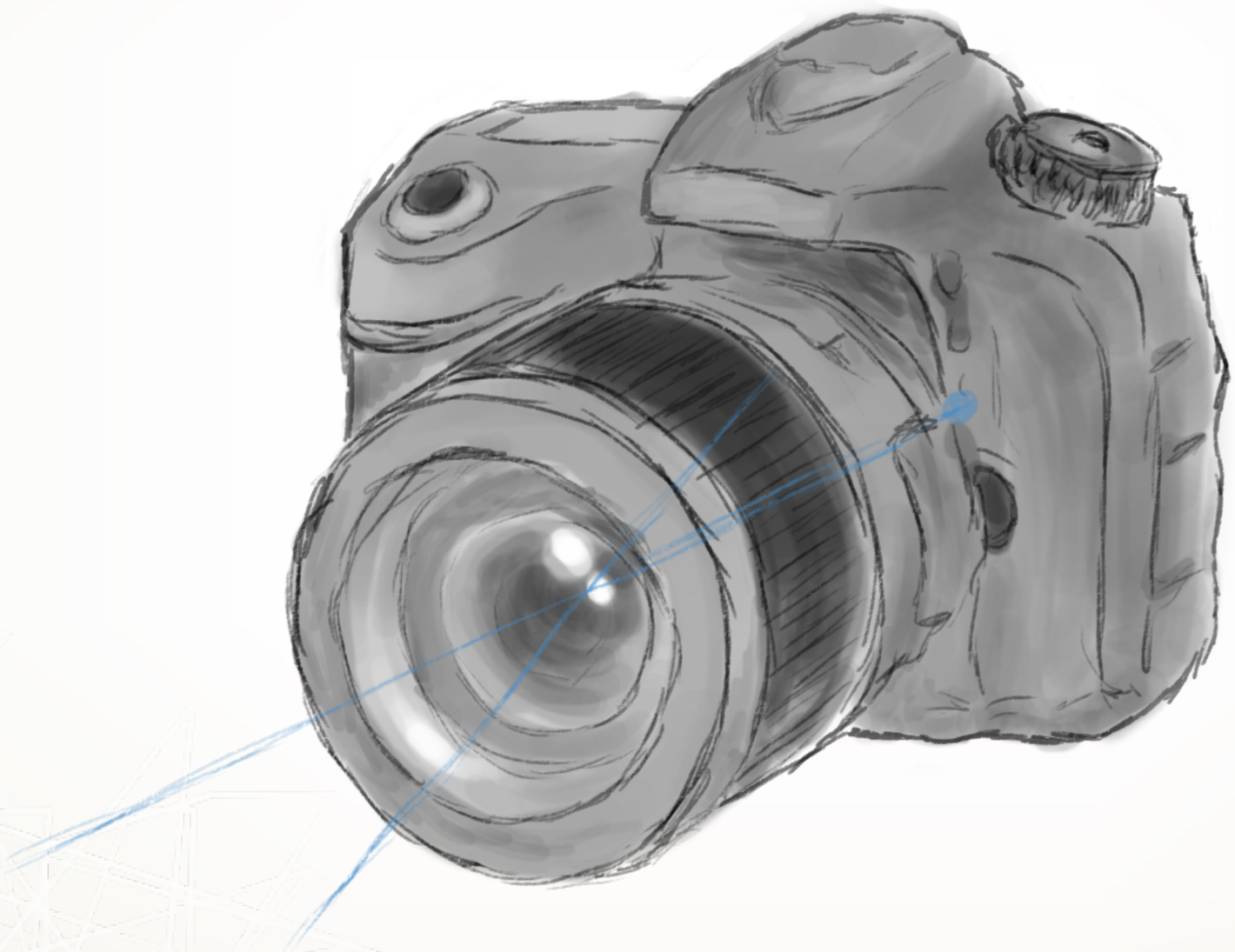
SIGGRAPH ASIA 2016 MACAO

Deep Joint Demosaicking and Denoising

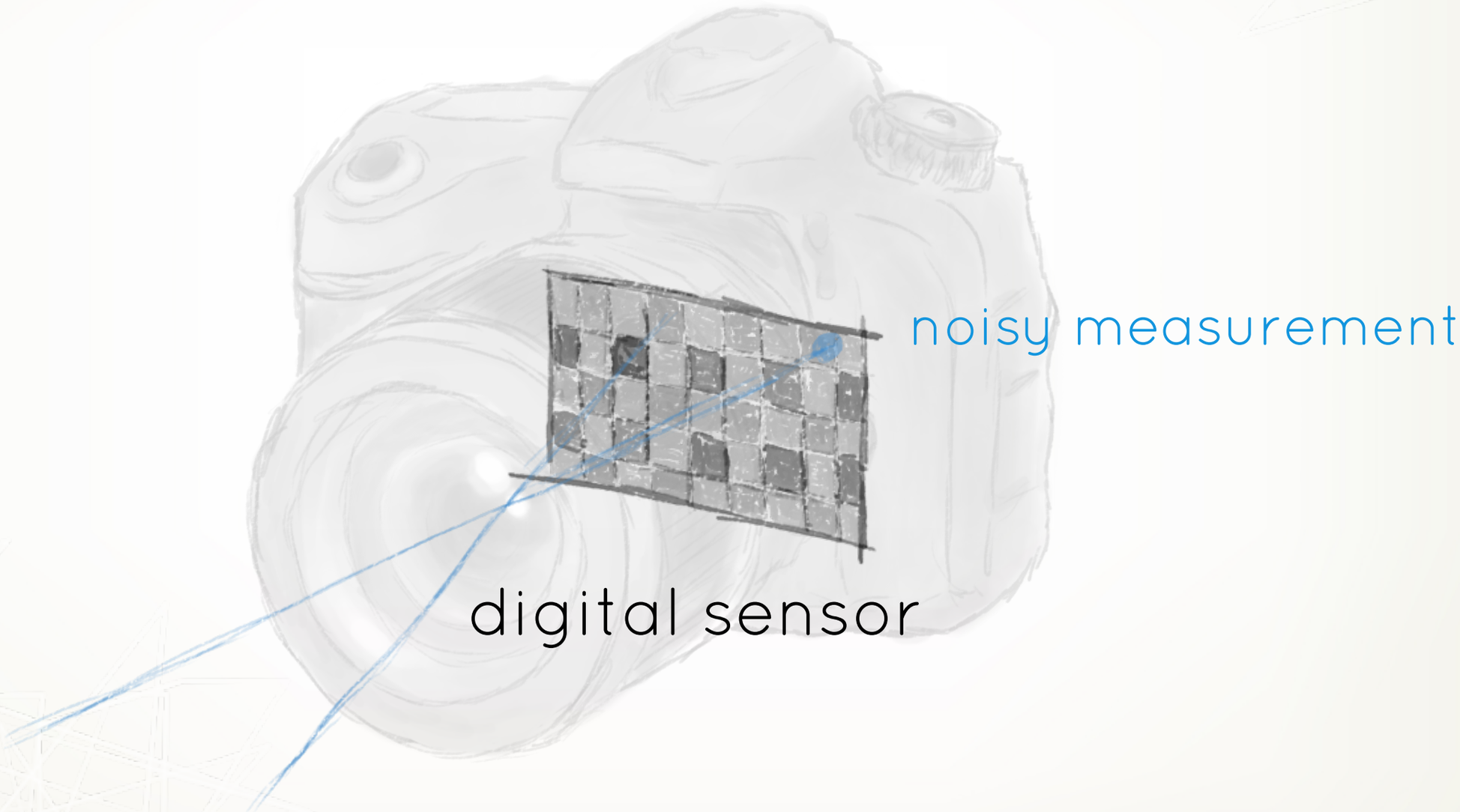
Michaël Gharbi[1] gharbi@mit.edu
&
G. Chaurasia[1] S. Paris[2] F. Durand[1]

MIT CSAIL Adobe

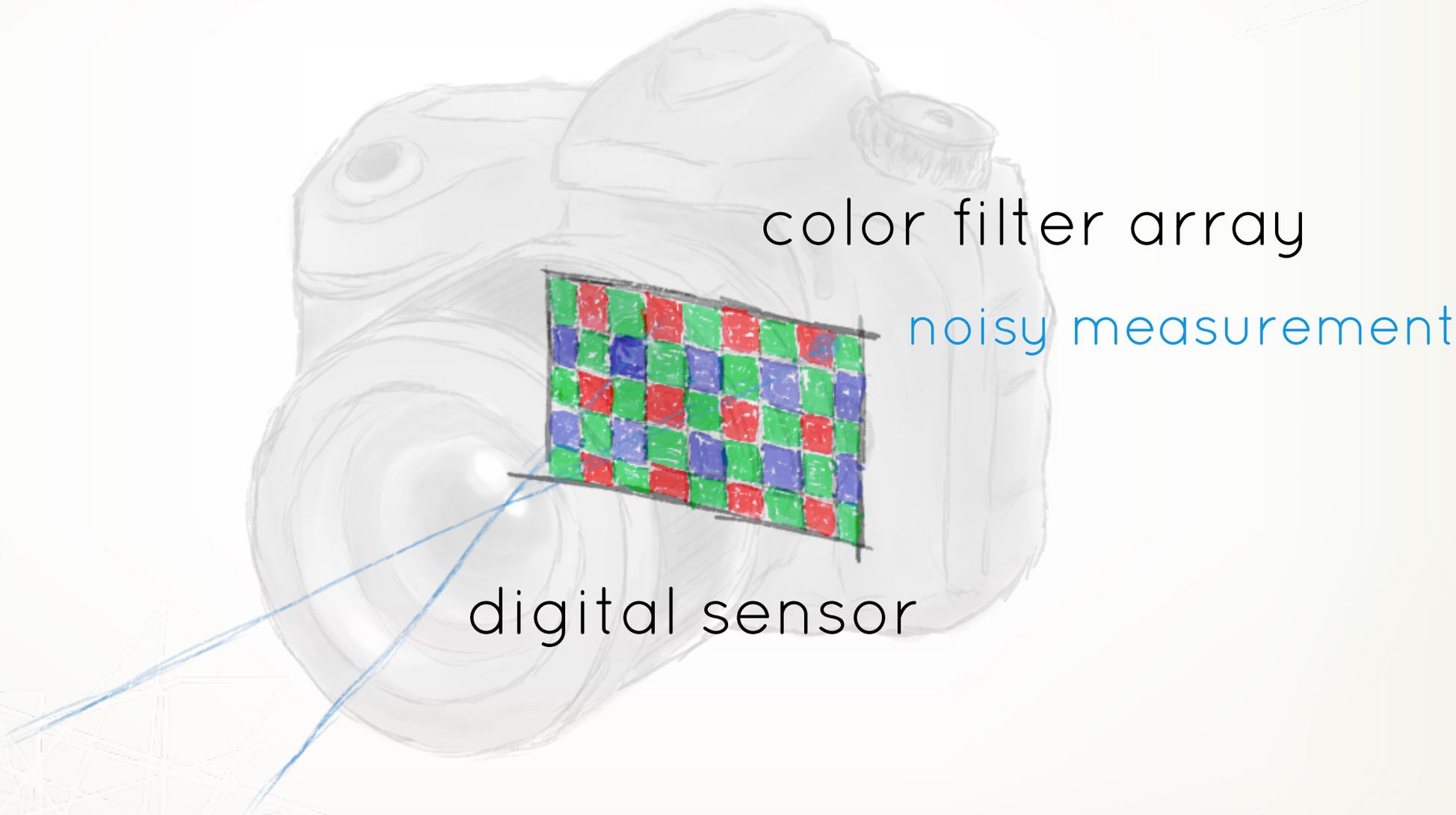
Digital imaging pipeline



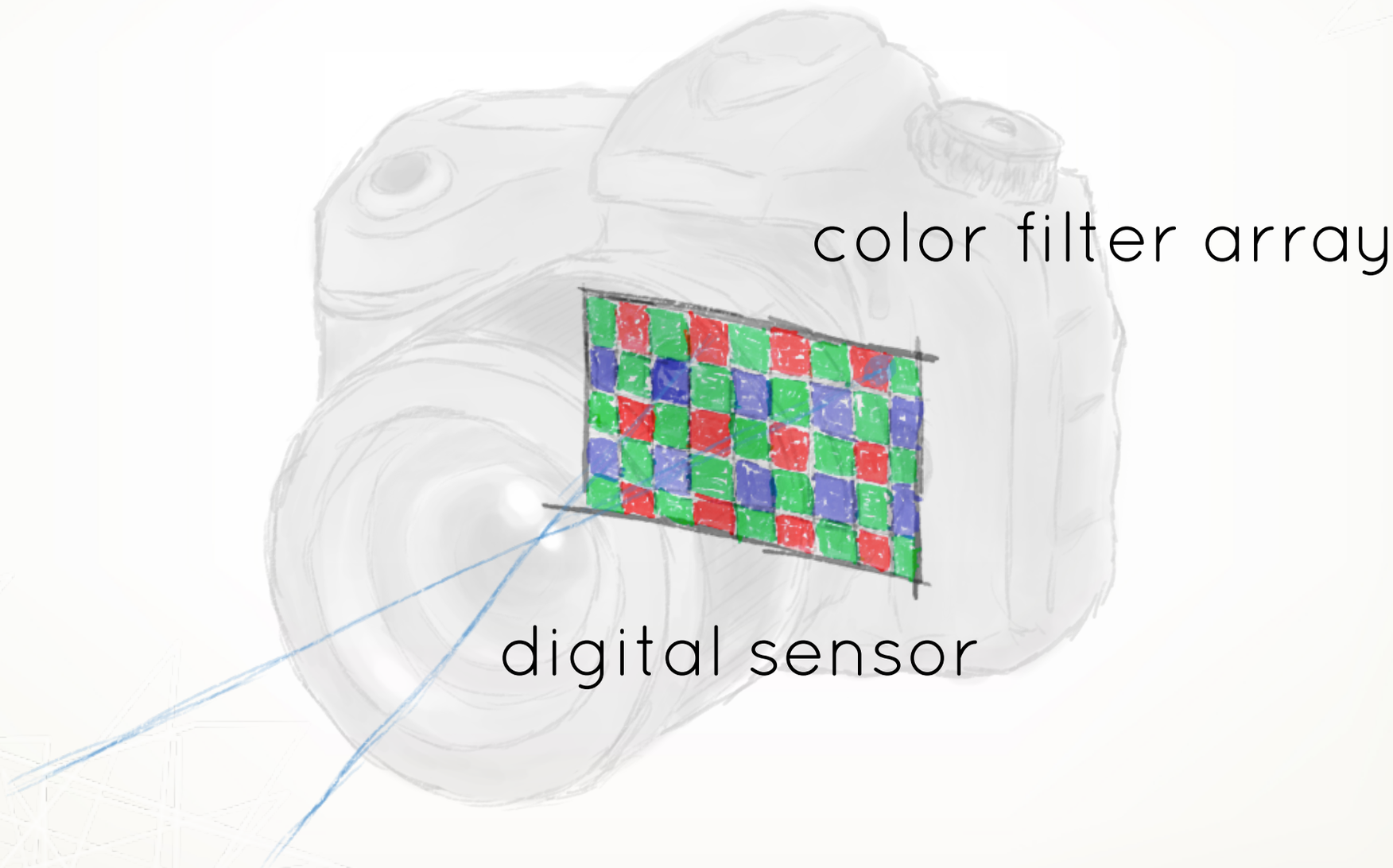
Digital imaging pipeline



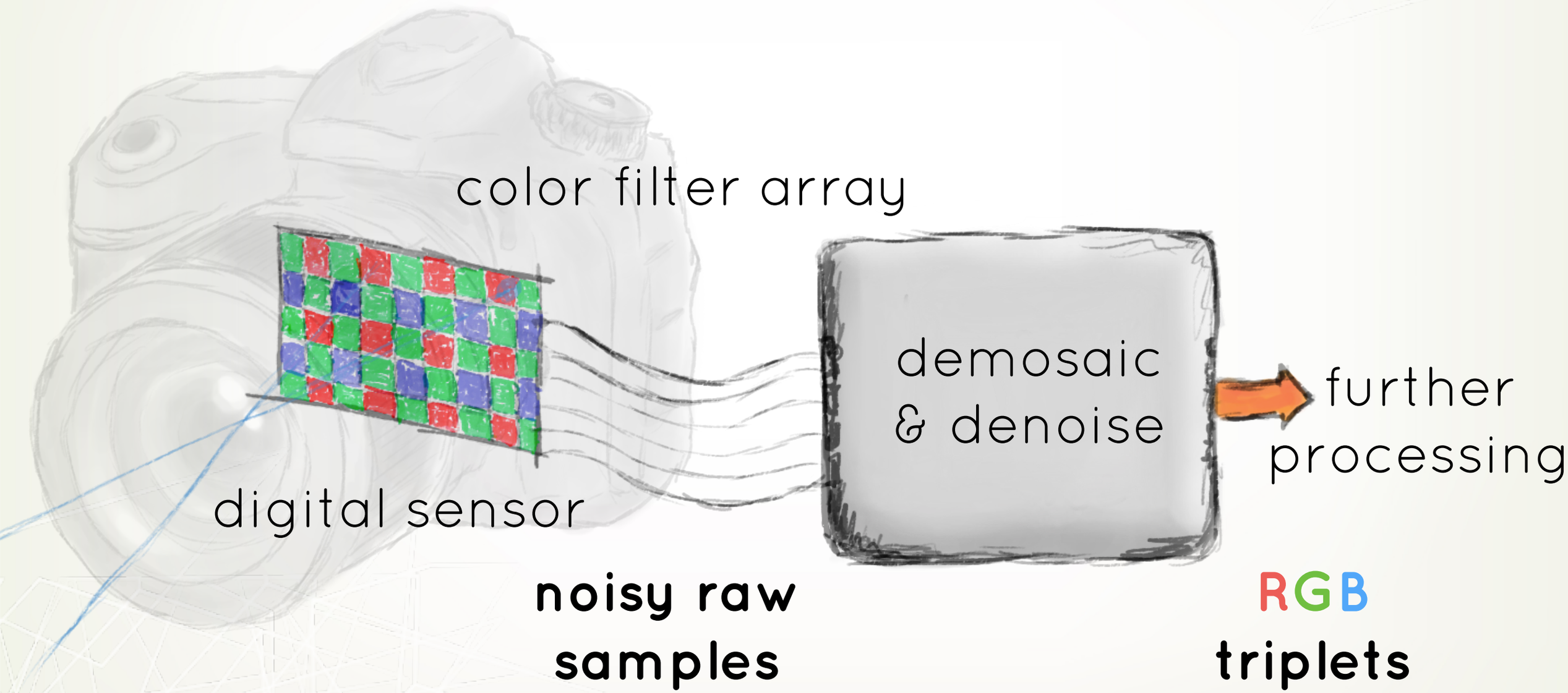
Digital imaging pipeline



Digital imaging pipeline



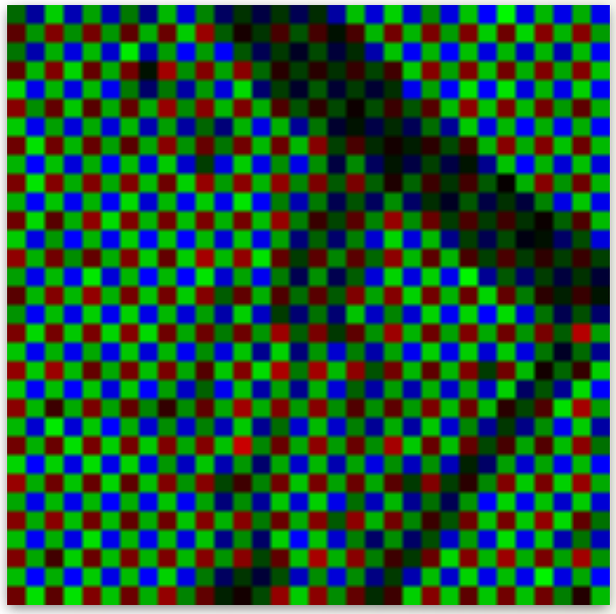
Digital imaging pipeline



Demosaicking-denoising:
critical first stage!

Ill-posed problem

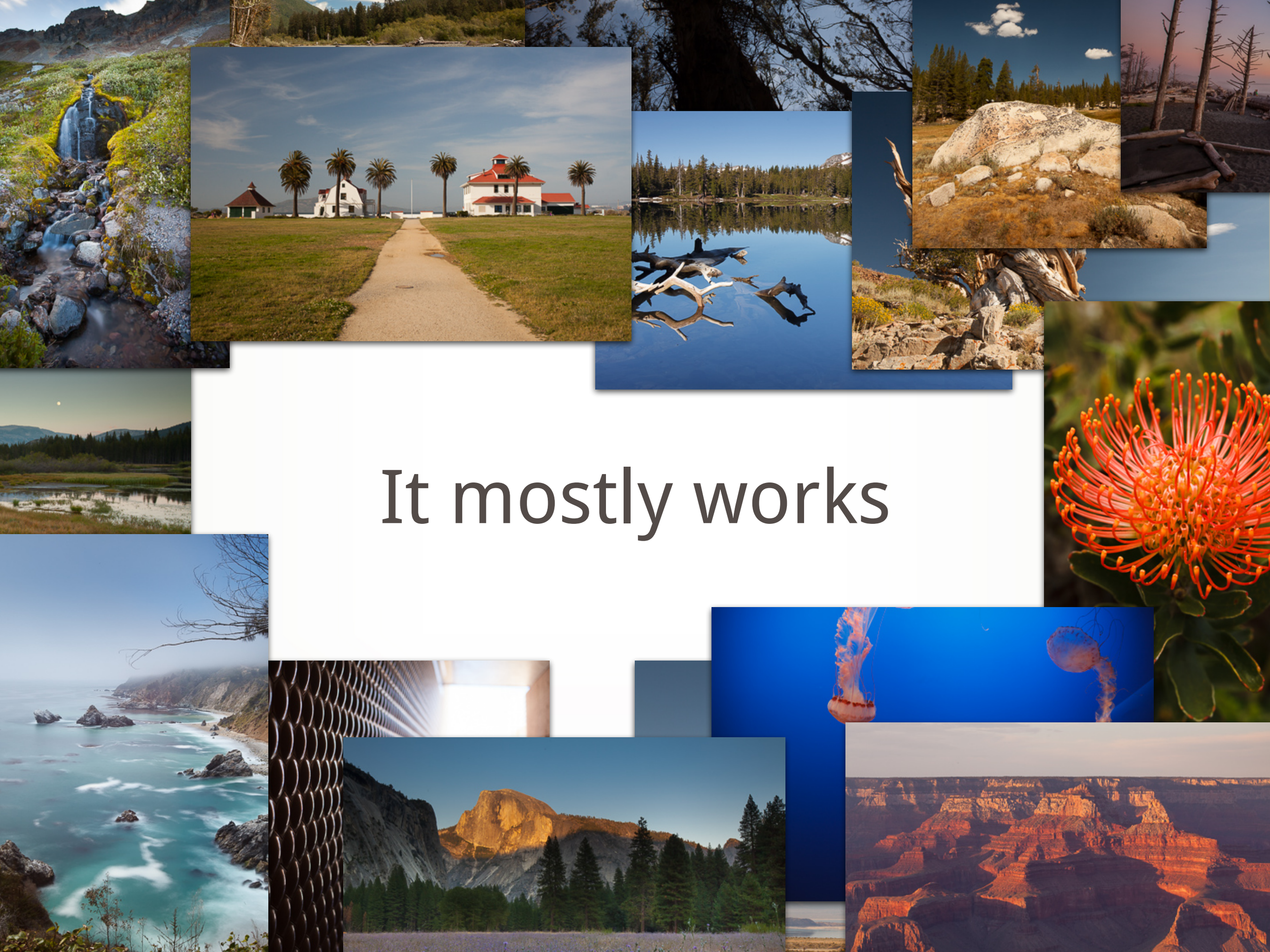
- **incomplete** color information: mosaic
  - 3 unknowns per input sample
- **noisy** sensor measurements
  - photon, thermal noise
- **mis-aligned** samples
  - 3 interdependent interpolations
  - spatial multiplexing (e.g. Bayer)



Previous work

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

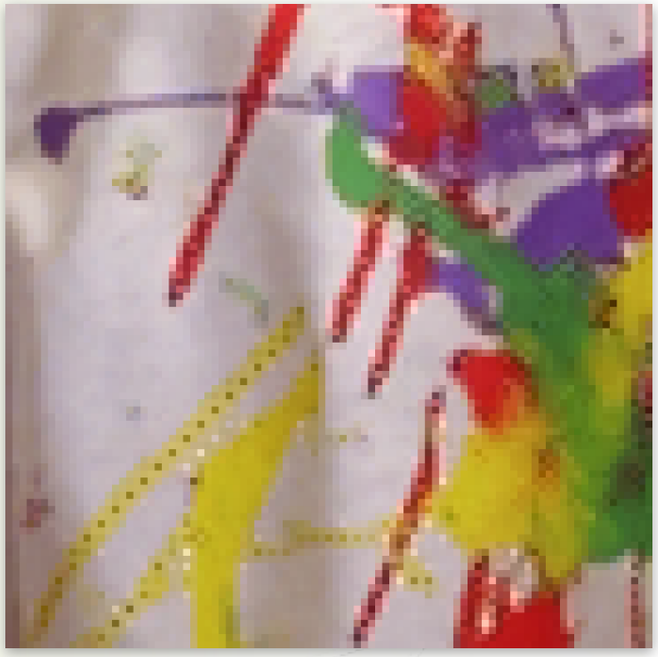
- **more accurate**, modern approaches:
  - joint demosaicking-denoising [Condat 2012]
  - non-local priors [Buades 2009, Zhang 2011]
  - global optimization [Heide 2014]
  - machine-learning [Kashabi 2014, Klatzer 2016]



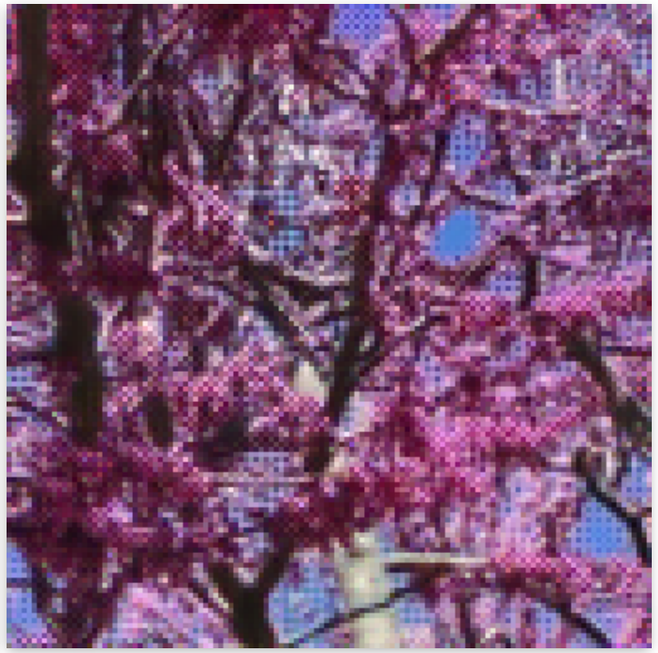
It mostly works

Artifacts on challenging images

previous methods struggle with them!

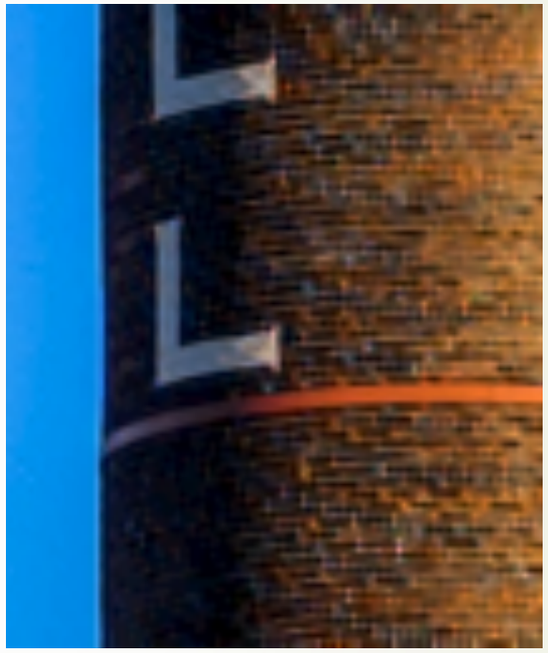


zippering
[Buades 2009]



discoloration
[Heide 2014]

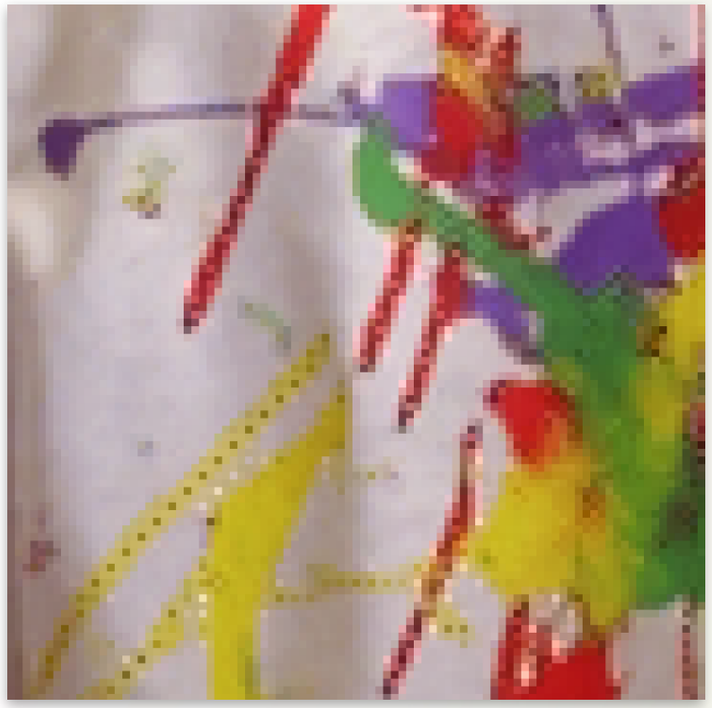
blur
[Getreuer 2011]



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

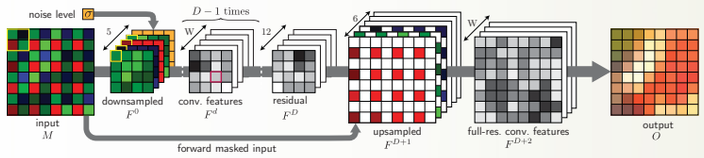
Three-step learning approach

to joint demosaicking-denoising

baseline database

1. train

baseline model (deep CNN)



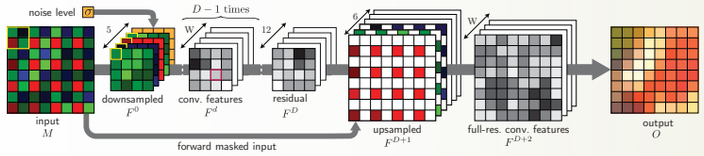
Three-step learning approach

to joint demosaicking-denoising

baseline database

1. train

baseline model (deep CNN)



2. mine hard cases

challenging database

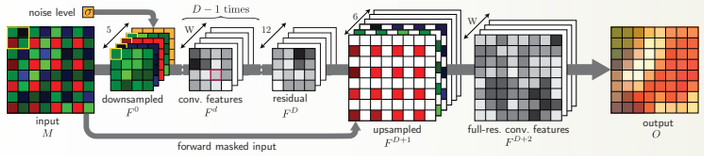
Three-step learning approach

to joint demosaicking-denoising

baseline database

1. train

baseline model (deep CNN)

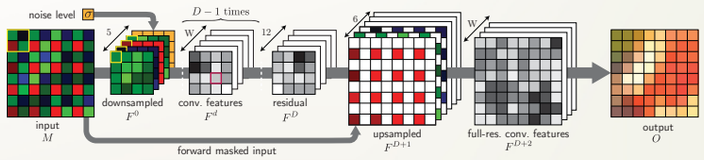


2. mine hard cases

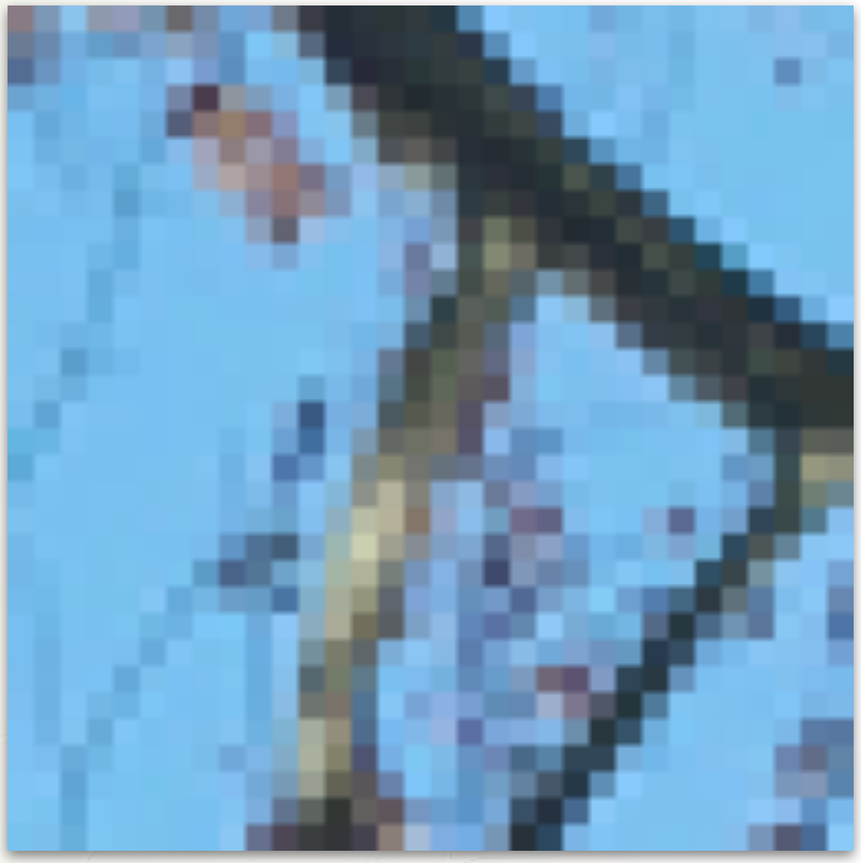
challenging database

3. re-train

final model



Pseudo ground truth

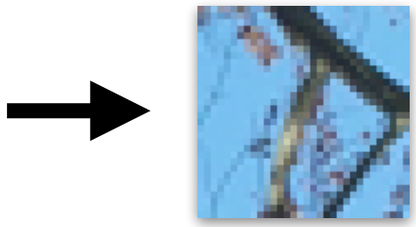


sRGB image from the web

Pseudo ground truth

becomes 1 pixel

sRGB image from the web



4x downsample

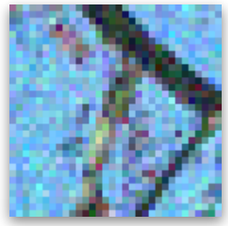
Pseudo ground truth

becomes 1 pixel

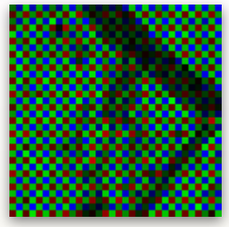
sRGB image from the web

4x downsample

artifacts and noise greatly reduced



Gaussian noise [Jeon 2013]



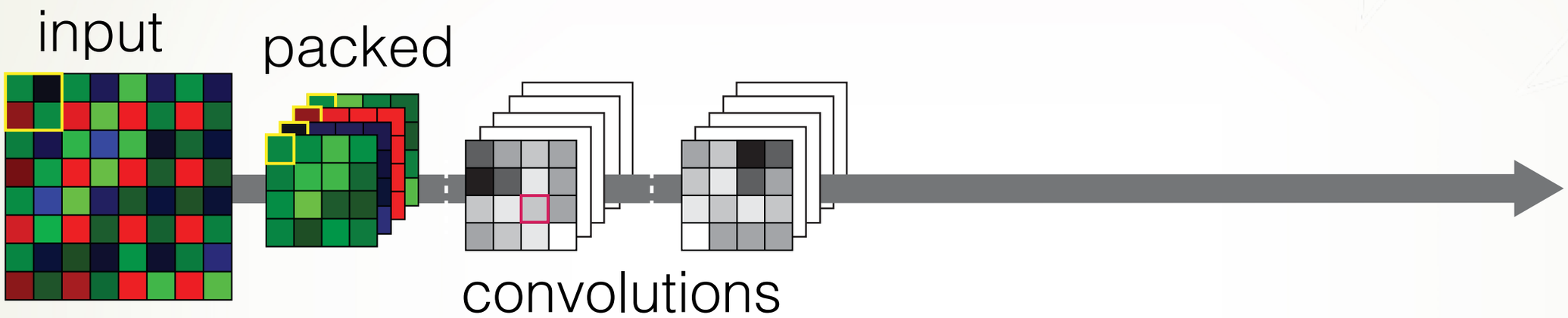
mosaick

Model architecture

Deep CNN architecture

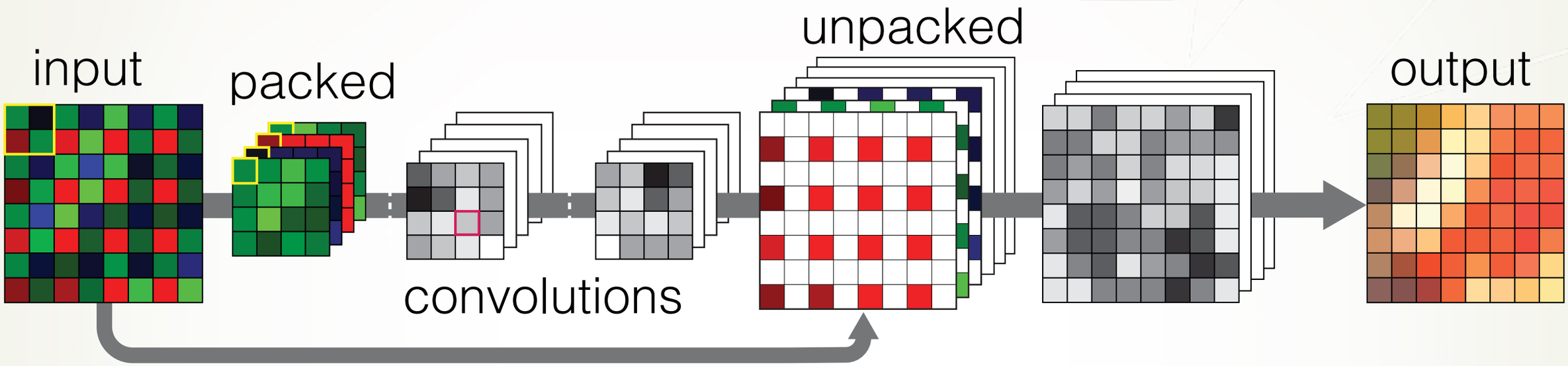
- translation invariance

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

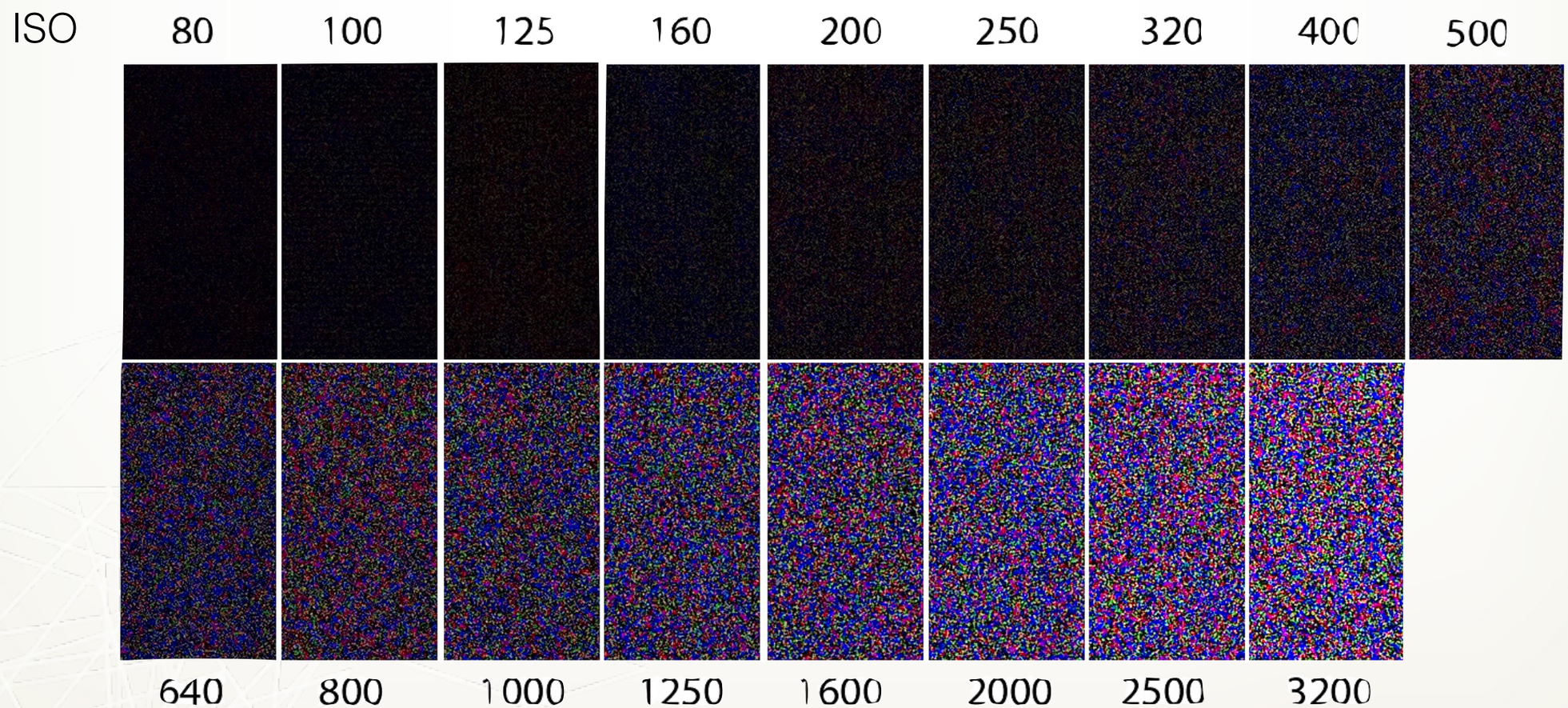
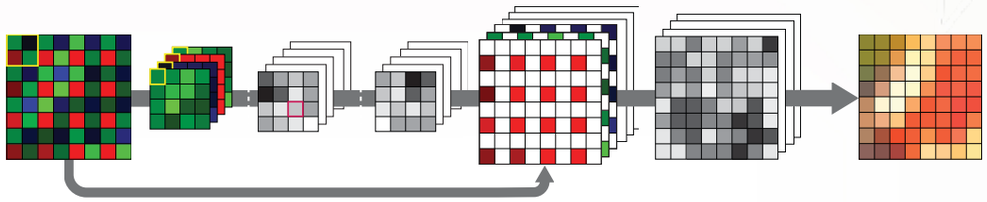


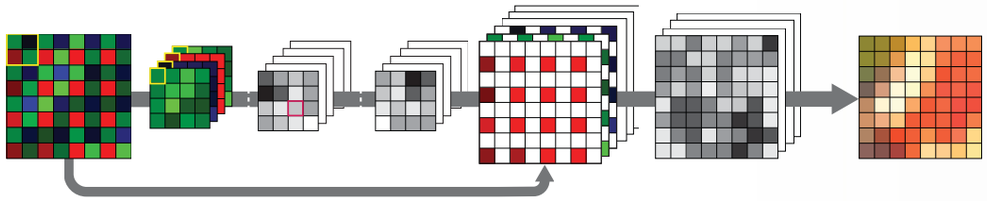
image blog.carlchapman.com

Naive: one model per noise level

noise level $\sigma = 0$



$\sigma = 4$

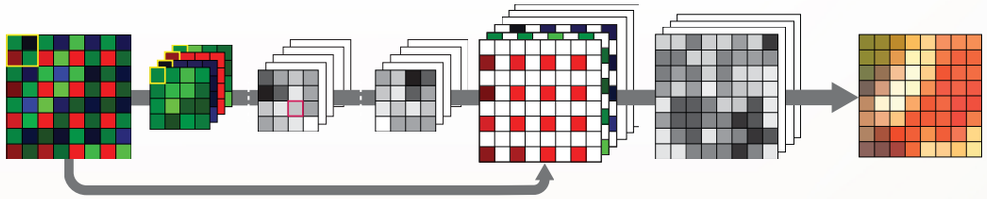


•

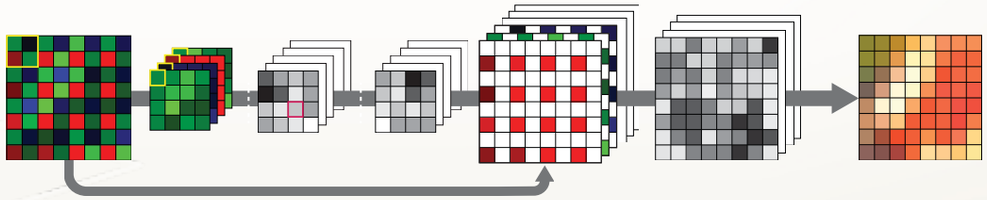
•

•

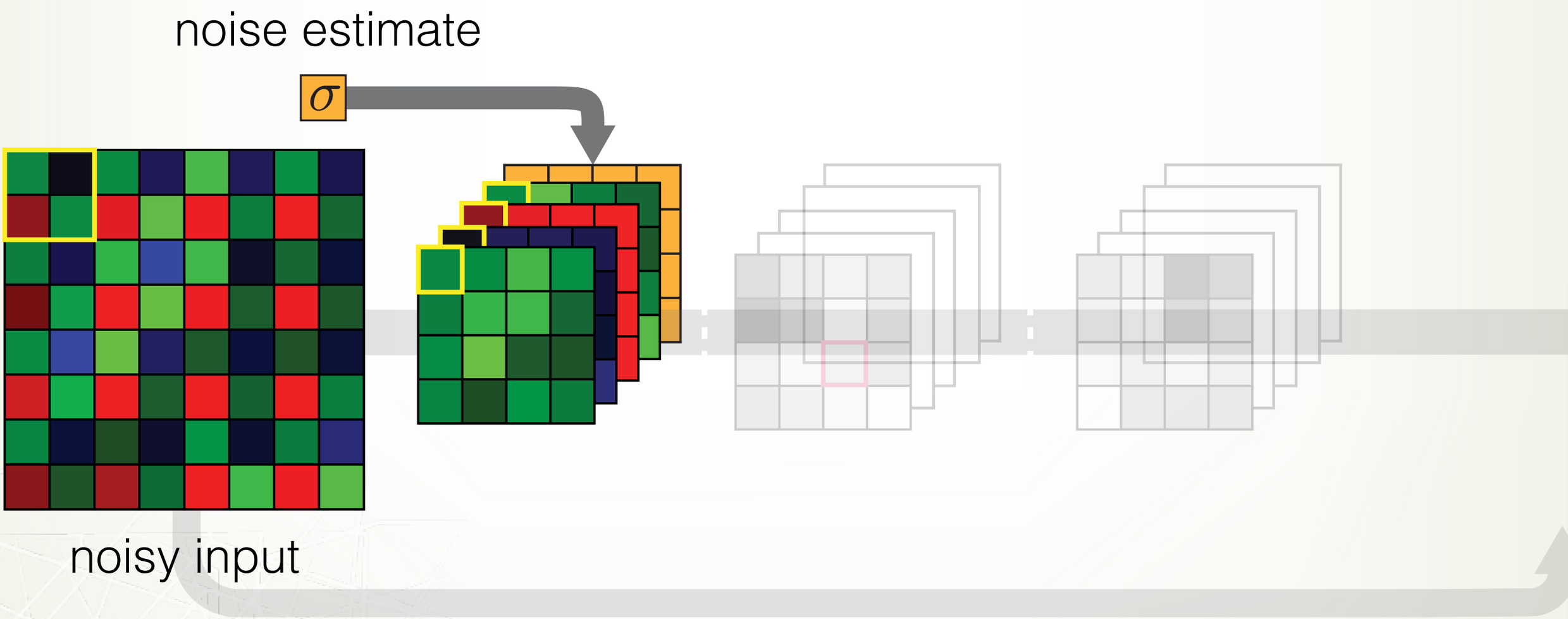
$\sigma = 16$



$\sigma = 20$

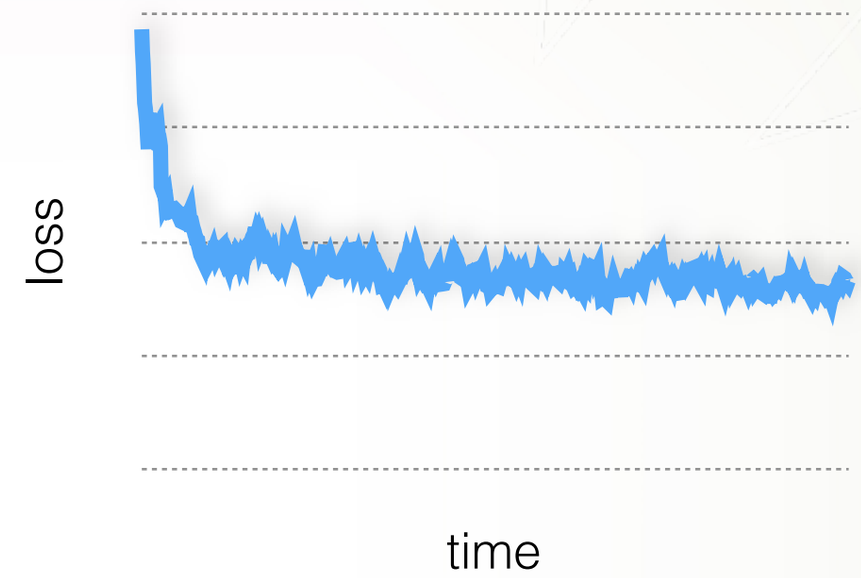


Parametrize by noise level



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




Imagenet baseline

Standard benchmarks look good...

(numbers too)

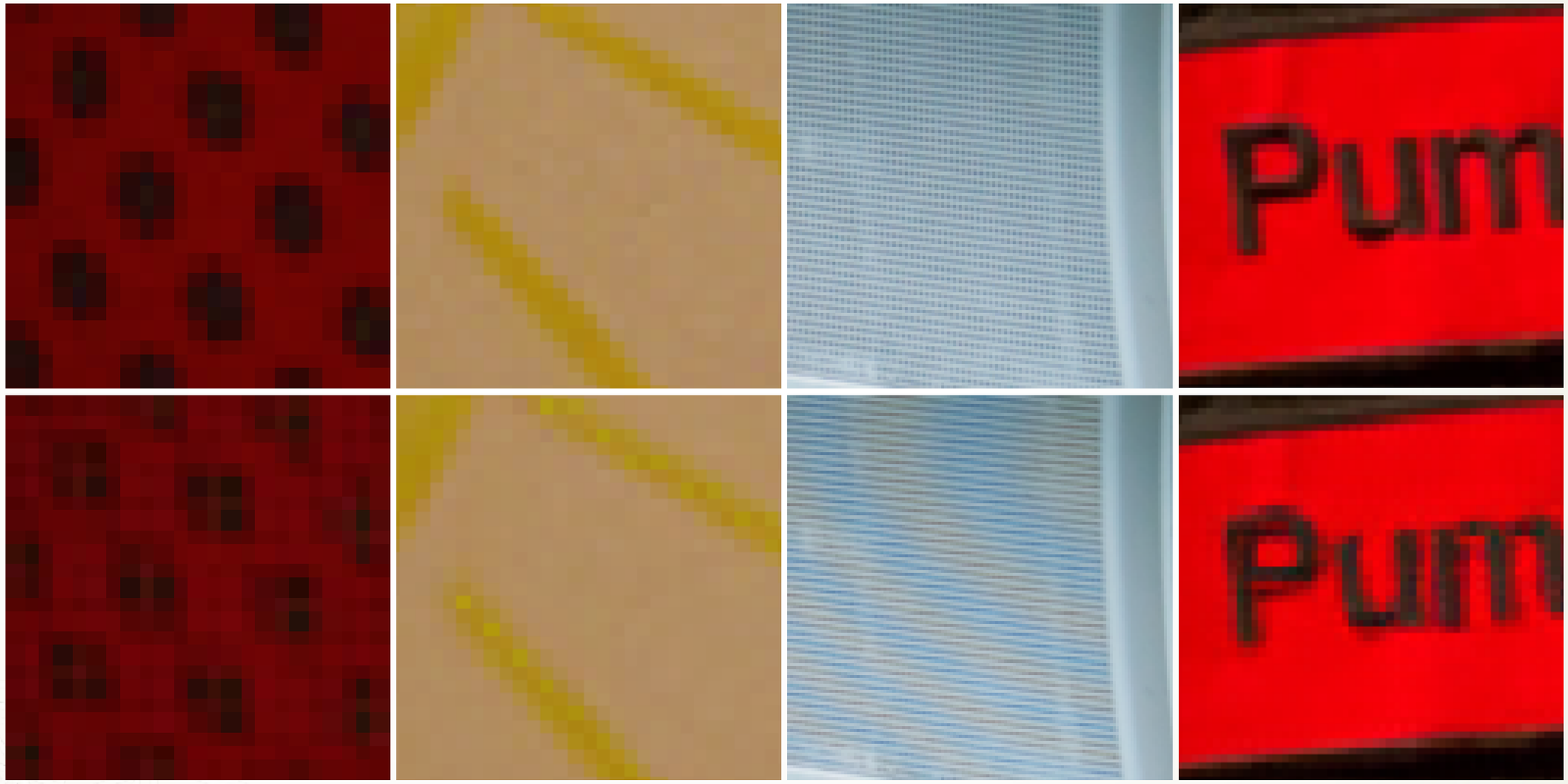



...but major artifacts remain

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?

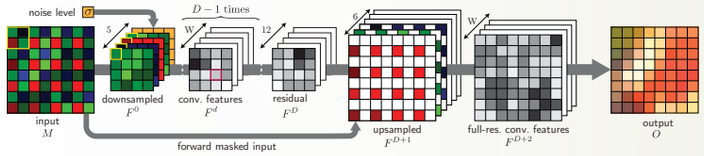
- too few challenging images
  - most patches are smooth [Levin 2012]
  - only 1 in 2,000 has artifacts
- metrics cannot detect artifacts [Sergej 2014]
  - SSIM, MSE: low correlation with human perception

Three-step learning approach

baseline database

1. train

baseline model (deep CNN)



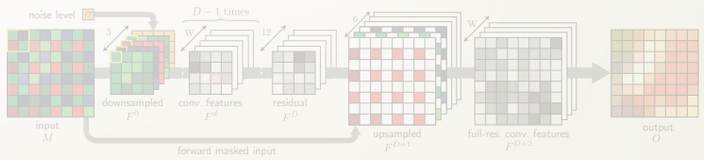
2. mine hard cases

new metrics

challenging database

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

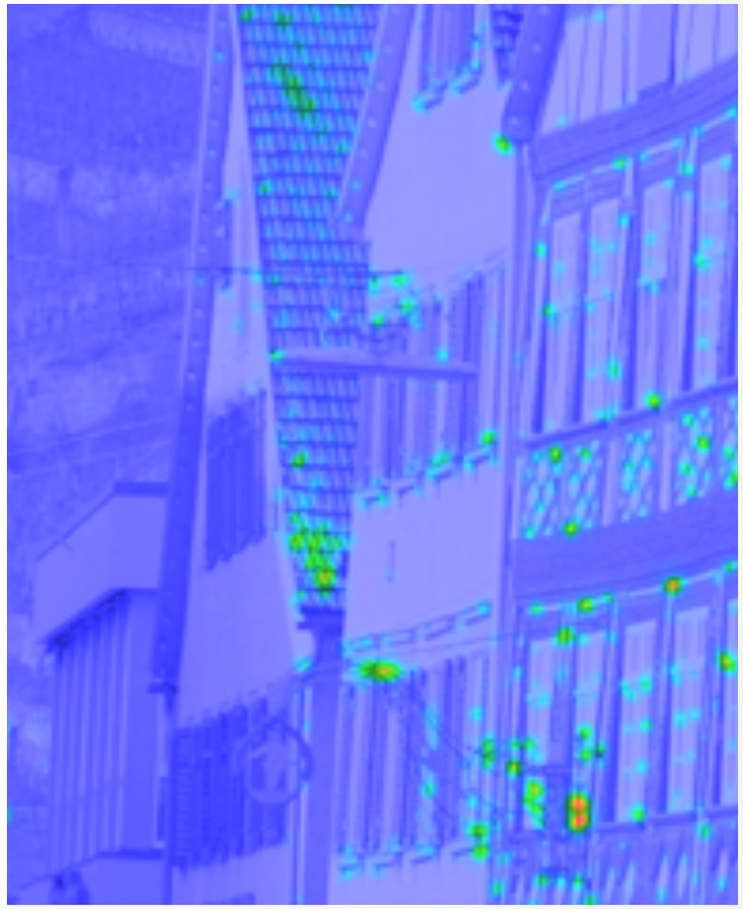
- **analyze** millions of photographs
  - 1 month of scraping, 100 computers
- **process** them with our baseline model
  - we now have before/after ground-truth
- **rejection-sampling**: keep hard cases
  - **2 criteria:** luminance errors, color moiré

Finding luminance errors

use HDR-VDP, perceptual model [Mantiuk 2012]

ref.

corrupted



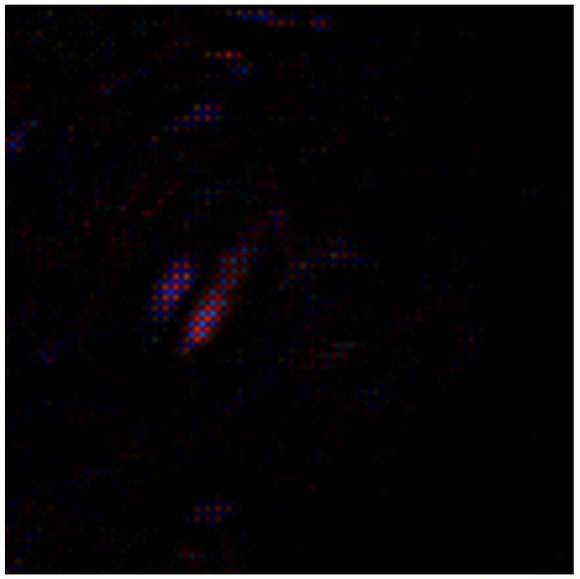
error map

Exposing color moiré

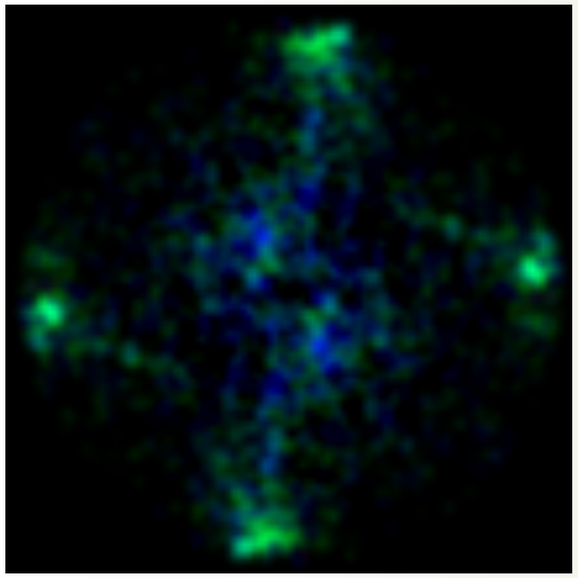
detect added low-frequency chroma

ref.

corrupted

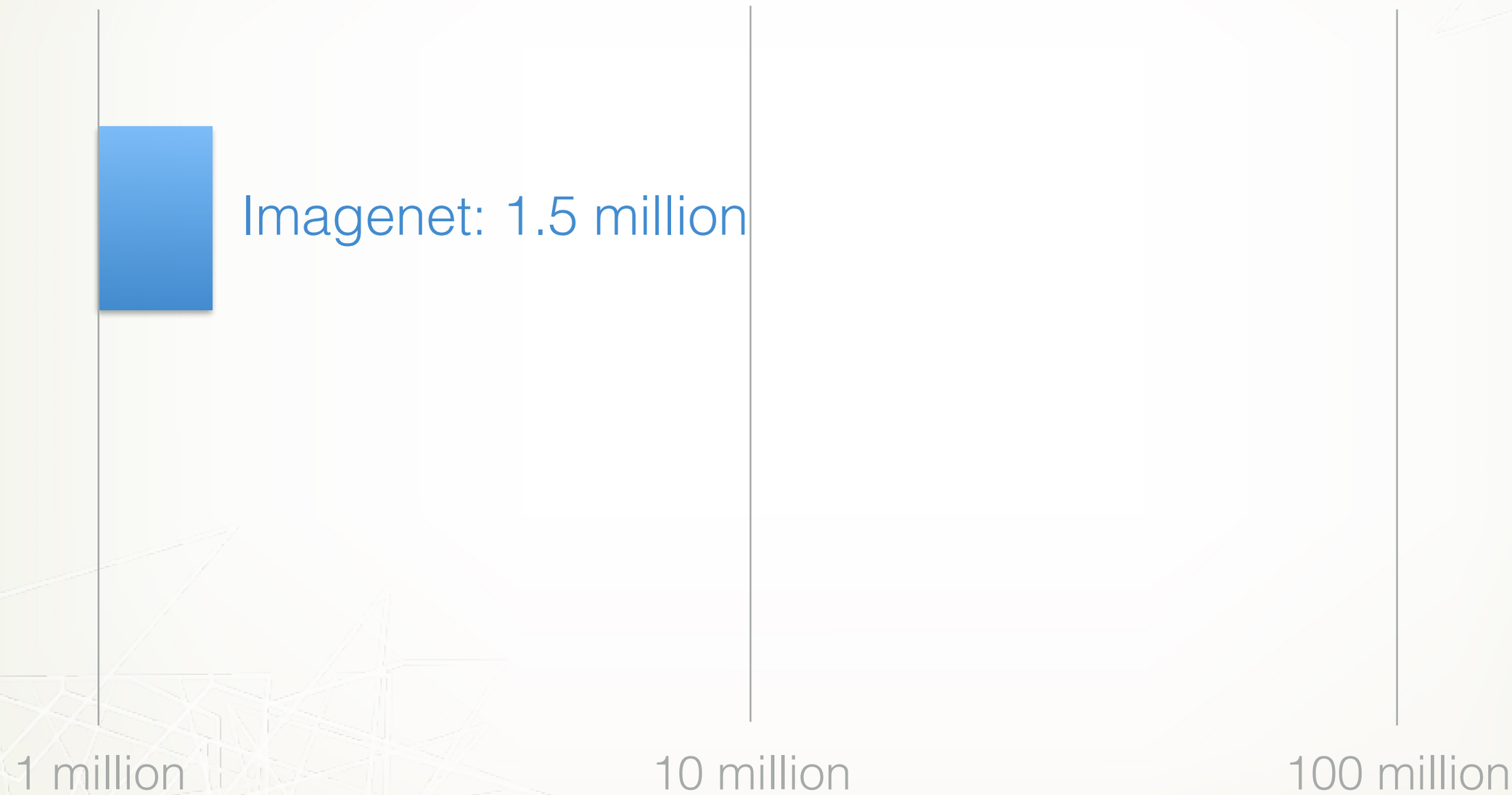


difference

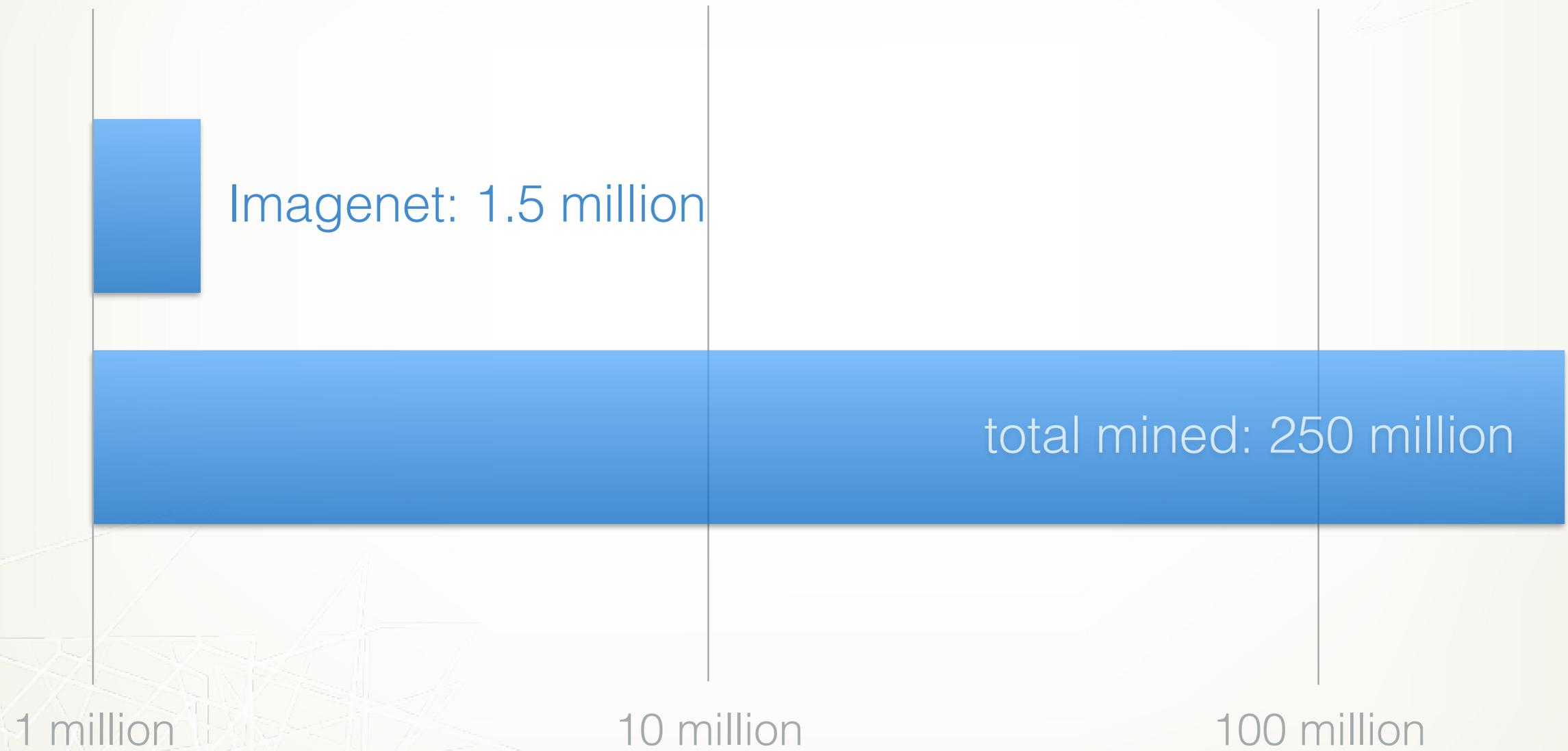


amplitude gain in Fourier domain

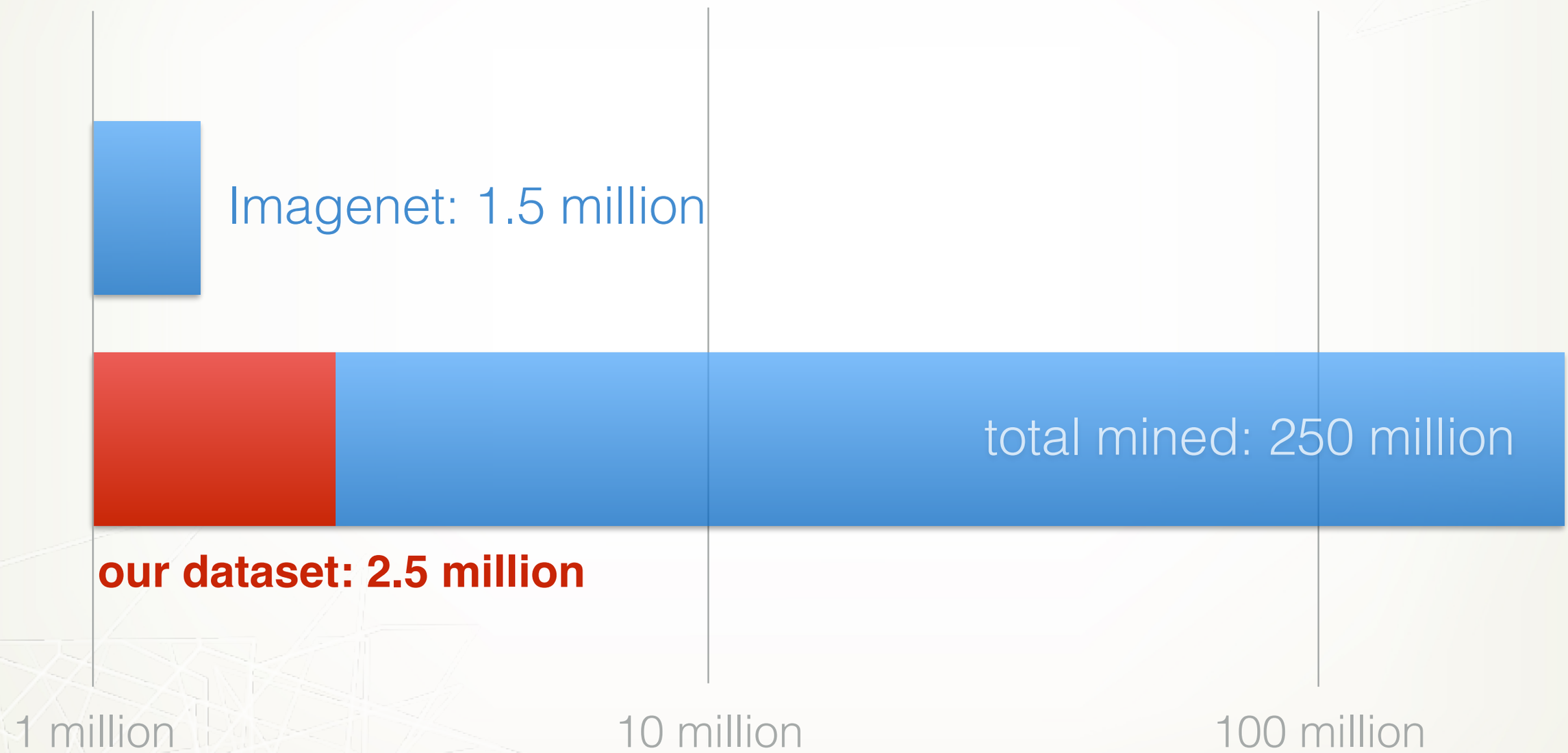
Challenging dataset



Challenging dataset



Challenging dataset



Results

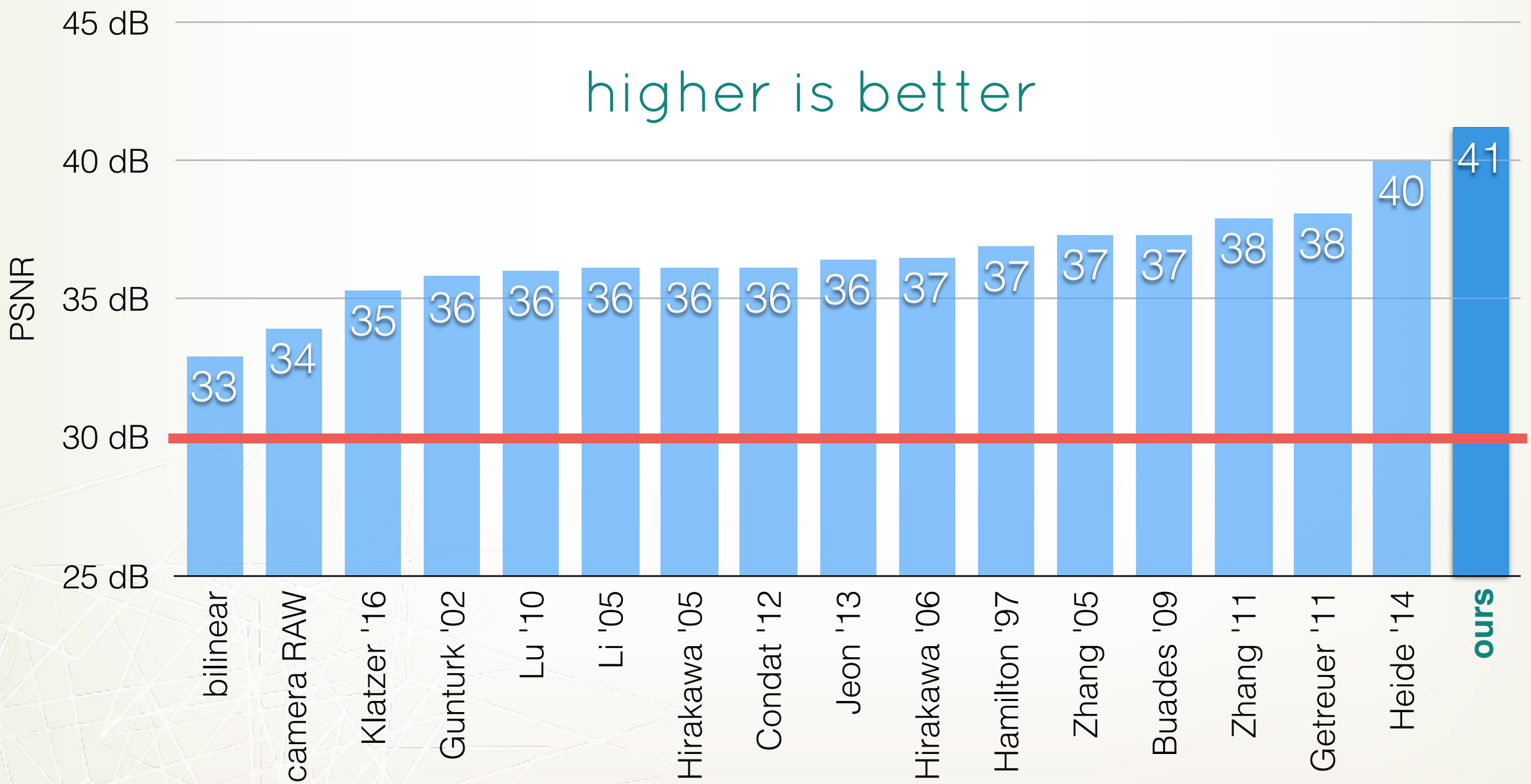
training our model on this new dataset

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

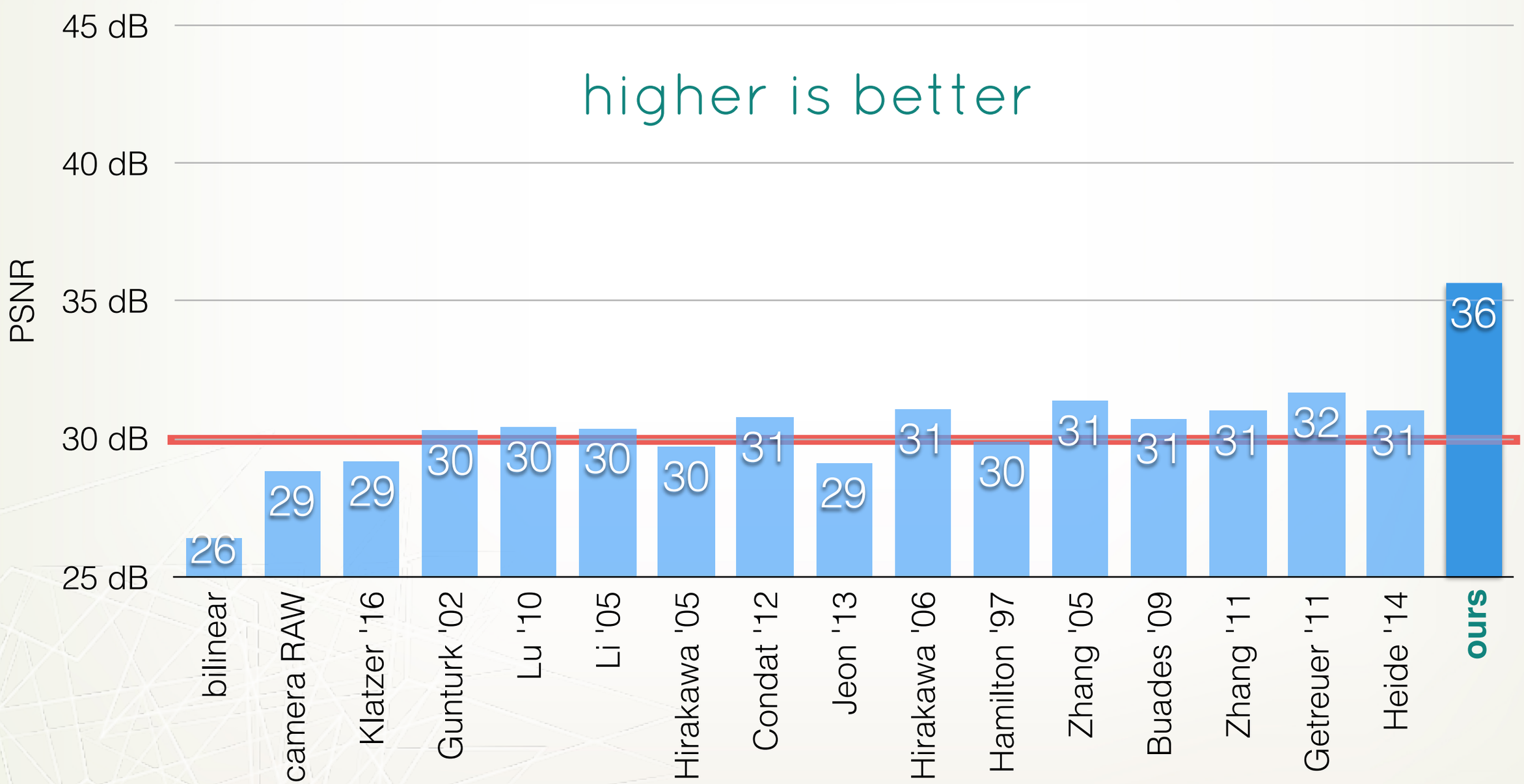
Demosaicking only

Kodak and McMaster datasets (noise free)

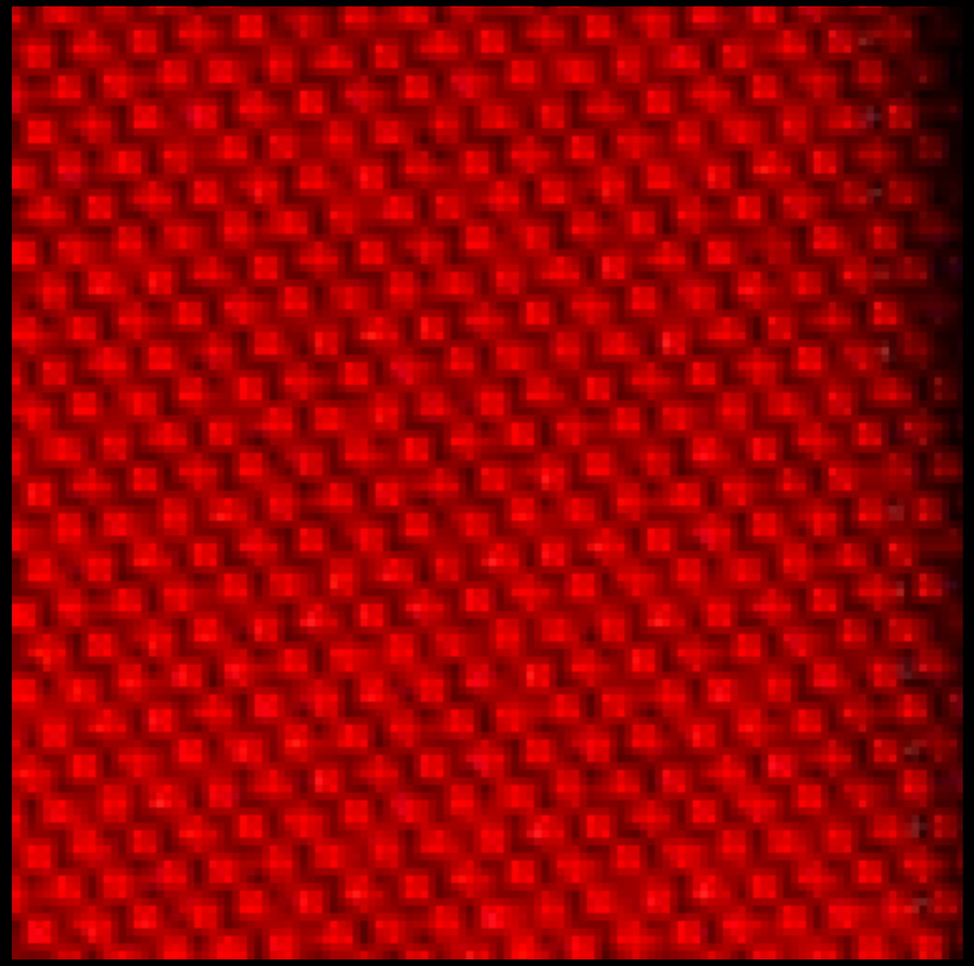


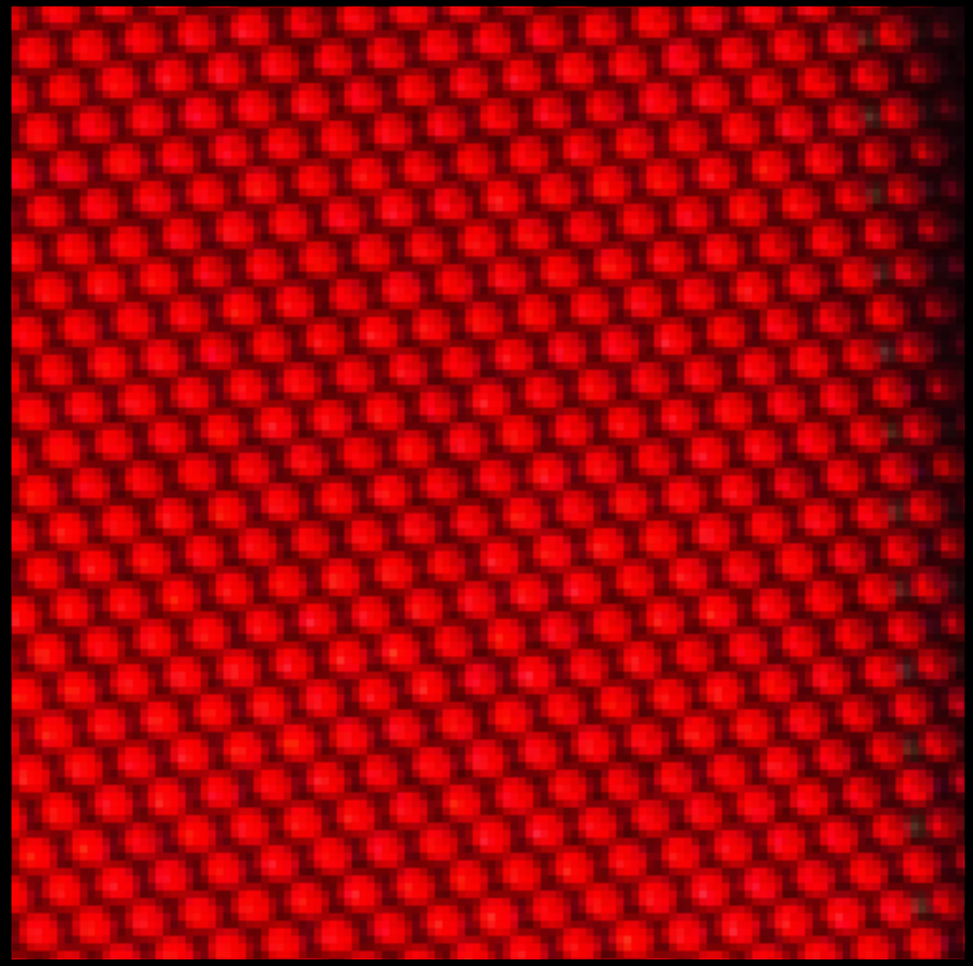
Demosaicking only

our test dataset



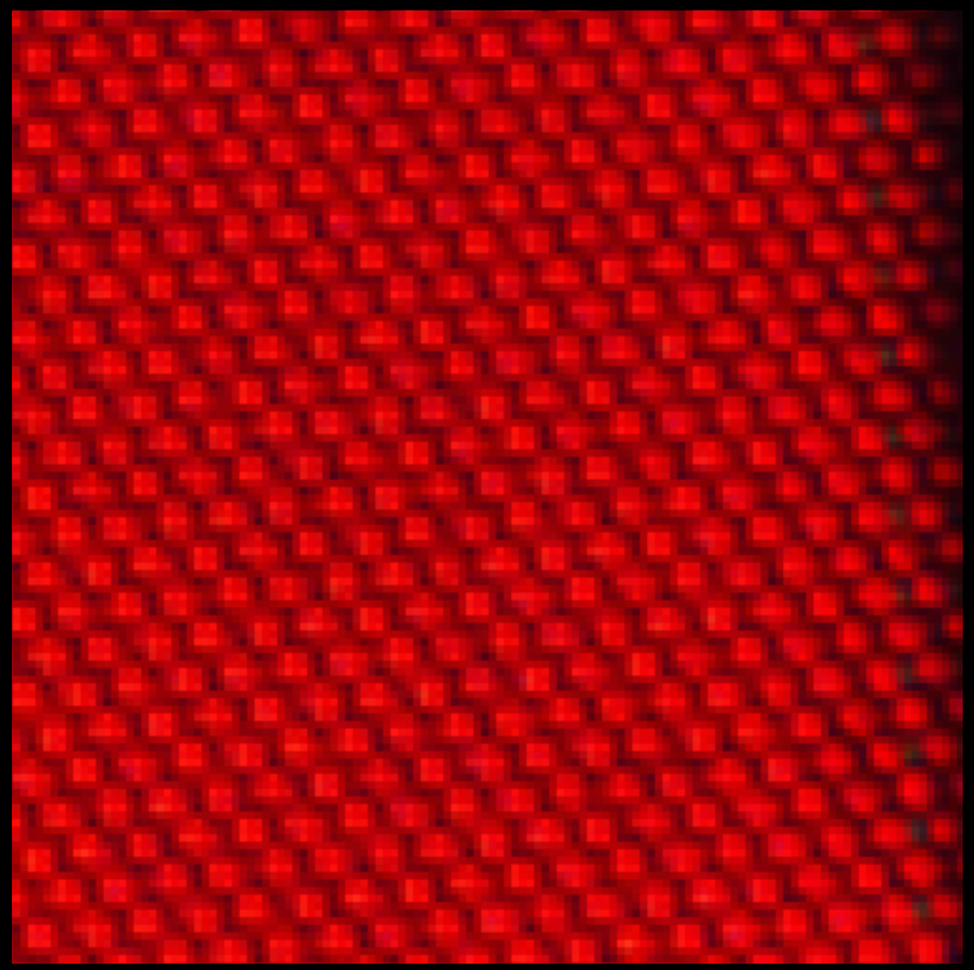
Photoshop
25 dB

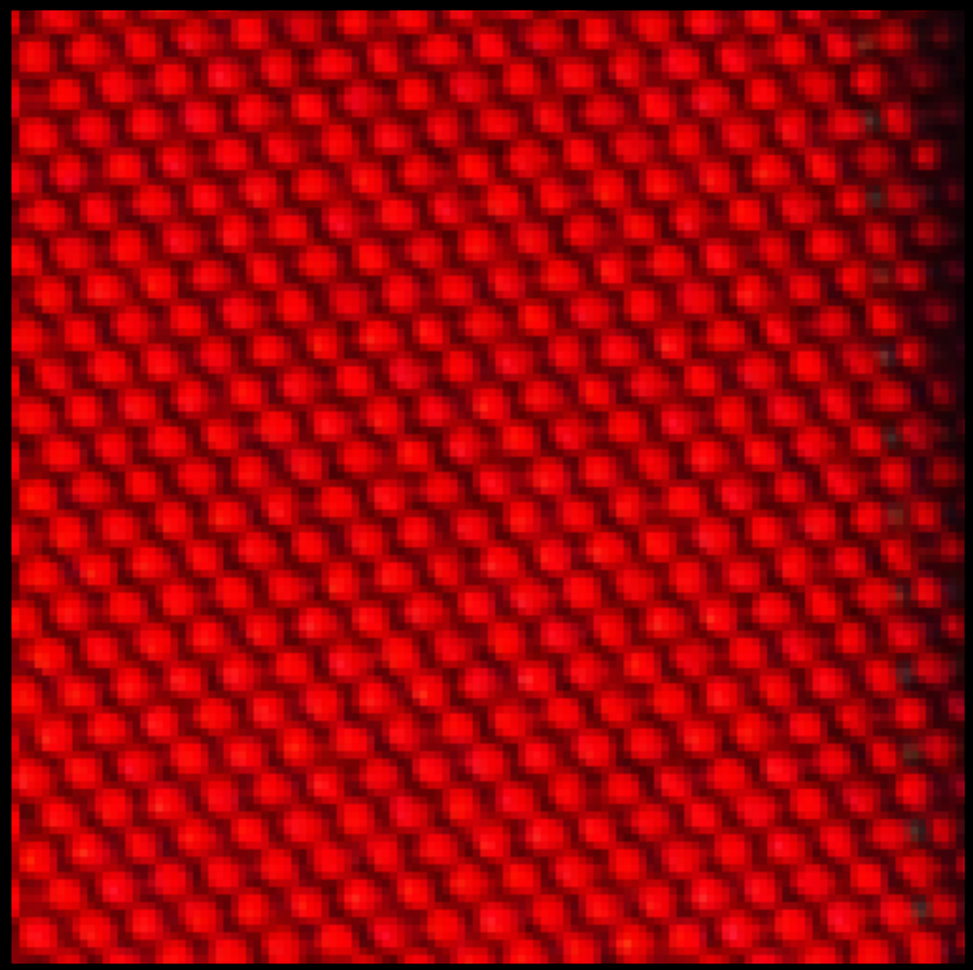




ref

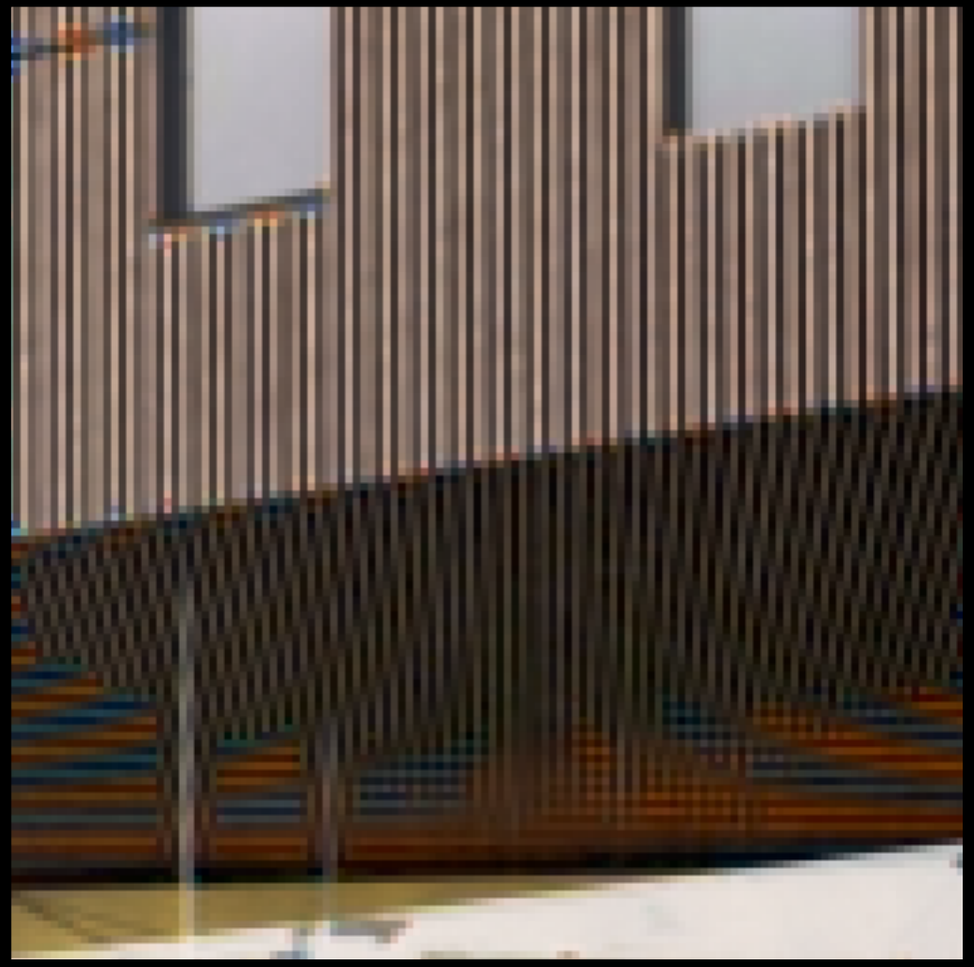
FlexISP
27 dB

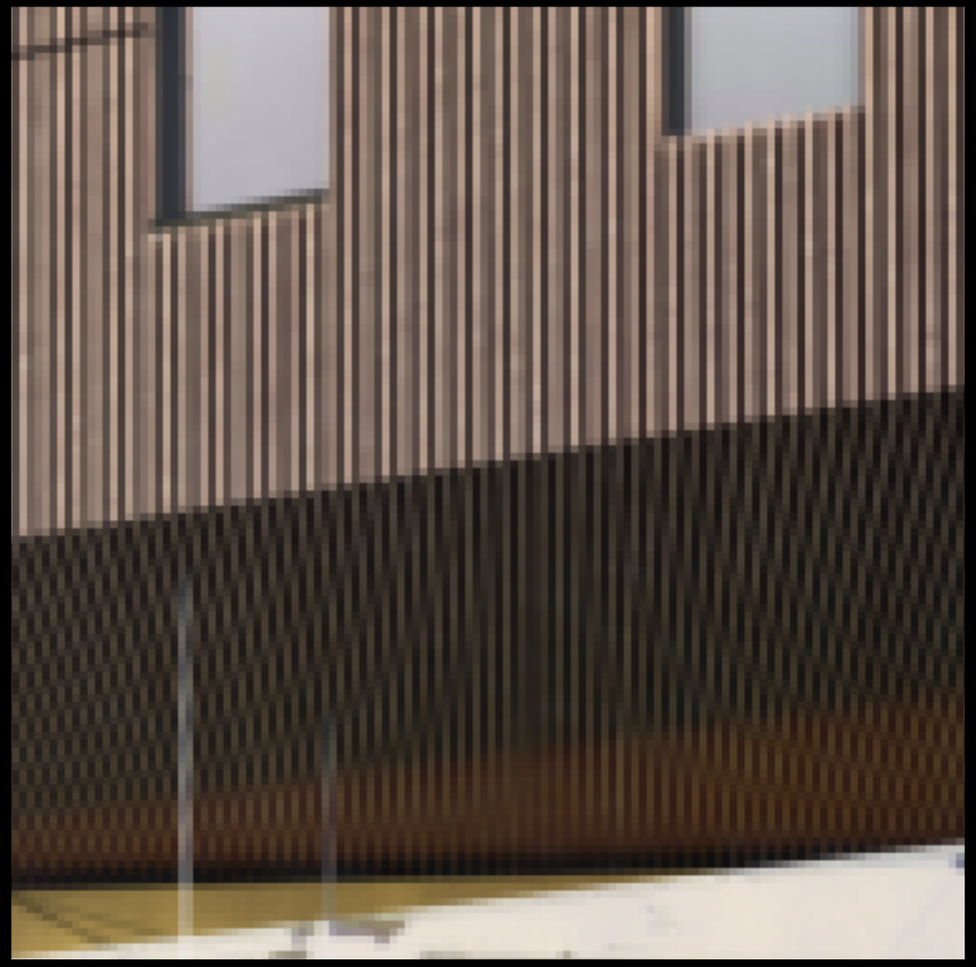




ours
31 dB

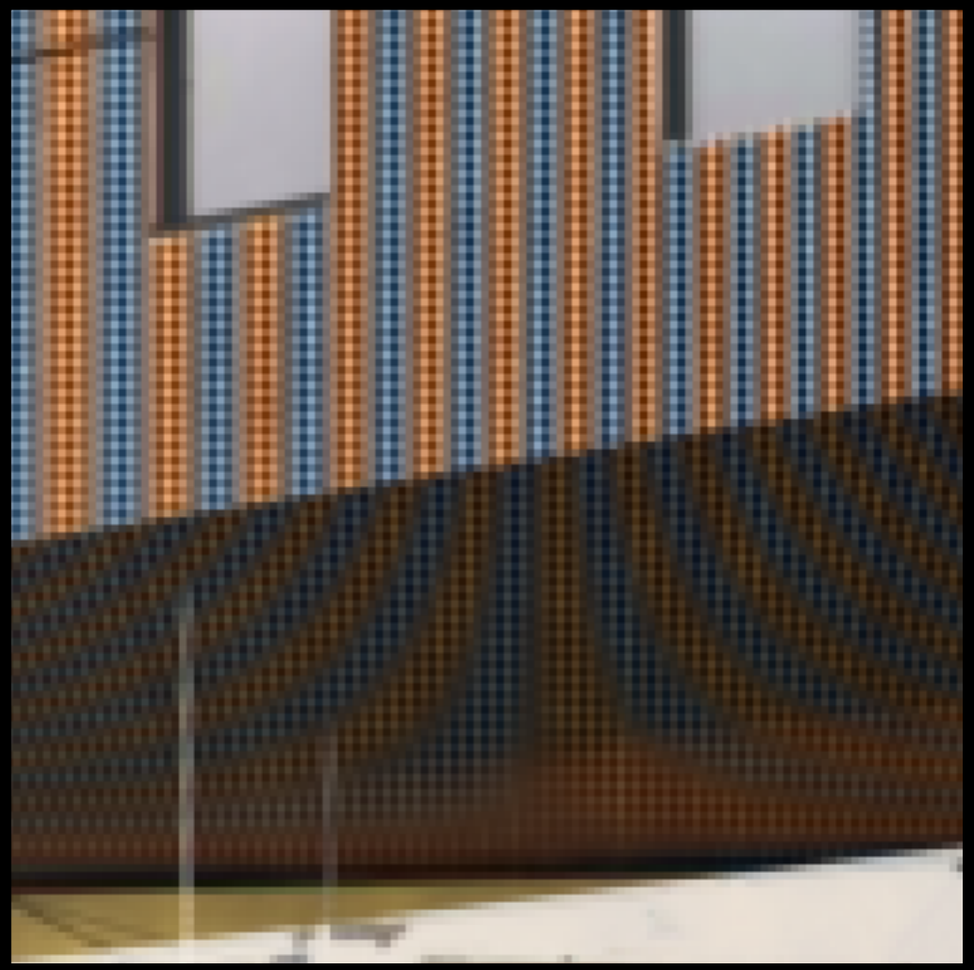
Photoshop
30 dB





ref

FlexISP
22 dB

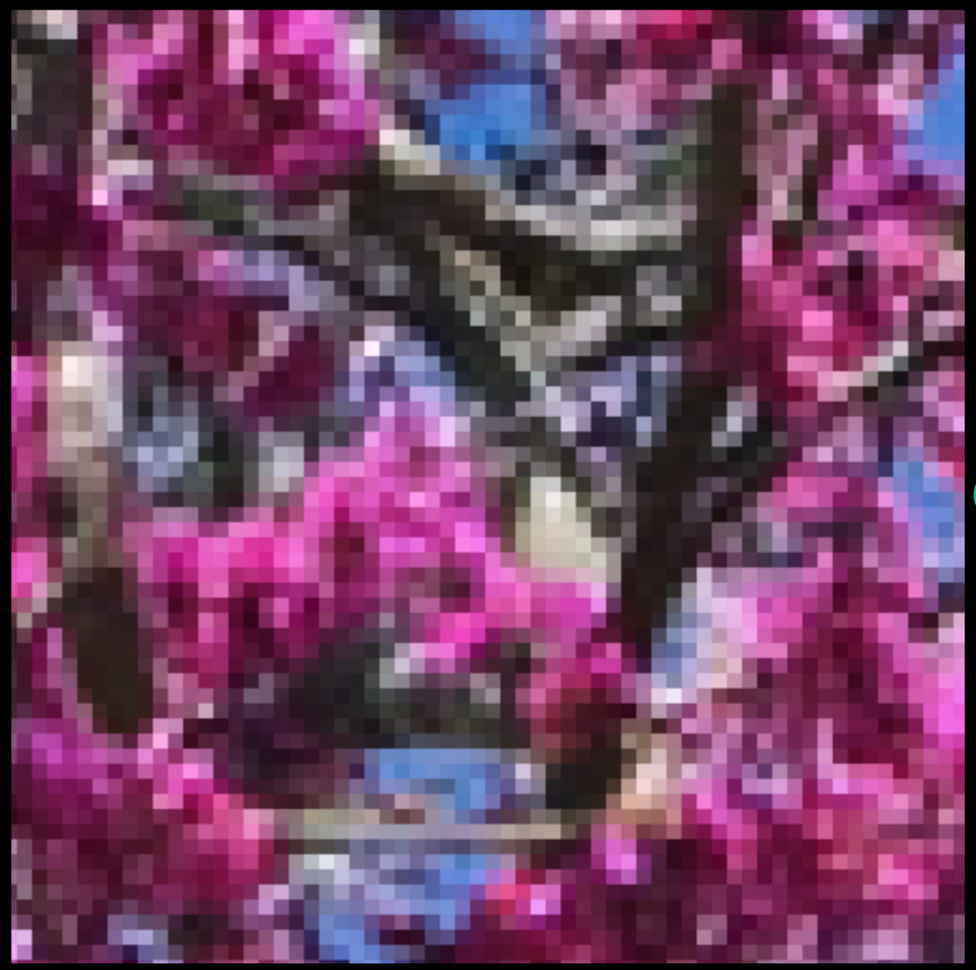


ours
40 dB

Photoshop
24 dB

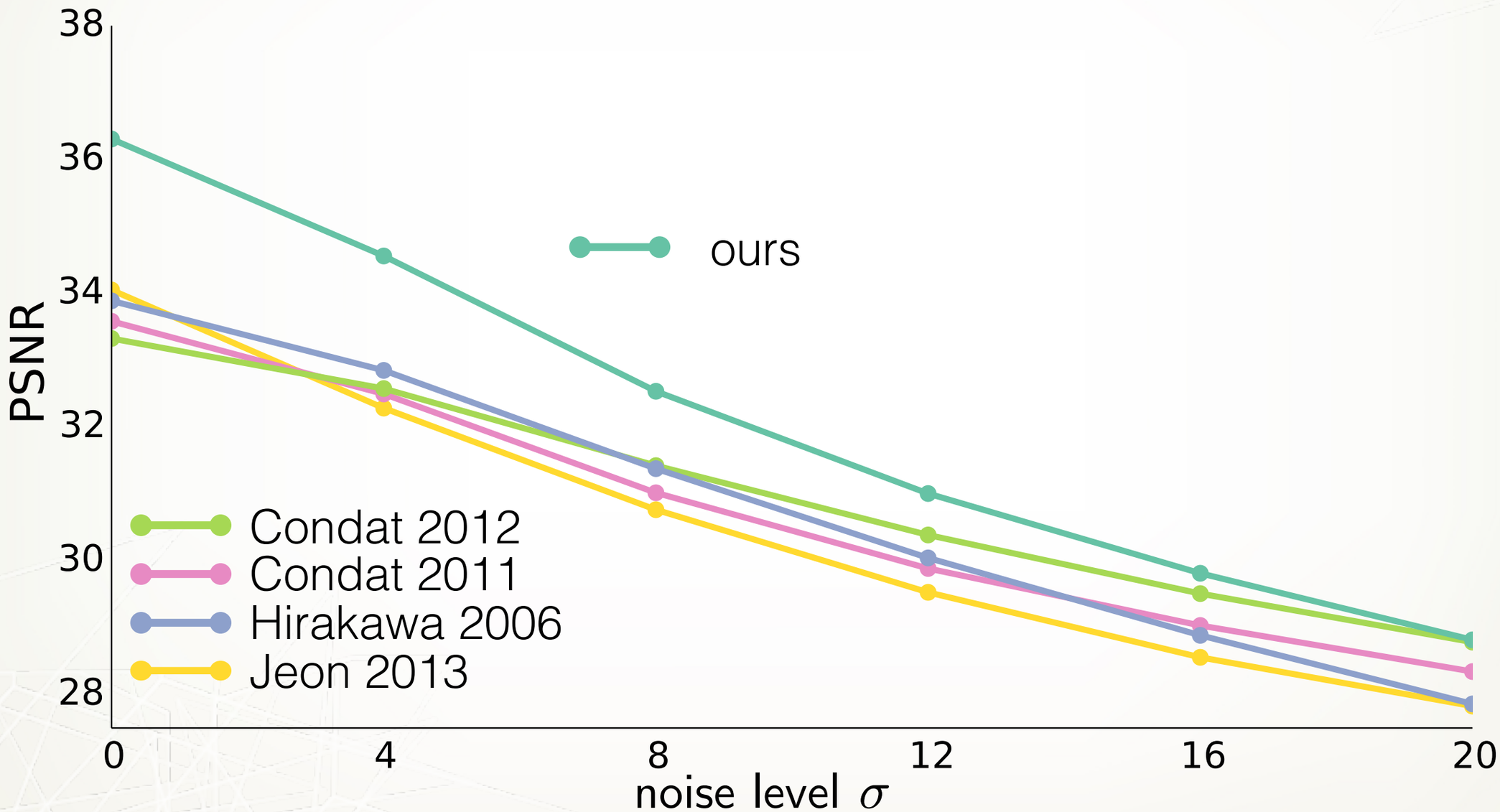
ref

FlexISP
22 dB

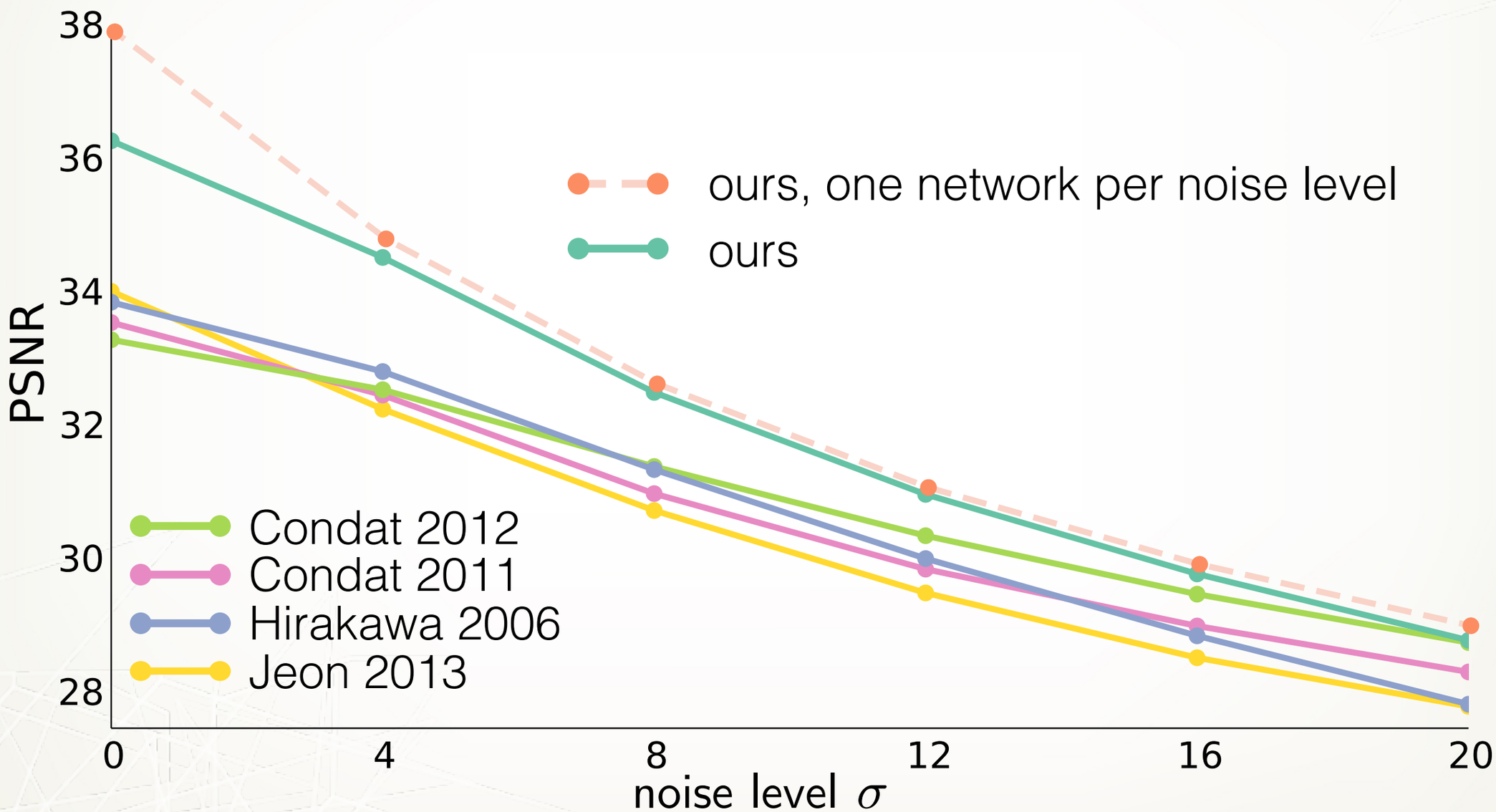


ours
27 dB

Joint demosaicking/denoising



Joint demosaicking/denoising



noisy
σ=4

ref

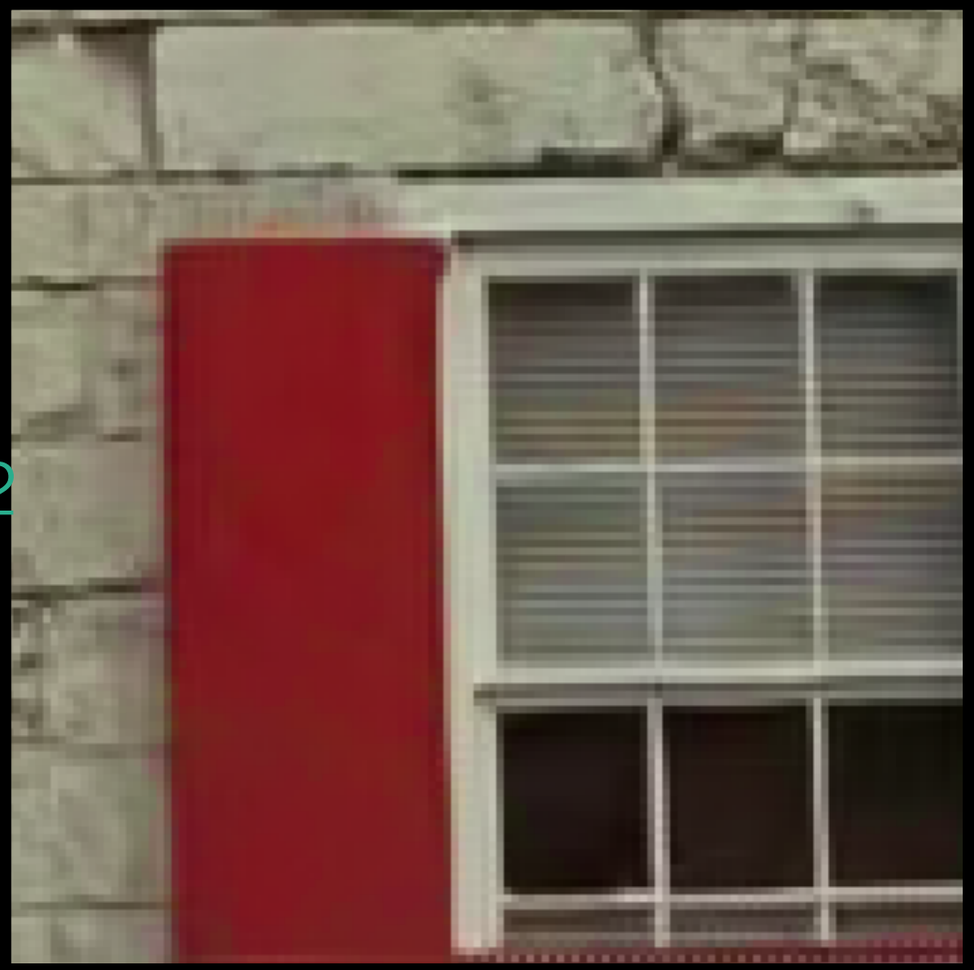
Condat 2012
27 dB

ours
31 dB

noisy
σ=12

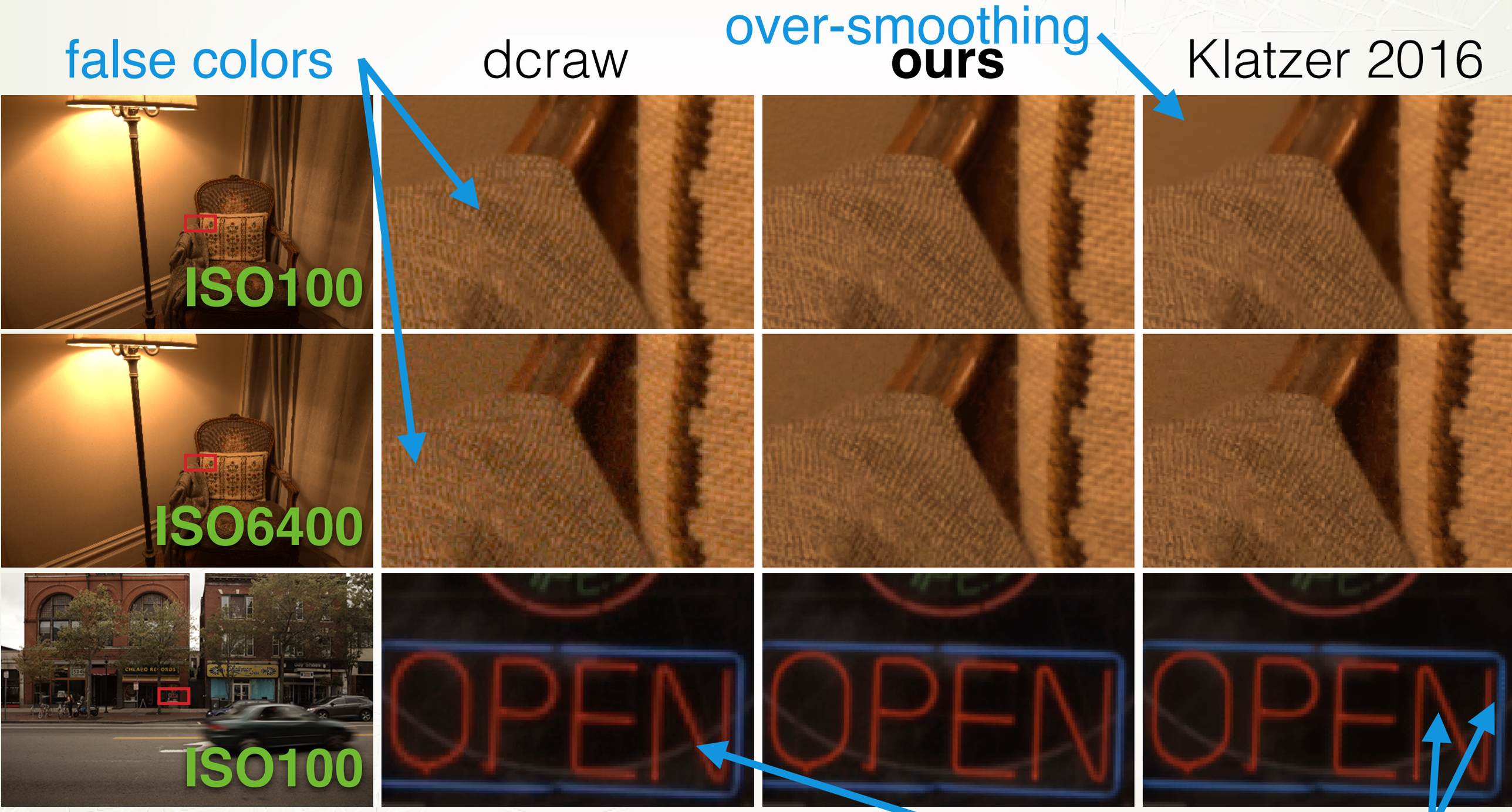
ref

Condat 2012
27 dB



ours
31 dB

Real RAW data

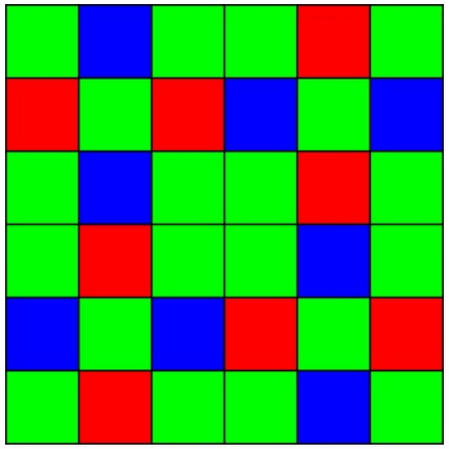


captured with a Canon 5D mark II

edge artifacts

Non-Bayer mosaic

RAW Fuji X-Trans



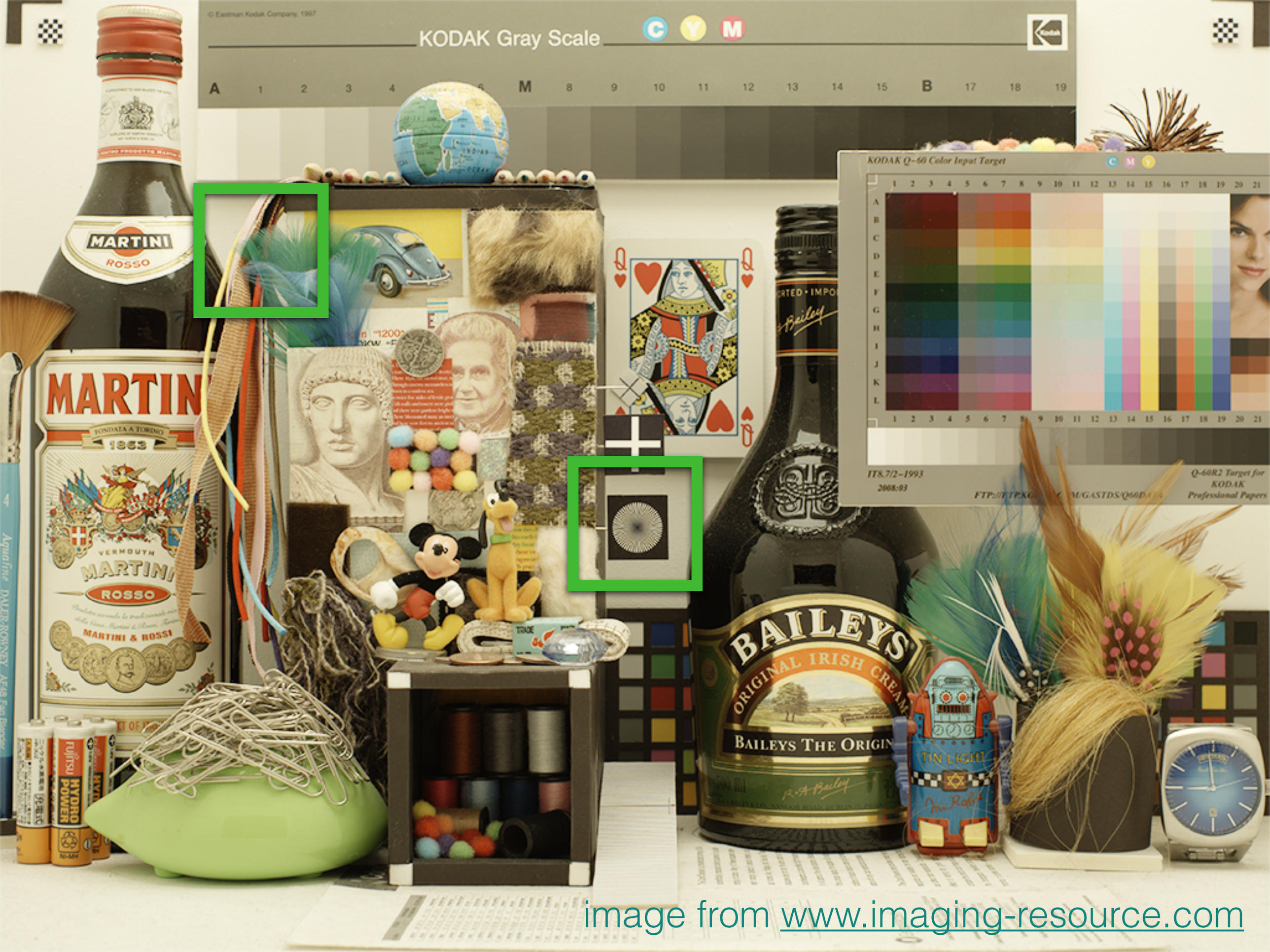
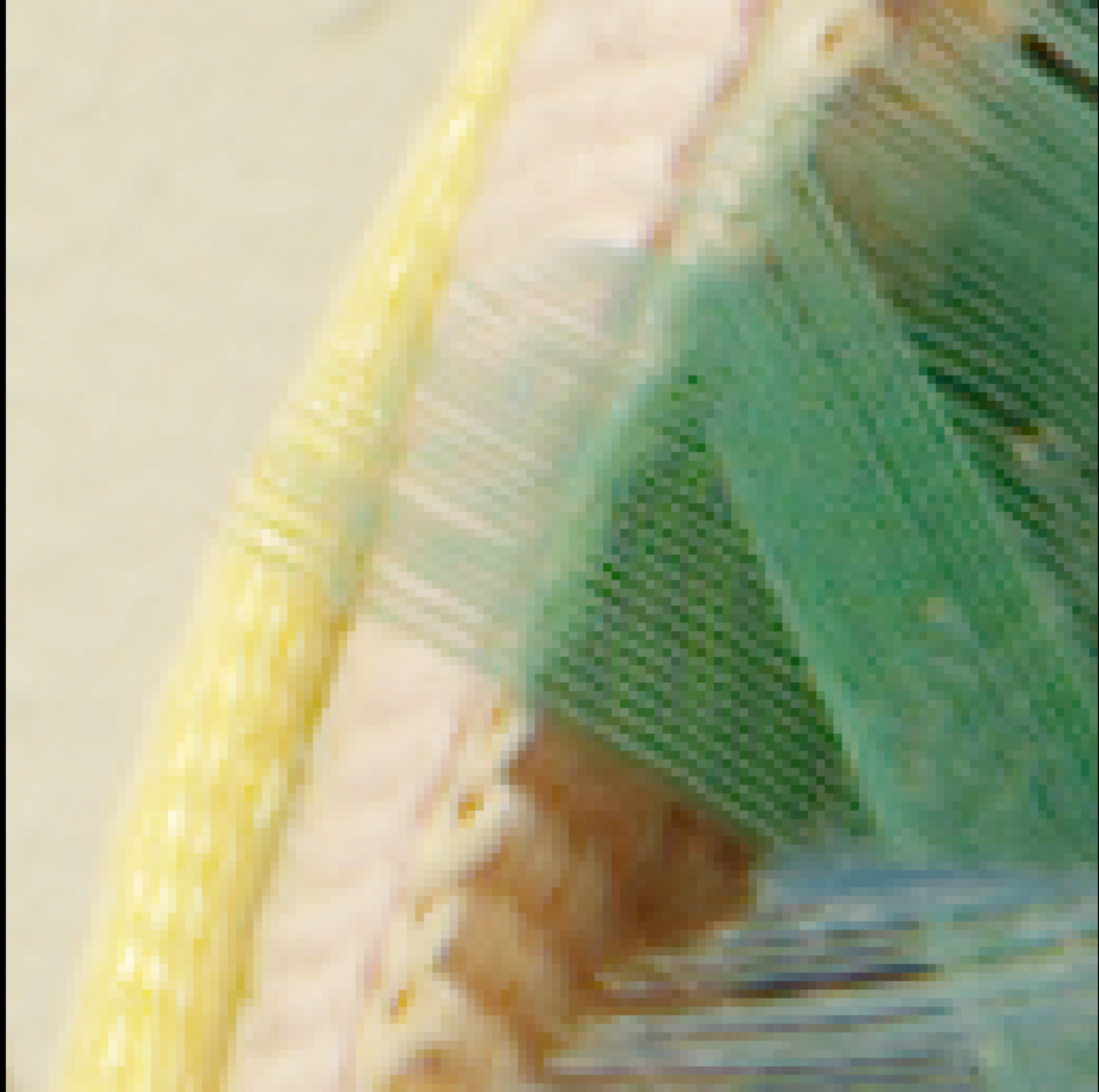


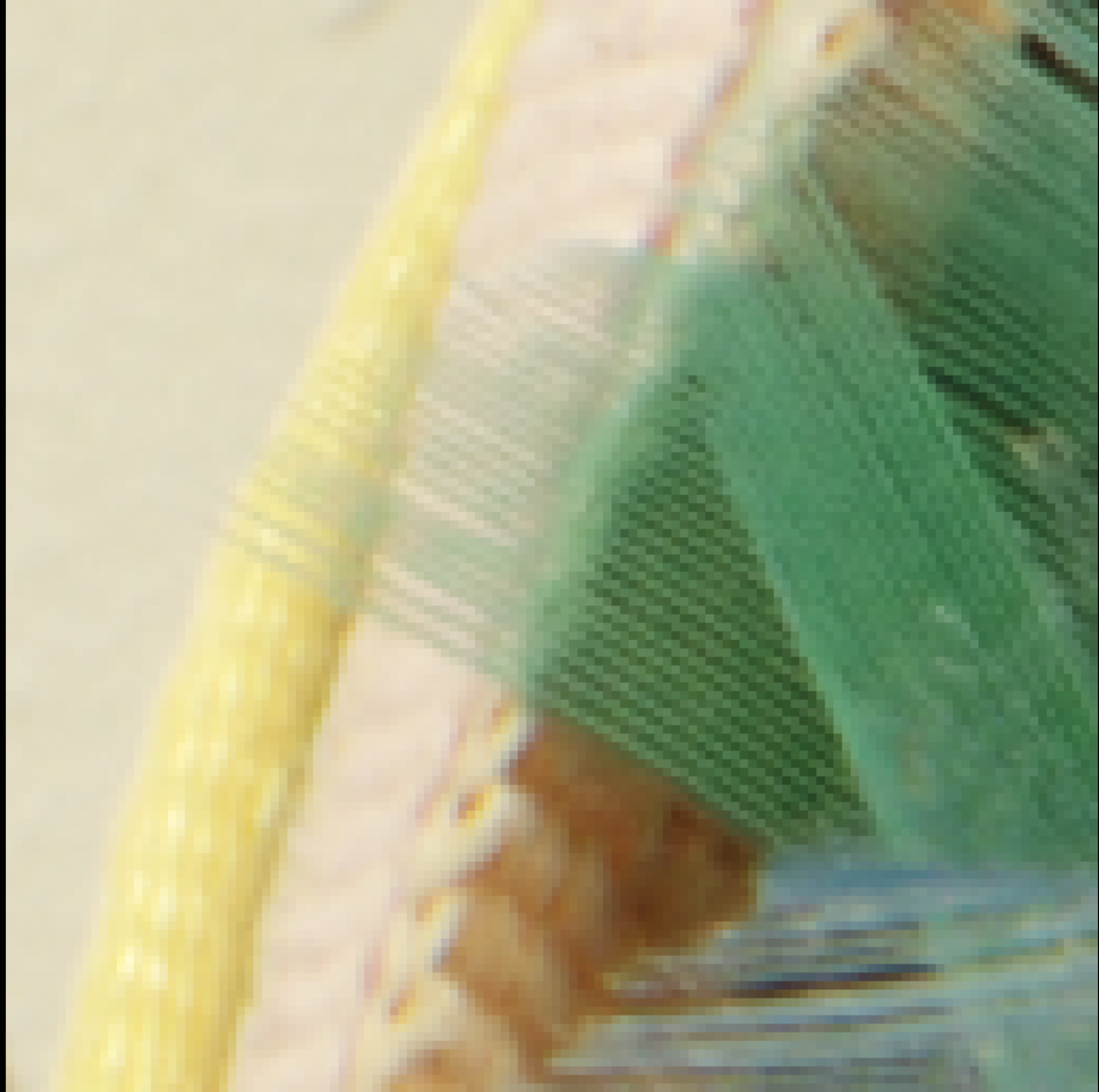
image from www.imaging-resource.com

Fuji X-Trans
dcraw

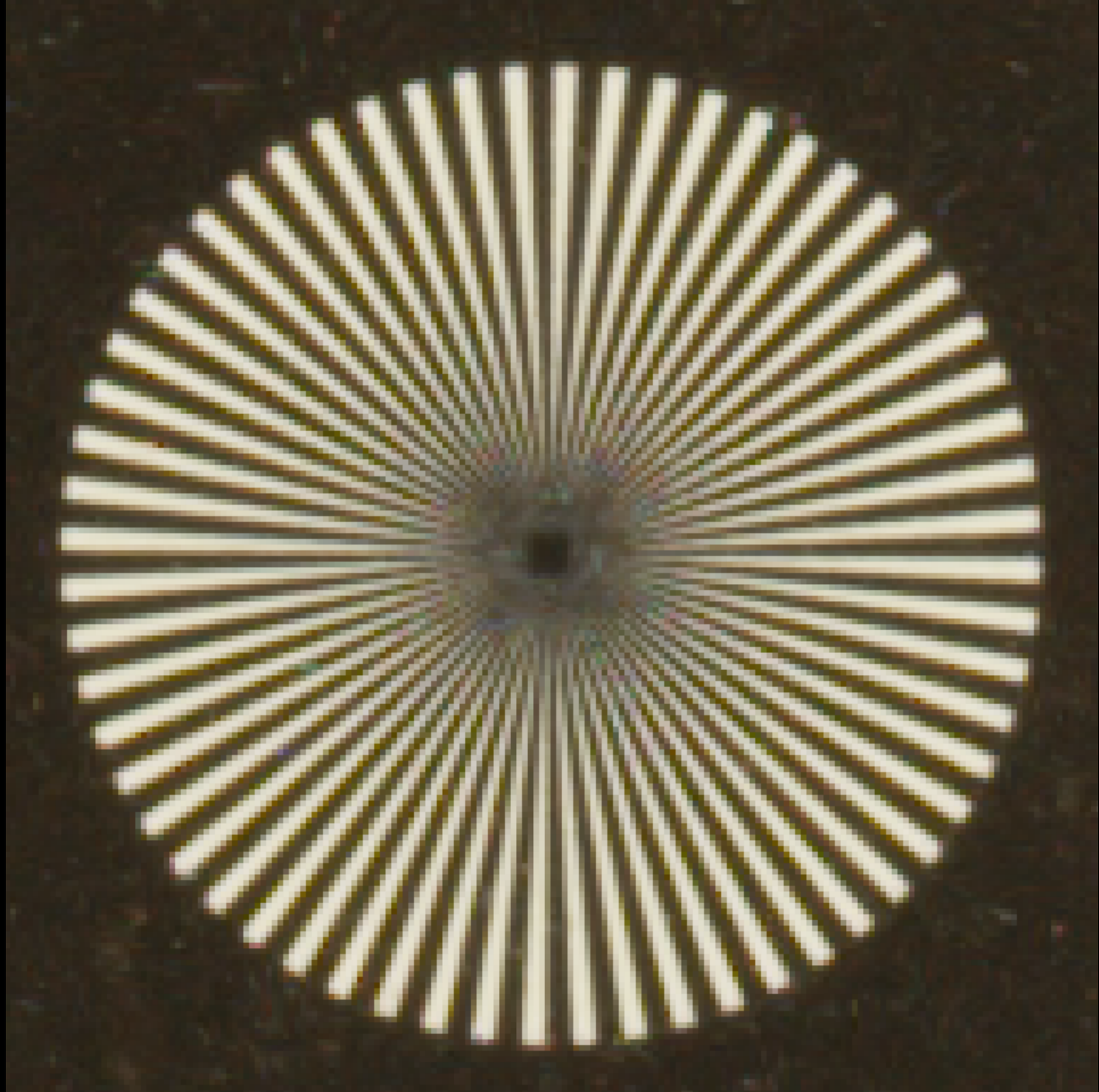


Fuji X-Trans

ours

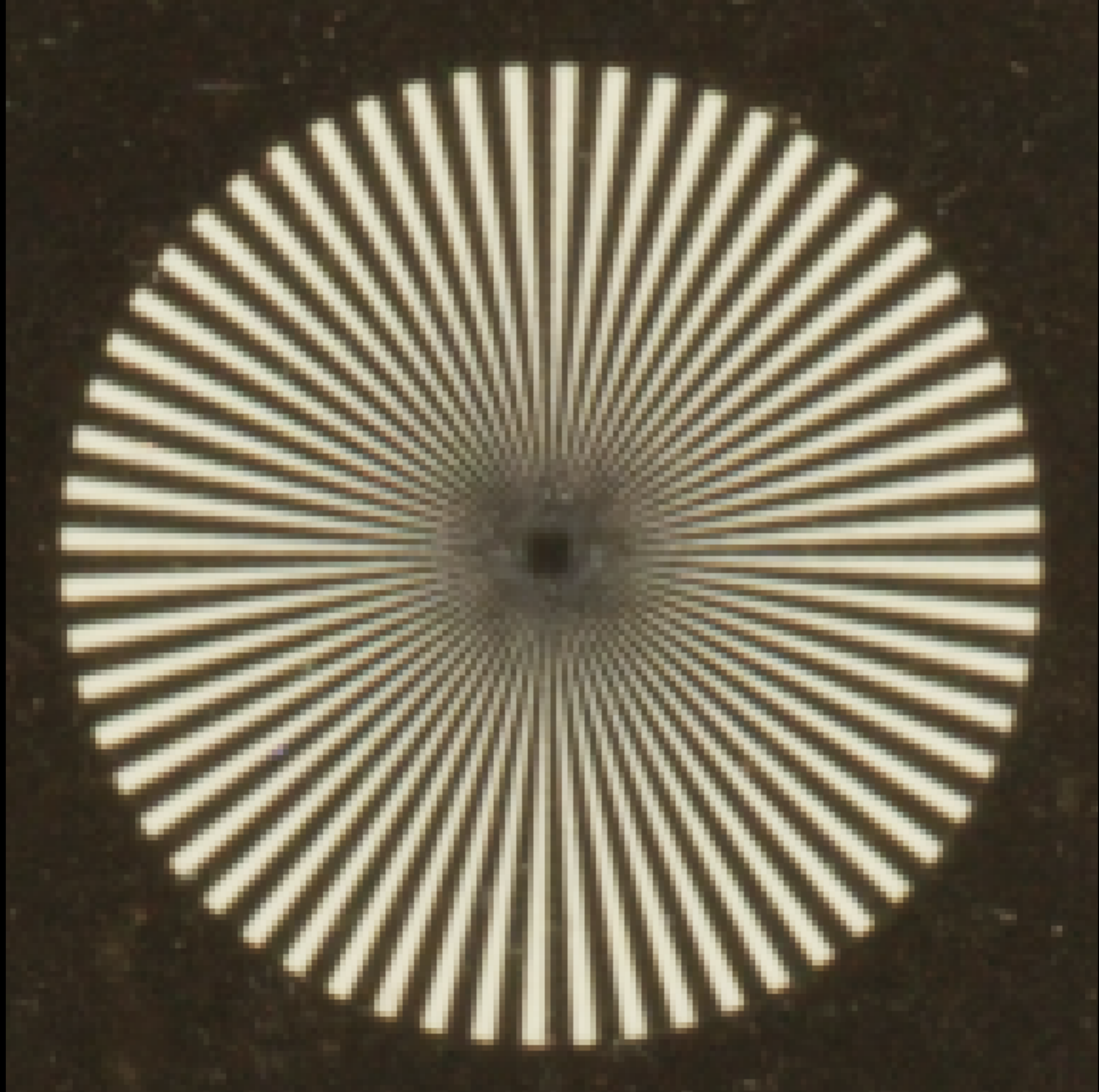


Fuji X-Trans
dcraw

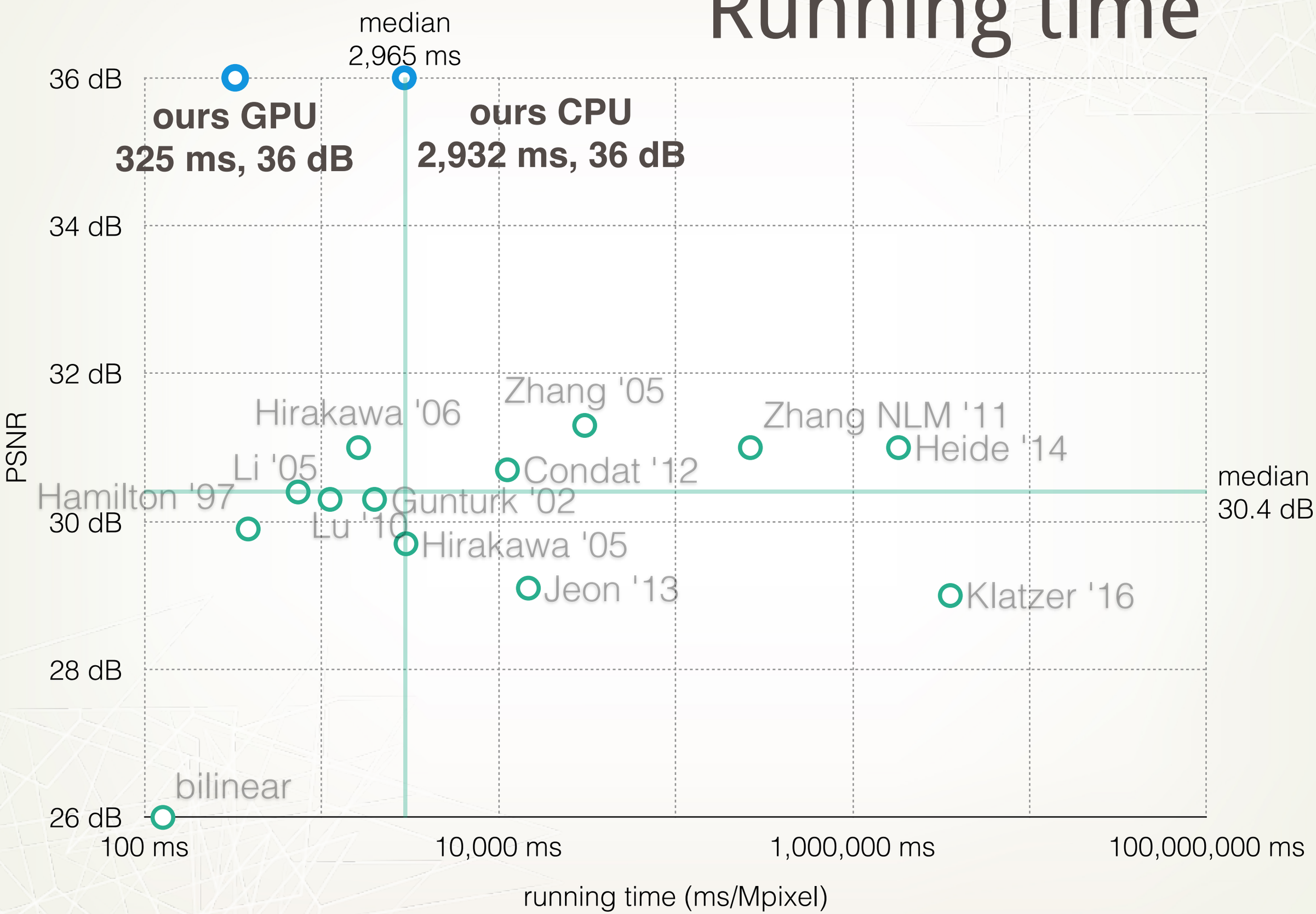


Fuji X-Trans

ours

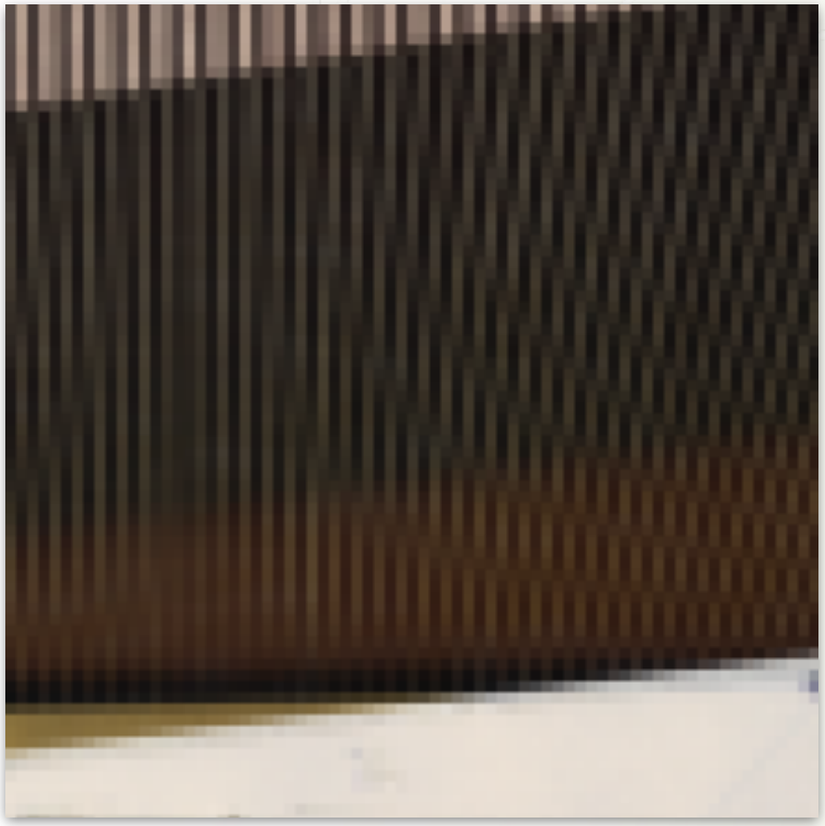


Running time



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