Deep Joint Demosaicking and Denoising

Michaël Gharbi\textsuperscript{1} gharbi@mit.edu
&
G. Chaurasia\textsuperscript{1} S. Paris\textsuperscript{2} F. Durand\textsuperscript{1}
Digital imaging pipeline
Digital imaging pipeline

digital sensor

noisy measurement
Digital imaging pipeline

digital sensor

color filter array

noisy measurement
Digital imaging pipeline

color filter array

digital sensor
Digital imaging pipeline

digital sensor

noisy raw samples
demosaic & denoise

RGB triplets

further processing

color filter array
Demosaicking-denoising: **critical** first stage!
Ill-posed problem

- **incomplete** color information: mosaic
  - 3 unknowns per input sample
- **noisy** sensor measurements
  - photon, thermal noise
- **mis-aligned** samples
  - 3 interdependent interpolations
  - spatial multiplexing (e.g. Bayer)
Previous work

- **faster**, traditional methods
  - sequential [Park 2009, Akiyama 2015]
  - filter design [Laroche 1994, Li 2008]
  - ad-hoc post-processing [Hirakawa 2005]

- **more accurate**, modern approaches:
  - joint demosaicking-denoising [Condat 2012]
  - non-local priors [Buades 2009, Zhang 2011]
  - global optimization [Heide 2014]
  - machine-learning [Kashabi 2014, Klatzer 2016]
It mostly works
Artifacts on challenging images

previous methods struggle with them!

zippering
[Buades 2009]
discooloration
[Heide 2014]
blur
[Getreuer 2011]
moiré
(Photoshop)
Rare but catastrophic failures

• **salient** artifacts
  • aliasing, zippering, blur...
  • all cameras prone to them

• **less than 1%** of the pixels

• **scarcity** *impedes progress*
  • good training data is hard to get
Our approach
Three-step learning approach
to joint demosaicking-denoising

1. train

Baseline database

Baseline model (deep CNN)
Three-step learning approach

to joint demosaicking-denoising

1. train

2. mine hard cases

baseline model
(deep CNN)

baseline database

challenging database
Three-step learning approach to joint demosaicking-denoising

1. train
   baseline model (deep CNN)

2. mine hard cases
   2. mine hard cases

3. re-train
   final model

baseline database

challenging database
Pseudo ground truth

sRGB image from the web
Pseudo ground truth

sRGB image from the web

downsample

becomes 1 pixel

4x downsample
Pseudo ground truth

sRGB image from the web

becomes 1 pixel

4x downsample artifacts and noise greatly reduced

Gaussian noise [Jeon 2013]

mosaick
Model architecture
Deep CNN architecture

- translation invariance
Deep CNN architecture

- translation invariance
- trainable stack of convolutions
  - large footprint non-linear filter
Deep CNN architecture

- translation invariance
- trainable stack of convolutions
  - large footprint non-linear filter
- predict difference from input: residual
- easier than synthesizing output from scratch
- similar to ResNet [He 2016]
Noise varies with ISO

- need to handle multiple noise levels
- noise characteristics **known** in advance
Naive: one model per noise level

noise level $\sigma = 0$

$\sigma = 4$

$\sigma = 16$

$\sigma = 20$
Parametrize by noise level

noise estimate

noisy input
Training procedure

• minimize L2 loss
  • Adam [Kingma 2014]
  • 1 week of training

• range of noise levels
  • train jointly on all noise levels
  • random noise variance per training image
  • fixed range $\sigma \in [0, 20]$
Baseline trained on Imagenet
1.5 million images
Standard benchmarks look good...
(numbers too)
...but major artifacts remain

Imagenet baseline

ground truth

trained on imagenet
Why does this happen?

- too few challenging images
  - most patches are smooth [Levin 2012]
  - only 1 in 2,000 has artifacts
Why does this happen?

• too few challenging images
  • most patches are smooth [Levin 2012]
  • only 1 in 2,000 has artifacts

• metrics cannot detect artifacts [Sergej 2014]
  • SSIM, MSE: low correlation with human perception
Three-step learning approach

1. Train
   - Baseline model (deep CNN)

2. Mine hard cases
   - New metrics

3. Re-train
   - Final model
Detecting artifacts

- **analyze** millions of photographs
  - 1 month of scraping, 100 computers
Detecting artifacts

- **analyze** millions of photographs
  - 1 month of scraping, 100 computers

- **process** them with our baseline model
  - we now have before/after ground-truth
Detecting artifacts

- **analyze** millions of photographs
  - 1 month of scraping, 100 computers
- **process** them with our baseline model
  - we now have before/after ground-truth
- **rejection-sampling**: keep hard cases
  - 2 criteria: luminance errors, color moiré
Finding luminance errors
use HDR-VDP, perceptual model [Mantiuk 2012]
Exposing color moiré

detect added low-frequency chroma

ref.  corrupted  difference  amplitude gain in Fourier domain
Challenging dataset

Imagenet: 1.5 million
Challenging dataset

Imagenet: 1.5 million

total mined: 250 million
Challenging dataset

Imagenet: 1.5 million

our dataset: 2.5 million

total mined: 250 million
Results

training our model on this new dataset
Evaluation

- new, separate test and validation sets
  - 2,000 images each
  - standard datasets are too easy
- more accurate than state-of-the-art
- faster than best previous work
  - though not real time yet
Demosaicking only
Kodak and McMaster datasets (noise free)

higher is better

PSNR

30 dB
35 dB
40 dB
45 dB

bilinear
camera RAW
Klatzer '16
Gunturk '02
Lu '10
Li '05
Hirakawa '05
Condat '12
Jeon '13
Hirakawa '06
Hamilton '97
Zhang '05
Buades '09
Zhang '11
Getreuer '11
Heide '14
ours

33
34
35
36
36
36
36
37
37
37
38
38
40
41
Demosaicking only

higher is better

| Method       | PSNR |
|--------------|------|
| Bilinear     | 26   |
| Camera RAW   | 29   |
| Klatzer '16  | 29   |
| Gunterk '02  | 30   |
| Lu '10       | 30   |
| Li '05       | 30   |
| Hirakawa '05 | 30   |
| Cond '12     | 31   |
| Jeon '13     | 29   |
| Hirakawa '06 | 31   |
| Hamilton '97 | 30   |
| Zhang '05    | 31   |
| Buades '09   | 31   |
| Zhang '11    | 31   |
| Getreuer '11 | 32   |
| Heide '14    | 31   |
| Ours         | 36   |
Photoshop
25 dB

FlexISP
27 dB

ours
31 dB

ref
Joint demosaicking/denoising
Joint demosaicking/denoising

![Graph showing PSNR vs. noise level for different methods.](image)

- ours, one network per noise level
- ours
- Condat 2012
- Condat 2011
- Hirakawa 2006
- Jeon 2013
noisy
$\sigma = 4$

Condat 2012
27 dB

ours
31 dB
noisy
\( \sigma = 12 \)

Condat 2012
27 dB

ours
31 dB
Real RAW data

false colors  dcraw  over-smoothing  ours  Klatzer 2016

captured with a Canon 5D mark II  edge artifacts
Non-Bayer mosaic
RAW Fuji X-Trans
Fuji X-Trans
dcraw
Fuji X-Trans
dcraw
Fuji X-Trans
ours
Running time

- **ours GPU**: 325 ms, 36 dB
- **ours CPU**: 2,932 ms, 36 dB

| PSNR   | Running time (ms/Mpixel) |
|--------|--------------------------|
| 36 dB  | median 2,965 ms          |
| 34 dB  |                          |
| 32 dB  |                          |
| 30 dB  |                          |
| 28 dB  |                          |
| 26 dB  | 100 ms                   |

Median: 30.4 dB
Limitations and future work

• better ground truth
  • e.g. moiré in the reference

• better metrics
  • HDR-VDP does not capture all the luminance errors
Conclusion

• fast, state-of-the-art demosaicking/denoising
• noise-parametrized network
• three-step process to mine challenging data
• code and data available online!

Michaël GHARBI gharbi@mit.edu
code & data: www.mgharbi.com