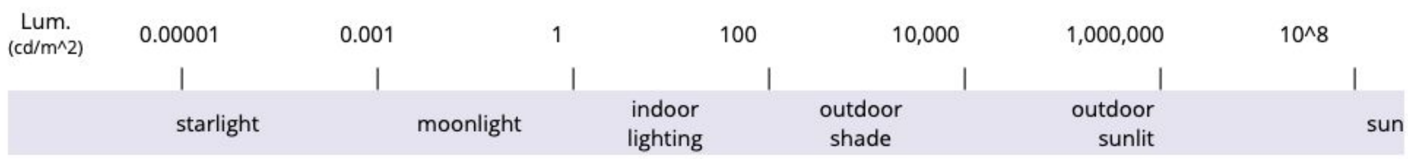
HDR imaging
using Deep Learning

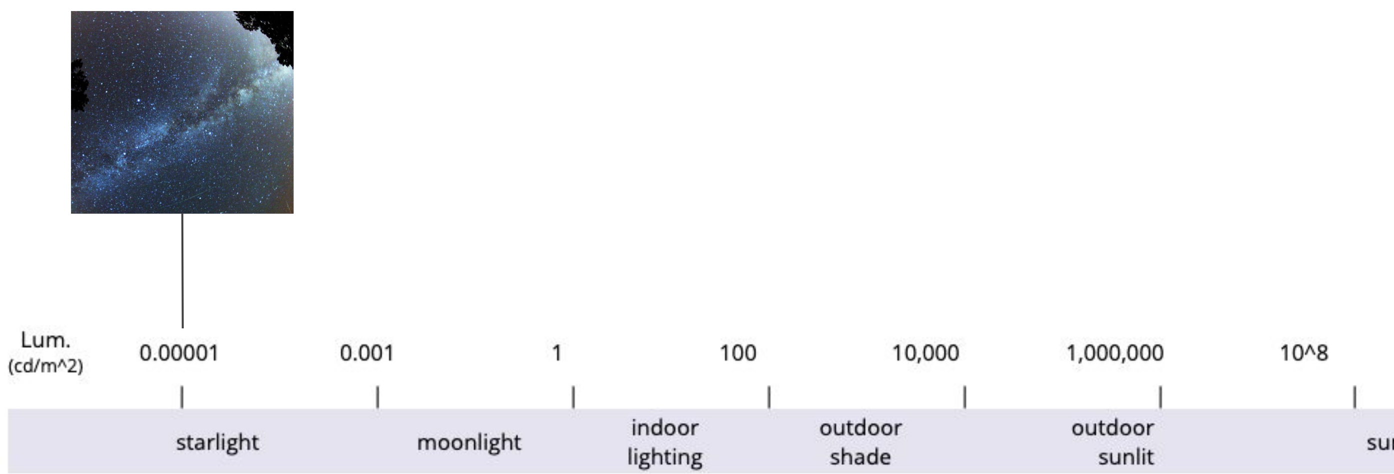
Mukul Khanna, IIT Gandhinagar

HDR

High Dynamic Range

Dynamic Range


Lum.
(cd/m^2)
0.00001
0.001
1
100
10,000
1,000,000
10^8
starlight
moonlight
indoor lighting
outdoor shade
outdoor sunlit
sun

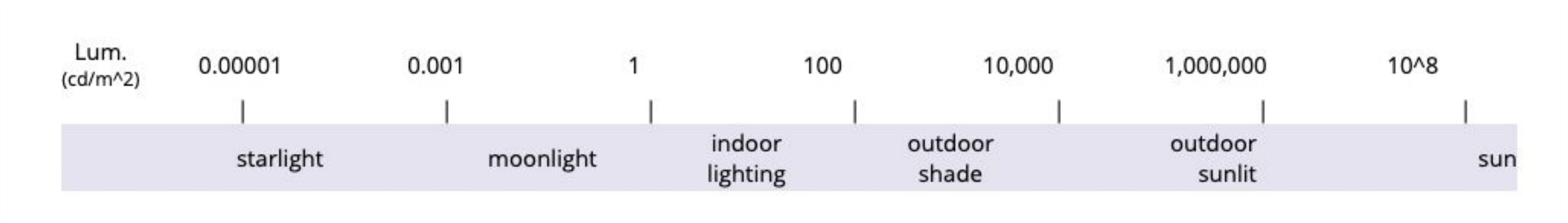

Lum.
(cd/m^2)
0.00001
0.001
1
100
10,000
1,000,000
10^8
starlight
moonlight
indoor lighting
outdoor shade
outdoor sunlit
sun


Lum.
(cd/m^2)
0.00001
0.001
1
100
10,000
1,000,000
10^8
starlight
moonlight
indoor
lighting
outdoor
shade
outdoor
sunlit
sun


Lum.
(cd/m^2)
0.00001
0.001
1
100
10,000
1,000,000
10^8
starlight
moonlight
indoor lighting
outdoor shade
outdoor sunlit
sun

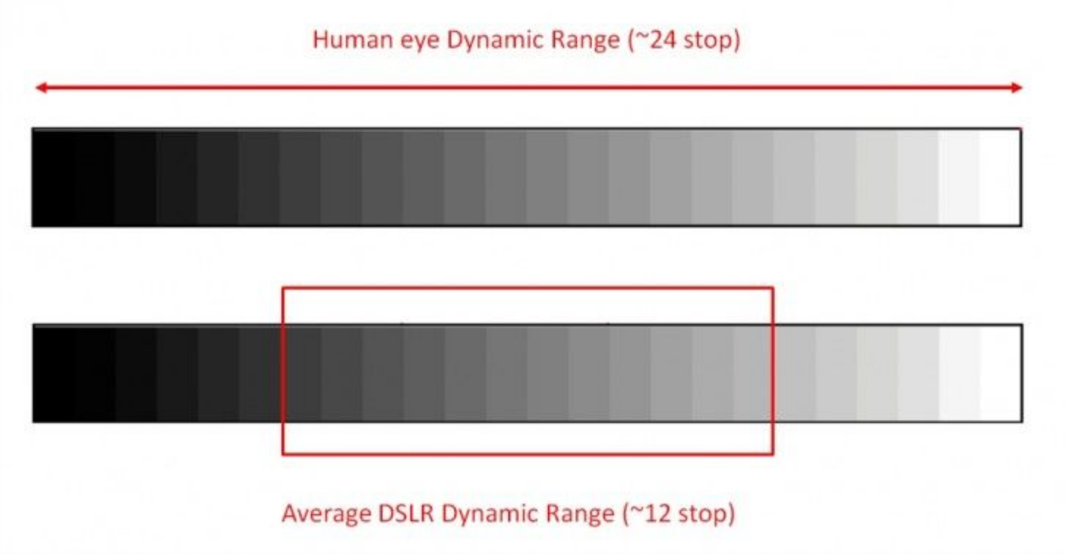
Introduction

- Common digital cameras can not capture the wide range of light intensity levels in a natural scene.



Introduction

- Common digital cameras can not capture the wide range of light intensity levels in a natural scene.



Introduction

- Common digital cameras can not capture the wide range of light intensity levels in a natural scene.
- This can lead to a loss of pixel information in under-exposed and over-exposed regions of an image, resulting in a low dynamic range (LDR) image.

Introduction

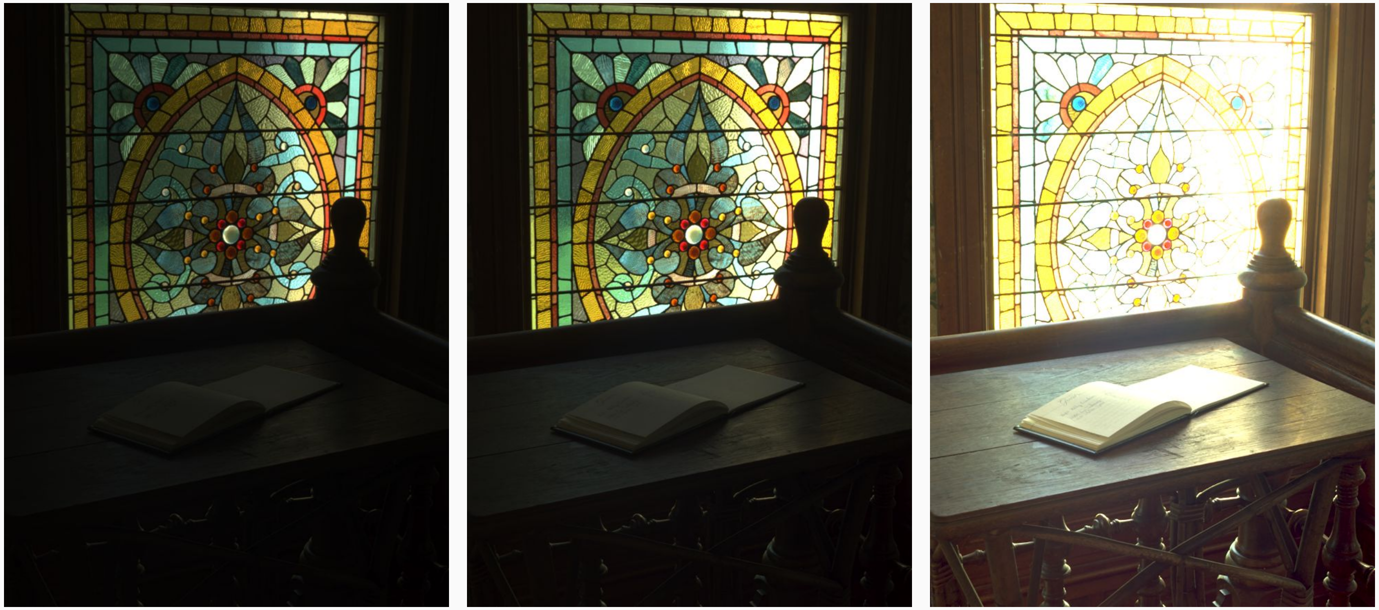
- Common digital cameras can not capture the wide range of light intensity levels in a natural scene.
- This can lead to a loss of pixel information in under-exposed and over-exposed regions of an image, resulting in a low dynamic range (LDR) image.



Courtesy: OpenHDR (viewer.openhdr.org)

Introduction

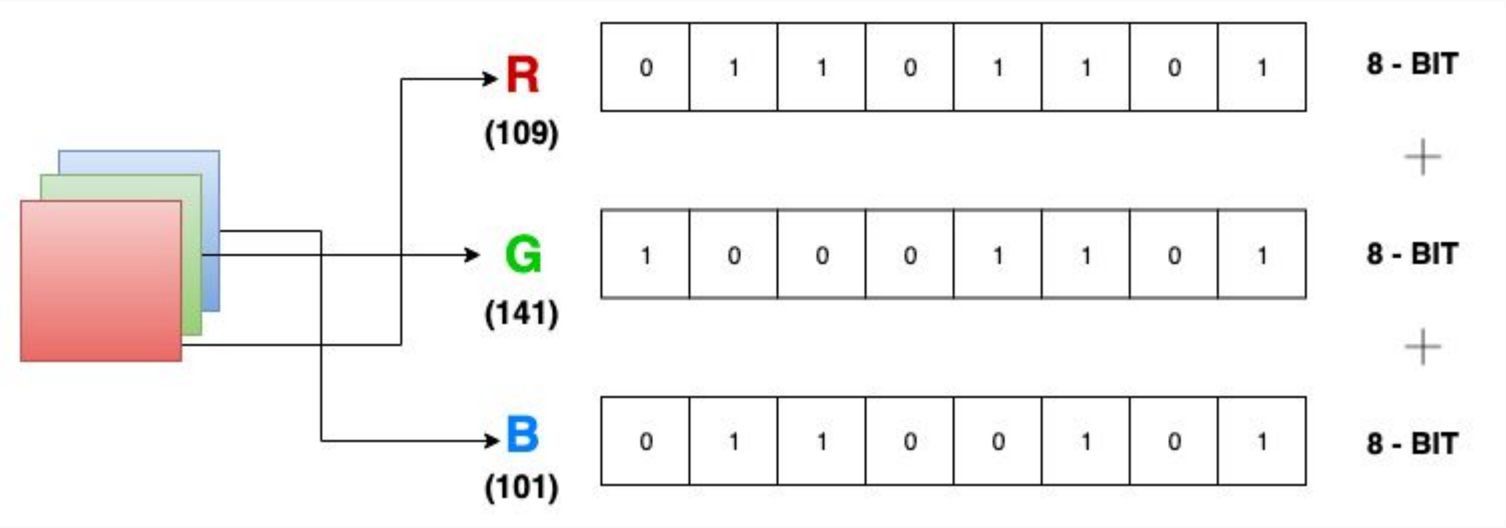
- To recover the lost information and represent the wide range of illuminance in an image, **High Dynamic Range (HDR)** images need to be generated.



HDR IMAGE ENCODING

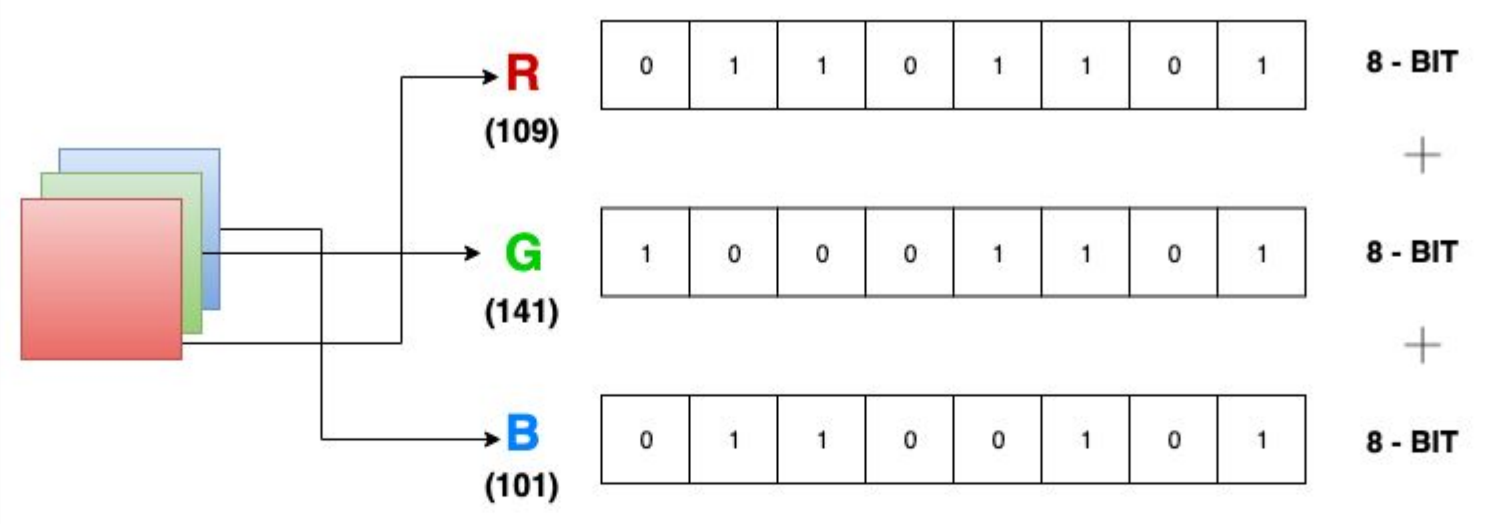
HDR image encoding

- Commonly, the images that we see on our phones and computers, are 8-bit (per channel) encoded RGB images.



HDR image encoding

- Each pixel's value is stored using 24-bit representations, 8-bit for each channel (R, G, B). Each channel of a pixel has a range of 0–255 intensity values.



HDR image encoding

- The problem with this encoding that it is not capable of containing the large dynamic range of natural scenes. It only allows a range of 0–255 (only integers) for accommodating the intensity range, which is not sufficient.

HDR image encoding

- The problem with this encoding that it is not capable of containing the large dynamic range of natural scenes. It only allows a range of 0–255 (only integers) for accommodating the intensity range, which is not sufficient.
- To solve this problem, HDR images are encoded using 32-bit floating point numbers, for each channel. This allows us to capture the wide uncapped range of HDR images.

HDR image encoding

- The problem with this encoding that it is not capable of containing the large dynamic range of natural scenes. It only allows a range of 0–255 (only integers) for accommodating the intensity range, which is not sufficient.
- To solve this problem, HDR images are encoded using 32-bit floating point numbers, for each channel. This allows us to capture the wide uncapped range of HDR images.
- There are various formats for writing HDR images, the most common being **.hdr** and **.exr**.

DISPLAYING HDR IMAGES

Displaying HDR images

- Most off the shelf display devices are incapable of delivering the wide uncapped range of HDR images.
- They expect the input source to be in the three-channel 24-bit (3x8) RGB format.
- Due to this reason, the wide dynamic range needs to be toned down to be able to accommodate it in the 0–255 range of RGB format.

Tone-mapping

- Tone mapping addresses the problem of strong contrast reduction from the scene radiance to the displayable range while preserving the image details and color appearance important to appreciate the original scene content.

HDR IMAGE GENERATION

APPROACHES

- Non-learning based
- Learning based

Non learning based approach

Non learning based approach

- Conventionally, HDR images are developed by merging images captured at different exposures.

Non learning based approach

- These images are merged using a software algorithm and are saved as a single HDR image, in a way that the best portions of each image make it to the final image.

Non learning based approach

- These images are merged using traditional image processing algorithms and are saved as a single HDR image, in a way that the best portions of each image make it to the final image.

Caveats

- Conventional approaches aren't robust enough when it comes to dynamic scenes with motion between the bracketed frames.

Caveats

- Conventional approaches aren’t robust enough when it comes to dynamic scenes with motion between the bracketed frames.
- They rely on Optical Flow to account for the motion between frames.

Caveats

- Conventional approaches aren’t robust enough when it comes to dynamic scenes with motion between the bracketed frames.
- They rely on Optical Flow to account for the motion between frames.

Caveats

- Conventional approaches aren’t robust enough when it comes to dynamic scenes with motion between the bracketed frames.
- They rely on Optical Flow to account for motion between frames.
- But Optical Flow is not accurate.

Caveats

- Conventional approaches aren't robust enough when it comes to dynamic scenes with motion between the bracketed frames.
- They rely on Optical Flow to account for motion between frames.
- But Optical Flow is not accurate.
- This can result in ghosting artifacts in the final image.

Caveats

- Conventional approaches aren't robust enough when it comes to dynamic scenes with motion between the bracketed frames.
- They rely on Optical Flow to account for motion between frames.
- But Optical Flow is not accurate.
- This can result in ghosting artifacts in the final image.

Learning based approach

Learning based approach

- Learning based approaches harness the capabilities of deep neural network architectures as function approximators to learn LDR to HDR representations.

Learning based approach

- Learning based approaches harness the capabilities of deep neural network architectures as function approximators to learn LDR to HDR representations.

Learning based approach

- Learning based approaches harness the capabilities of deep neural network architectures as function approximators to learn LDR to HDR representations.
- Such networks can do better due to -
    - improved learning based flow mechanisms
    - hallucinating HDR content in saturated regions when LDR input is limited
    - optimised, quick, low-memory alternative

Learning based approach

- Learning based approaches can be broken down into two types -

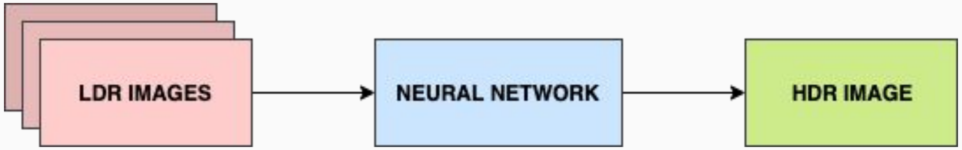
Learning based approach

- Learning based approaches can be broken down into two types -
    - Single LDR input

Approaches - learning based

- Learning based approaches can be broken down into two types -

    - Single LDR input
    - Multiple LDR inputs



Learning based - multiple LDR inputs

- Multiple exposure input

Learning based - multiple LDR inputs

- Multiple exposure input

Learning based - multiple LDR inputs

- Multiple exposure input
- More dynamic range is provided to the network

Learning based - multiple LDR inputs

- Multiple exposure input
- More dynamic range is provided to the network
- Explicit mechanism is required for motion compensation

Learning based - multiple LDR inputs

- Multiple exposure input
- More dynamic range is provided to the network
- Explicit mechanism is required for motion compensation
- Better results

Learning based - multiple LDR inputs

- Multiple exposure input
- More dynamic range is provided to the network
- Explicit mechanism required for motion compensation
- Better results
- But input is a constraint

Single LDR input approaches

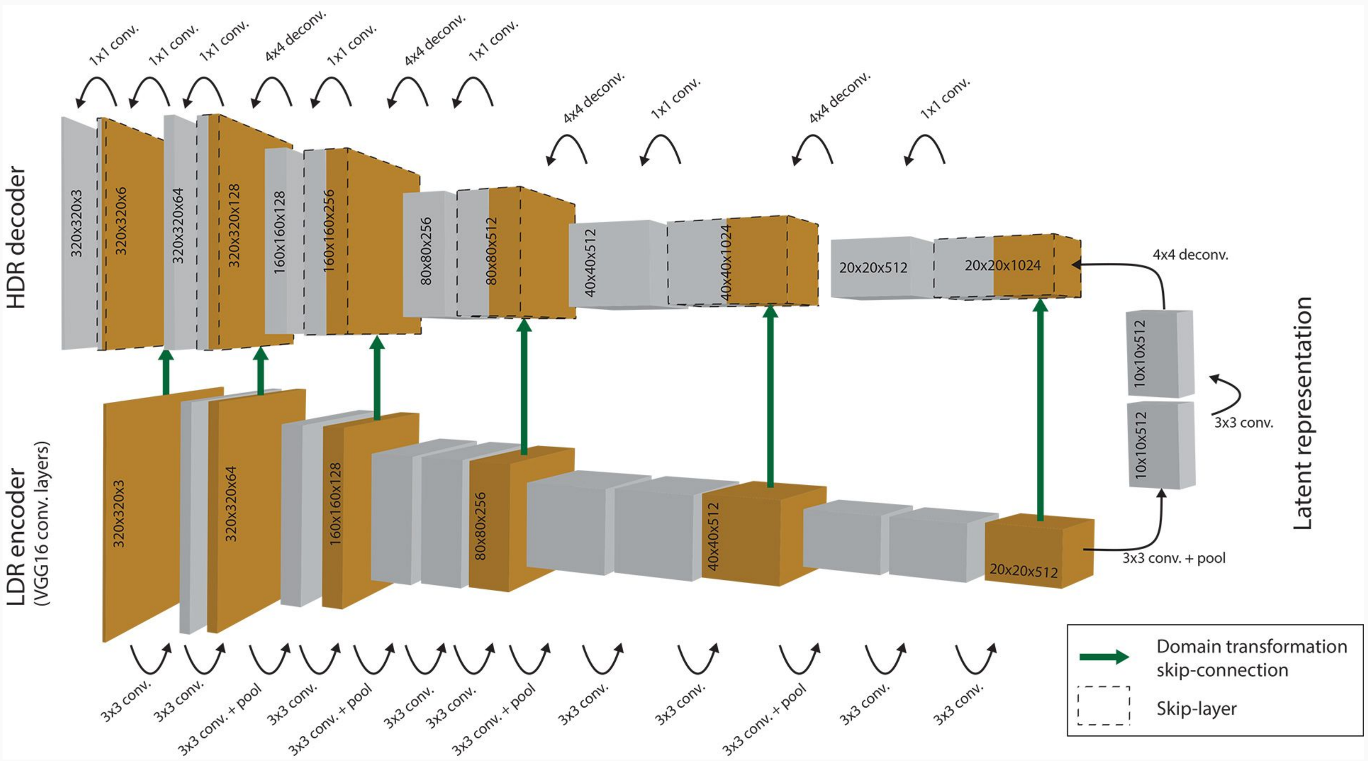
Learning based - single LDR input

- More challenging scenario
- Limited dynamic range information input
- More important for real life situations
- Heavily relies on ability of deep CNNs to hallucinate content in saturated image regions.

Related work

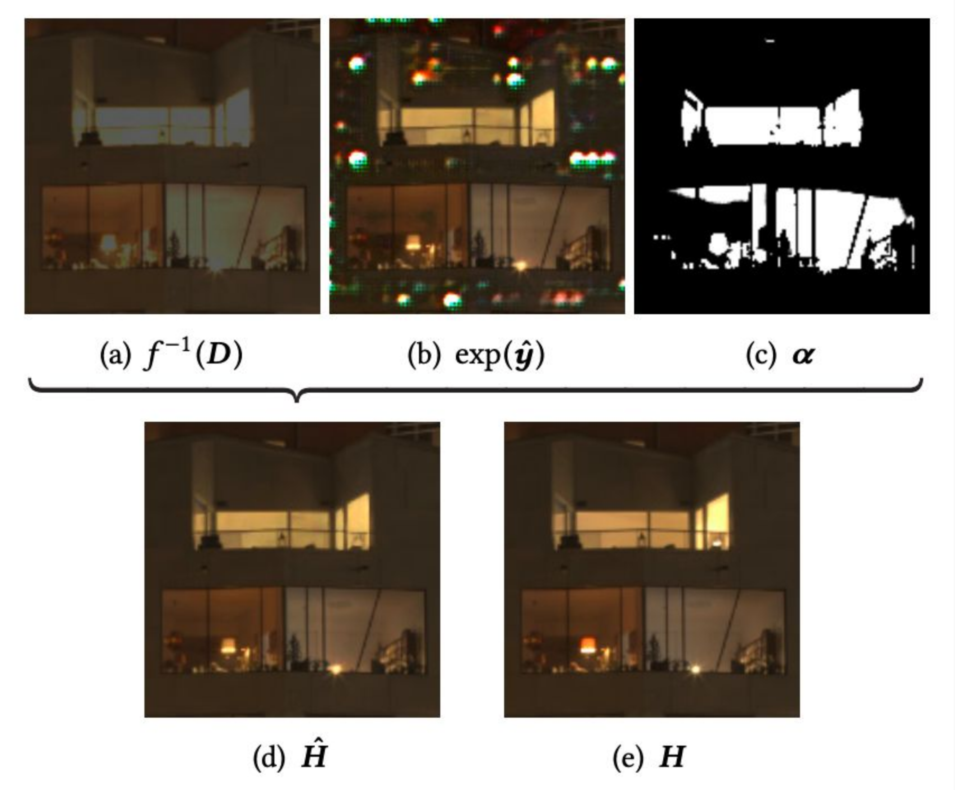
HDRCNN

G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, “Hdr image reconstruction from a single exposure using deep cnns,” ACM Transactions on Graphics (TOG), vol. 36, no. 6, p. 178, 2017



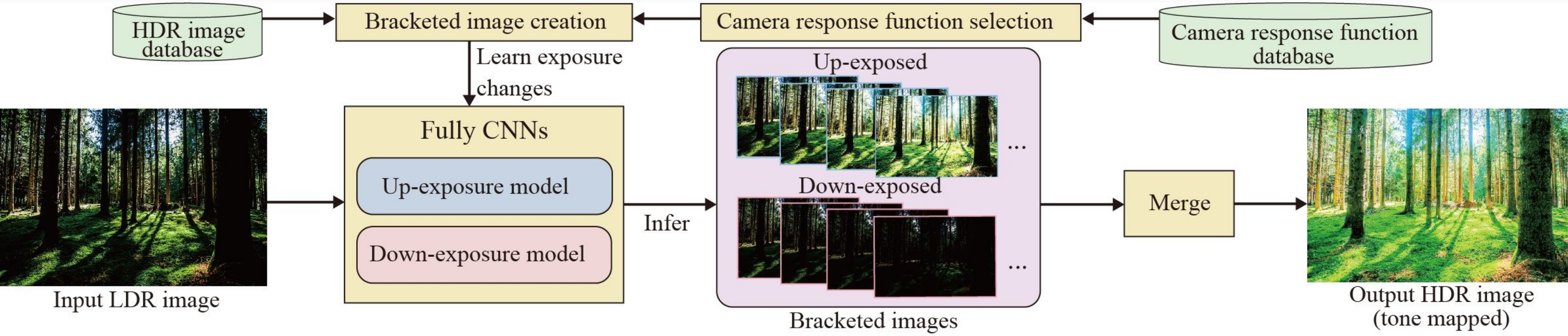
HDRCNN

G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, “Hdr image reconstruction from a single exposure using deep cnns,” ACM Transactions on Graphics (TOG), vol. 36, no. 6, p. 178, 2017



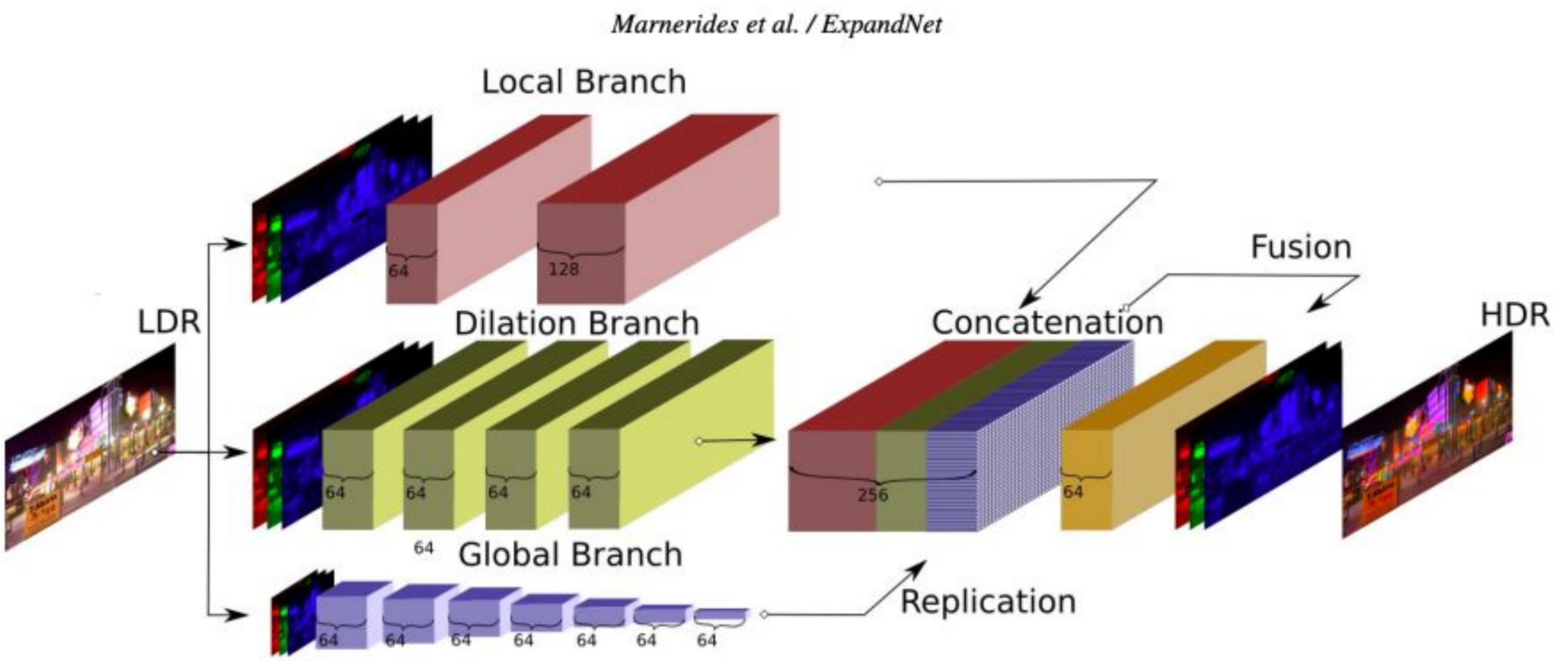
Deep reverse tone mapping

Y. Endo, Y. Kanamori, and J. Mitani, “Deep reverse tone mapping.,” ACM Trans. Graph., vol. 36, no. 6, pp. 177–1, 2017.



ExpandNet

D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, “Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content,” in Computer Graphics Forum, vol. 37, pp. 37–49, Wiley Online Library, 2018.



Caveats

- Not end-to-end trainable

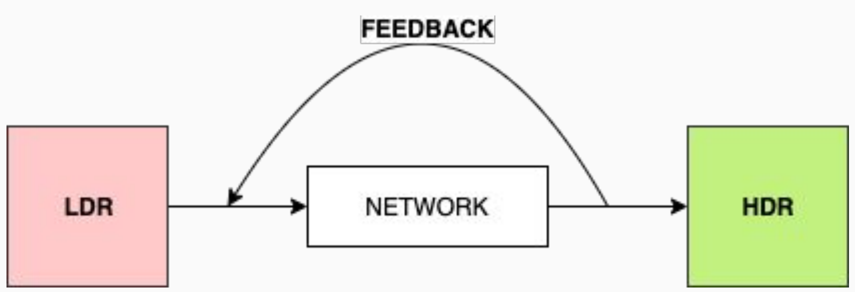
    OR/AND
- Only overexposed regions are recovered

    OR/AND
- High network parameter count

Our approach

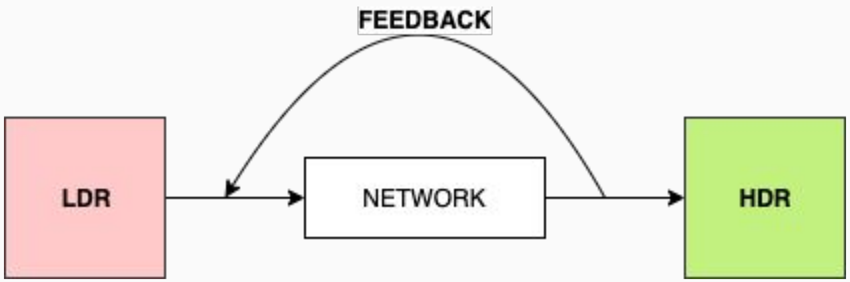
Feedback networks

- Feedback systems are adopted to influence the input based on the generated output.

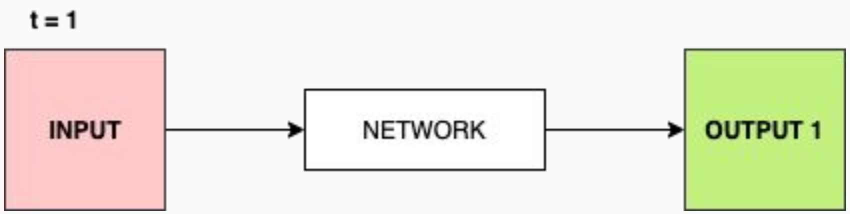


Feedback networks

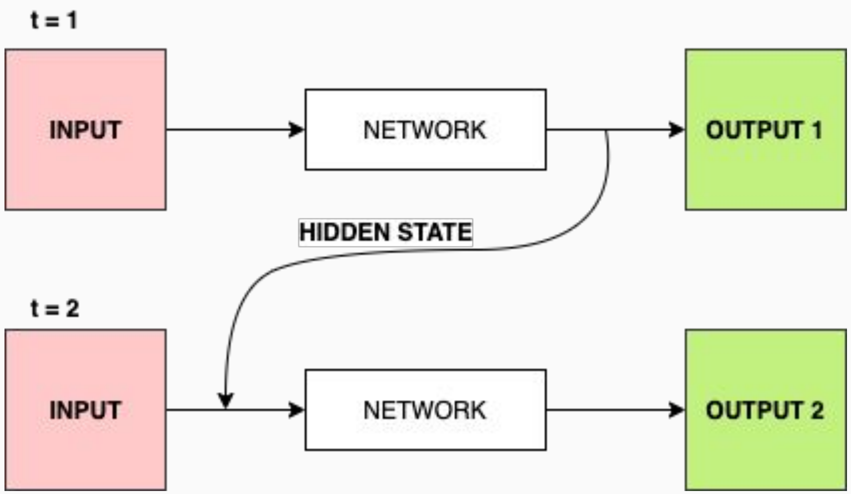
- Feedback systems are adopted to influence the input based on the generated output.
- Initial low level features are guided by the high level features using a hidden state of a Recurrent Neural Network over n iterations.



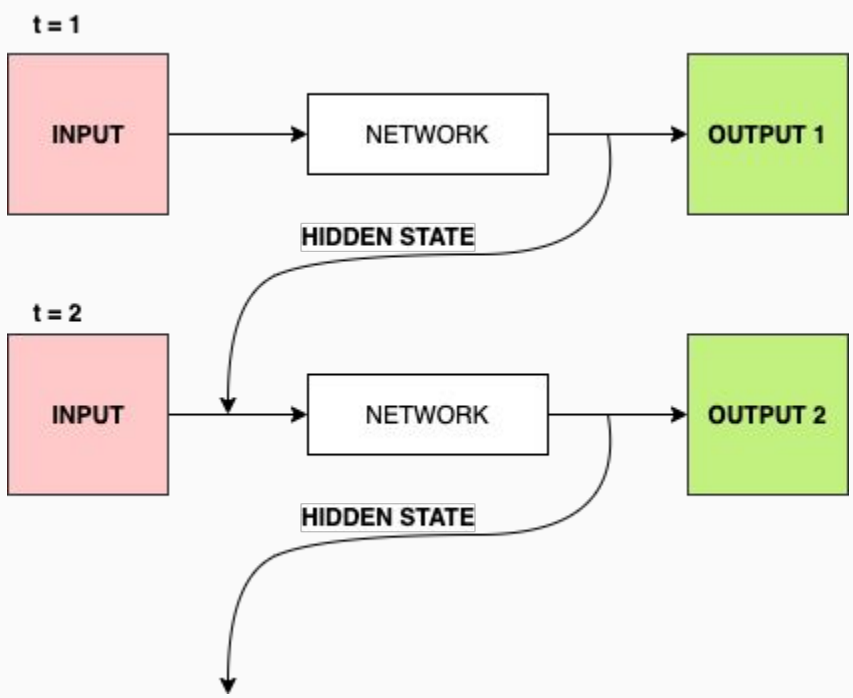
Feedback networks



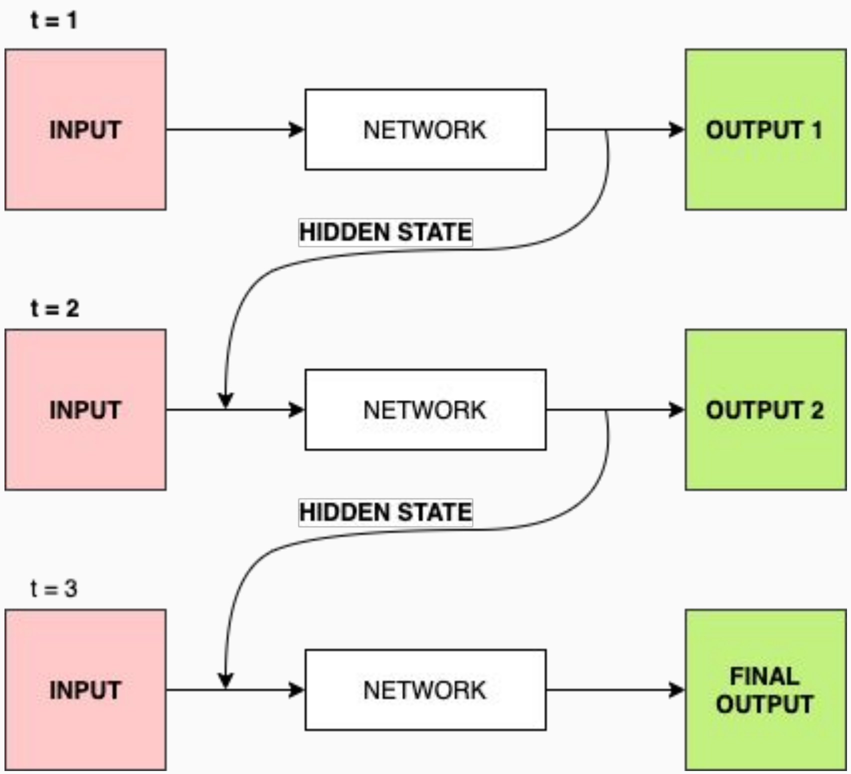
Feedback networks



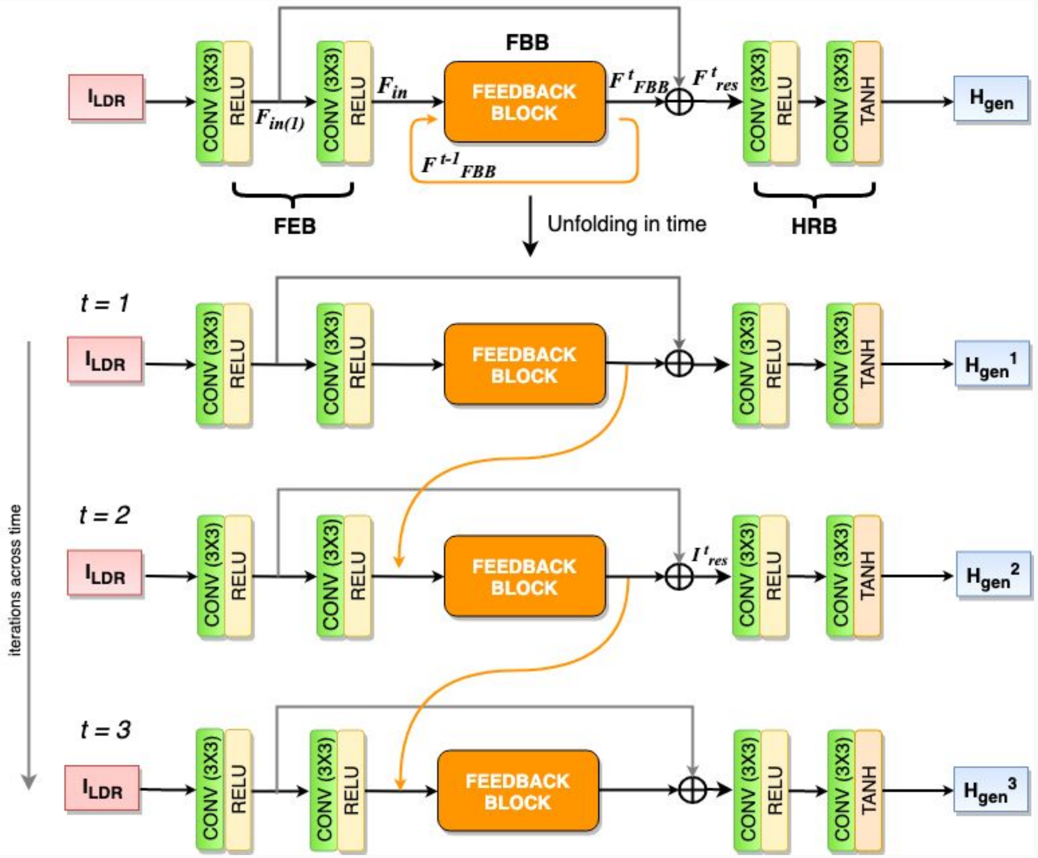
Feedback networks



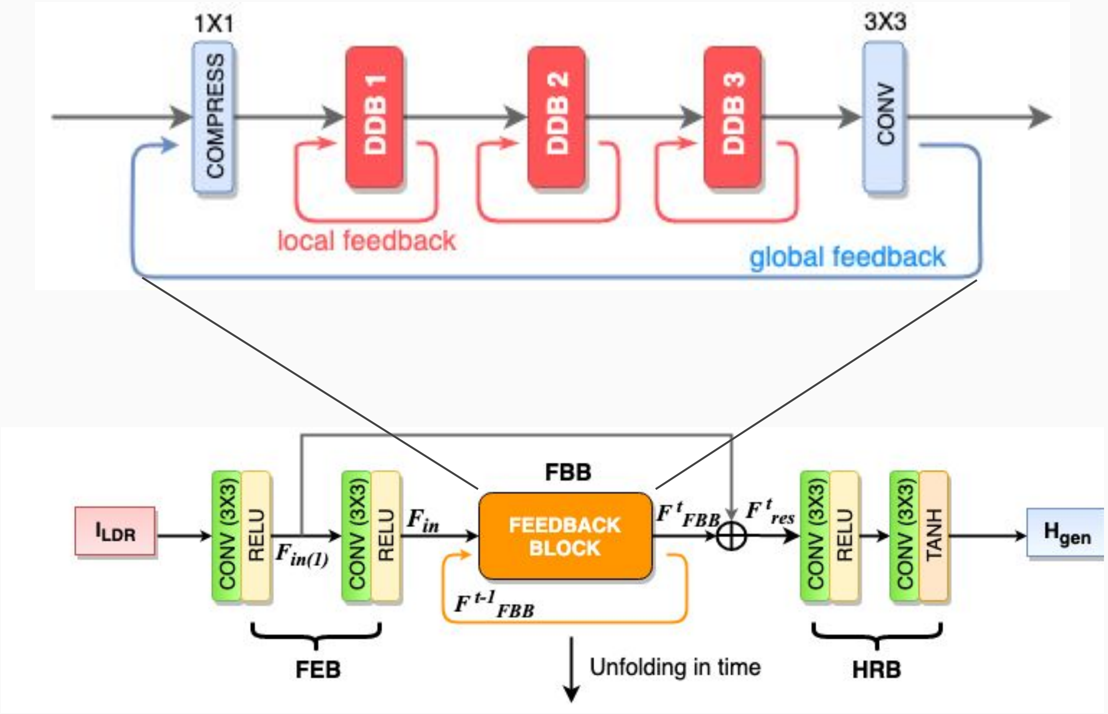
Feedback networks



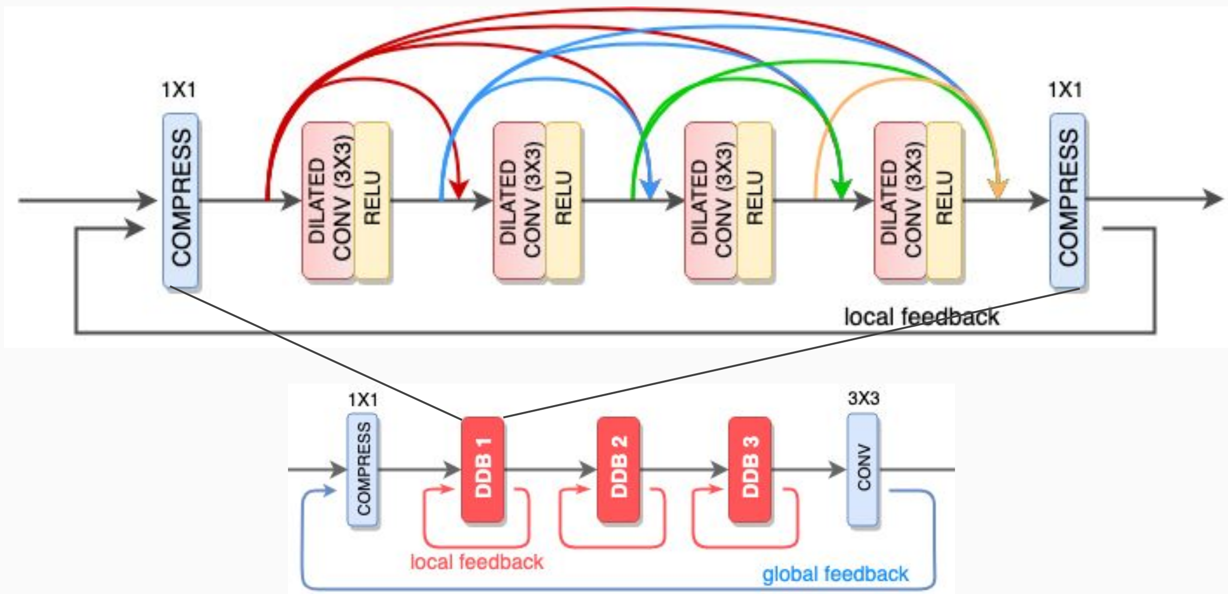
Model architecture

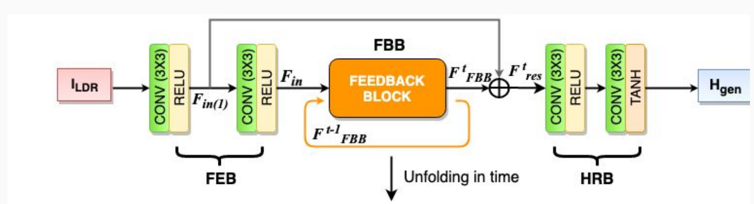


Feedback block

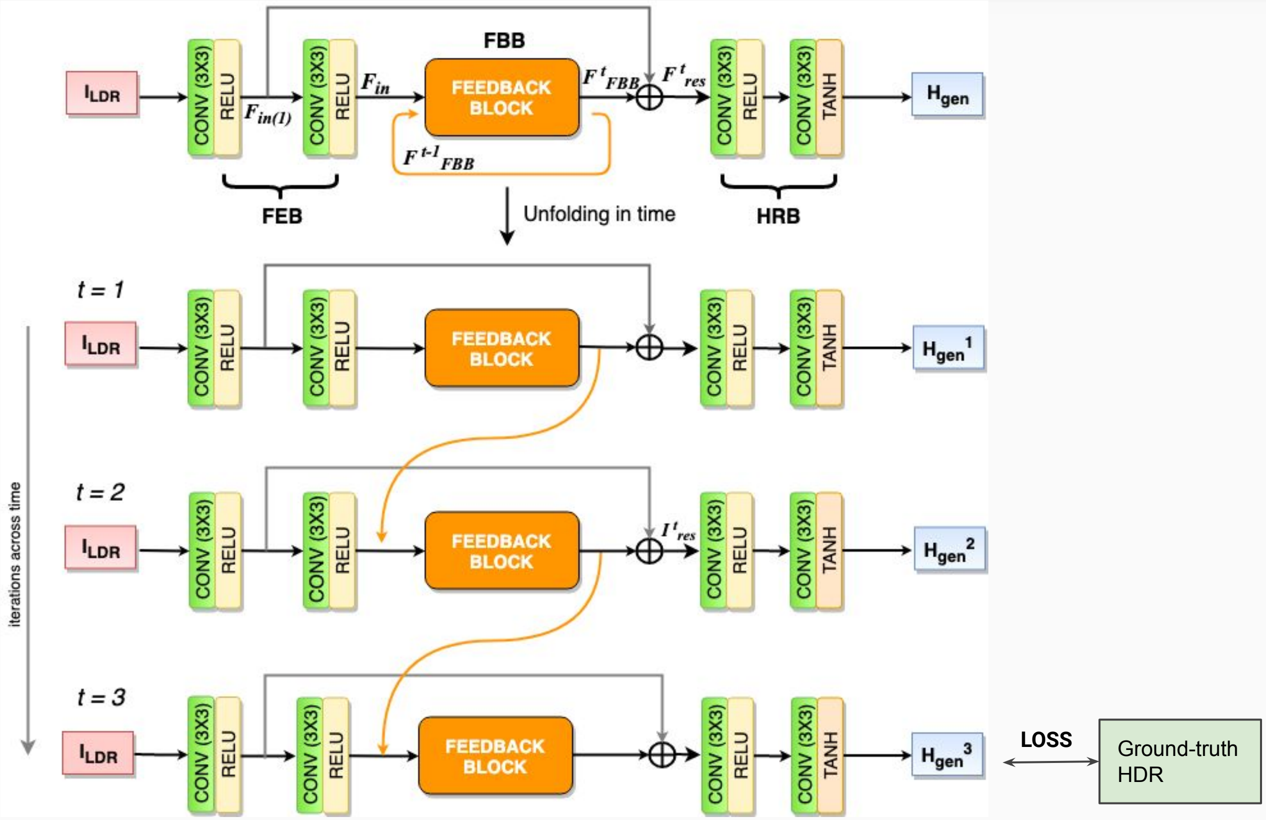


Dilated Dense Block (DDB)





Loss function



Loss function

- Loss calculated directly on HDR images is misrepresented due to the dominance of high intensity values of images with a wide dynamic range.

Loss function

- Loss calculated directly on HDR images is misrepresented due to the dominance of high intensity values of images with a wide dynamic range.
- Therefore, we tonemap the generated and the ground truth HDR images to compress the wide intensity range before calculating the loss.

Loss function

- Loss calculated directly on HDR images is misrepresented due to the dominance of high intensity values of images with a wide dynamic range.
- Therefore, we tonemap the generated and the ground truth HDR images to compress the wide intensity range before calculating the loss.
- We use the μ-law for tonemapping -

$$T(H_{gen}^{t}) = \frac{log(1 + \mu H_{gen}^{t})}{log(1 + \mu)}$$

Loss function

- Loss calculated directly on HDR images is misrepresented due to the dominance of high intensity values of images with a wide dynamic range.
- Therefore, we tonemap the generated and the ground truth HDR images to compress the wide intensity range before calculating the loss.
- We use the μ-law for tonemapping.
- **L1 loss and Perceptual loss** ($\lambda = 0.1$)

$$\mathcal{L} = \mathcal{L}_p + \lambda \mathcal{L}_{L1}$$

Experiments

Datasets

The performance of the network was evaluated over two datasets-

Datasets

The performance of the network was evaluated over two datasets-

- CityScene dataset
  - 128 x 64 size
  - Training set - 39,460 LDR-HDR image pairs
  - Testing set - 1,672 pairs

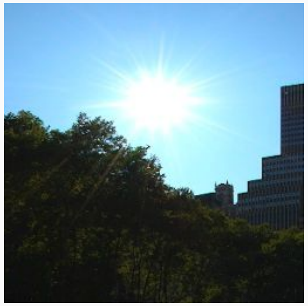
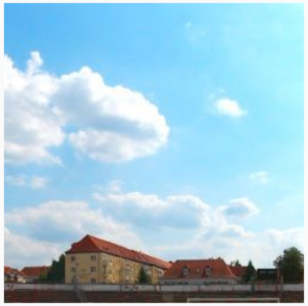
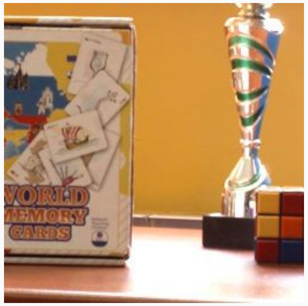
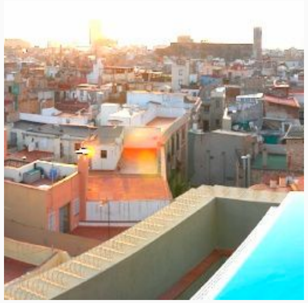




Datasets

The performance of the network was evaluated over two datasets-

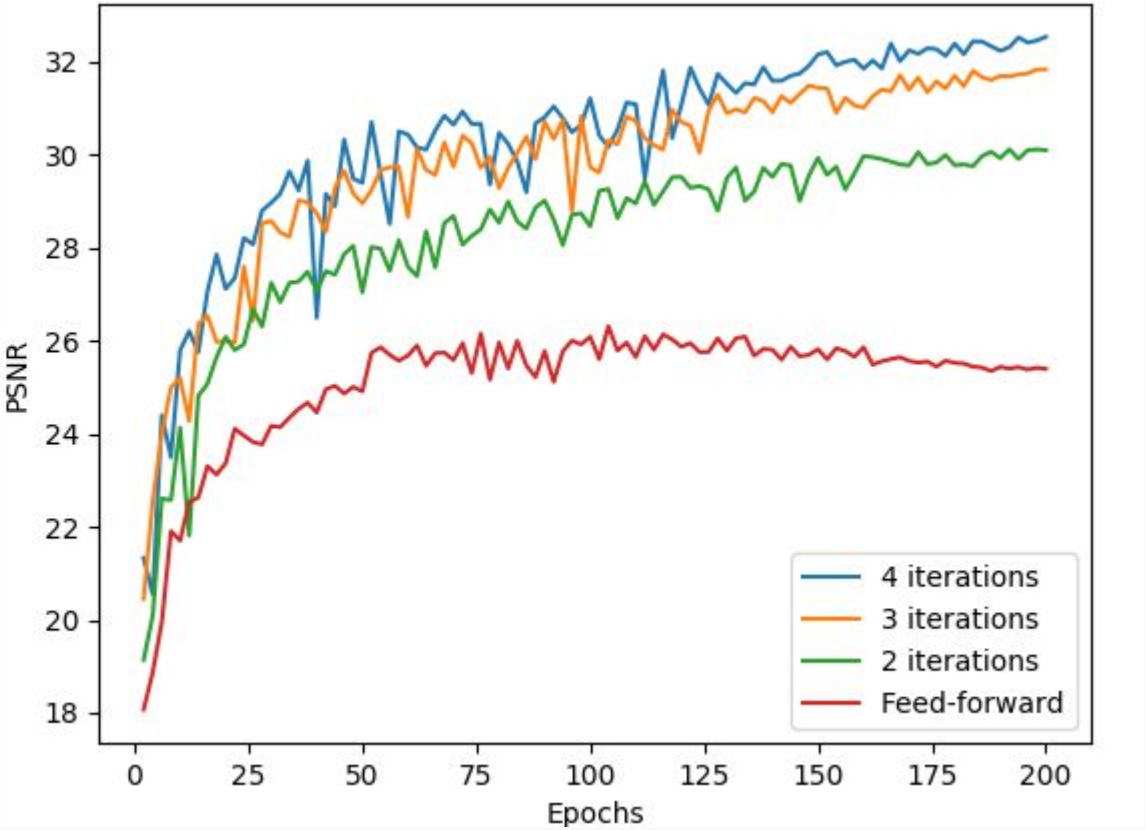
- Curated dataset
    - 256 x 256 size
    - Training set - 11,262 LDR-HDR image pairs
    - Testing set - 500 image pairs (512 x 512)


Evaluation metrics

- PSNR score (db) - Peak Signal-to-Noise Ratio
- SSIM score - Structural Similarity Index
- HDR-VDP2 Q-score

Feedback mechanism analysis



Results

Qualitative evaluation

LDR

GENERATED

GROUND TRUTH

Qualitative evaluation

LDR

GENERATED

GROUND TRUTH

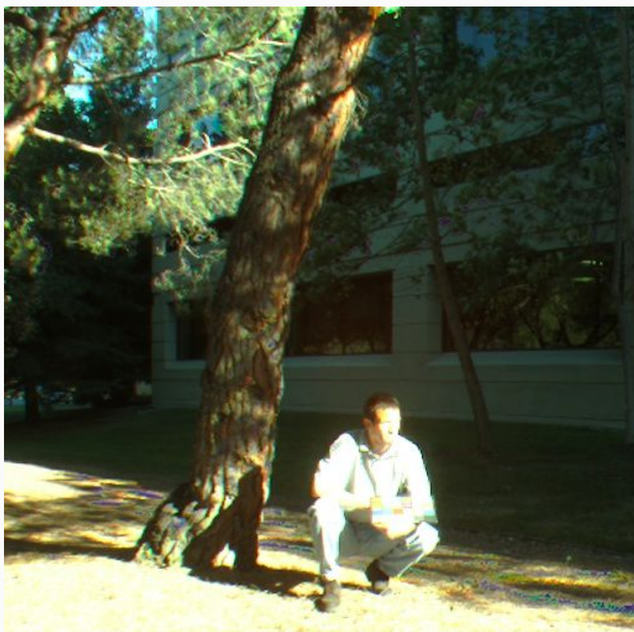
Qualitative evaluation

LDR

GENERATED

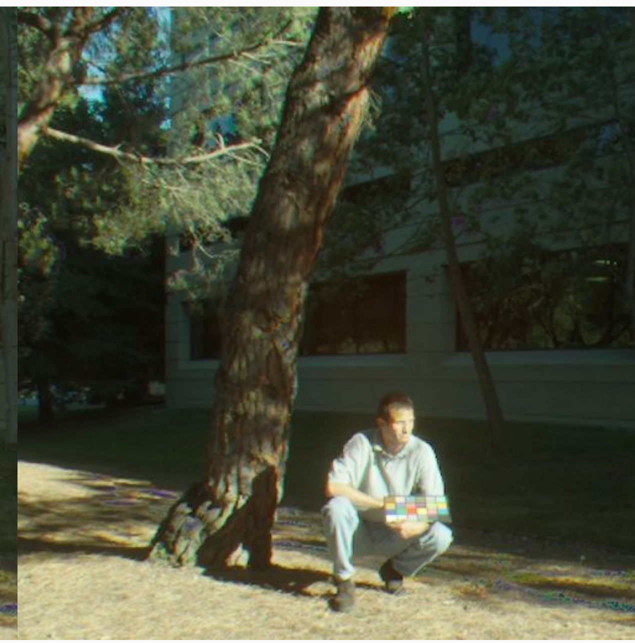
GROUND TRUTH

Qualitative evaluation



LDR

GENERATED



GROUND TRUTH

Qualitative comparisons

LDR | **DRTMO** | **FHDR** | **GROUND TRUTH**

Qualitative comparisons

LDR HDRCNN FHDR GROUND TRUTH

Qualitative evaluation

LDR

GENERATED

GROUND TRUTH

Qualitative comparisons

LDR **DRTMO** **FHDR** **GROUND TRUTH**

Qualitative comparisons

LDR HDRCNN FHDR GROUND TRUTH

Quantitative evaluation

Methods	City Scene Dataset			Curated HDR Dataset		
	PSNR	**SSIM**	**Q-score**	**PSNR**	**SSIM**	**Q-score**
AKY [14]	15.35	0.44	35.40	9.58	0.20	33.47
KOV [15]	16.77	0.59	35.31	12.99	0.41	29.87
HDRCNN [1]	13.21	0.38	54.34	12.13	0.34	55.32
DRTMO [3]	-	-	-	11.4	0.28	58.85
DRHT [4]	-	0.93	61.51	-	-	-
FHDR/W	25.39	0.89	63.21	16.94	0.74	65.27
FHDR	**32.54**	**0.95**	**67.18**	**20.3**	**0.79**	**70.97**

HDR VIDEO

HDR VIDEO GENERATION

HDR Video generation

- LDR video -> HDR video

HDR Video generation

- LDR video -> HDR video
  - Single exposure LDR to HDR
  - Multiple exposure LDR sequences to HDR

HDR Video generation

- LDR video -> HDR video
  - Single exposure LDR to HDR
  - Multiple exposure LDR sequences to HDR

- Temporal coherency is crucial because of vulnerability of neural networks to produce highly varied outputs for minutely different inputs.

HDR Video generation

- LDR video -> HDR video
  - Single exposure LDR to HDR
  - Multiple exposure LDR sequences to HDR

- Temporal coherency is crucial because of vulnerability of neural networks to produce highly varied outputs for minutely different inputs.

- RNNs, LSTMs to propagate temporal information across sequences.

HDR Video generation

- LDR video -> HDR video
  - Single exposure LDR to HDR
  - Multiple exposure LDR sequences to HDR

- Temporal coherency is crucial because of vulnerability of neural networks to produce highly varied outputs for minutely different inputs.

- RNNs, LSTMs to propagate temporal information across sequences.
- Adversarial training using temporal discriminators.

Conclusion

Conclusion

- HDR content is important.

Conclusion

- HDR content is important.
- Deep learning helps - outperforms traditional approaches, again.

Thank you

Mukul Khanna

mukul18khanna@gmail.com