Seeing in Extra Darkness Using a Deep-Red Flash

Jinhui Xiong[1*], Jian Wang[2*], Wolfgang Heidrich[1], Shree Nayar[2]
[1]KAUST, [2]Snap Research ([*] denotes equal contribution)

Paper

Code

[Code] github.com/vccimaging/Deep-Red-Flash
[Paper] vccimaging.org/Publications/Xiong2021Seeing/Xiong2021Seeing.pdf

CVPR VIRTUAL JUNE 19-25

Motivation

Existing approaches for low-light imaging:

- No-flash: burst photography, data driven (deep-learning based)

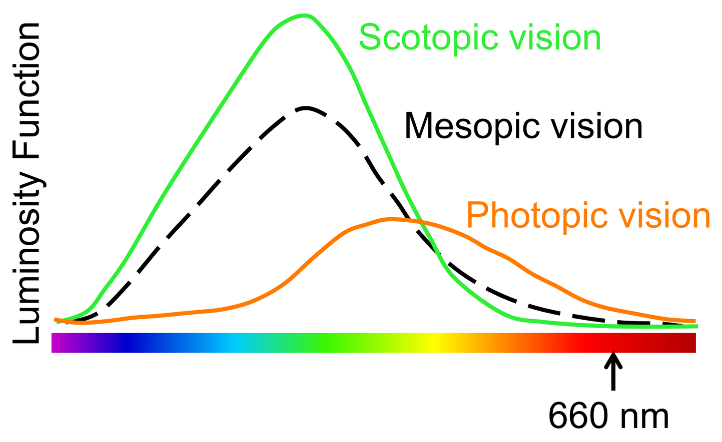
Fail when camera raw data suffers from poor signal-to-noise ratio. They still need adequate illumination. Difficult for low-light video reconstruction.

- Flash-based: white flash, invisible flash (like near-infrared (NIR))
  - White flash dazzles human eyes, not user-friendly.
  - RGB cameras do not have sensitivity to the invisible spectrum, thus an additional camera for NIR sensing is required [1].

Our Solution

A deep-red (660 nm) flash is proposed for low-light photography and videography. The design of this flash is in the consideration of the human visual system.

Photopic vision (cones) and scotopic vision (rods) are the vision of the eye under well-lit conditions (above 10 cd/m^2) and low-light conditions (below 10^{-3} cd/m^2), respectively. A combination of cones and rods forms *mesopic vision*.



Perceived Brightness

In extra dim conditions (around 10^{-2} cd/m^2), using white flash is over two orders of magnitude brighter than using the red flash as perceived by the human eye when the camera receives the same amount of signals.

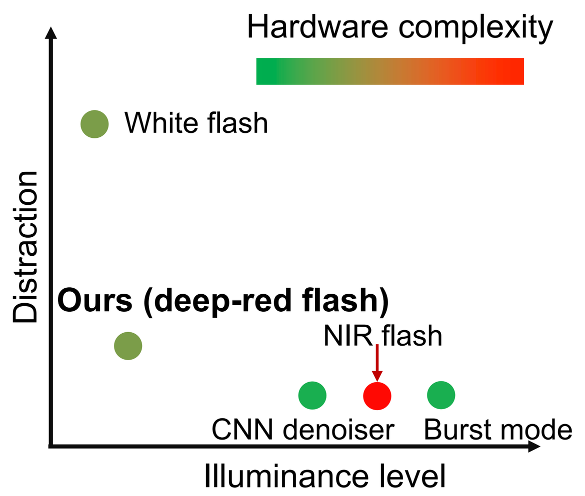
Luminance (cd/m^2)	Brightness gain
0.005	1/353.47
0.01	1/338.45
0.05	1/25.22
0.1	1/15.98

Night Vision

When exposed to light like a burst of white flash, the rods photobleach and lose their sensitivity to light [2]. Night vision is preserved using deep-red light.

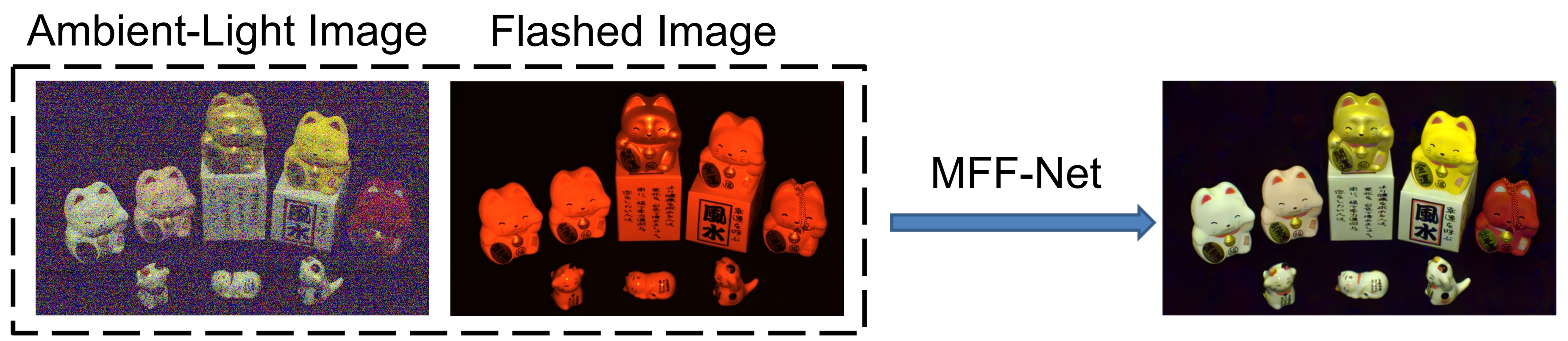
Hardware

A single RGB camera is capable of capturing both red-flash and no-flash image pairs, unnecessary to use an extra camera.



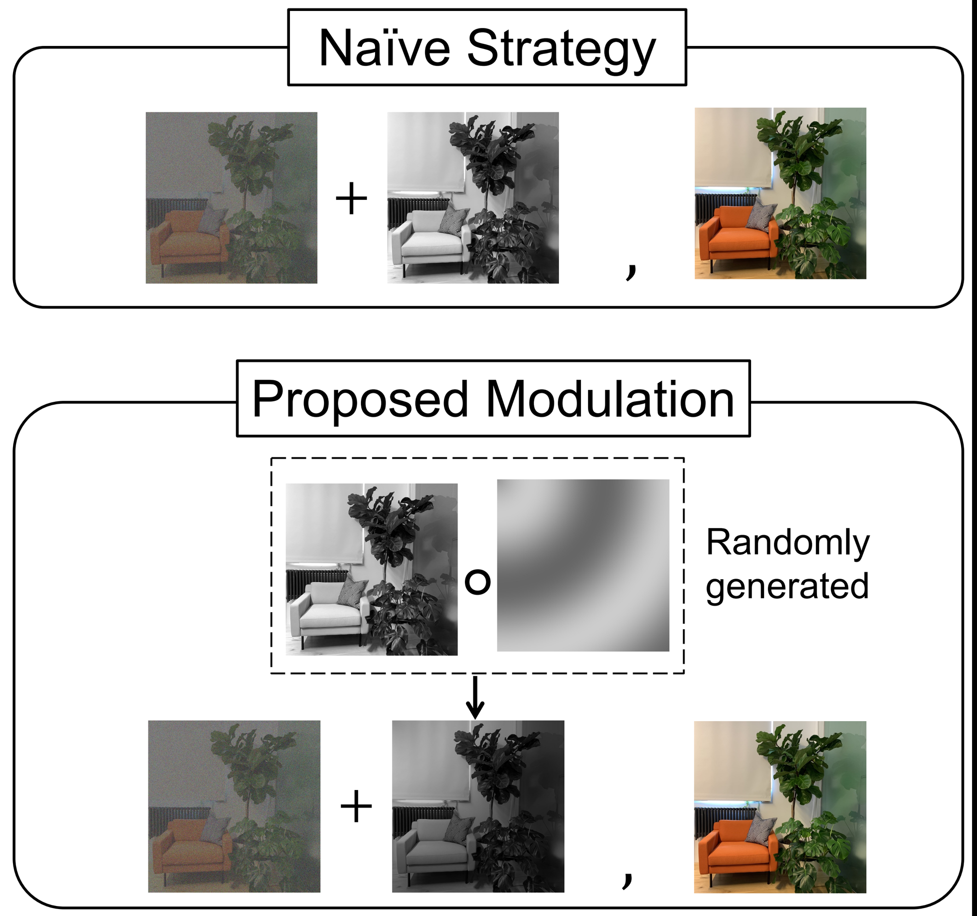
Photography

We trained a neural network, named as *mesopic flash fusion* network (MFF-Net), to fuse the ambient-light image and the flashed image.



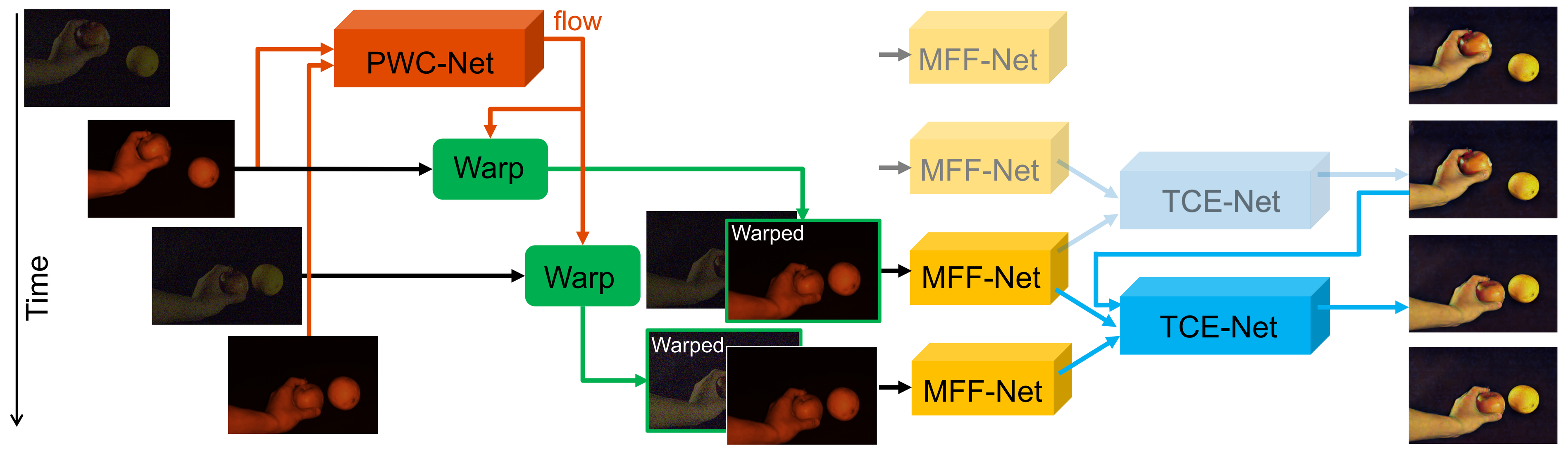
Training Data Generation

A low-frequency sinusoidal function with random period, amplitude and phase is applied as a modulator during training. Applying this modulation function to the flashed image decorrelates the intensity in the flashed and output images, while still retaining their edge correlation. This helps the network learn to exploit the edges rather than rely on the intensity in the flashed image.



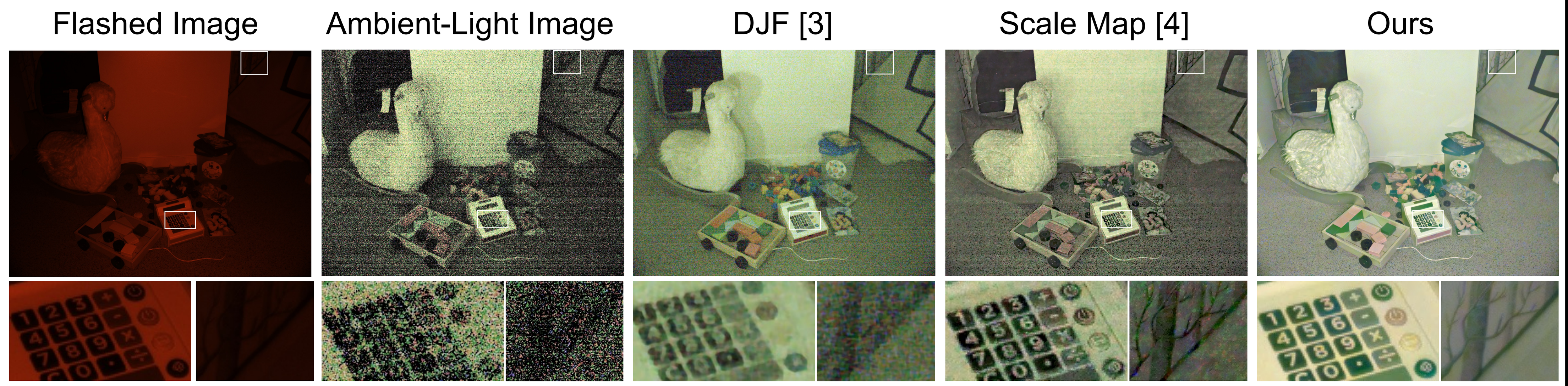
Videography

The input for videography is a sequence of interlaced ambient-light (no-flash) and flashed image pairs. We propose an effective pipeline for video reconstruction, which yields robust image alignment in the dark while preserving the original frame rate.



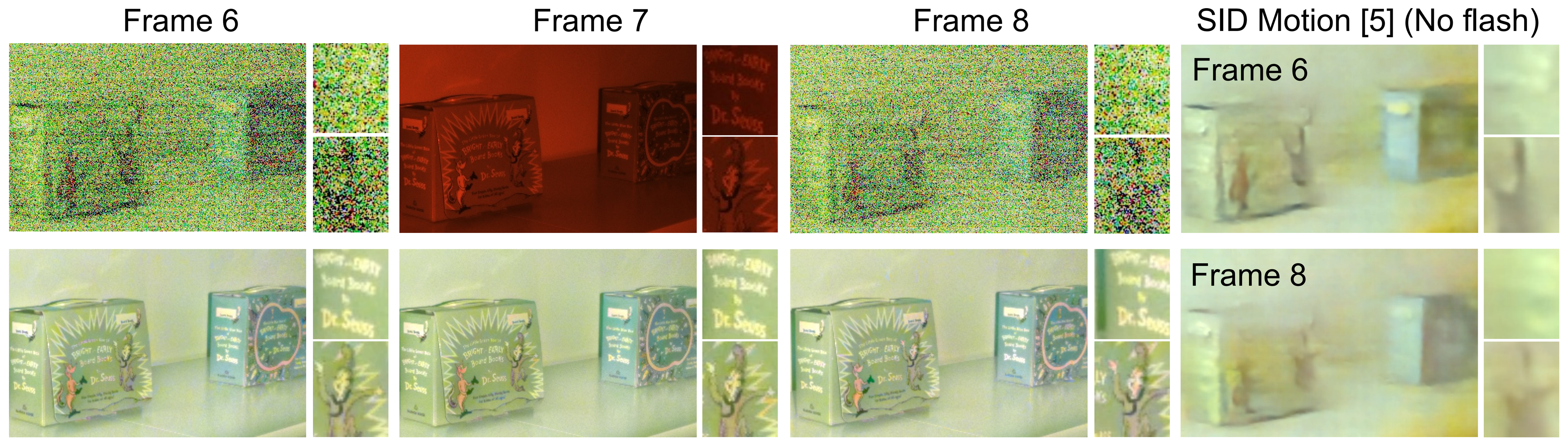
Results

Compared to alternative guided filtering methods:



The compared methods fail to retrieve fine image details in the rather challenging situation, whereas our method produces clean outputs with fine image features recovered.

Video reconstruction:



In extra dim environments, which are beyond the capability of no-flash methods, it is essential to use flashes to capture interested scene details. A flash can easily boost the light level by 100~1000X.

References

[1] Jian Wang, et al. "Stereoscopic dark flash for low-light photography". In ICCP, 2019.
[2] Mathew Alpern. "Effect of a bright light flash on dark adaptation of human rods". In Nature, 1971.
[3] Yijun Li, et al. "Joint image filtering with deep convolutional networks". In PAMI, 2019.
[4] Xiaoyong Shen, et al. "Multispectral joint image restoration via optimizing a scale map". In PAMI, 2015.
[5] Chen Chen, et al. "Seeing motion in the dark". In ICCV, 2019.