

ITM-CNN: Learning the Inverse Tone Mapping from Low Dynamic Range Video to High Dynamic Range Displays using Convolutional Neural Networks



Soo Ye Kim Dae Eun Kim Munchurl Kim
 School of Electrical Engineering
 Korea Advanced Institute of Science and Technology
 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
 {sooyekim, kimde, mkimee}@kaist.ac.kr



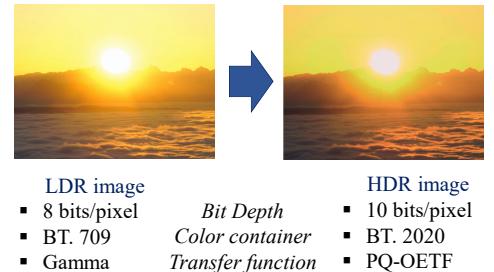
Summary

Inverse Tone Mapping (ITM)

- Until now, inverse tone-mapping (ITM) problems focused on rendering the HDR images on professional monitors
- The advent of HDR TVs open up new horizons for the consumption of high quality HDR multimedia contents
- Unfortunately, the HDR results of previous methods exhibit noise or lack details when viewed on HDR TVs

Research Objectives

- We first present the ITM problem for producing HDR images suitable for commercial HDR TVs
- A novel image-decomposition-based fully end-to-end CNN architecture, which restores lost details and local contrast
 - CNN composed of three parts: LDR decomposition, feature extraction and HDR reconstruction
- Pre-training strategies using guided filter decompositions
 - By decomposing the image, the CNN is able to focus on either the overall intensity or finer details of the input
- As a result, our method produces highly realistic images on HDR TVs, when compared to conventional methods

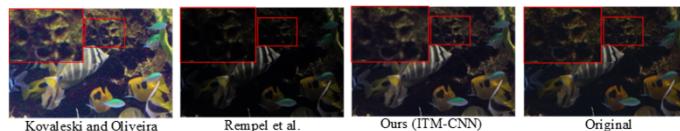


Introduction

Dynamic range



Previous methods

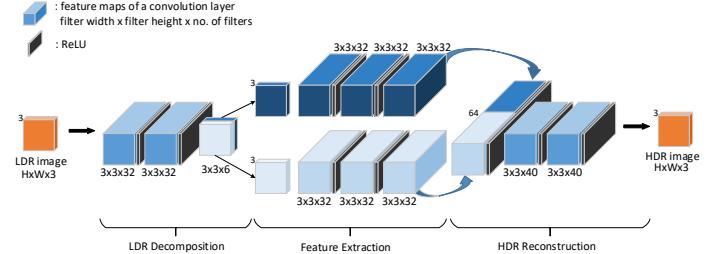


- Noise amplification in dark regions or lack of details

Training Strategy

Network architecture

- Composed of three parts: LDR decomposition, feature extraction, HDR reconstruction



Training procedures

Procedure	(i)	(ii)	(iii)	(iv)
Order	Decom.	Feat.	Recon.	Decom.
1 st	-	-	-	-
2 nd	-	-	-	-
3 rd	■	■	■	■
PSNR of Y	46.71	46.28	47.11	47.27
PSNR of YUV	48.80	48.15	48.98	49.21

Decom.: LDR decomposition part, Feat.: feature extraction part, Recon.: HDR reconstruction part

■: guided filter based pre-training, ■: simple training

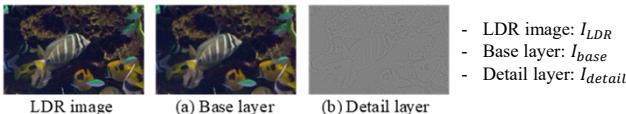
Proposed training strategy

1. Guided filter based pre-training of the *feature extraction* and *HDR reconstruction* parts
2. Pre-training of the LDR decomposition part
3. Main training phase of the whole network

Input Decomposition

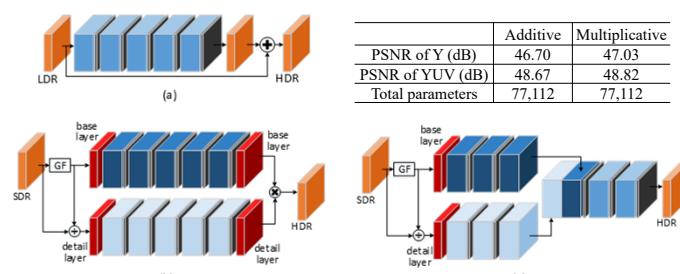
Input decomposition using guided filter

- Decomposing the input lets the feature extraction layers to concentrate on the different decompositions (e.g. overall intensity or finer details)
- I_{base} : guided-filtered result of the input LDR image, I_{LDR}
- $I_{detail} = I_{LDR} \oslash I_{base}$



- LDR image: I_{LDR}
- Base layer: I_{base}
- Detail layer: I_{detail}

Different architectures using image decompositions

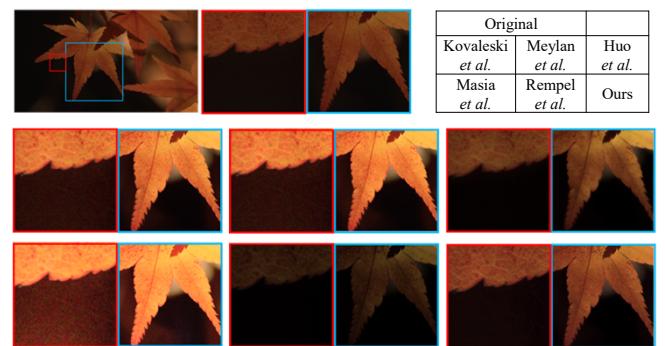


Structure	(a)*	(a)	(b)	(c)*	(c)
Layer	Number of filter channels (input, output)				
1	3.32	3.32	3.45	3.32	3.32
2	32.32	32.32	45.45	32.32	32.32
3	32.32	32.32	45.48	32.32	32.32
4	32.32	32.32	48.45	32.32	32.32
5	32.32	32.32	45.45	32.32	32.32
6	32.3	32.3	45.3	32.3	32.3
Total parameters	38,592	38,592	77,760	77,184	77,328
PSNR of Y	45.46	46.36	46.36	46.84	46.73
PSNR of YUV	47.28	48.25	48.39	48.65	48.87

(a)*: (a) without residual learning

(c)*: (c) with element-wise multiplication after layer 3

Benchmark



Applications

- ITM-CNN is a powerful means to solve the lack of HDR contents using legacy LDR videos
- Our image decomposition training strategy is extensible to other CNN methods