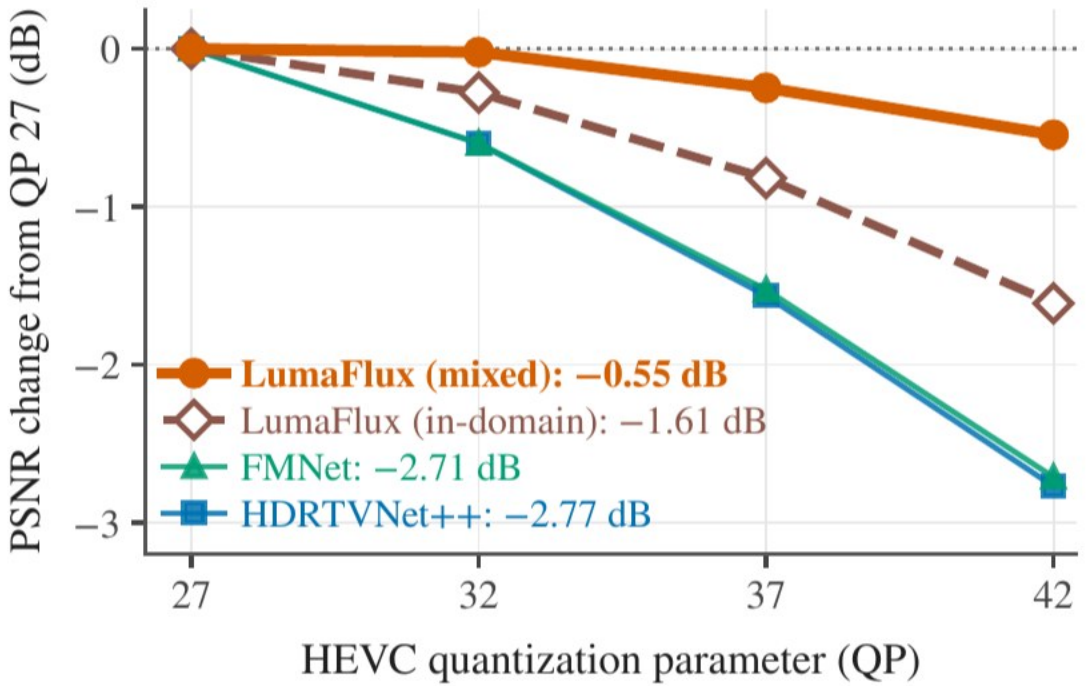
Row	PGA	Gate	PCM	Coup.	T-mod	RQS	PSNR (Δ vs g)	ΔE_{ITP} (Δ vs g)
(c)	•				•	•	28.57 (–0.62)	33.48 (+5.66)
(d)	•	•			•	•	**29.56 (+0.37)**	28.38 (+0.57)
(f) seed-0 repeat of (g)	•	•	•	•	•	•	28.67 (–0.52)	35.00 (+7.18)
(g) full design, ours	•	•	•	•	•	•	29.19 (ref.)	**27.81 (ref.)**
(h) linear spline	•	•	•	•	•	lin.	28.74 (–0.45)	32.16 (+4.35)
(i) no RQS	•	•	•	•	•		19.50 (–9.69)	261.78 (+233.97)
(j) no T-mod	•	•	•	•		•	28.38 (–0.81)	33.90 (+6.09)

Component ablations: 20k steps on a 55-video pool, the same 82 held-out frames. A dot marks a component that is on; Gate is the spectral gate, T-mod the timestep-and-layer modulation, lin. a linear spline. Higher PSNR and lower ΔE_{ITP} are better; deltas are relative to row (g). Rows (a), (b) and (e) are in the backup.

- No RQS, row (i), falls 9.69 dB.
- A linear spline, row (h), costs 0.45 dB.
- The spectral gate, (c) to (d), lifts 28.57 to 29.56 dB.
- No timestep-layer modulation, row (j), costs 0.81 dB.

> Rows (f) and (g) are the same configuration, 0.52 dB apart, so smaller gaps are noise.

Heavier compression barely moves LumaFlux



PU21-PSNR change from QP 27 (QP is the HEVC quantization parameter) on the 117 HDRTV1K pairs. Flatter is better.

PU21-PSNR (dB)	QP 27	QP 32	QP 37	QP 42	Δ
HDRTVNet++	**35.90**	**35.30**	**34.34**	**33.13**	−2.77
FMNet	35.56	34.97	34.04	32.85	−2.71
LumaFlux (mixed-corpus)	27.29	27.27	27.04	26.75	**−0.55**
LumaFlux (in-domain)	33.94	33.67	33.12	32.33	−1.61

Fixed-QP track, 117 HDRTV1K pairs. Δ is QP 42 minus QP 27; closer to zero is better. Mixed-corpus is the main checkpoint; in-domain is the HDRTV1K-specific one. Bold best, underline second per column. ΔE_{ITP} and HDR-VDP-3 are in the backup.

From QP 27 to 42 the mixed-corpus model loses 0.55 dB. HDRTVNet++ loses 2.77 and FMNet 2.71.

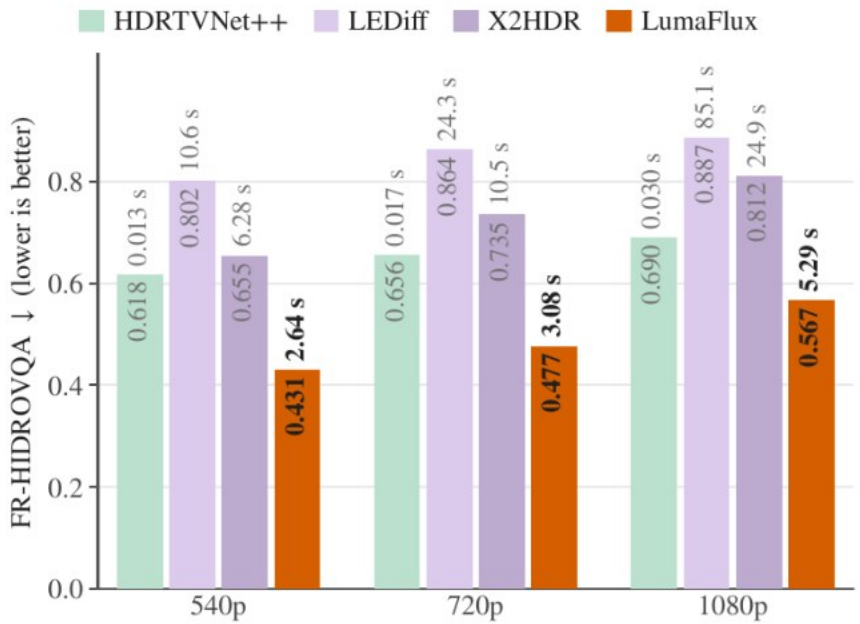
The CNNs still score higher at every QP on their home set.

What it costs

5.28 s
per 1080p frame, 8 steps, one GH200

27.06 GB
peak memory at 1080p

71.32M
trainable of 12.48B parameters



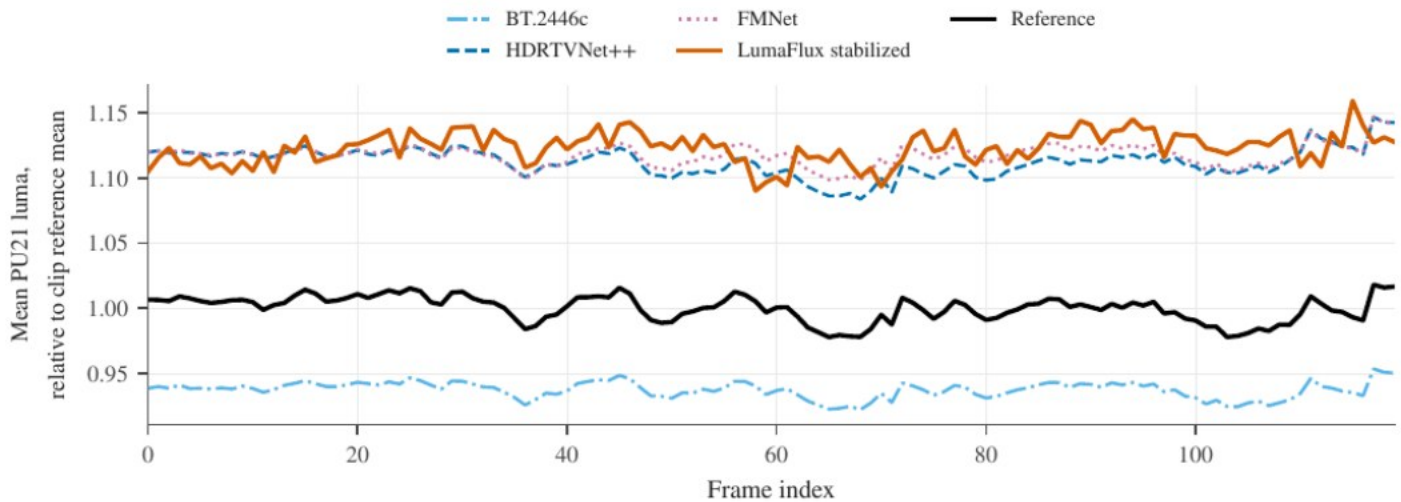
FR-HIDROVQA (lower is better) on 30 expert-track frames. LumaFlux 0.431, 0.477 and 0.567 at 540p, 720p and 1080p, in 2.64, 3.08 and 5.29 s.

Method	Steps	Params (M)	Trainable (M)	s/frame	TFLOPs	Mem. (GB)
HDRTVNet++	1	0.59	0.59	0.857	0.5	2.31
ITM-LUT*	1	0.60	0.60	0.367	–	0.18
PromptIR + Reinhard	1	35.59	35.59	1.708	10.9	17.86
LEDiff (512^2)	50	1370.23	859.53	4.449	149	6.37
X2HDR (512^2)	30	17020.02	149.42	4.244	1,023	33.77
LumaFlux, 8 steps	8	12484.77	71.32	5.282	2,067	27.06
LumaFlux, 40 steps	40	12484.77	71.32	24.993	10,100	27.06

One GH200, batch 1, complete 1080p exports; LEDiff and X2HDR at their registered 512^2 protocol (scaled by pixel count to 1080p: 35.19 and 33.57 s). *ITM-LUT runs a compiled kernel the tracer cannot observe. Lower is better for s/frame, TFLOPs and memory. Full table in the backup.

2.4 to 16.1× faster than X2HDR and LEDiff. HDRTVNet++ stays 173 to 201× faster than us.

Video without temporal training



Per-frame mean PU21 luma, four clips, first 120 frames, relative to each reference mean. Only the stabilized run is drawn.

Method	Bonfire	02c3	055e	0392	0237	Mean
BT.2446c	7.813	1.411	0.450	**0.063**	**0.124**	1.972
HDRTVNet++	3.005	0.207	0.147	0.442	1.257	**1.012**
FMNet	3.853	**0.202**	**0.074**	0.579	1.124	1.166
LumaFlux	**1.992**	0.691	2.500	0.968	0.569	1.344
LumaFlux + stabilization	2.109	0.469	1.676	0.879	0.362	1.099

Temporal-energy deviation $|\text{flicker}_{\text{pred}} - \text{flicker}_{\text{ref}}|$ in 10^6 PU21-energy units on five clips (02c3 to 0237 are CHUG content IDs). Smaller is closer to the reference; bold marks the closest method per column. The reference row is zero by definition.

- Reuse one bridge noise ε for every frame.
- Smooth the spline over time, $\bar{\theta}_f = 0.8\,\bar{\theta}_{f-1} + 0.2\,\theta_f$.
- Signed excess flicker falls 53.3%. Raw flicker stays within 3.27% of the reference.

> Nothing is trained for video, so there are no temporal layers, optical flow or fine-tuning.

Lifting generated video to HDR



Two of six LTX-Video clips. Right column: gray is at or below 100 nit, color is above it.

- Text-to-video gives SDR; LumaFlux lifts it frame by frame.
- Peaks of 739 to 957 nit, 7.4 to 9.6× the SDR ceiling.
- 5.82 s per frame, same checkpoint, no retraining.

Limitations

- **The VAE is an SDR codec.** The spline cannot restore what clipping removed.
- **Oversaturation.** Mean ITP chroma is 60.65 against 49.80 for the references.
- **Native HDRTV1K.** In-domain CNNs reach 38.11 dB where we reach 23.56.
- **Speed.** 5.282 s per 1080p frame against 0.367 to 1.708 s single-pass.

> Still ahead: a larger primary track and a controlled subjective study.

Thank you

Code: github.com/shreshthsaini/LumaFlux

Project page: shreshthsaini.github.io/LumaFlux

Paper: arXiv:2604.02787

Related work of mine

- **LumaGuide**, training-free HDR generation. Preprint, 2026
- **Seeing Beyond 8bits**, HDR-UGC video quality. CVPR 2026
- **Rectified CFG++** for flow based models. NeurIPS 2025
- **BrightRate**, quality of user-generated HDR video. WACV 2026, oral, best student paper
- **HIDRO-VQA**, contrastive learning for HDR video quality. WACV 2024

TEXAS The University of Texas at Austin

The University of Texas at Austin LIVE Laboratory for Image & Video Engineering

YouTube