THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING

SIPL

Signal and Image Processing Lab

TECHNION Israel Institute of Technology

Work in Progress

Deep Learning Based Image Processing for a Smartphone Camera

Alexey Golub and Yanay Dado, Supervised by Dr. Meir Bar-Zohar

Introduction

- Many smartphone cameras suffer from poor image quality in comparison to DSLR cameras
- Various deep learning-based methods to improve this quality have been proposed
- Some methods try to improve existing photos, while others attempt to replace the entire image processing pipeline

Image artifacts: blurring (above) and flattening (below)

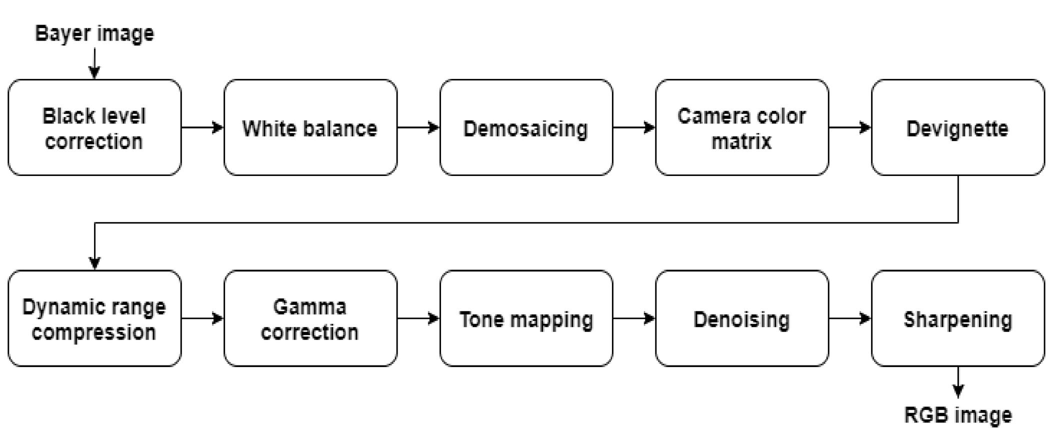
Goals

- Given a network (from paper [1]) that replaces a smartphone image processing pipeline, improve its results
- The general improvement approach was investigation of different possible loss functions

Challenges

- Many possible loss function candidates
- Many ways to measure performance (MSE, PSNR, SSIM, MS-SSIM)
- Both local and global image quality needs to be improved

The ISP Pipeline

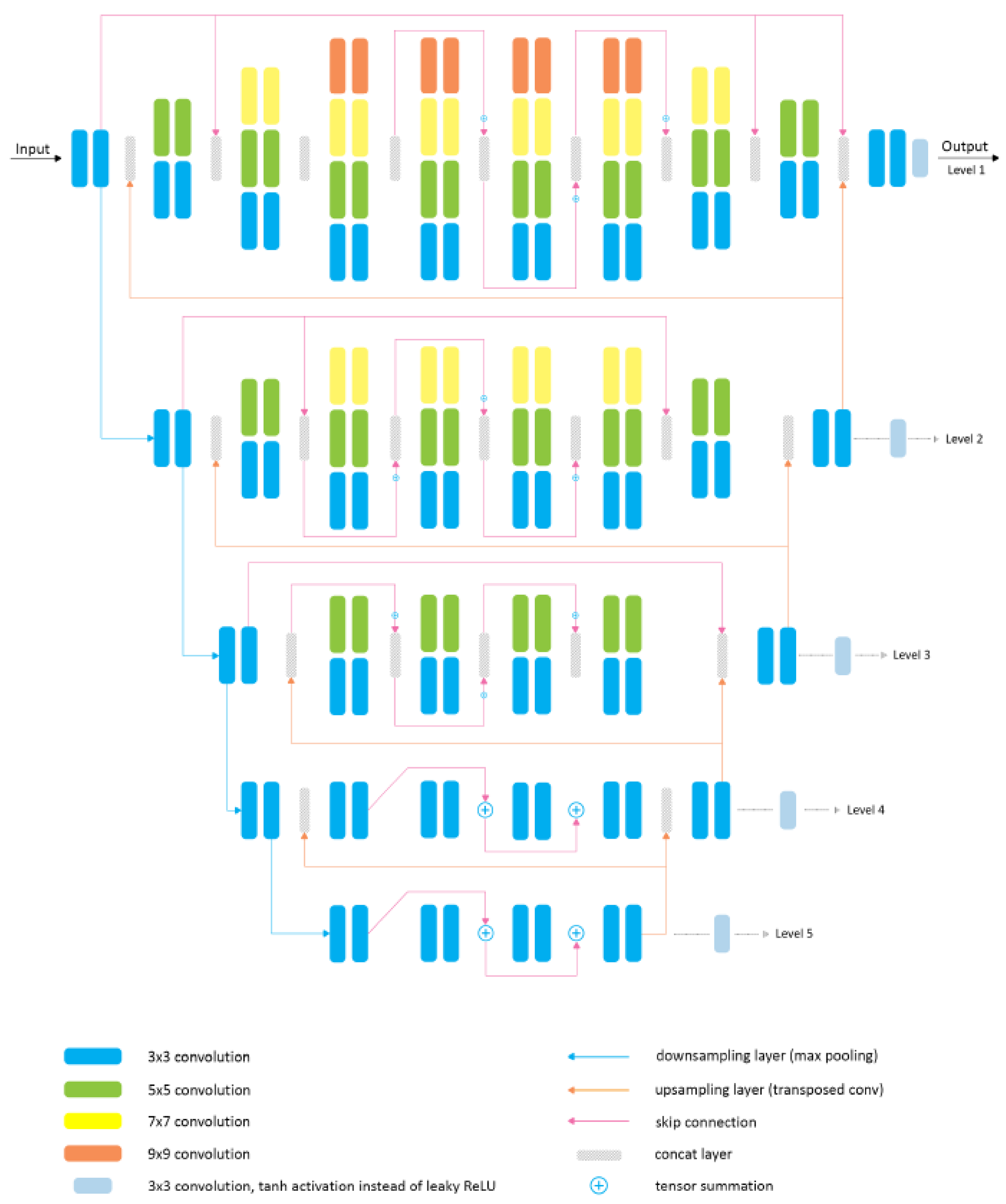


A generic ISP pipeline and its stages

- ISP stands for “image signal processing”
- This is the process that transforms the raw image sensor data (known as the “Bayer image”) to the color image on the screen
- Includes many different stages, as seen above
- Entirely classical process (i.e., no deep learning involved)
- Some possible image errors (like zippering artifacts) stem from this process
- Different companies have their own processes

PyNET Network

- Convolutional layer-based network
- Consists of 6 layers; each layer is trained separately
- Replaces the ISP pipeline entirely: input is Bayer images, output is RGB images
- Target images are DSLR camera images
- Requires alignment between input Bayer image and target DSLR image
- Achieves PSNR of 21.19 dB and multi-scale SSIM of 0.862 (from [1])



PyNET network structure

Loss Functions

- Three types of loss functions in the original network:
  - MSE loss – responsible for global image aspects such as color and brightness
  - VGG-19 loss – responsible for preserving finer object details and shapes
  - MS-SSIM loss – responsible for preserving local image aspects and increasing the dynamic range of the result images
- Most successful new loss function was LAB loss (adapted from paper [2]):
  - Consists of an L1 loss term operating on the images in LAB space and a MS-SSIM loss term on the luma channel
  - Alpha parameter controls strength of each term
  - Further improved by adding a VGG-19 loss term

The LAB loss function

$$(1-\alpha)\left\|Lab(I)-Lab(\hat{I})\right\|_1+\alpha[1-MSSSIM(L(I),L(\hat{I}))]$$

Results

- The best overall (numerical + visual) performance was achieved by using the LAB loss function with an added VGG-19 loss term
- PSNR was 22.02 dB and multi-scale SSIM was 0.856 (avg. over test dataset)

Target DSLR image (aligned to the smartphone image)

Result image produced with the original loss function
(PSNR: 20.73 dB, SSIM: 0.778, MS-SSIM: 0.862)

Result image produced with the LAB loss function
(PSNR: 20.8 dB, SSIM: 0.784, MS-SSIM: 0.865)

Conclusions

- Successful improvement of PyNET performance
- Results indicate that image quality comparable to that of DSLR cameras may be achievable
- However, the network is very complex and slow to train
- Further work may be done on decreasing network footprint and simplifying its structure

[1] Andrey Ignatov, et al. “Replacing Mobile Camera ISP with a Single Deep Learning Model”

[2] Eli Schwartz, et al. “DeepISP: Towards Learning an End-to-End Image Processing Pipeline”

June 2021