

A Torch Without Light: Low-Light Imaging for Mobile phones

Dipak Bange, Abhishek Gaikwad, Tejas Gajare, Aditya Khadse, Swati Shinde

Abstract - Photography used to be a hobby that required equipment such as a professional camera. Today, photography has evolved to be a daily activity conducted on an unprecedented scale due to the adoption of camera into smartphones. Mobile phone cameras are on the way to completely replace other forms of camera due to their portability and quality. Millions of images are captured on mobile devices across the globe. These images are clear and crisp. But all these images are captured in daylight. Images taken in low illumination essentially turn out to be too dark to be comprehensible. Research shows that current solutions to this problem work for dim to moderate light level but fail in extreme low light. There are certain problems involved with these techniques. Firstly, image denoising relies on image priors limiting the situations on what it will work on. Other deep learning techniques work on synthetic data and cannot be proficient on real data. Secondly, Low light image enhancement assumes that images already contain a good representation of scene content. This paper proposes to capture low illumination images and transform them to high quality images using end to end fully convolutional neural network trained on our data set of raw images shot in low aperture and their corresponding high aperture raw images. As an outcome, we will be able to transform images to high quality and identify objects.

Index Terms – Introduction, Existing Methods, Dataset, Using Transfer Learning, Training, Results, Conclusion, Future Work, References

Keywords: Artificial Intelligence, Computer Vision, Convolution Neural Network, Low-light Photography.

Abbreviations:

ARW: Sony Alpha Raw
BM3D: Block-Matching and 3D Filtering
DSLR: Digital Single-Lens Reflex
DNG: Digital Negative Specification
ISO: International Organization of Standardization
RGGB: Red Green Green Blue

I. INTRODUCTION

Mobile phones have become a staple for every person's needs. Further, cameras in mobile phones have started to compete against professional cameras developed only for photography. But due to limited hardware and space constraints in a mobile phone, cameras are unable to outperform the professional cameras. The sensors used in mobile phones are far inferior to the ones found in

Revised Manuscript Received on October 15, 2019

* Correspondence Author

Dipak Bange, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: tejasgajare24@gmail.com

Abhishek Gaikwad, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: gaikwadabhishek1997@gmail.com

Tejas Gajare*, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: tejasgajare24@gmail.com

Aditya Khadse, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: akk5597@gmail.com

Dr. Swati V. Shinde, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: swaatii.shinde@gmail.com

professional cameras. Of the few spaces that mobile cameras lack, is low light images. Mobile cameras have been using flash for the past few years to enable the user to capture images in low light. But the images captured do not truly see what the human eyes had intended to see.

We aim to enable mobile cameras to be able to compete with the professional cameras through the use of artificial intelligence. The techniques we aim to use are based on software and hence do not add to the cost of building the mobile phone. Our method abandons the traditional image processing pipeline to move to a more data driven pipeline that is built using examples of similar past experiences rather than the image priors, which are details of images that are known prior to the image even being captured. Abandoning the traditional pipeline helps us to achieve a more data driven approach.



Figure 1: Image Captured in Extreme Low Light vs. Enhanced Image produced from our Convolutional Neural Network

II. EXISTING METHODS

A. Manual Control

Digital cameras have had an important feature that has been taken for granted today. They have the ability to change their ISO on the fly. ISO here refers to the camera sensor's sensitivity to light. A high ISO means the sensor would be more sensitive to light and would increase the overall brightness of the image. This is the most basic method of improving visibility of a scene in low-light images.

Another manual control is to increase the exposure of the camera sensor. It can be controlled by two parameters: shutter speed and aperture. Shutter speed is the amount of time that the shutter remains open to capture the scene. A low shutter speed means that the shutter stays open for very small amount of time. The amount of light that enters the sensor is proportional to the time that the sensor is able to capture a light. For example, an image taken with shutter speed as 1/800 seconds would be darker than an image of the same scene taken with shutter speed as 1 second. A high exposure helps to get a brighter image as more photons can be captured by the camera sensor. A camera works similar to the human eye. There exists an opening that allows light to enter a sensor that captures the light and convert it to an image. Aperture can be compared with the size of the pupil of the

human eye. It controls the amount of light that can enter and be captured on the camera's sensor. A high aperture allows more light to enter and improve visibility in low-light images. Low-light images captured this using traditional pipeline is not perceptually good as it has a high amount of amount of noise. Several denoising techniques are used to further improve the visibility within the image. One of the most used techniques is 3D transform-domain filtering (BM3D).

B. BM3D

In 2007, K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian proposed a novel image denoising strategy based on an enhanced sparse representation in transform domain [1]. The idea was to use 2D image fragments found using matching of reference blocks and group them into 3D data arrays. They called these 3D data arrays groups. Further, they used collaborative filtering for the 3D groups. The filtering was divided into 3 successive steps: 3D transformation of a group, shrinkage of the transform spectrum and an inverse 3D transformation. The estimated 2D fragments were then returned to their locations. Due to multiple matching 2D fragments, they would be required to aggregate to take advantage of the redundancy.

BM3D has outperformed other recent denoising techniques when used on real low light images [2]. BM3D fails to outperform other data driven techniques as they still rely on processed image data rather than raw sensor data. The method also relies on one image of the scene. Another method tackled low-light imaging using burst photography.

C. Burst photography for High Dynamic Range and low-light imaging on mobile cameras

In 2016, Hasinoff et al [3] proposed a computational photography pipeline that can capture, align and merge a burst of frames to reduce noise and increase dynamic range. The pipeline takes in the burst of raw frames. It then aligns the burst to merge. Then it applies color and tone mapping to produce a single full-resolution output.

The cameras in mobile phones allow auto exposure adjustment as the user moves the camera around. To get improved results in high dynamic range, they developed their own algorithm for adjusting the auto exposure within the mobile camera.

The major problem with burst images is that if the scene contains a moving object, it is at a different position in every frame. It becomes difficult to find the best position for the object. Hence, burst images introduce blurred objects, or ghosting if the objects move a lot faster than the shutter speed. To overcome this, they used temporal mean with alignment and merged it robustly to achieve almost zero blur compared to the temporal mean.

D. Learning to See in The Dark

At CVPR 2018, C Chen et al [4] proposed a pipeline for processing low-light images based on end-to-end training of a convolutional neural network on short exposure images and corresponding long exposure images. The neural network would be trained on raw sensor data rather than output images. This improves performances over other methods since this method does not rely on image priors, which are assumptions made about the image and its data prior to obtaining the image.

This method works by first training a fully convolutional neural network with short exposure images, which are captured in extreme low-light images. The architecture used in the neural network is U-net [5]. Further they used 2 cameras to form the dataset required for training the neural

network. The See-in-the-Dark dataset has images from Sony α 7S II and Fujifilm X-T2. There are 5094 short exposure images with their reference long exposure images.

III. DATASET

We collected a dataset of 592 images made up of 165 unique scenes. The dataset was collected using the camera of OnePlus 6 (16 MP Sony Exmor IMX519 (f/1.7, OIS + EIS) + 20 MP Sony Exmor IMX376K (f/1.7)). The images were stored in RAW format supported by smartphones, which is DNG format. The camera has a Bayer filter and has a black level of 63. The Bayer filter has pattern of RGGB. Every unique scene has at least one short exposure image and a corresponding long exposure image. We captured the long exposure image by trying to capture the scene by changing the values until we reached an image that looked perceptually good. The objective of collecting the dataset was to enable training of a model for the specific task of processing images captured on a smartphone. This dataset can be considered as the target task.

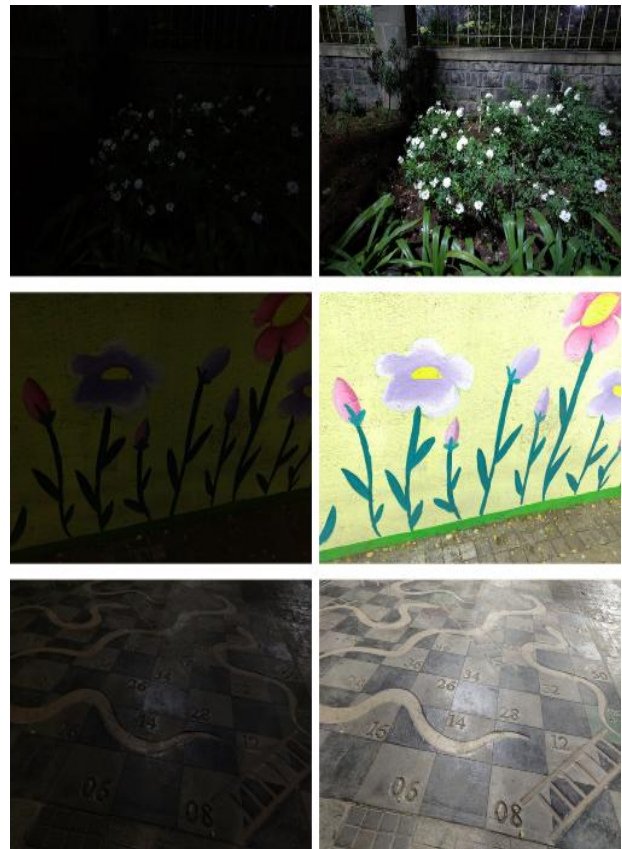


Figure 2: Left Column shows input images and Right column shows the corresponding output images

Every unique scene has at least one short exposure image and a corresponding long exposure image. We captured the long exposure image by trying to capture the scene by changing the values until we reached an image that looked perceptually good.

The objective of collecting the dataset was to enable training of a model for the specific task of processing images captured on a smartphone. This dataset can be considered as the target task.

IV. USING TRANSFER LEARNING

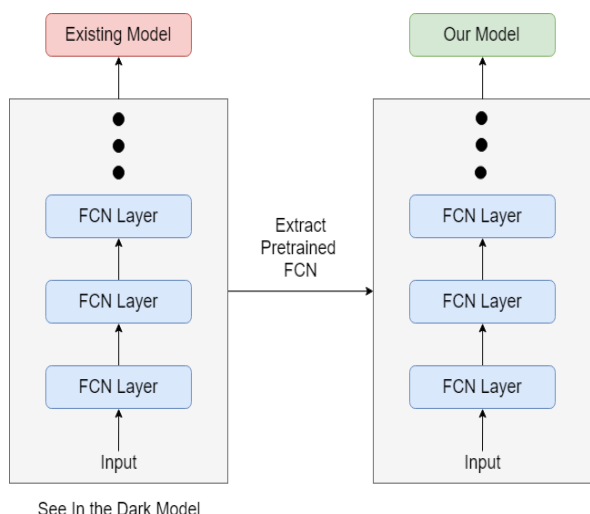


Figure 3: Transfer learning model

Our current solution is based on the pretrained model trained on raw images on Sony *a7S II* camera (in .arw format) with black level of 512 and Bayer filter has pattern of RGGB. Using Transfer Learning [6] over this pretrained model we trained the network for the dataset of raw images (in .dng format) captured on OnePlus 6 to increase the accuracy and colors of the image for low light condition on mobile phones.

V. TRAINING

We train the fully convolutional neural network [7] using the L1 loss and the Adam optimizer. During training, the input to the network is the raw data of the short-exposed image and the ground truth is the corresponding long-exposure image in sRGB space (processed by libraw, a raw image processing library). We train one network for OnePlus 6. The amplification ratio is set to be the exposure difference between the input and reference images (e.g., x100, x250, or x300) for both training and testing. In each iteration, we randomly crop a 512x512 patch for training and apply random flipping and rotation for data augmentation. The learning rate is initially set to 10⁻⁴ and is reduced to 10⁻⁵ after 2000 epochs. Training proceeds for 5000 epochs.

VI. METHOD

The working of the experiment can be broadly classified in three phases:

A. Preprocessing:

The mobile application captures raw (DNG format) image. Rawpy is used to open the image. The image is in Bayer filter (R, G, G, B arranged in a 2x2 matrix for complete image) format and hence is converted into 4- channels (RGBG) of numpy array. Black level (63 for One Plus 6) is subtracted from the channels and divided by 216-1 for normalizing the values. The ground truth image converted to its 16-bit counterpart using Rawpy and pixel values are normalized.

B. Processing

After pre-processing, the 4 channels are passed to the fully convolutional neural network. Fully convolutional networks (FCNs) are a rich class of models that address many pixel wise tasks. FCNs for semantic segmentation dramatically

improve accuracy by transferring pre trained classifier weights, fusing different layer representations, and learning end-to-end on whole images. End-to-end, pixel-to-pixel operation simultaneously simplifies and speeds up learning and inference [7]. U-net architecture is used for the network as it is more effective [4]. A random block of 512 x 512 is selected from image and fed to the network. The network then uses max pooling, ReLU and up sampling functions throughout the network [5]. Amplification ratio is fed to the network in order to define how much brighter the image must be. The ratio we chose is depends on ground truth and input image exposure. The network outputs the processed image directly in sRGB space.

C. Post-Processing

Image pixel values converted in range 0-255. The image (sRGB space) obtained after processing is the enhanced version of the input image converted to .png extension. It is sent back again from the computer to the mobile phone where it can be viewed in the application. In our experiment have not used any post processing image enhancing algorithm as we wanted to deal with raw images as they are.

D. Server-Side Processing

To provide user a seamless experience of capturing a visually perceivable image, our system shifts the system load onto server side rather than on client's mobile phone. Computationally intensive tasks are executed onto the server. The RAW image captured by the user is transferred to the server where the trained model exists. This file transfer is performed using two methods,

- 1) Peer-to-Peer File, Transfer
- 2) Http Multipart Upload

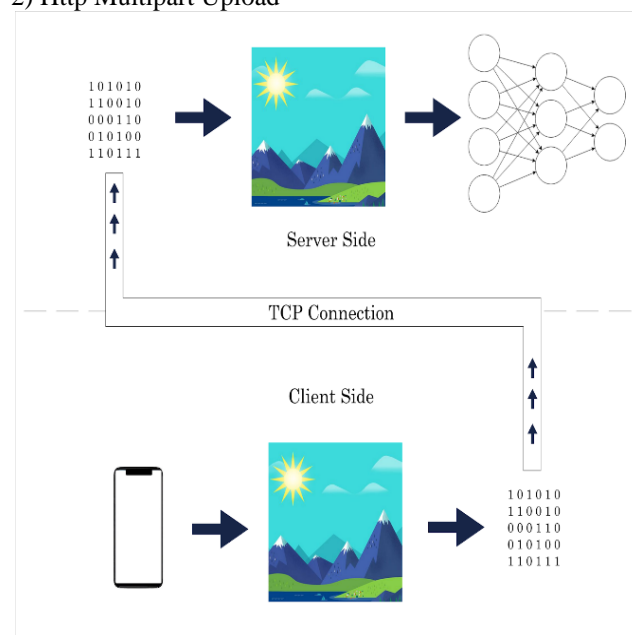


Figure 4: File Transfer from mobile device to Server-side for further processing

VII. RESULTS



Figure 5: Input image shot on mobile phone (.dng) vs. Output of pretrained model on digital SLR cameras



Figure 6: Input image shot on mobile phone (.dng) vs. Output from model trained on mobile phone cameras



Figure 7: Input image shot on mobile phone (.dng) vs. Output from model trained on DSLR cameras and then on mobile phone cameras using transfer learning



Figure 8: Actual Output (ground truth) vs. Our final output.

VIII. CONCLUSION

Imaging in the dark, at sub-lux conditions presents a challenge that traditional pipeline fails to overcome. In this system, we present a methodology through which a visually enhanced image is retrieved from an image which is captured in extreme low light conditions. Computational processing of low-light images proves to be inefficient and far less effective for the extreme low light conditions under consideration. A data driven approach is implemented to overcome this scenario. Specifically, we train deep neural networks to learn the image processing pipeline for low light raw data, including color transformations, demosaicing, noise reduction, and image enhancement.

Current system that is presented in the "Learning to see in the dark" paper [4] is the basis of our product. The deep neural network implemented in the aforementioned paper was trained using raw images captured through DSLR. This network is able to enhance images that are captured with the similar setup. To design a mobile application that shall enhance images captured from its comparatively smaller sensor, we enact the method of Transfer Learning. Through this methodology we train the existing trained model with our dataset captured through a mobile device. This provides us with a model suitable to satisfy our use case of enhancement of mobile camera images. We present a mobile application through which users can capture images in extreme low light conditions and enhance them into visually per perceivable images. This application is based on client-server architecture. The mobile application will upload the raw image onto the server for processing and the corresponding enhanced image will be retrieved.

IX. FUTURE WORK

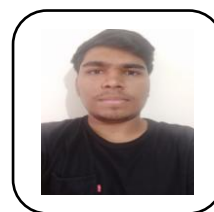
Current limitation of the presented pipeline is that the amplification ratio must be chosen externally. It would be useful to infer a good amplification ratio from the input, akin to Auto ISO. Furthermore, we currently assume that a dedicated network is trained for a given camera sensor. Our preliminary experiments with cross-sensor generalization are encouraging, and future work could further study the generalization abilities of low-light imaging networks. Another opportunity for future work is runtime optimization. The presented pipeline takes 0.66 seconds to process full-resolution One plus 6 images; this is not fast enough for real-time processing at full resolution, although a

low-resolution preview can be produced in real time. We expect future work to yield further improvements in image quality, for example by systematically optimizing the network architecture and training procedure.

REFERENCES

1. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering IEEE Transactions on Image Processing", 16(8), 2007.
2. V. Jain and H. S. Seung, "Natural image denoising with convolutional networks", NIPS, Feb 2008.
3. S. W. Hasino, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras", ACM Transactions on Graphics, 35(6), 2016.
4. Chen, Qifeng Chen, Jia Xu, Vladlen Koltun, "Learning to See in the Dark", CVPR, 2018.
5. Olaf Ronneberger, Philipp Fischer, Thomas Bro, "U-Net - Convolutional Networks for Biomedical Image Segmentation", May 2015.
6. Lisa A. Burke, Holly M. Hutchins, "Training Transfer: An Integrative Literature Review", Sept 2007.
7. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", CVPR, 2015.

AUTHORS PROFILE



Dipak Bange has completed his B.E. from Pimpri Chinchwad College of Engineering, Pune in Computer Engineering, class of 2019. He is currently working as Software Developer in different technologies at Coditation Systems Pvt. Ltd. He is very passionate about machine learning and done various projects in machine learning. He has keen interest in Soft Computing and Deep Learning. He and his team received "Runners' Up" best project award at CSI Regional Project Competition for 'A Torch Without Light'. He has been a finalist in 'Smart India Hackathon' conducted at Noida and also at 'International Collegiate Programming Contest (ICPC) 2019' Regionals, Gwalior.



Abhishek Gaikwad has graduated from Pimpri Chinchwad College of Engineering, Pune (PCCOE) in Computer Engineering in the year 2019. He is currently working as Software Developer at "Teachers Insurance and Annuity Association (TIAA)". He is passionate about technology and has a keen interest in Artificial Intelligence and Computer Vision. He and his team received "Runners' Up" for best project award at "CSI Regional Project Competition" for "A Torch Without Light". He has been a finalist for "Smart India Hackathon 2018" and also at "International Collegiate Programming Contest (ICPC) Regionals". He was the 'Chairperson' for "PCCOE ACM Student Chapter" and 'President' of "Computer Engineering Students'



A Torch Without Light: Low-Light Imaging for Mobile phones

Association", which organised various technical and community activities for the students of his institute.



Tejas Gajare has graduated from Pimpri Chinchwad College of Engineering, Pune (PCCOE) in Computer Engineering in the year 2019. He is currently working as Software Developer at Yardi Systems. His interests include Robotics and Machine Learning. He and his team received "Runners' Up" best project award at CSI Regional Project Competition for 'A Torch Without Light'. He has been a finalist in

'Smart India Hackathon' conducted at Noida and also at 'International Collegiate Programming Contest (ICPC) Regionals, Gwalior. He has also developed software in domain of Nutrition Care in Pune (Patent Pending). He was a part of Team Automotons of PCCOE where they stood 12th (Asia-Pacific) in ABU Robocon 2017.



Aditya Khadse has graduated from Pimpri Chinchwad College of Engineering, Pune in 2019. He received 'Most Innovative Student of the Year 2019' for outstanding contribution in the field of research and innovation in PCCOE. He currently works as a Systems Engineer at Digitate, which is a venture of Tata Consultancy Services (TCS) Ltd. He was "Webmaster" for

PCCOE ACM Student Chapter and various other clubs in his college. He and his team received "Runners' Up" best project award at CSI Regional Project Competition for 'A Torch Without Light'. He has been a finalist in 'Smart India Hackathon 2018' conducted at Noida and also at 'International Collegiate Programming Contest (ICPC) 2019' Regionals, Gwalior.



Dr. Swati V. Shinde has completed Ph.D. in Computer Science and Engineering, from Swami Ramanand Teertha Marathwada University, Nanded, ME degree in Computer Engineering from Bharti Vidyapeeth, Pune, in 2006 and BE(Computer Science & Engineering) degree from Swami Ramanand Teertha Marathwada University, Nanded in 2001. She

has total 18 years of teaching experience and currently she is working as a Professor in Pimpri Chinchwad College of Engineering, Pune. Her research interests include Machine Learning, Deep Learning, Soft Computing, Artificial Neural Network and Fuzzy Logic. She has published 59 research papers in reputed conferences and journals. Two of these are Published in the prestigious Scienedirect- Elsevier journals. She has filed three research patents. She also has received the research grant of SPPU University, Pune, International FDP grant by SPPU and also the conference grant by AICTE. She is the certified trainer and Ambassador of NVIDIA'S Deep Learning Institute. She conferred with the "Dr. APJ Abdul kalam Women Achievers Award" by Iitech Bangalore. She received the three Best Paper Awards in different IEEE Conferences. She got awarded by Indo Global Engineering Excellence Award by Indo Global Chamber of Commerce Industries and agriculture. She has guided more than 40 UG and PG Projects in the different domains of Computer Science and Engineering and Information Technology.