Image Enhancement For Unconstrained Environments

Sougato Bagchi

Oct 23rd, 2023

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Outline

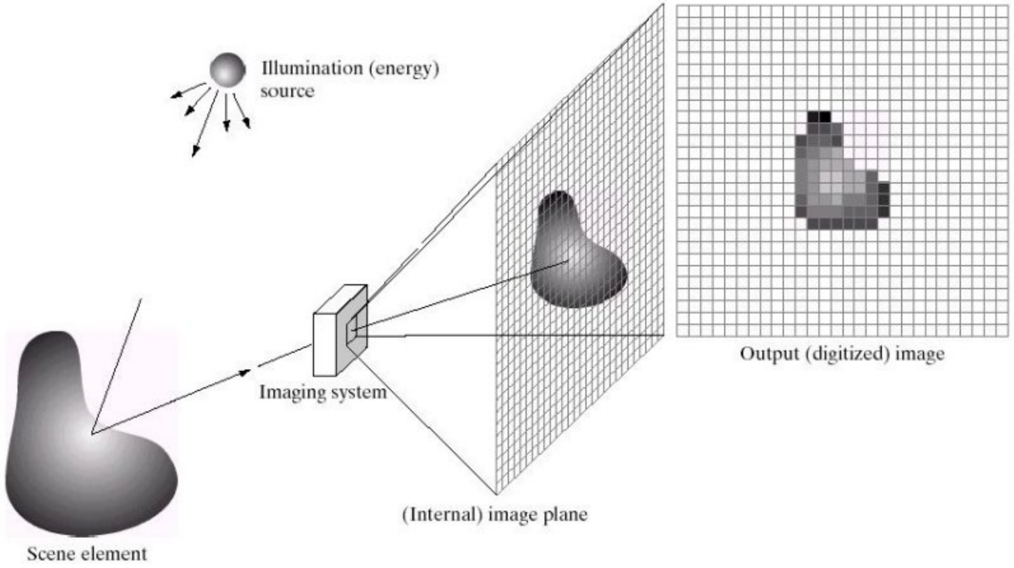
- Understanding the digital camera’s image acquisition process
- Noise characteristics, and its sources with denoising
- How to improve image illumination and handle color degradation
- Create a synthetic data which incorporates these factors for training a Super Resolution Generative Adversarial Network (SRGAN)
- Understand a model's behavior on different datasets

How are digital images formed?

Images are formed due to the photoelectric effect

Incident rays with unique energy levels, form patterns of charged pixels on the sensor.

After we get the unique pattern, we apply different mathematical operations for getting a final image.

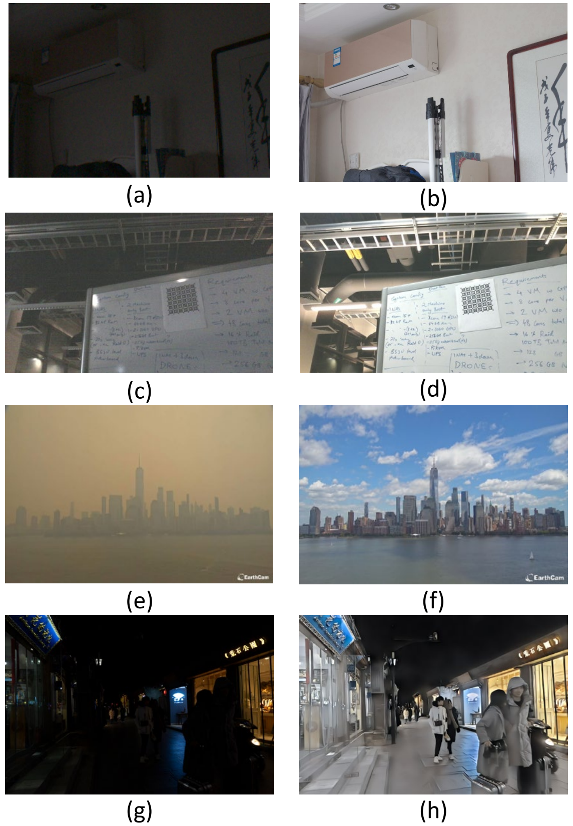


An example of the digital image acquiring process

E Woods, Richard, and Rafael C Gonzalez. "Digital image processing." (2008).

Non-Ideal Images vs Ideal Images

- non-ideal images can be due to extreme noise, improper illuminations and decolorization
- shared images from is in order of LOL, Pepper, EarthCam, VE-LOL-H datasets
- each row contributes to an image pair from a particular dataset
- left image: non ideal
- right image: ideal



How do we improve feature detection for a non-ideal image?

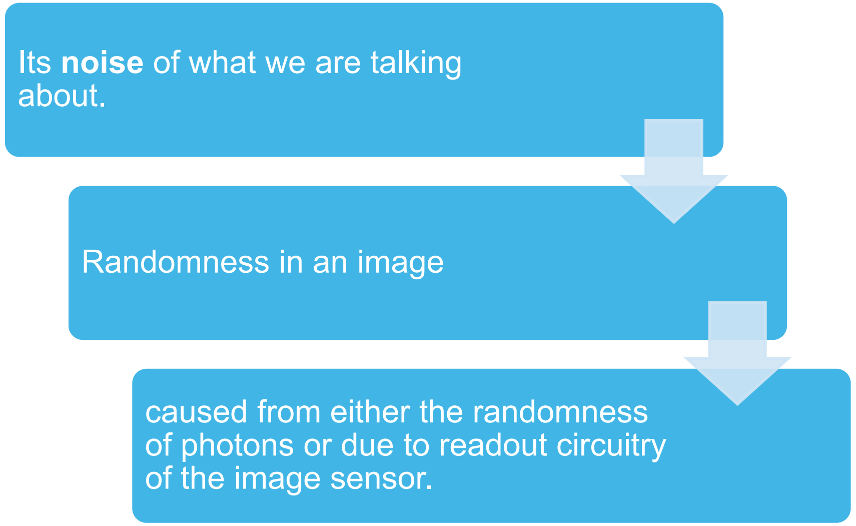
Image Acquisition Process

- Which factors affect an image ?
- Understand the image acquisition pipeline

Retinex Model

- Image can be distributed into two components
- $S = R \circ I$
- S: source image; R: réflectance; I: Illumination

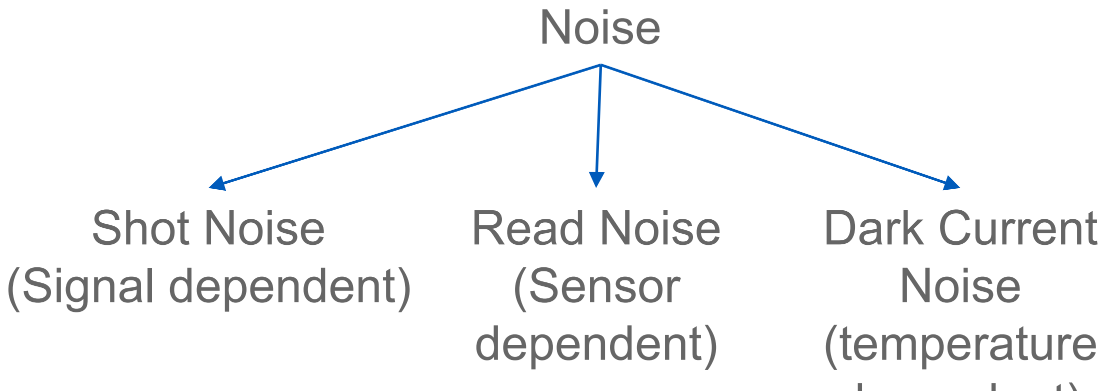
Noise



Its **noise** of what we are talking about.

Randomness in an image

caused from either the randomness of photons or due to readout circuitry of the image sensor.

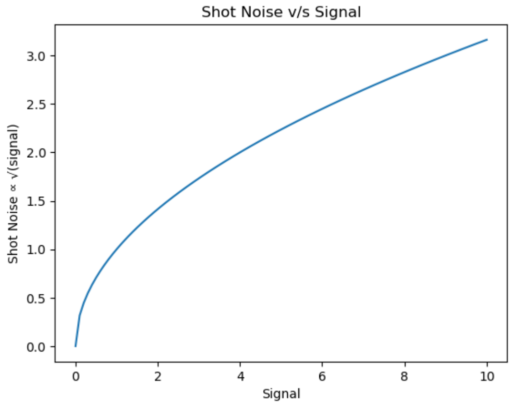


Noise

Shot Noise (Signal dependent)

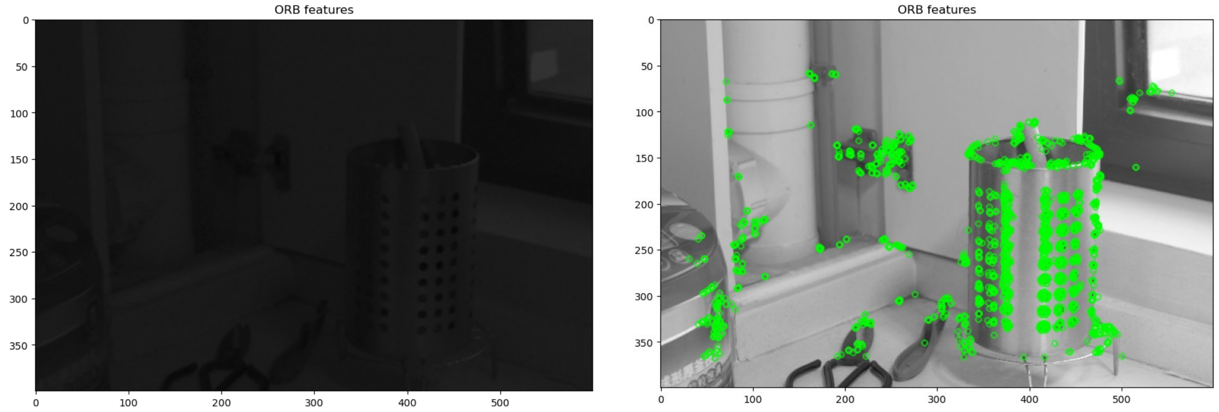
Read Noise (Sensor dependent)

Dark Current Noise (temperature dependent)

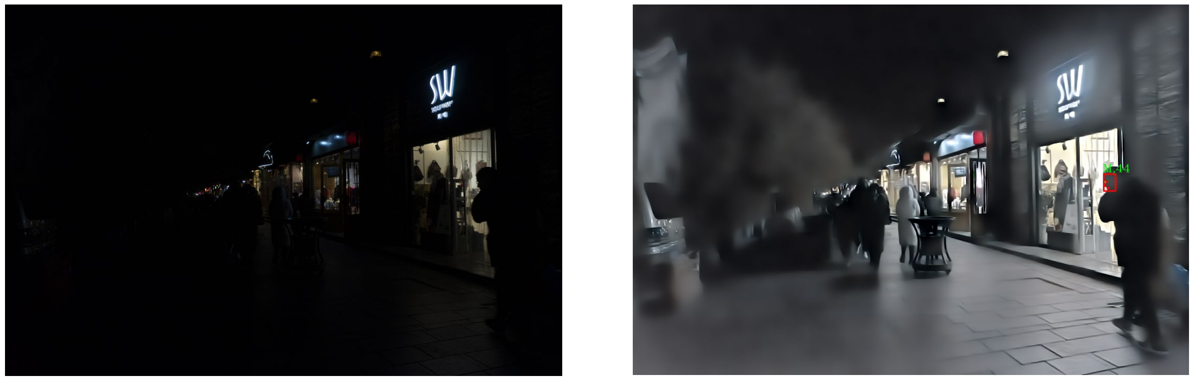


Characteristics curve of shot noise

Impact of non-ideal images



(a) No feature detected in low light conditions (b) Features detected in bright light conditions



(a) face detection algorithm completely failed

(b) face detection algorithm able to detect at least 1 face

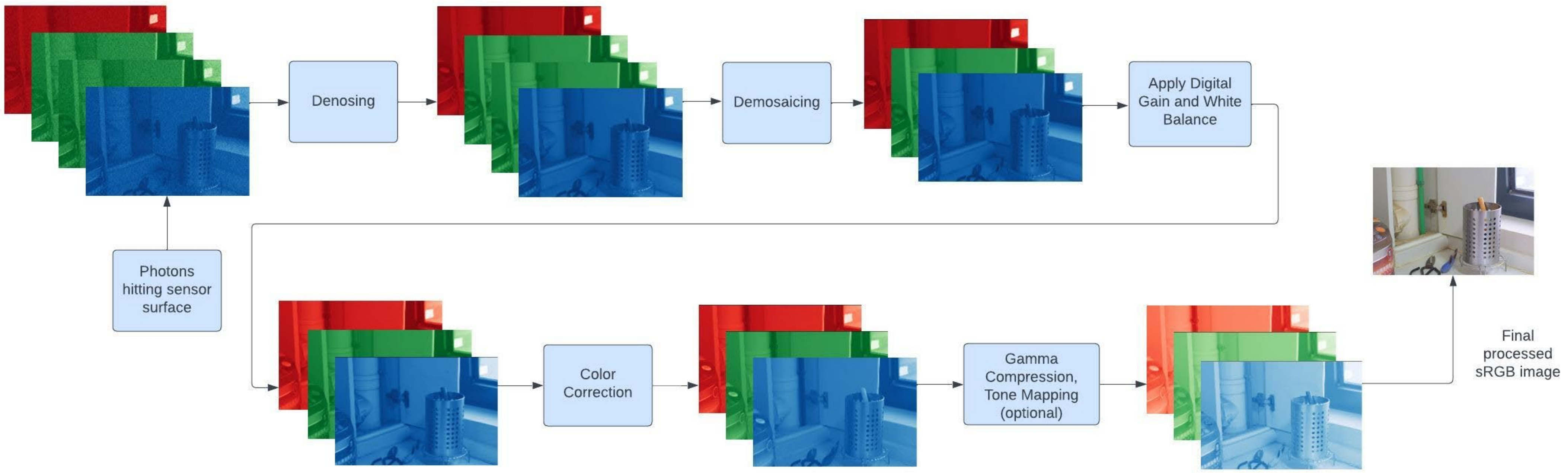
Comparing the detected ORB features, we found that

- Low light images hampers feature detection
- Tough to infer which leads to severe degradation of a particular application’s performance.
- Image src. : LOL dataset

RetinaFace [1] face detection network used on the VE-LOLH-H dataset

[1] J. Deng, J. Guo, E. Ververas, I. Kotsia and Z. S, "Retinaface: Single-shot multi-level face localisation in the wild.," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5203-5212).*, 2020.

Images Acquisition pipeline



Denosing

Demosaicing

Apply Digital Gain and White Balance

Photons hitting sensor surface

Color Correction

Gamma Compression, Tone Mapping (optional)

Final processed sRGB image

[2] Brooks, Tim, et al. "Unprocessing images for learned raw denoising." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

Prior works in image enhancement

Traditional Process

- Works only for a small set of data
- Tailor made solutions
- Considered denoising, and image illumination improvement as separate problems.

Noise Model

- Noise considered as Gaussian (not ideal)
- Denoising techniques, were dataset specific

Illumination improvement

- Tailor made modifications
- Ex: Y channel histogram modifications after image being transformed to YCbCr color space
- Results varied with datasets

Recent Approach

Noise Modeling

Neural networks trains on **paired data**, with one being noisy and the other being captured in proper circumstances.

Difficult to capture image pairs for the same scene.

Image Enhancement

Neural network trains on **illumination invariant color maps and reflectance maps**

Also needs **paired dataset** where the pairs are captured in different illumination scenario.

But we lack in good image pairs for training. Capturing is tedious.

Deep Neural Networks need data

- These models need paired data.
- Capturing a huge number of real-paired data is nearly impossible
- Previous studies on creating datasets used techniques such as lowering the brightness of well-exposed images to generate underexposed ones [4] or taking pictures with a short exposure time.
- Images created with these methods doesn’t model the noise properties well

[4] C. Wei et al., "Deep retinex decomposition for low-light enhancement," in *arXiv preprint arXiv:1808.04560 (2018)*.

Proposed approach

Create a framework which can:

- Generate low-light image from the ideal bright images with shot & read noise added. This creates image pair which is much needed for the image enhancement neural networks.
- Use of a GAN based network named LLFLOW [3], which incorporates the Modified Retinex Theory.
- Cross-Dataset evaluation for robustness
- Model characterization on different applications

[3] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI Conference on Artificial Intelligence.*, 2022.

LLFlow (Low Light Image Enhancement with Normalizing Flow)

Modified Retinex Theory

- $S = R \circ I + n$ | n : noise

Reflectance (R)

- Avoids direct comparison of x_l and x_{ref}
- Compares the color map between x_l and x_{ref} | x : image
- $D(\, g(\, C(x_l)\,),\, \mathrm{x}_{ref}\,)$
- $C(x) = \frac{x}{\mathrm{mean}_c(x)}$ | mean_c calculates the mean value of each pixel among the RGB channels

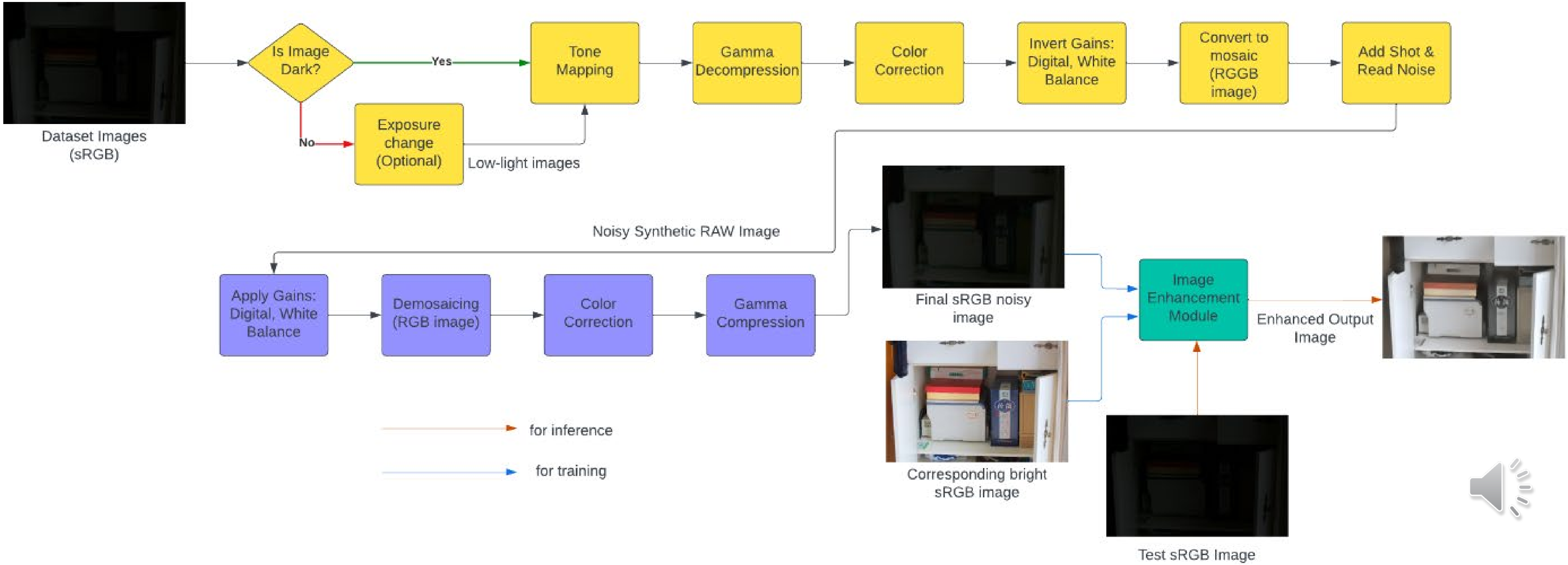
Illumination (I)

- Incorporates Histogram Equalization
- Increases Global Contrast

Noise (n)

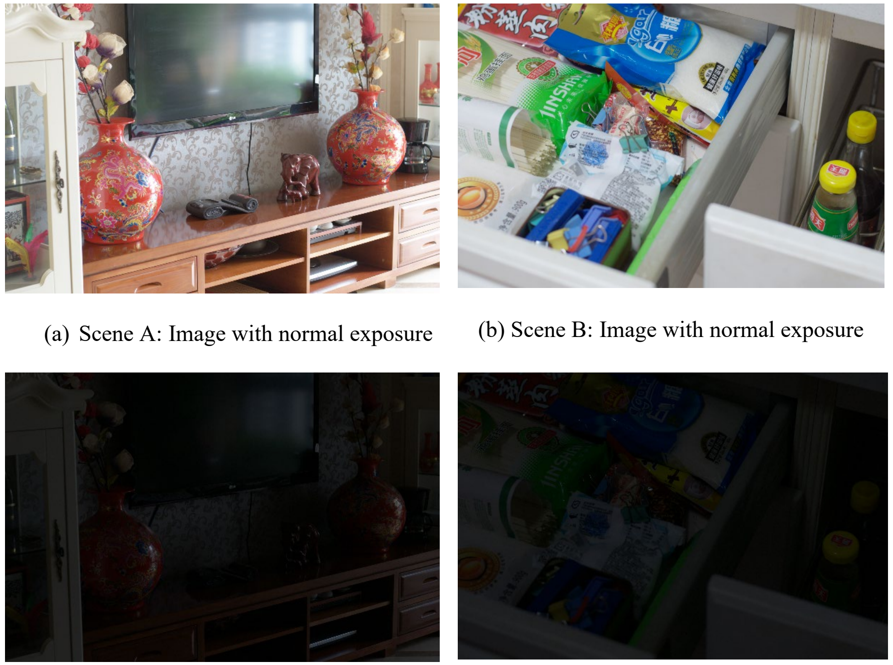
- A noise map is generated
- $N(x) = \max(abs(\nabla_x C(x)), abs(\nabla_y C(x))$
- ∇_x & ∇_y denotes the gradient maps in the x & y directions
- $\max(\mathrm{x}, \mathrm{y})$ function returns the max value between x & y at pixel channel level

Proposed Approach



LOL dataset

- 500 image pairs with low-light as well as normal-light images.
- Parameters like exposure time and ISO changed to get a low-light image



(a) Scene A: Image with normal exposure

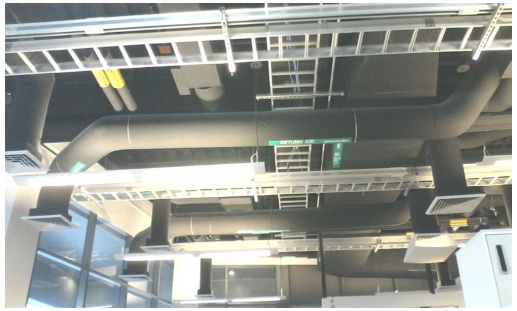
(b) Scene B: Image with normal exposure

(c) Scene A: Image with low exposure

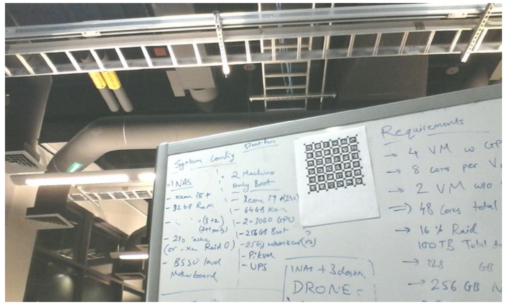
(d) Scene B: Image with low exposure

Pepper Data

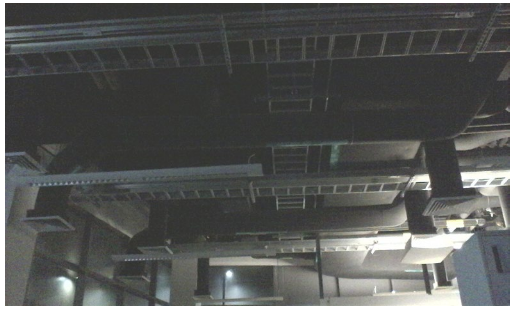
- 12 image pairs captured from the pepper robot's inbuilt camera.



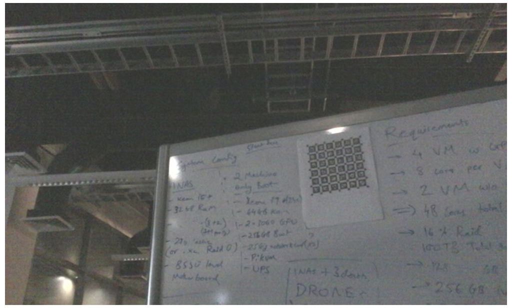
(a) Scene A: Image with normal illumination



(b) Scene B: Image with normal illumination



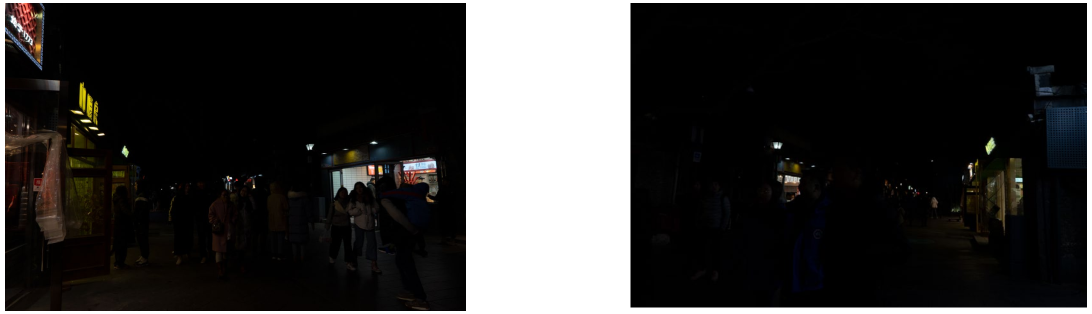
(c) Scene A: Image with low illumination



(d) Scene B: Image with low illumination

VE-LOL-H dataset [4]

- 10,940 images in total
- Downscaled to a resolution of 1080x720 from 6Kx4K
- Captured in busy streets on low-light scenario
- Sony α6000 and Sony α7 E-mount cameras used
- Images have human face, with manual annotations



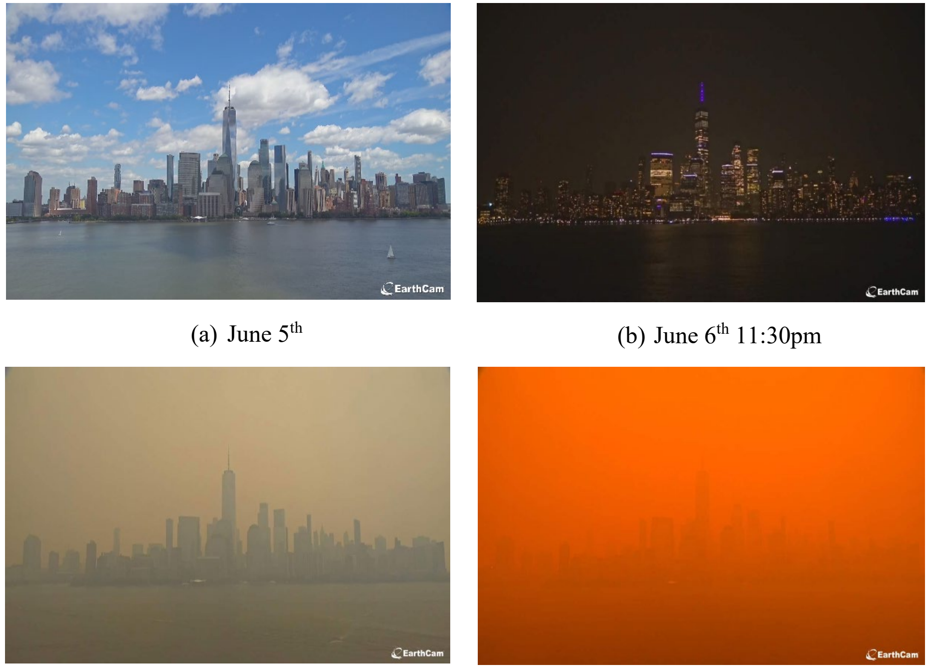
(a) (b)

Images from the VE-LOL-H dataset

[4] J. Liu, D. Xu, W. Yang, M. Fan and H. Huang, "Benchmarking Low-Light Image Enhancement and Beyond," in *International Journal of Computer Vision (2021)*, 2020.

EarthCam Data

- Earthcam pictures of New York City from June 6th to June 8th
- Captured the change in texture of the image due to the Canadian Wildfire
- Helped in better characterization of these GANs



(a) June 5th

(b) June 6th 11:30pm

(a) June 7th 12pm

(b) June 7th 2pm

https://www.earthcam.com/usa/newyork/worldtradecenter

Experiments

- LOL datasets modified in different ways
- Models trained on these variants for performance comparison (Ablation study)
- 500 image pairs: 485 train - 15 for test.

Performance of different variants of the LLFlow network

- Selected best models based on mean performance
- Narrowed down the choice using standard deviation

Model type	PSNR (↑)	SSIM (↑)	LPIPS [5] (↓)
custom v2_RAW_noisy	$\mu = 20.21$ $\sigma = 2.640$ $\sigma^2 = 6.973$	$\mu = 0.70$ $\sigma = 0.105$ $\sigma^2 = 0.011$	$\mu = 0.37$ $\sigma = 0.081$ $\sigma^2 = 0.006$
custom v3_RAW_noisy	$\mu = 20.43$ **$\sigma = 1.128$** **$\sigma^2 = 1.27$**	**$\mu = 0.743$** **$\sigma = 0.072$** **$\sigma^2 = 0.005$**	$\mu = 0.406$ **$\sigma = 0.070$** **$\sigma^2 = 0.005$**
pretrained	**$\mu = 20.71$** $\sigma = 2.909$ $\sigma^2 = 8.46$	$\mu = 0.70$ $\sigma = 0.182$ $\sigma^2 = 0.033$	**$\mu = 0.496$** $\sigma = 0.311$ $\sigma^2 = 0.096$
custom v2_ no_ccm_wb_gain	$\mu = 19.45$ $\sigma = 3.294$ $\sigma^2 = 10.851$	$\mu = 0.71$ $\sigma = 0.096$ $\sigma^2 = 0.009$	$\mu = 0.36$ $\sigma = 0.076$ $\sigma^2 = 0.006$
custom v3_no_ccm_wb_gain	$\mu = 19.46$ $\sigma = 1.914$ $\sigma^2 = 3.666$	$\mu = 0.72$ $\sigma = 0.052$ $\sigma^2 = 0.003$	$\mu = 0.41$ $\sigma = 0.066$ $\sigma^2 = 0.004$

Performance of the LLFlow models being tested on different variants of LOL datasets.

[5] Z. Richard, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric.," in *Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.*.

Comparing the type of noise {summarize the outcome}

Model type	Eval dataset	PSNR (↑)	SSIM (↑)	LPIPS (↓)
custom v3_raw_noisy	LOL	19.23	**0.78**	**0.34**
custom v3_LOL_AWGN	LOL	**20.16**	0.77	0.45

Model trained on Additive White Gaussian Noise (AWGN) gets outperformed in majority of the metrics, namely SSIM and LPIPS.

Performance

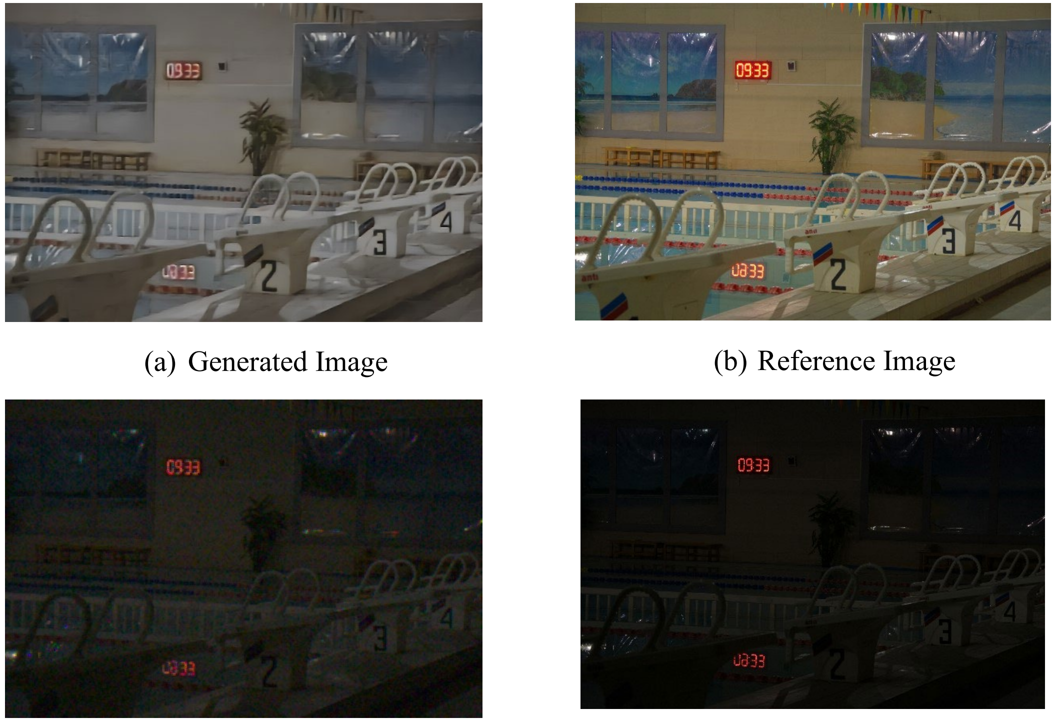
Model type	Eval data	PSNR (↑)	SSIM (↑)	LPIPS (↓)
custom v3_raw_noisy	Pepper	**13.56**	0.53	**0.49**
pretrained	Pepper	12.45	**0.54**	0.57
custom v2_ no_ccm_wb_gain	Pepper	13.04	0.44	0.52
custom v3_RAW_noisy	EarthCam	11.71	0.53	**0.68**
pretrained	EarthCam	**13.31**	**0.63**	0.69
custom v2_ no_ccm_wb_gain	EarthCam	12.31	**0.63**	0.73

- Our pepper dataset is meant for low-light image enhancement task (it contains proper paired datasets)
- “custom v3_raw_noisy” model outperforms the pretrained one in PSNR and LPIPS
- This model is performed according what we concluded using the LOL dataset performance comparison based on standard deviation and variance.

Model Outputs

LOL dataset

- Output from the model “custom v3_RAW_noisy”



(a) Generated Image

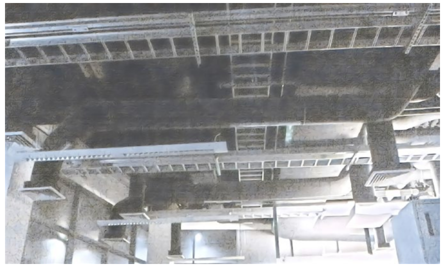
(b) Reference Image

(c) Noisy Low Light Image

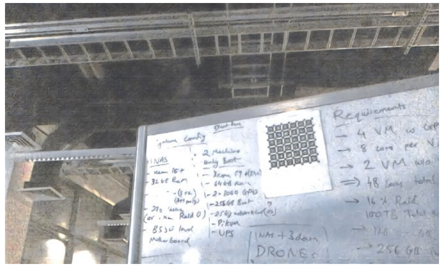
(d) Paired Low light Image

Pepper Data

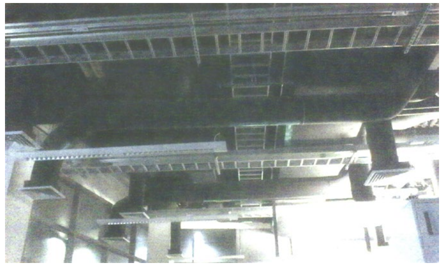
- Image (a) & (b) are generated from the "**custom v3_raw_noisy**" model
- Image (c) & (d) are the outputs from the "**pretrained**" model



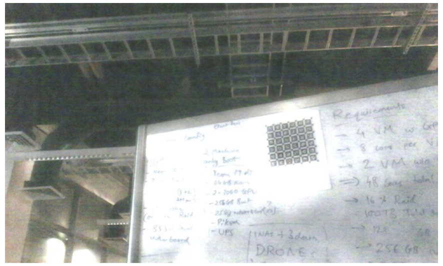
(a)



(b)



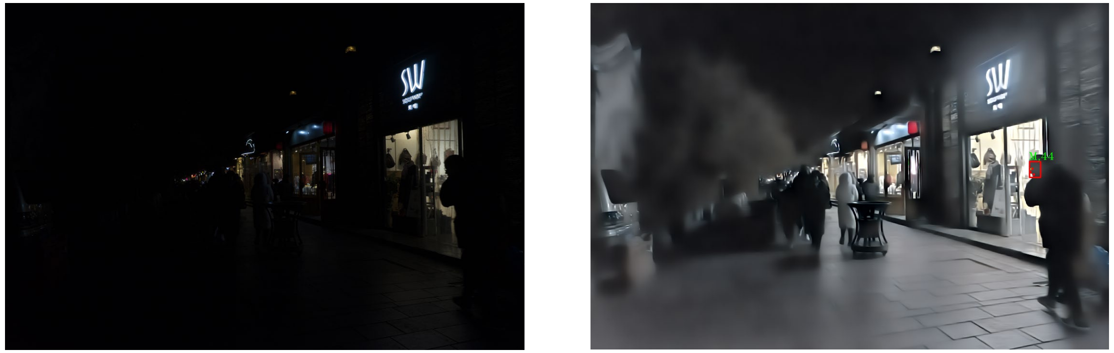
(c)



(d)

VE-LOL-H dataset

- Face detection algorithm (RetinaFace [4]) used for drawing bounding boxes
- The enhanced had 37.5% of increase in face detection than the original images.



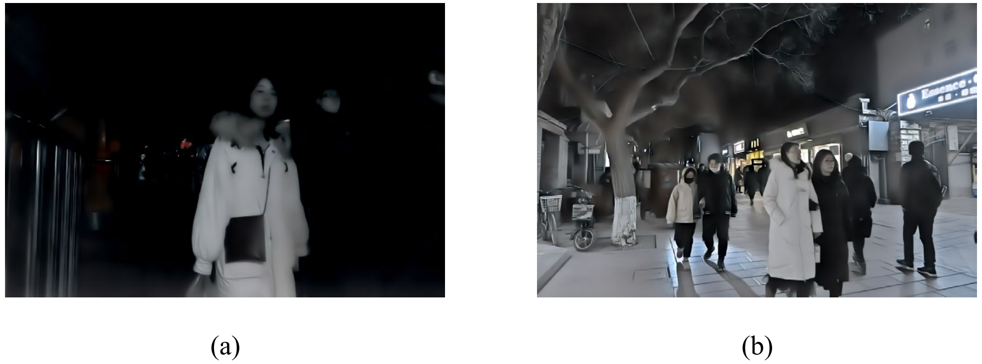
(a) (b)

Image (a) is from the VE-LOL-H dataset, where it's completely unprocessed whereas the right image (b) is the same image passed via our image enhancement network.

[4] Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5203-5212)

Face recognition on VE-LOL-H

- Face detection did improve
- Face recognition needs intricate facial details, which is lost



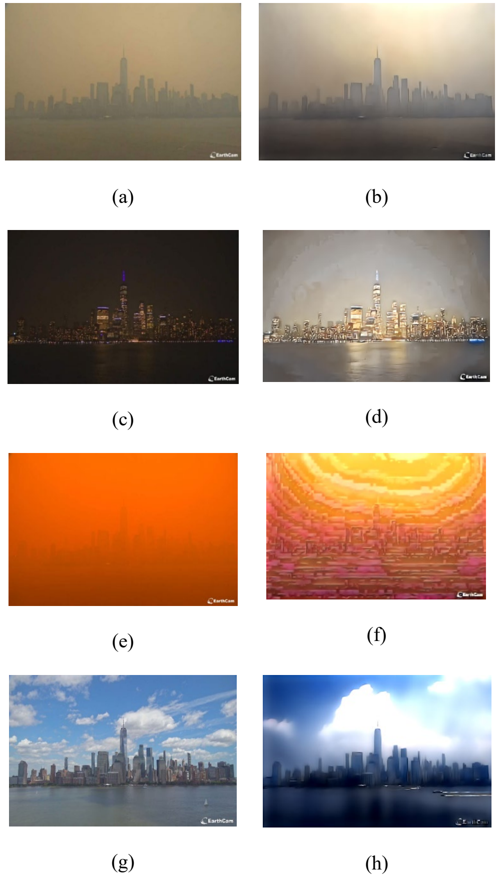
(a)

(b)

Typical output from our image enhancement network. The faces are generally smoothened while the illumination of these images is being adjusted by the network.

EarthCam dataset

- Images on the left side are the unaltered EarthCam image
- Images on the right are the corresponding outputs from the “**custom v3_raw_noisy**” model
- Performing as its supposed to only for pair (a)-(b), (c)-(d)
- The 3rd pair is severely affected due to the lack of details in the image.
- The 4th pair performs poorly due to lack of contextual awareness and gets hallucinated.



Comparison of the EarthCam image of New York city from June 5th to 7th

Conclusion

- Adding shot + read noise while training helped in improving the model’s image enhancement performance.
- Still the datasets lack in modeling all the possible real-life circumstances.
- Ways to improve the model performance:
    - Collect massive amount of data
    - Contextual awareness

Thanks

University at Buffalo
Department of Computer Science
and Engineering
School of Engineering and Applied Sciences