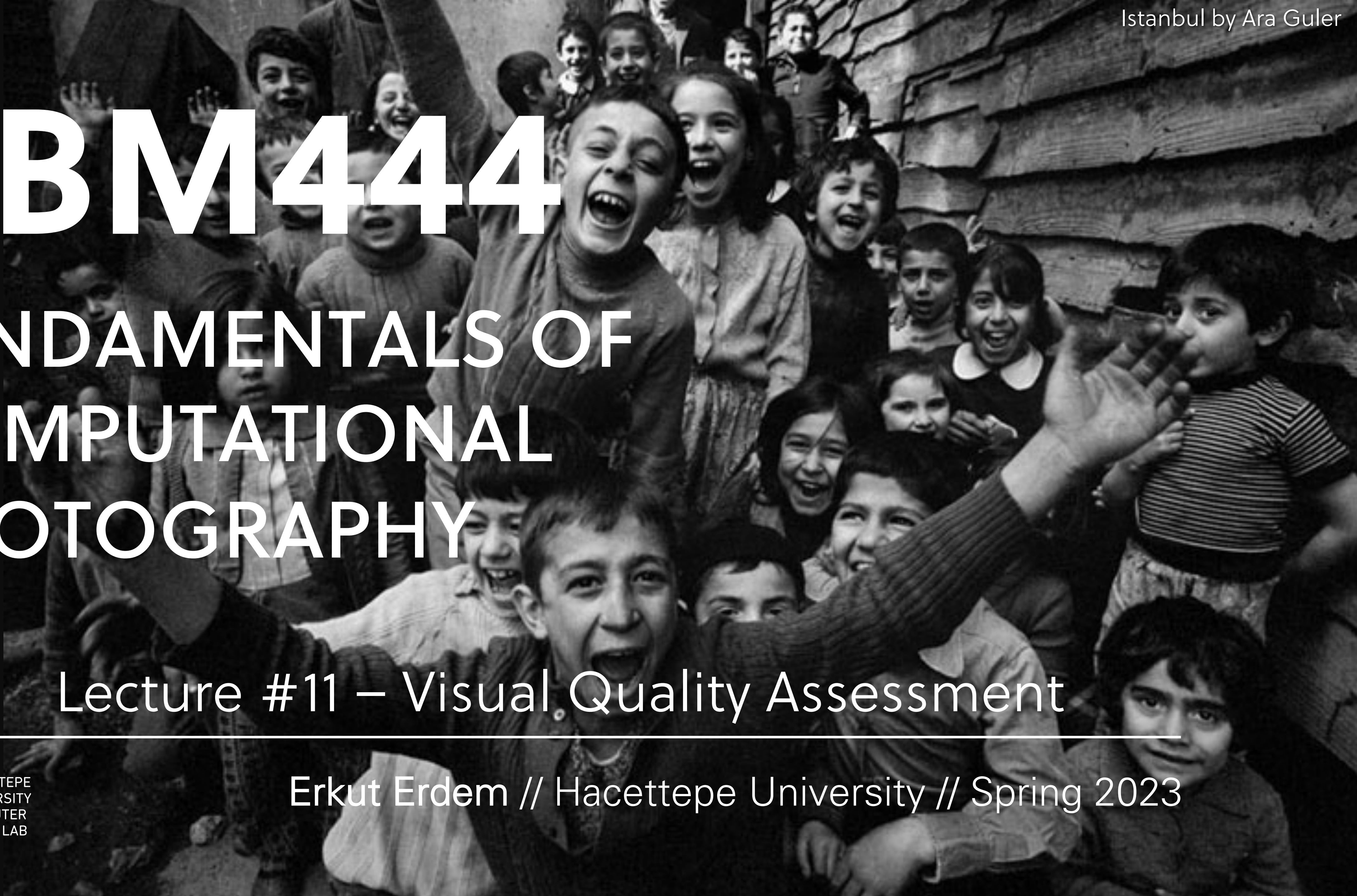


Istanbul by Ara Guler

BBM444

FUNDAMENTALS OF COMPUTATIONAL PHOTOGRAPHY

Lecture #11 – Visual Quality Assessment

Erkut Erdem // Hacettepe University // Spring 2023

HACETTEPE UNIVERSITY COMPUTER VISION LAB

Today's Lecture

- Introduction about image quality assessment (IQA)
- Full-reference IQA models
- No-reference IQA models
- The Perception-Distortion Tradeoff
- What makes a great picture?

Disclaimer: The material and slides for this lecture were borrowed from

—Alexei Efros's CS194-26/294-26 "Intro to Computer Vision and Computational Photography" class

—Kede Ma and Yuming Fang's "Image Quality Assessment in the Modern Age" tutorial at ACM MM 2021

Introduction about image quality assessment

What is Image Quality Assessment?

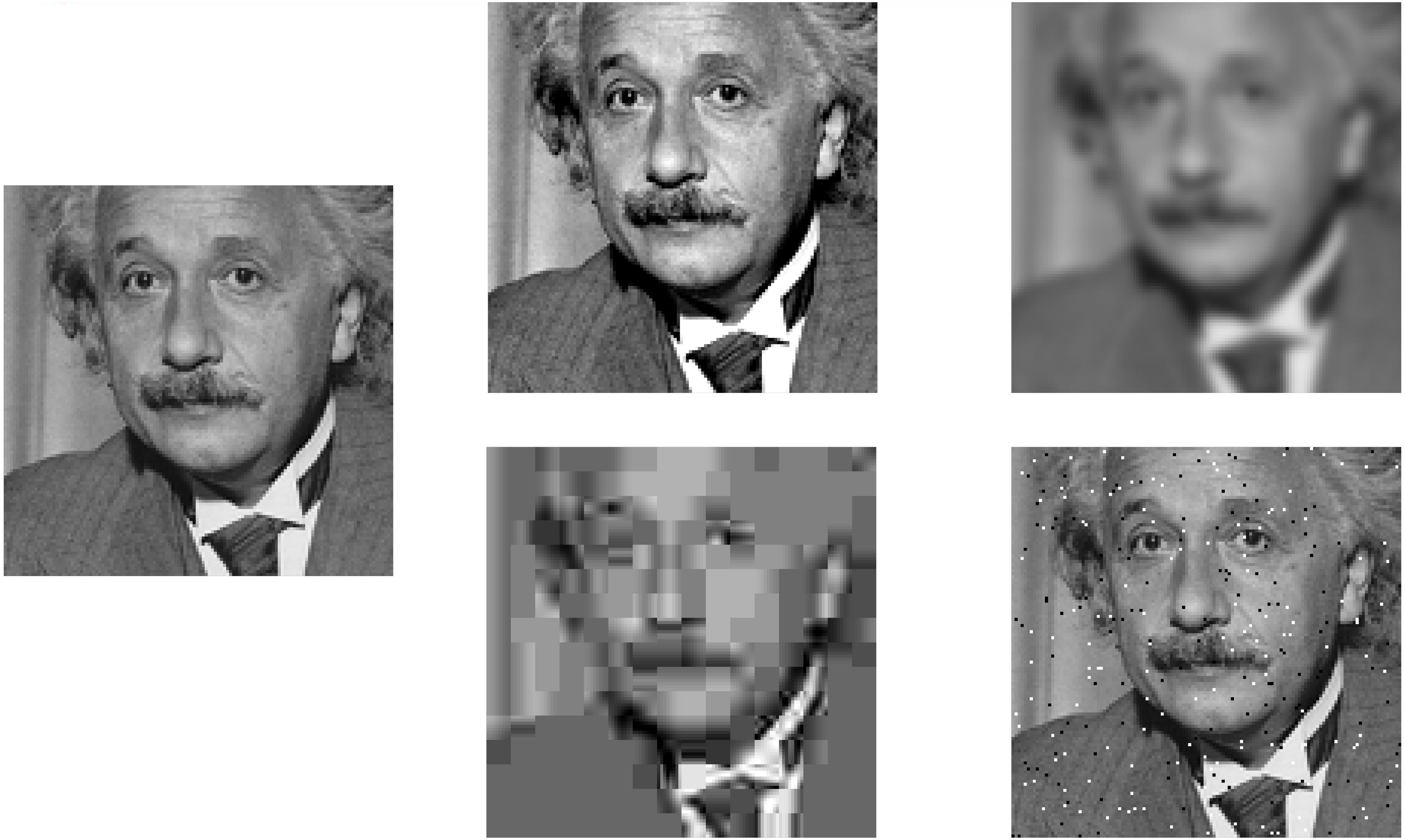


Image Restoration (IR) and Image Quality Assessment (IQA)

- Image Restoration (IR) aims at recovering a high-quality image from its degraded observation.
- Image Quality Assessment (IQA) methods were developed to measure the distortion/perceptual-quality of images.
- IQA methods are widely used to evaluate IR algorithms, e.g., PSNR, SSIM and Perceptual Index (PI).

Slide credit: Gu et al.

Synthetic and Authentic Distortions

Synthetic Distortions: Simulated by Pristine Image

Pristine image

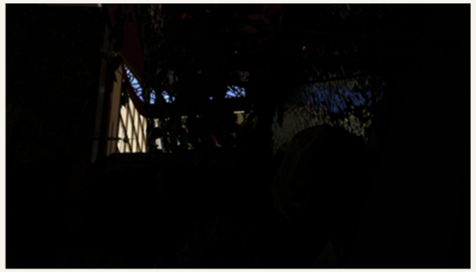
BLUR: level 4

JPEG: level 4

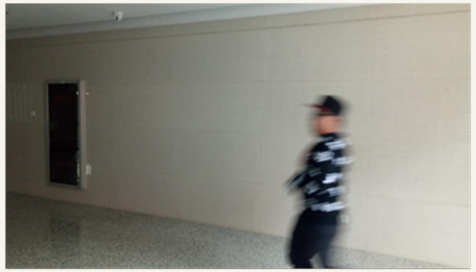
JP2K: level 4

Realistic Distortions: Captured from Mobile Devices

Smartphone Photography



Under-expoure



Motion blurring

Mixture distortions

Visual Quality Assessment

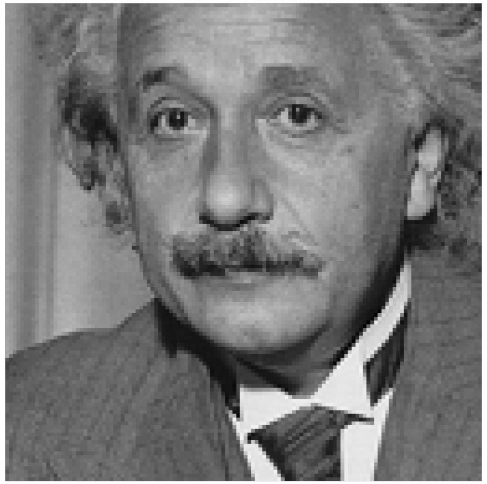
- Subjective quality assessment
  - Reliable and accurate quality prediction of visual content
  - Time-consuming, laborious and expensive
  - Not applicable in practical applications
- Objective quality assessment
  - Predict perceived visual quality automatically
  - Applicable in practical applications

Subjective Image Quality Assessment

- Absolute category rating (ACR)
  - Single stimulus method
  - Test images are presented one at a time without reference information
  - Voting time: less or equal to 10 seconds depending on the voting method
  - Presentation time: 10 seconds depending on the test image content
  - Five-level or nine-level scale overall rating
- Absolute category rating with hidden reference (ACR-HR)
  - The only difference from the ACR method: a reference version of each test image must be included as the test stimulus, which is termed as a hidden reference condition

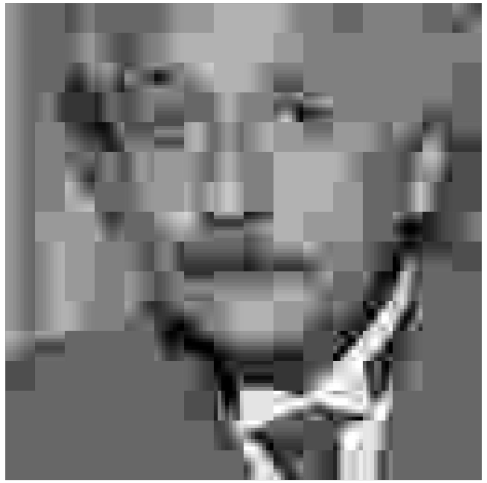
Subjective Image Quality Assessment

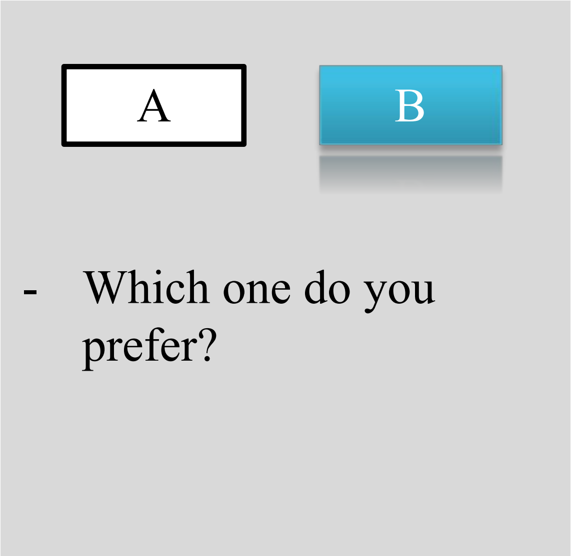
- Degradation category rating (DCR)
  - Double stimulus method
  - Test images are presented in pairs: one is reference image, while the other is distorted image
  - Voting time: less or equal 10 seconds depending on voting method
  - Presentation time: 10 seconds depending on the image content
  - Five-level scale overall rating



Subjective Image Quality Assessment

- Pair comparison (PC)
  - Double stimulus method
  - Two test images from two different systems are presented in pair from the same reference image
  - Participants are asked to provide the judgment on which one is preferred in the test pair
  - All possible pairs are compared (N stimuli → N(n-1)/2 pairs)
  - (optional) Convert paired comparison data to scale values





LIVE Dataset

- Reference images: 29. Distorted images: 779.
- Distortion types: 5 (fast fading, Gaussian blur, JP2K, JPEG, white noise)



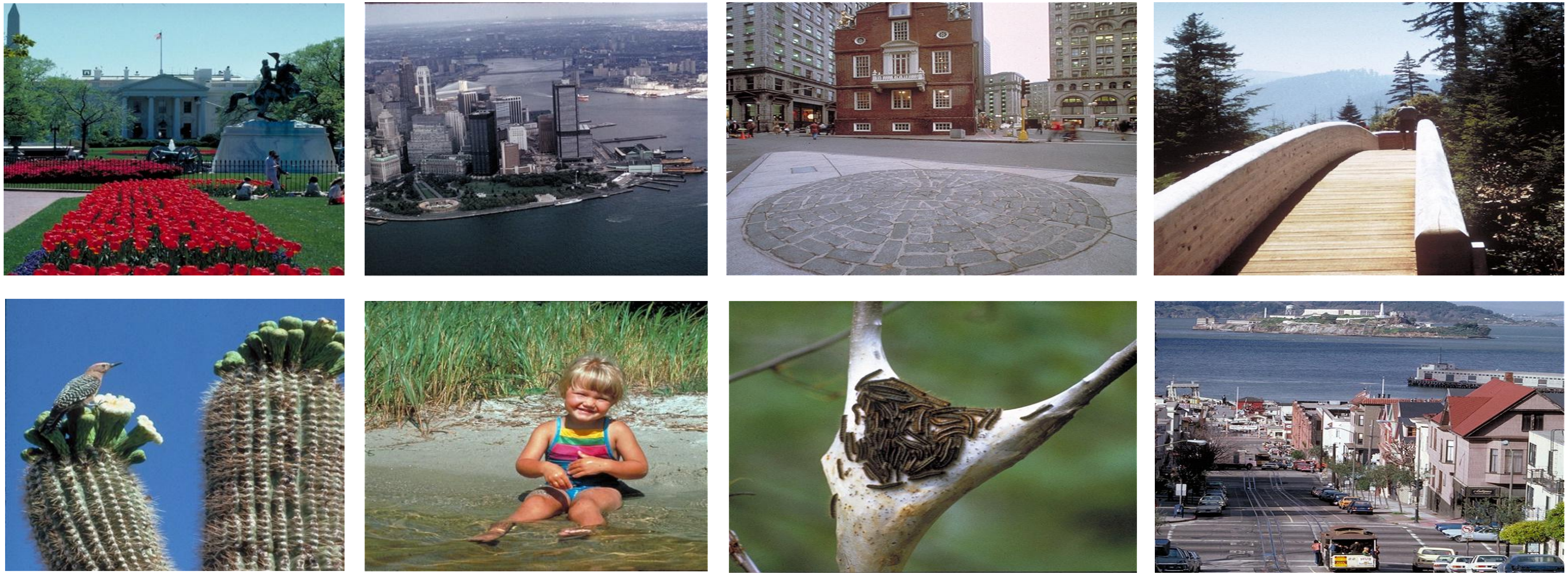




H. R. Sheikh, M. F. Sabir and A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE T-IP, 2006

CSIQ Dataset

- Reference images: 30. Distorted images: 866.
- Distortion types: 6 (JPEG, JP2K, Gaussian blur, white noise, contrast change, pink noise)



E. C. Larson and D, M. Chandler, Most apparent distortion: Full-reference image quality assessment and the role of strategy, J Electronic Imaging, 2010

TID2013 Dataset

- Reference images: 25. Distorted images: 3000.
- Distortion types: 24 (fast fading, Gaussian blur, JP2K, JPEG, white noise, etc.)

N. Ponomarenko, O. Ieremeiev, et al., Color image database TID2013: Peculiarities and preliminary results, in European Workshop on Visual Information Processing, 2013

KADID-10K Dataset

- Reference images: 81. Distorted images: 10125.
- Distortion types: 25 (Gaussian blur, JP2K, JPEG, white noise, motion blur, etc.)



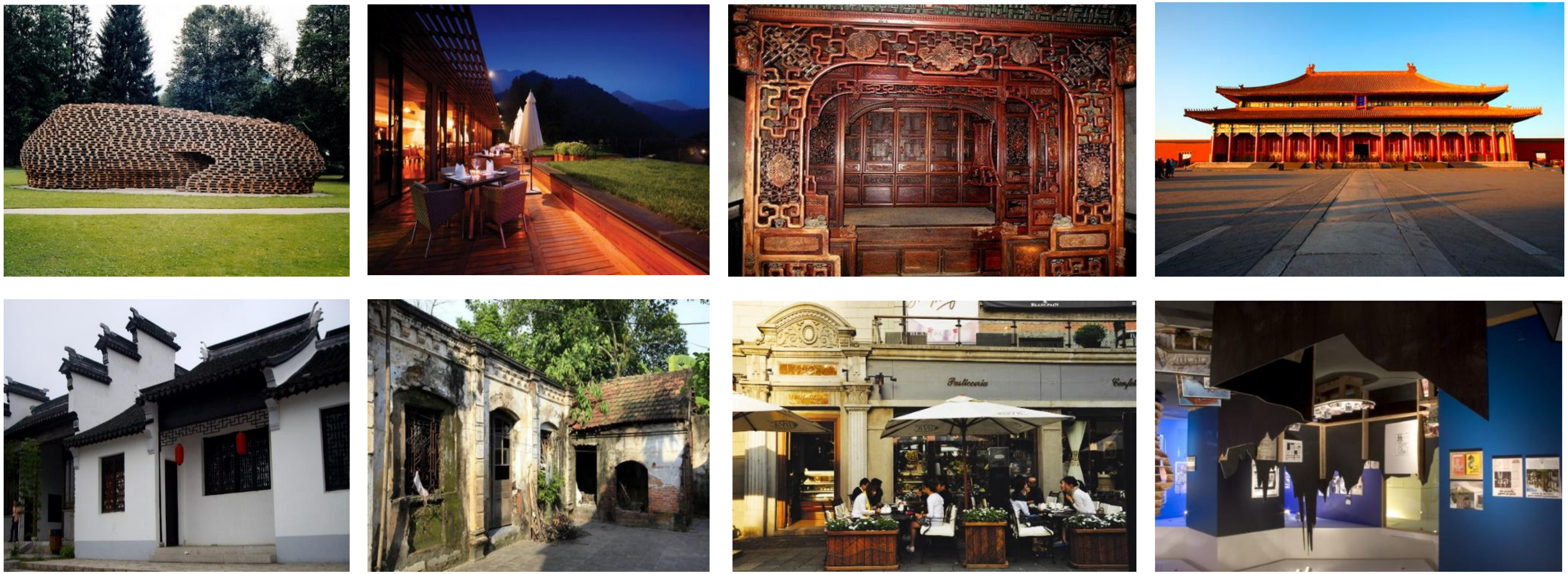





H. Lin, V. Hosu and D. Saupe, KADID-10K: A large-scale artificially distorted IQA database, in 2019 Eleventh International Conference on Quality of Multimedia Experience, 2019

Waterloo Exploration Dataset

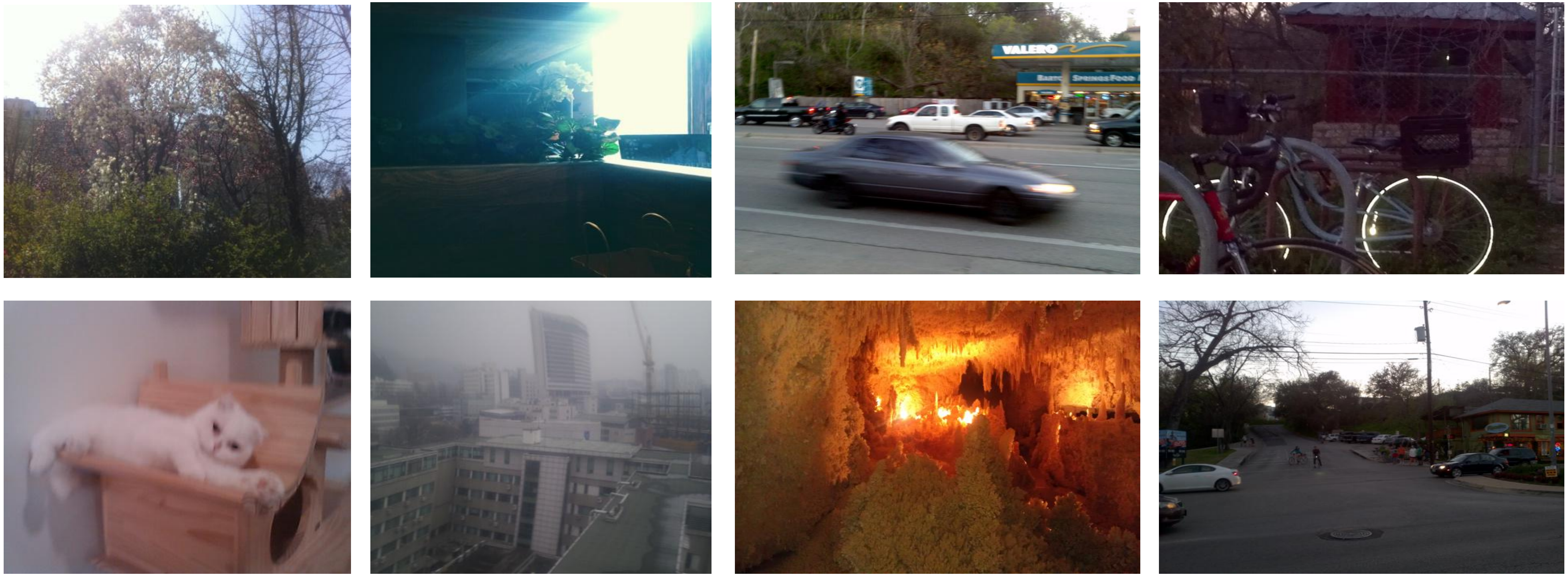
- Reference images: 4744. Distorted images: 94880.
- Distortion types: 4 (Gaussian blur, JP2K, JPEG, White noise.)



Kede Ma, et al., Waterloo exploration database: New challenges for image quality assessment models, IEEE T-IP, 2017

LIVE Challenge Dataset – Authentic Distortion

- Distorted images: 1162.
- Distortion types: Complex.



D. Ghadiyaram and A. C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, IEEE T-IP, 2015

KonIQ-10K Dataset – Authentic Distortion

- Distorted images: 10073.
- Distortion types: Complex.

V. Hosu, H. Lin, T. Sziranyi and D. Saupe, KonIQ-10K: An ecologically valid database for deep learning of blind image quality assessment, IEEE T-IP, 2020

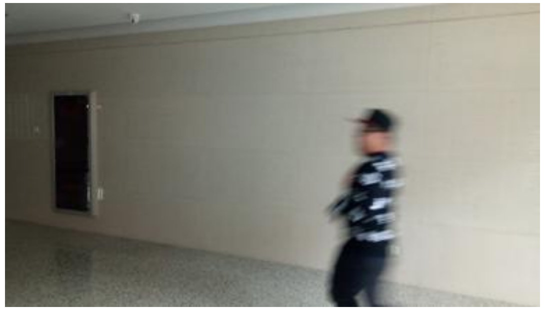
SPAQ Dataset – Authentic Distortion

- Distorted images: 11125 (taken by 66 smartphones with 11 manufacturers).
- Distortion types: Complex.


Under-exposure.


Over-exposure


Contrast reduction


Moving object blurring


Sensor noise


Out-of-focus

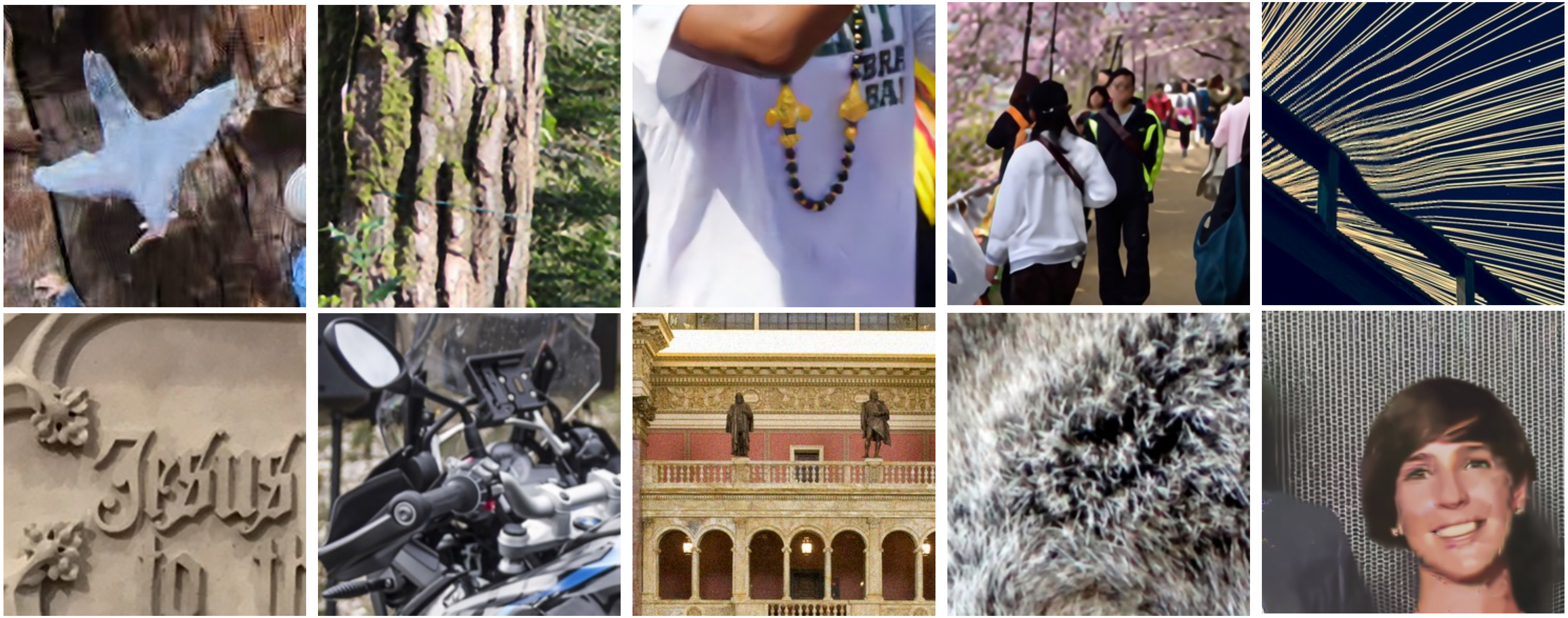

Camera motion blurring


Mixture distortions

Y. Fang, H. Zhu, Y. Zeng, K. Ma, Z. Wang, Perceptual Quality Assessment of Smartphone Photography, CVPR 2020

PIPAL Dataset

- Reference images: 250. Distorted images: 29000.
- Distortion types: 40 (GAN-based image restoration methods).



J. Gu, H. Cai, H. Chen, X. Ye, J. Ren, C. Dong, PIPAL: a Large-Scale Image Quality Assessment Dataset for Perceptual image Restoration, ECCV 2020

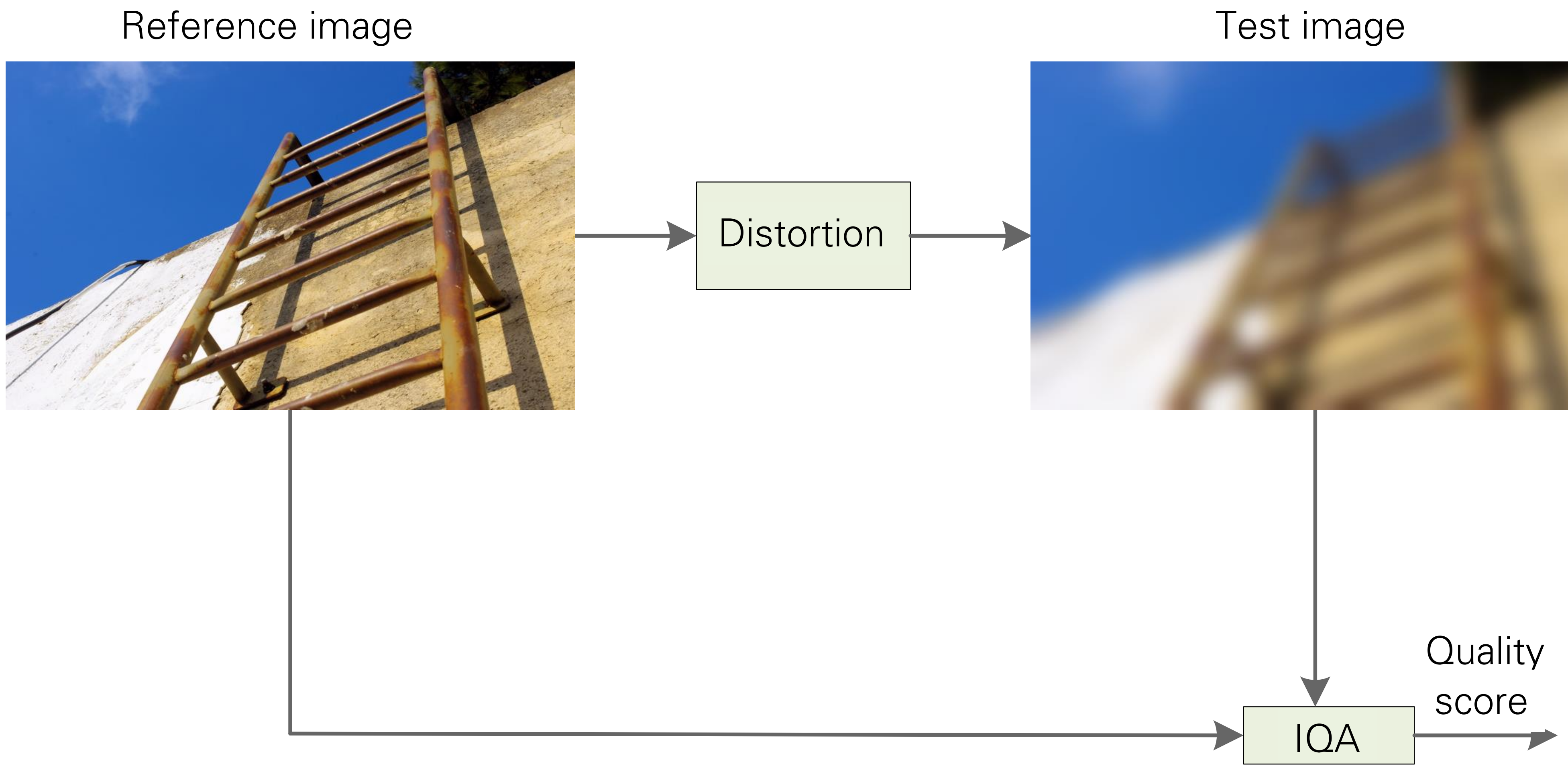
Objective Image Quality Assessment

- **Goal:** Build computational models that accurately predict human perception of image quality

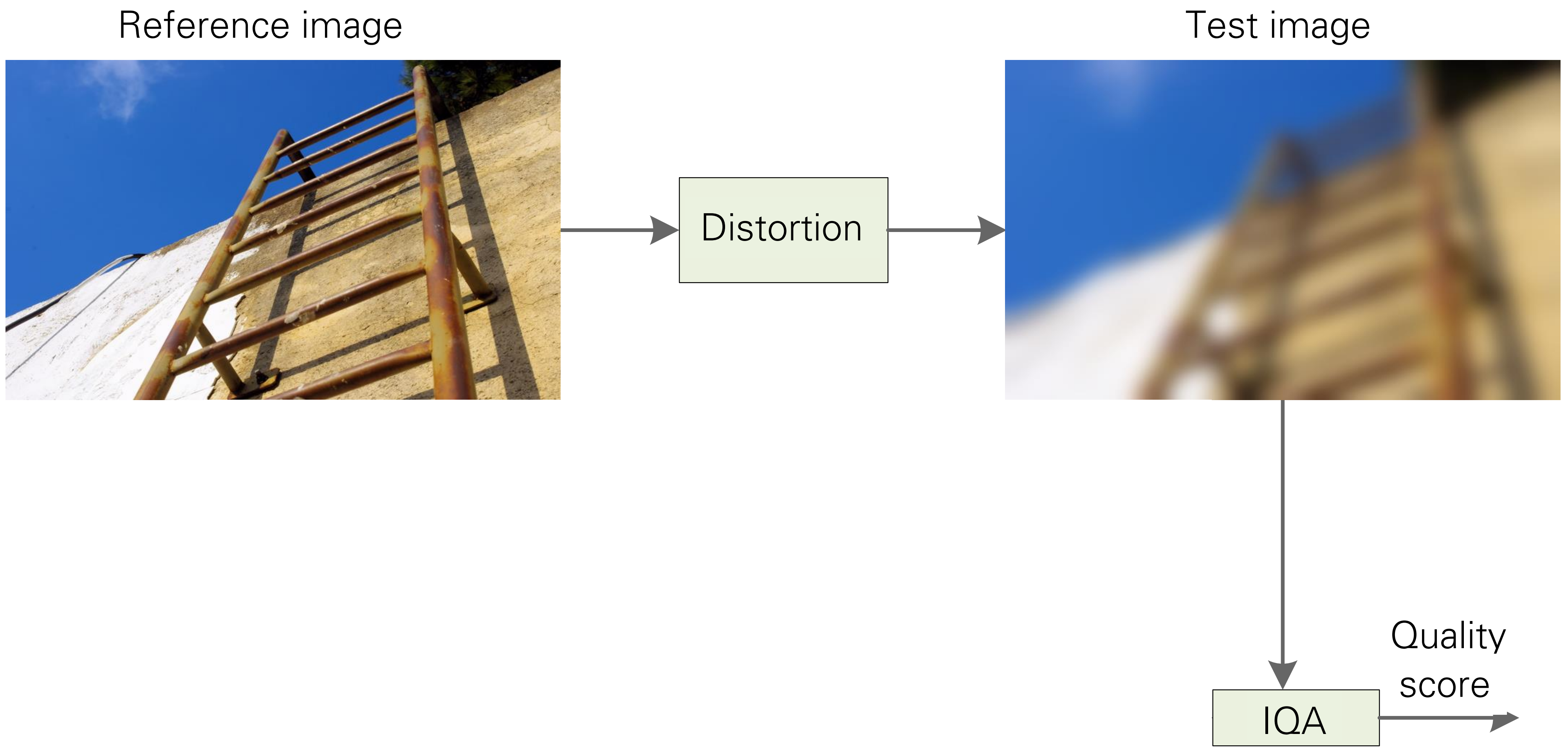
- Two categories:

1. Full-reference IQA
2. No-reference IQA

Full-Reference IQA



No-Reference IQA (Blind IQA - BIQA)



Full-reference IQA:
From Mean Squared Error
to Structural Similarity (and More)

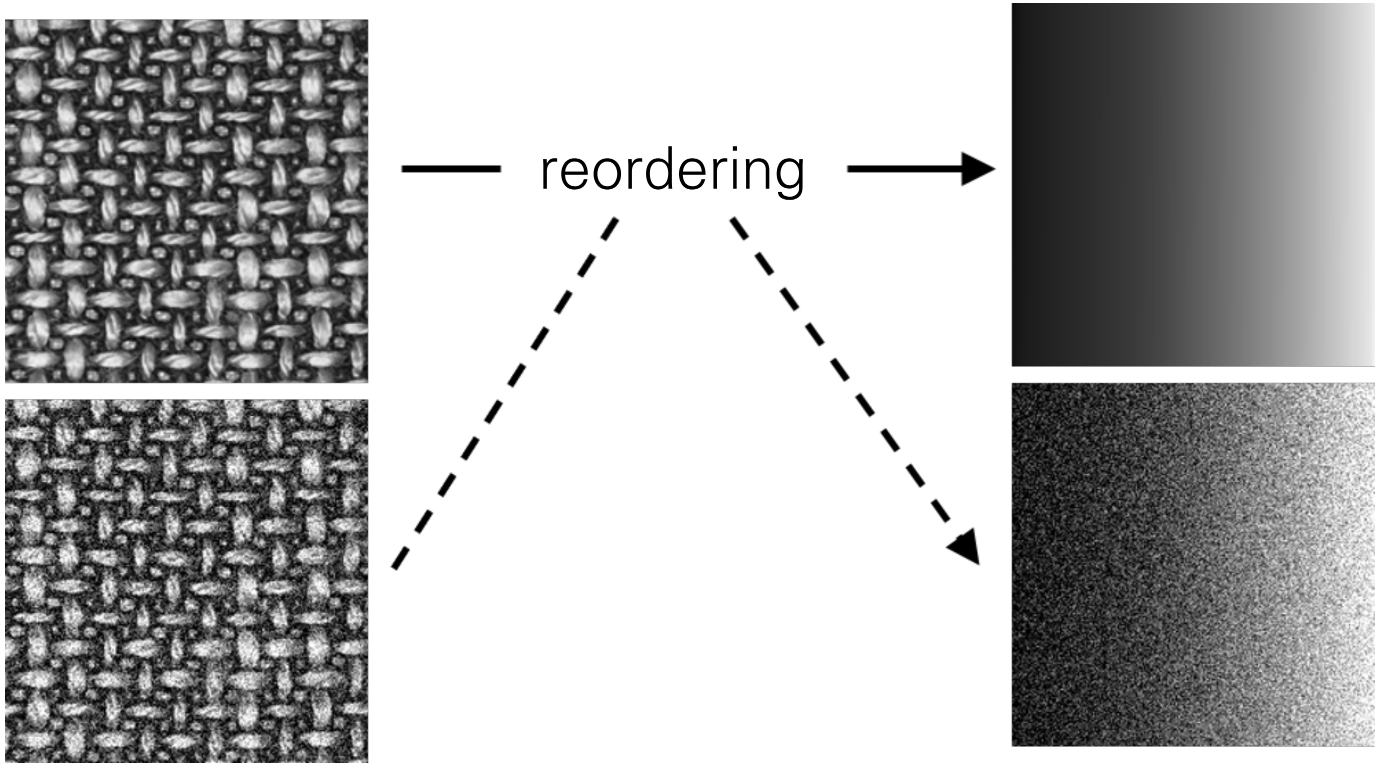
What is Wrong with MSE?

Image Credit: Berardino

What is Wrong with MSE?

$$\mathrm{MSE}(x, y) = \boxed{\frac{1}{N}\sum_{i=1}^{N}}(x_i - y_i)^2$$

Don’t care about pixel ordering



MSE= 1600, SSIM=0.637

MSE= 1600, SSIM=0.042

Image credit: Wang

What is Wrong with MSE?

$$\mathrm{MSE}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(\boxed{x_i - y_i})^2$$

Care about pixel difference, not the underlying signals

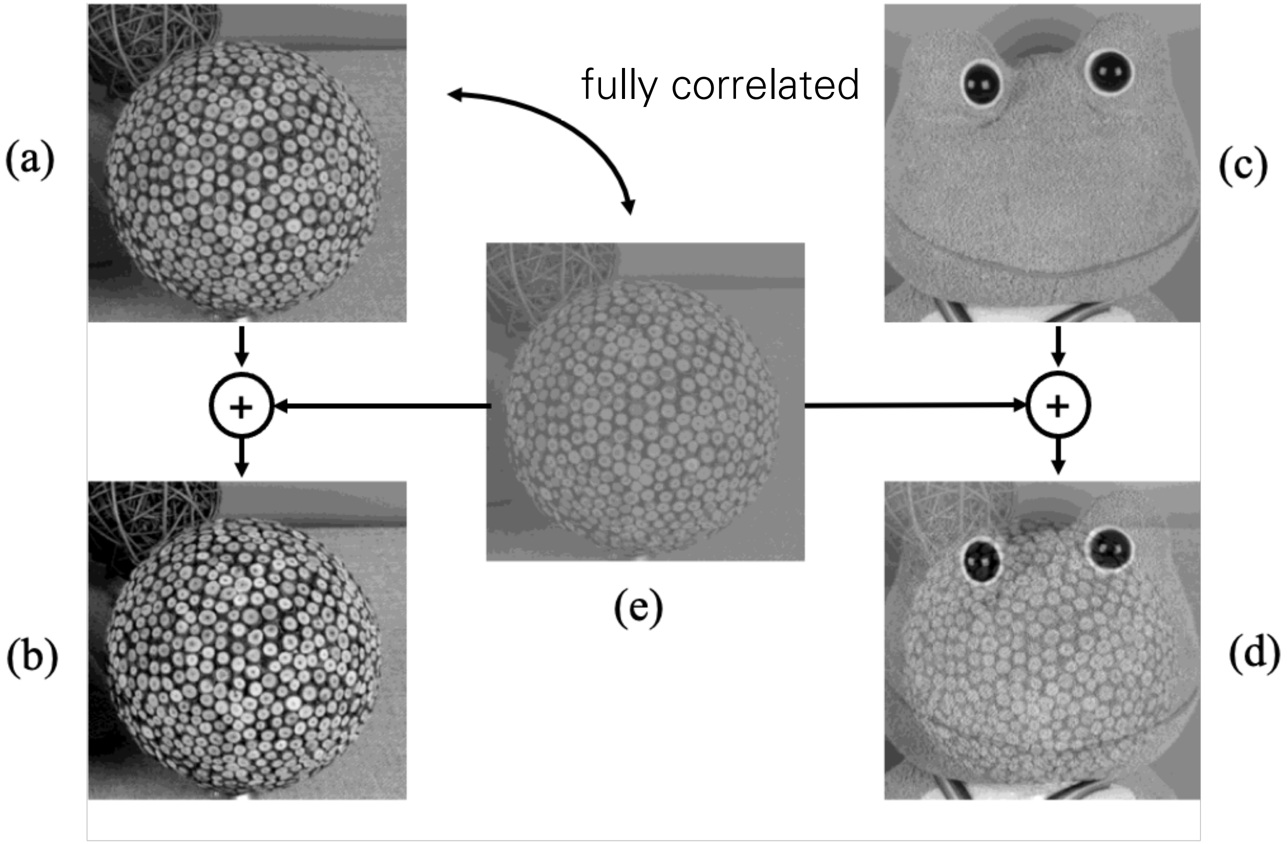
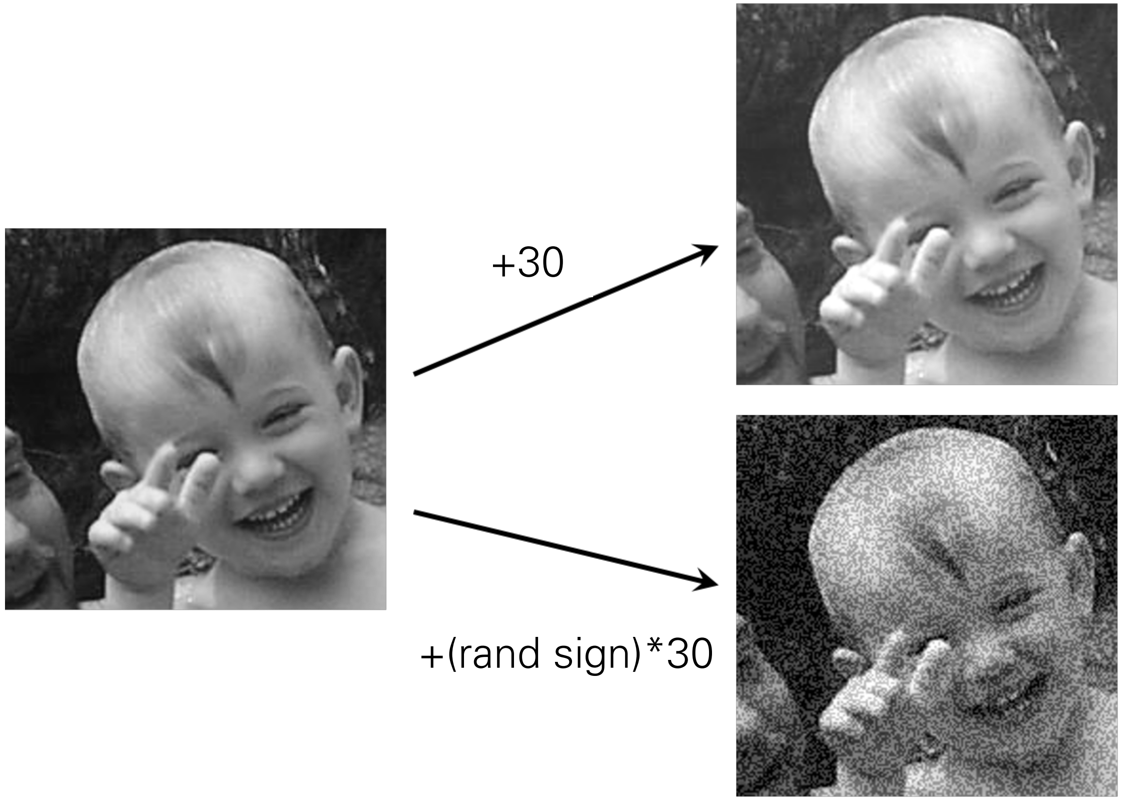


Image credit: Wang

What is Wrong with MSE?

$$\mathrm{MSE}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^{\boxed{2}}$$

Don’t care about the sign of pixel difference



MSE= 900,

SSIM=0.933

MSE= 900,

SSIM=0.247

Image credit: Wang

What is Wrong with MSE?

- MSE (or the more general Minkowski metric) implicitly assumes that errors are statistically independent
  - True, if spatial dependencies are eliminated prior to computation
  - No easy task as natural images are highly structured (i.e., spatially correlated)

- Possible solution?
  - Learn a "perceptual" transform f: $D(x, y) = \frac{1}{N} \sum_{i=1}^{N} (f(x)_i - f(y)_i)^2$

- Question: What are the desirable properties of f ?

Structural Similarity (SSIM)

- Assumption: The human visual system is highly adapted to extract structural information from the viewing field
- Methodology: A measure of structural information change provides a good approximation to perceived image distortion
- Questions:
    - How to define structural (and nonstructural) distortions?
    - How to separate structural and nonstructural distortions?

The SSIM Index
[Wang et al., 2004]

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Original image

Distorted image

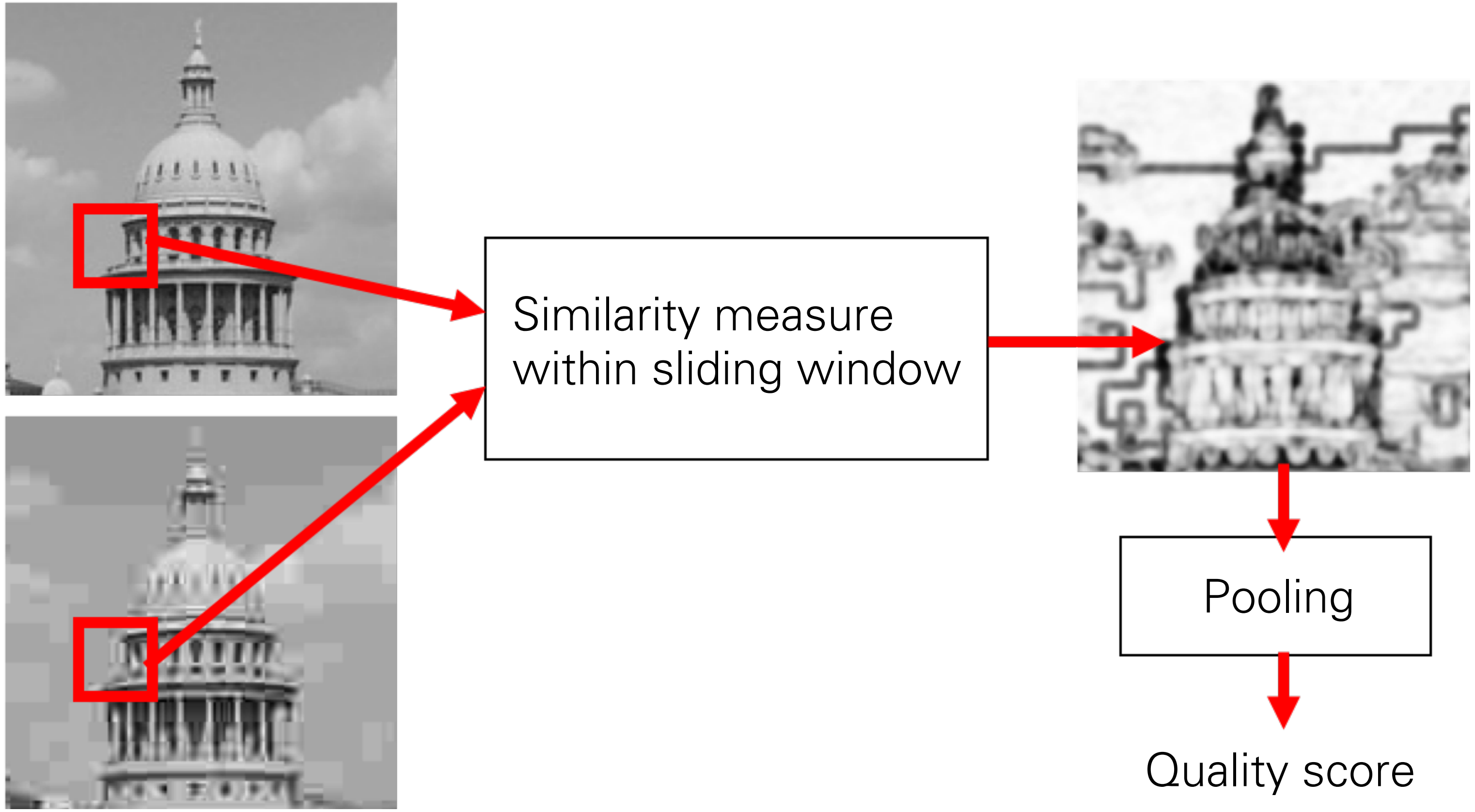


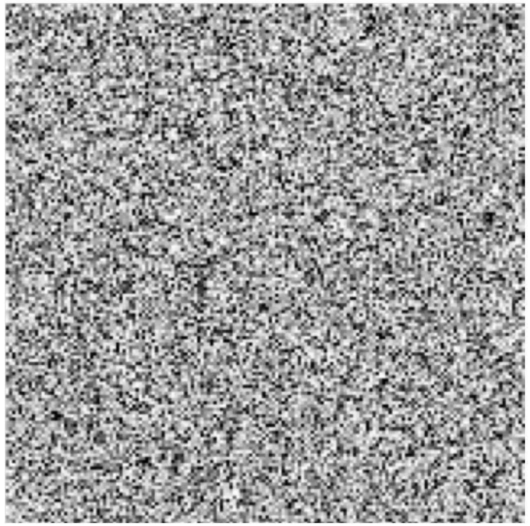
Image credit: Wang

Quality Map

Gaussian noise corrupted image

Original image

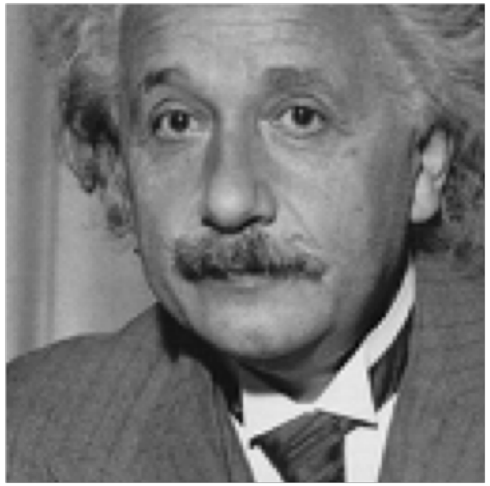
SSIM map



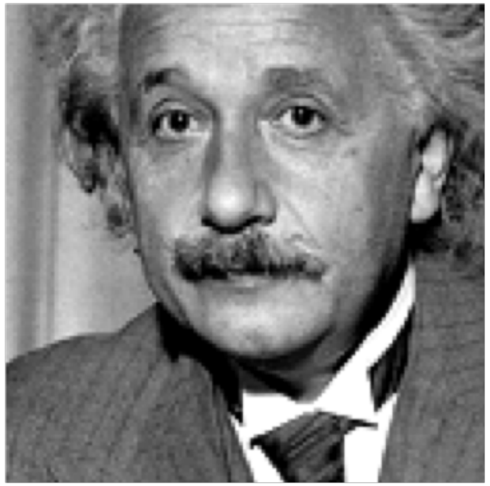
Absolute error map

Image credit: Wang

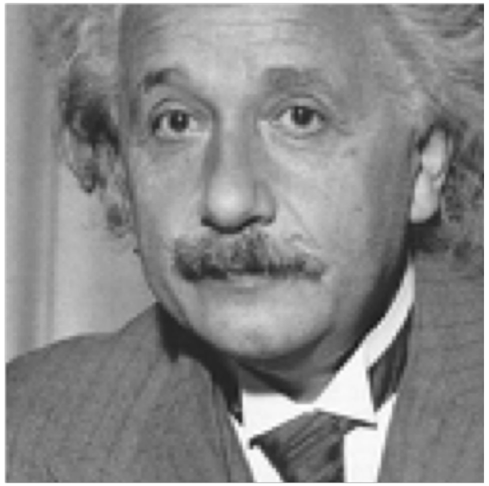
SSIM vs MSE



MSE=0, SSIM=1

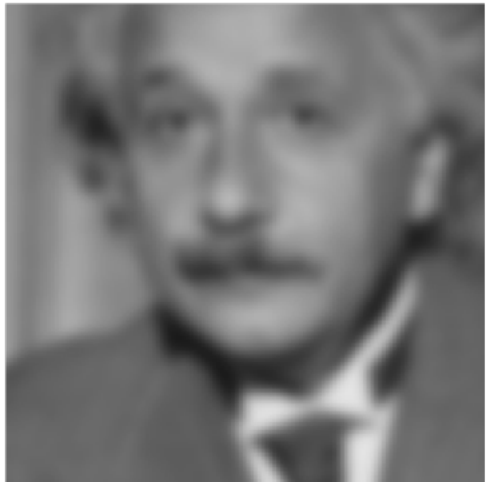


MSE=309, SSIM=0.93

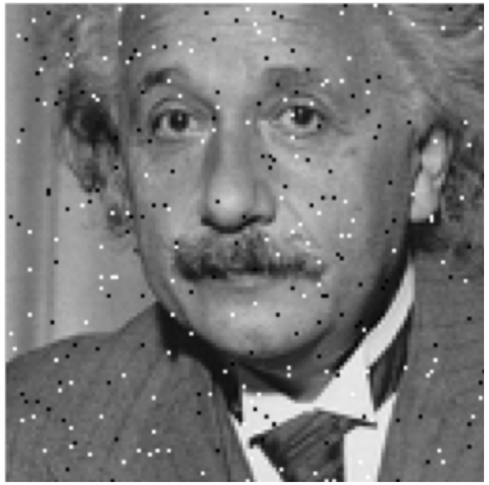


MSE=309, SSIM=0.99

MSE=309, SSIM=0.58



MSE=308, SSIM=0.64



MSE=309, SSIM=0.73

Image credit: Wang

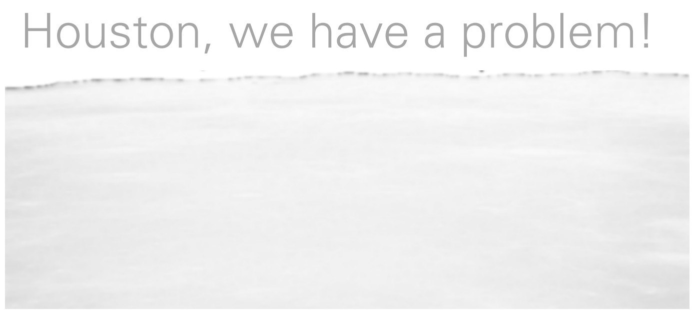
What is Wrong with SSIM?

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + \boxed{C_1})(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + \boxed{C_1})(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Normalization is sensitive to low intensities

Original image

Distorted image



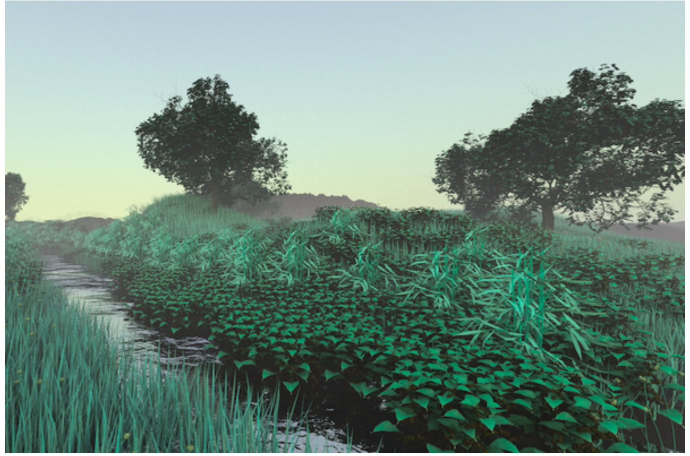
SSIM map

Image credit: Nilsson and Akenine-Möller

What is Wrong with SSIM?

$$\mathrm{SSIM}(\boxed{\mathrm{c2g}}(x), \boxed{\mathrm{c2g}}(y)) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$ Don't consider chrominance

Original image



Distorted image

SSIM map

Image credit: Nilsson and Akenine-Möller

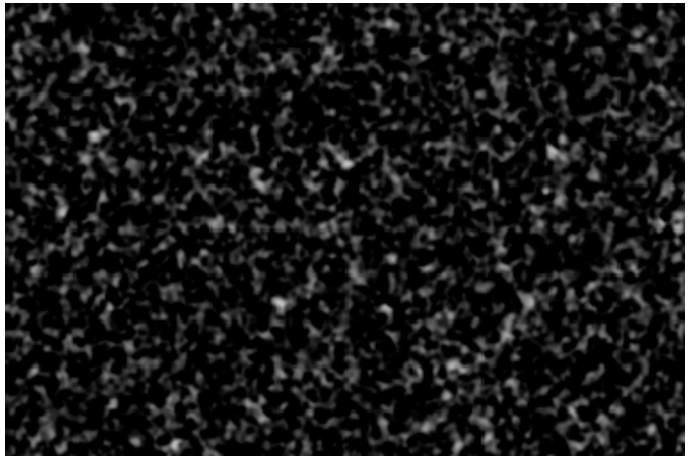
What is Wrong with SSIM?

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Rely on point-by-point comparison

Original image

Distorted image



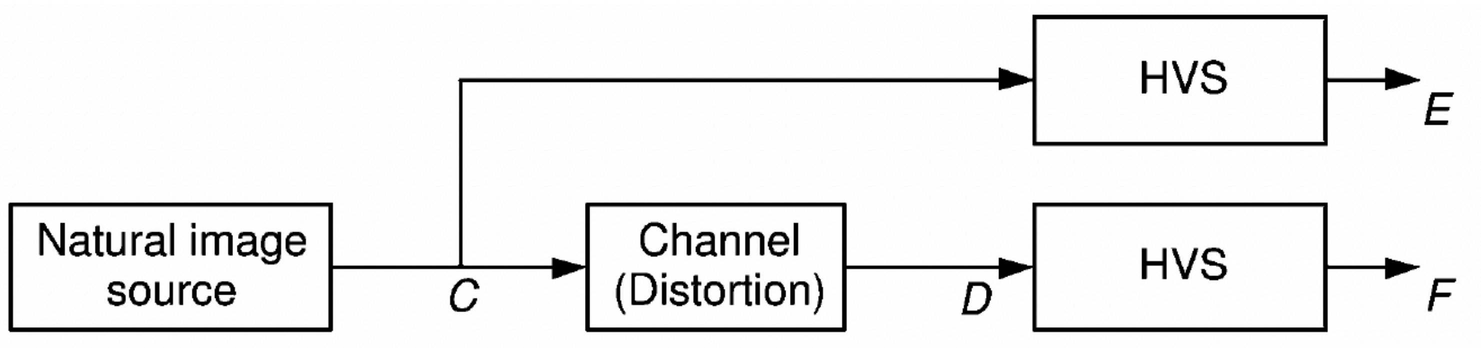
SSIM map

More Generally

- Not accurate enough
  - MS-SSIM, IW-SSIM, VIF, MAD, FSIM, VSI, NLPD, LPIPS, DISTS, ...
- Not computationally efficient enough
  - PAMSE, GMSD, ...
- Not misalignment-aware
  - Adaptive linear system, CW-SSIM, GTI-IQA
- Not color-aware
  - Adaptive linear system, FSIM_c, LPIPS, PieAPP, DISTS, ...
- Not texture-aware
  - STSIM, NPTSM, VGG Gram, LPIPS, DISTS, A-DISTS, ...

Visual Information Fidelity (VIF)
[Sheikh and Bovik, 2006]

- An information-theoretical approach
- Quantifies the amount of information preserved in the distorted image
- Works when the "distorted" image is visually superior to the reference



$$\mathrm{VIF} = \frac{MI(C;F)}{MI(C;E)}$$

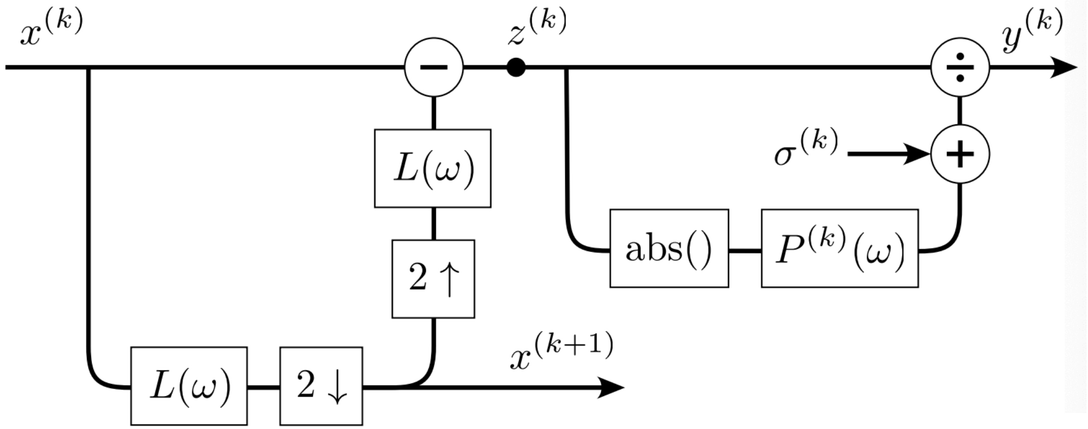
Image credit: Sheikh and Bovik

Most Apparent Distortion (MAD)
[Larson and Chandler, 2010]

- A multi-strategy approach
- A detection based strategy for near-threshold distortions
  - Look past the image and look for the distortions
- An appearance based strategy for clearly visible distortions
  - Look past the distortions and look for the image content

Normalized Laplacian Pyramid Distance (NLPD)
[Laparra et al., 2016]

- An error visibility method that models the early visual system
- Local luminance subtraction and local gain control
- The SOTA method for high-dynamic-range image tone mapping



$$\mathrm{NLPD}(x, \tilde{x}) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\sqrt{N^{(k)}}} \|y^{(k)} - \tilde{y}^{(k)}\|_2$$

Image credit: Laparra

Learned Perceptual Image Patch Similarity (LPIPS)
[Zhang et al., 2018]

- Investigate a wide range of network architectures and vision tasks
- Demonstrate the effectiveness of deep features in designing IQA models

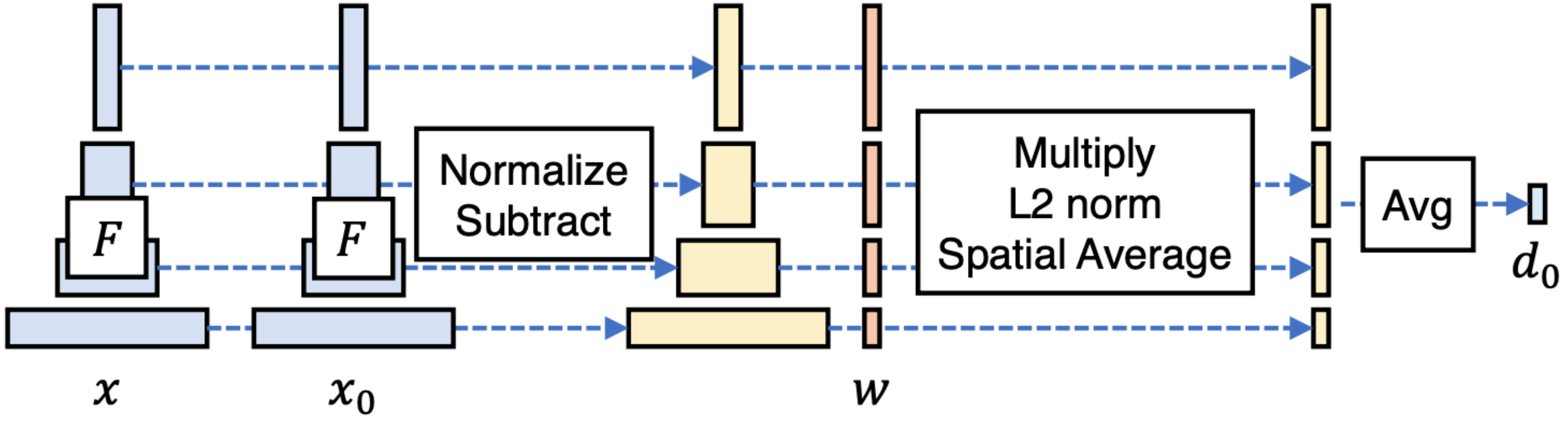
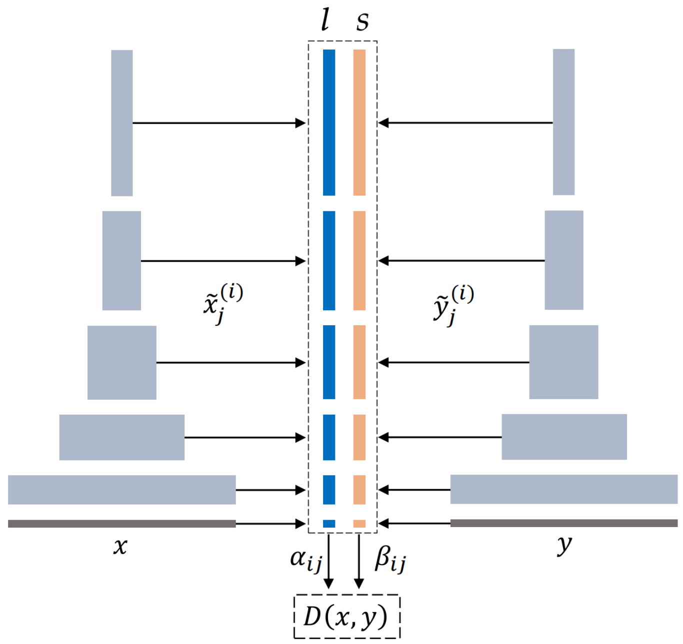


Image credit: Zhang

Deep Image Structure and Texture Similarity (DISTS) [Ding et al., 2020]

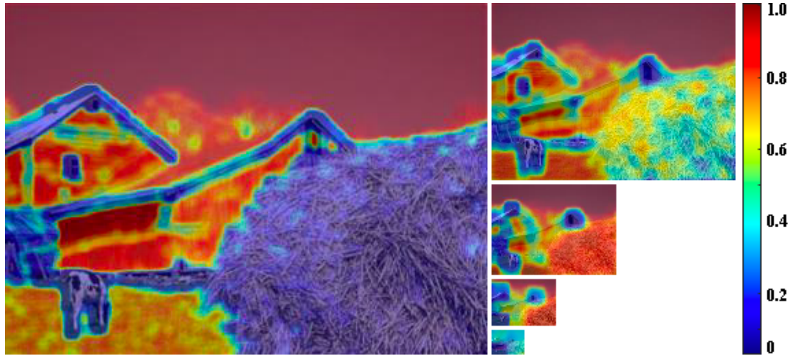
- Based on an injective mapping function built from a variant of VGG
- SSIM-like global structure and texture similarity measurements
- Robust to texture resampling and mild geometric transformations



$$\mathrm{DISTS}(x, y) = 1 - \sum_{i=0}^{m} \sum_{j=1}^{n_i} \left(\alpha_{ij} l(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) + \beta_{ij} s(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) \right)$$

Locally Adaptive DISTS
[Ding et al., 2021]

- Rely on the dispersion index to localize texture regions at different scales



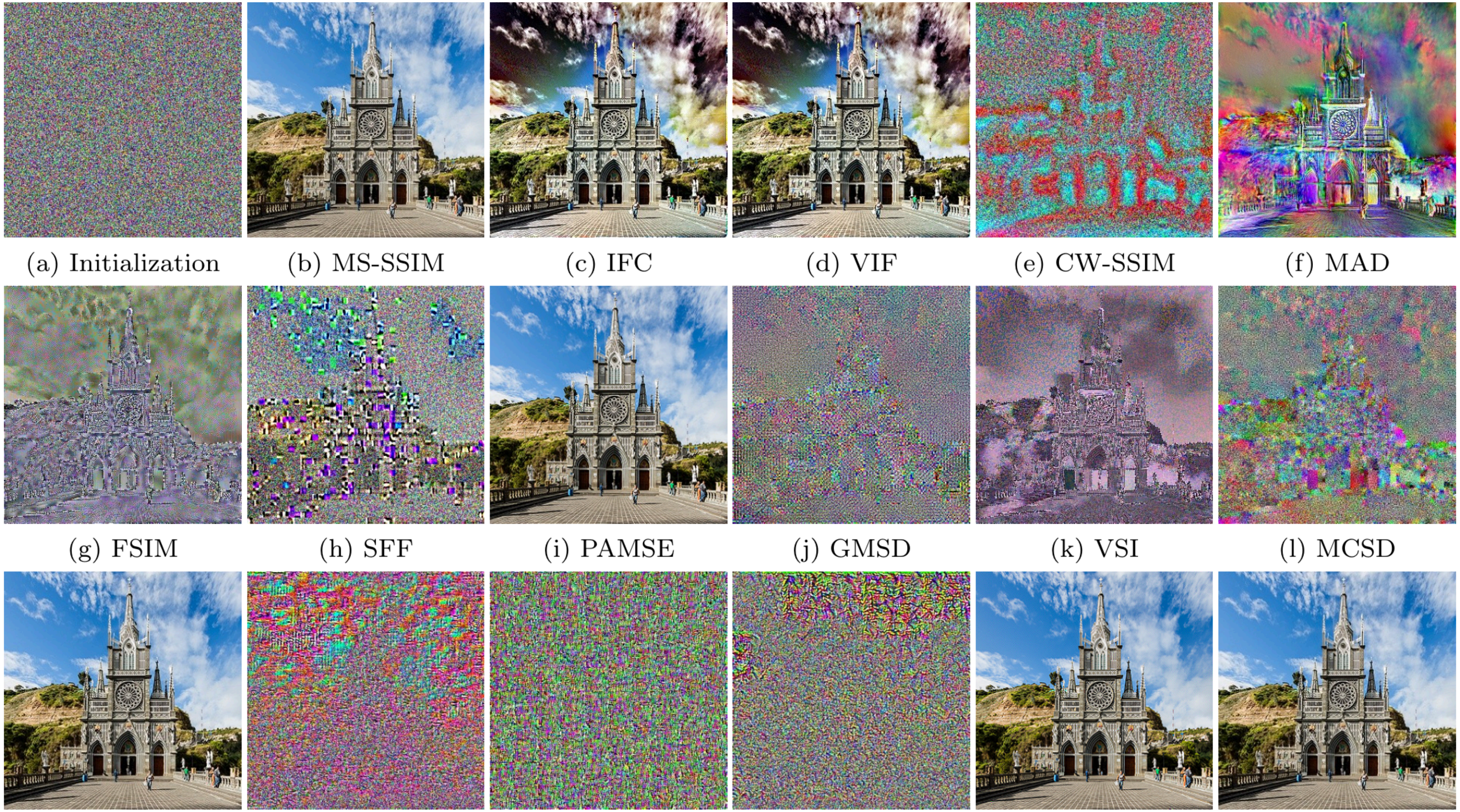
$$\text{A-DISTS}(X, Y) = 1 - \frac{1}{N}\sum_{i=0}^{M}\sum_{j=1}^{N_i} S\left(\tilde{X}_j^{(i)}, \tilde{Y}_j^{(i)}\right)$$

$$S(\tilde{X}_j^{(i)}, \tilde{Y}_j^{(i)}) = \frac{1}{K_i}\sum_{k=1}^{K_i}\left(\tilde{p}_k^{(i)} l\left(\tilde{x}_{j,k}^{(i)}, \tilde{y}_{j,k}^{(i)}\right) + \tilde{q}_k^{(i)} s\left(\tilde{x}_{j,k}^{(i)}, \tilde{y}_{j,k}^{(i)}\right)\right)$$

Full-Reference IQA: An Embarrassing Fact

Reference Image Recovery

$$y^{\star} = \arg\min_{y} D(x, y)$$



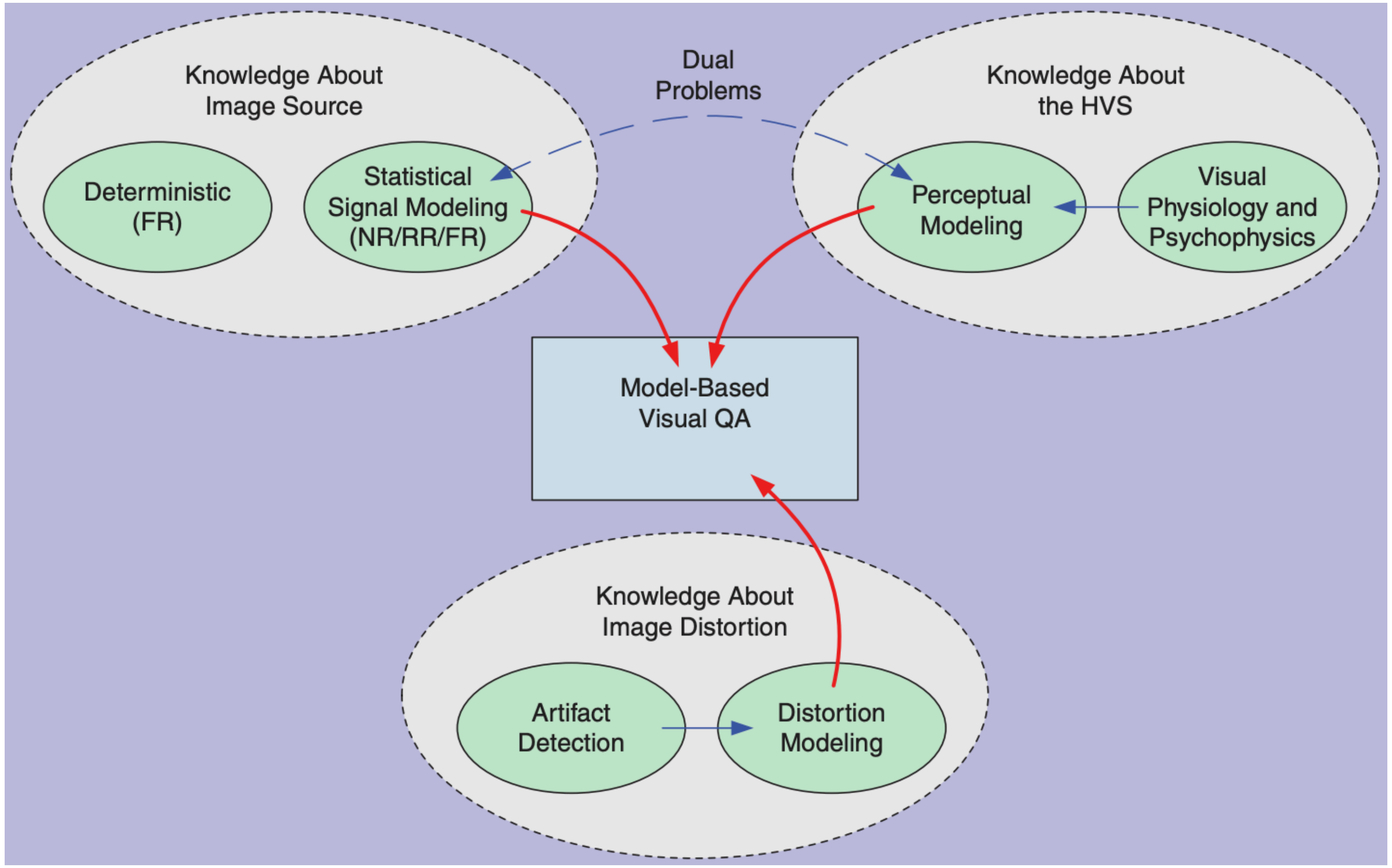
(a) Initialization (b) MS-SSIM (c) IFC (d) VIF (e) CW-SSIM (f) MAD

(g) FSIM (h) SFF (i) PAMSE (j) GMSD (k) VSI (l) MCSD

(m) NLPD (n) GTI-CNN (o) DeepIQA (p) PieAPP (q) LPIPS (r) DISTS

No-Reference IQA:
From Natural Scene Statistics to Learning based Approaches

Knowledge Map



Question: Do we really wish to leverage knowledge about image distortions?

Image credit: Wang

Natural Scene Statistics (NSS) based Approaches

- Assumption: Natural images exhibit strong statistical regularities, and reside in a tiny portion of the whole image space
- Methodology: A measure of violation from such statistical regularities provides an approximation to the unnaturalness (i.e., quality) of the image
  1. Handcraft statistical features from the image
  2. Summarize the extracted features using probability distributions (e.g. generalized Gaussian)
  3. Input the fitted parameters to a regression method (e.g, SVM) or compare the fitted distribution to a "reference" distribution

NSS based Approaches

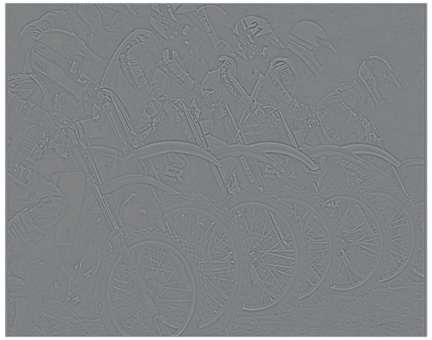
- Edge intensity/spread, sample entropy, BRISQUE, NIQE, IL-NIQE, ...
  - Spatial domain
- Frequency domain
  - DFT (blur kernel, phase congruency), DCT (BLIINDS-II), ...
- Wavelet domain
  - Local phase coherence, DIIVINE, LBIQ, ...

Natural Image Quality Evaluator (NIQE)
[Mittal et al., 2013]

- Without reliance on human ratings
- Without exposure to distorted images
- Widely used in real-world image processing


(a)


(b)


(c)


(d)


(e)

$$\mathrm{NIQE} = \sqrt{(\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2)}$$

Image credit: Mittal

(Deep) Learning based Approaches

- Methodology: Joint optimization of feature extraction and quality prediction
- Challenge: the large number of parameters to be optimized and the small number of human ratings as supervisory signals

(Deep) Learning based Approaches

- Attempt 1: Fine-tune models from other vision tasks (e.g., object recognition)
  - [Bianco, 2018], DB-CNN, UNIQUE, HyperIQA, MetaIQA, ...
- Limitation:
  - Lose the opportunity to search for the optimal and (possibly simpler) network architecture

(Deep) Learning based Approaches

- Attempt 2: Train no-reference models using image patches
  - CORNIA, [Kang et al., 2014], HOSA, DeepIQA, ...
- Limitation:
  - Local quality generally depends on global context
  - How to obtain a single global score for an image

(Deep) Learning based Approaches

- Attempt 3: Quality-aware pretraining followed by fine-tuning
  - Leverage distortion information
  - MEON, RankIQA, DB-CNN, ...
- Leverage full-reference models
  - dipIQ, [Kim et al., 2018], [Ma et al., 2019]
- Limitation: Difficult to extend to authentic image distortions

Evaluation of IQA Models

Standard Approach

Main Steps

1. Select a set of images from the image domain of interest
2. Collect the MOS for each image via psychophysical experiments (i.e., subjective user studies)
3. Compare the goodness of fit among the competing IQA models (i.e., sort by average performance)
   - Spearman rank correlation coefficient
     - prediction monotonicity
   - Pearson linear correlation coefficient
     - prediction linearity
   - Mean squared error
     - prediction accuracy

$$\mathrm{SRCC} = 1 - \frac{6\sum_i d_i^2}{M(M^2-1)}$$

$$\mathrm{PLCC}(x, y) = \frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_i (x_i - \mu_x)^2}\sqrt{\sum_i (y_i - \mu_y)^2}}$$

$$\mathrm{MSE}(x, y) = \frac{1}{M}\sum_i (x_i - y_i)^2$$

Caveats

- Sampling bias due to the extremely sparse distribution of the selected samples in the image space
  - i.e., the curse of dimensionality
- Algorithmic bias due to potentially overfitting the selected samples
  - The dataset creation precedes the algorithm development
- Subjective bias due to potentially cherry-picking test results

The Perception-Distortion Tradeoff

Perceptual Image Restoration

- The invention of Generative Adversarial Networks (GANs) greatly improves the perceptual performance

Ground Truth

Less distortion
PSNR-oriented

Photo-realistic
GAN-based

Slide credit: Gu et al.

Gap Between IQA Metric and Human Judgment

- Increasing inconsistency between high numerical performances (PSNR, SSIM, PI, etc.) and perceptual performance.

Ground Truth
PSNR / SSIM

PSNR-oriented

GAN-based

Slide credit: Gu et al.

Gap Between IQA Metric and Human Judgment

- Before 2018, Evaluation Using PSNR/SSIM

Ground Truth
PSNR / SSIM

23.52 / 0.7056
Good in PSNR, SSIM

19.86 / 0.5530
Preferred by Human

Slide credit: Gu et al.

Gap Between IQA Metric and Human Judgment

- After 2018, Evaluation Using PI/NIQE

Ground Truth
PI / NIQE

3.80 / 6.47
Good in PI, NIQE

4.30 / 6.90
Preferred by Human

Slide credit: Gu et al.

PI and NIQE are suggested in Y.Blau, and T. Michaeli. The perception-distortion tradeoff. CVPR 2018

The Perception-Distortion Tradeoff

- How to evaluate image restoration methods?
- Distortion and perceptual quality are at odds with each other.
- The lower the distortion of an algorithm, the more its distribution must deviate from the statistics of natural scenes.

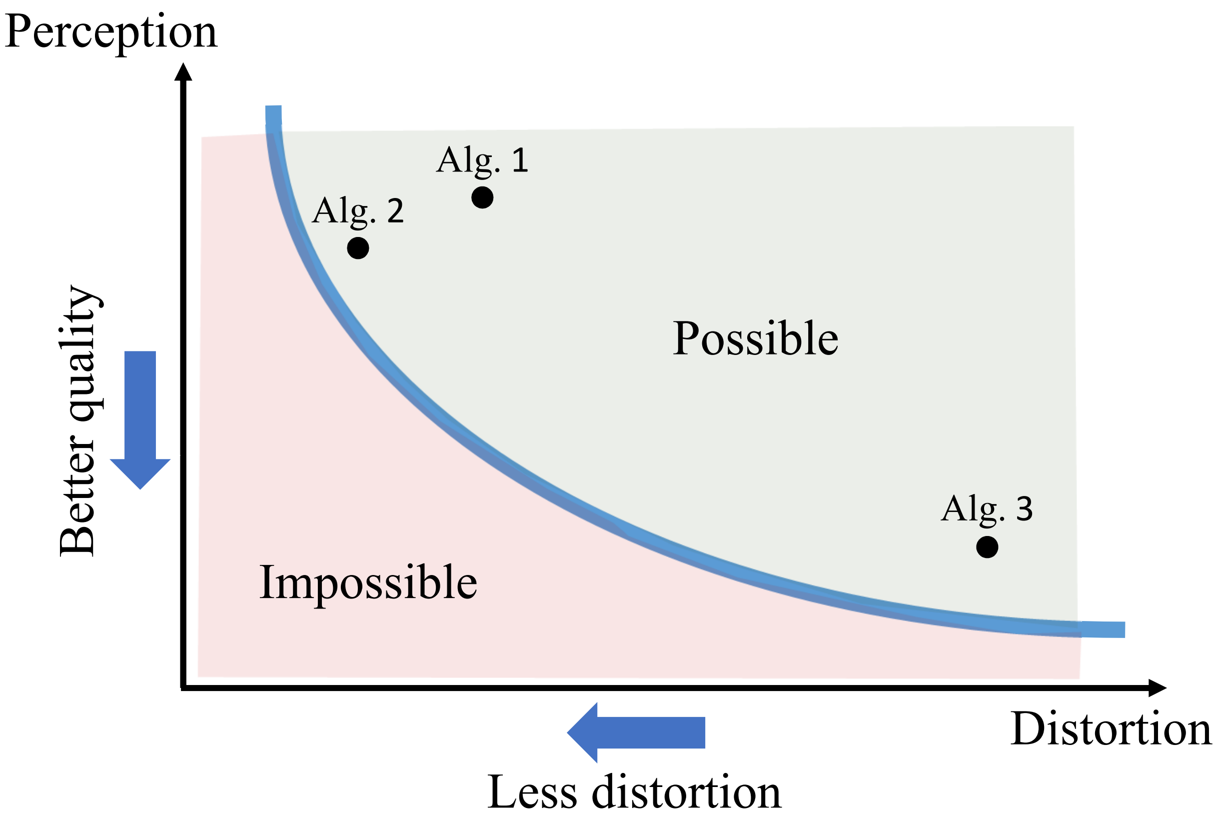


Image credit: Blau and Michaeli

What Makes a Great Picture?

Image Quality vs. Image Aesthetics

- Quality assessment deals with measuring low-level degradations such as noise, blur, compression artifacts, etc.
- Aesthetic prediction quantifies semantic level characteristics associated with emotions and beauty in images.

Photography 101: the **where** and **when**

- Composition
  - Framing
  - Rule of Thirds
  - Leading Lines
  - Textures and Patterns
  - Simplicity

- Lighting
  - Light Direction
  - Color coordination / balance
  - Sunny vs. cloudy
  - “Golden Hour”
  - B&W to focus attention
  - (sur) realism

Framing

"Photography is all about framing. We see a subject -- and we put a frame around it. Essentially, that is photography when all is said and done."

-- from photo.blorge.com

Frame serves several purposes:

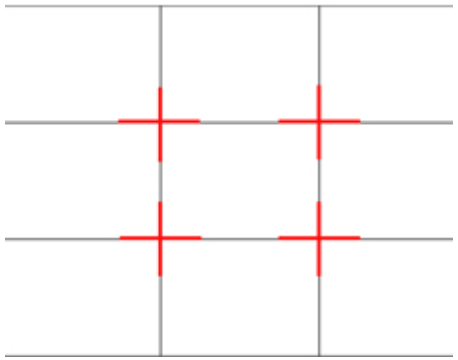
- 1. It gives the image depth
- 2. If used correctly, framing can draw the eye of the viewer of an interest to a particular part of the scene.
- 3. Framing can bring a sense of organization or containment to an image.
- 4. Framing can add context to a shot.

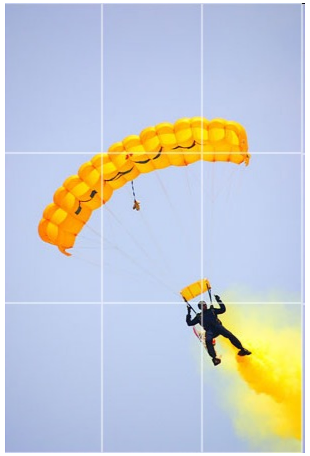
http://digital-photography-school.com/blog/frame-your-images/

Examples of nice framing

http://flickr.com/photos/paulosacramento/226545698/
http://flickr.com/photos/chrisbeach/13868545/
http://flickr.com/photos/74531485@N00/929270814/
http://flickr.com/photos/freakdog/223117229/
http://flickr.com/photos/cdm/253805482/

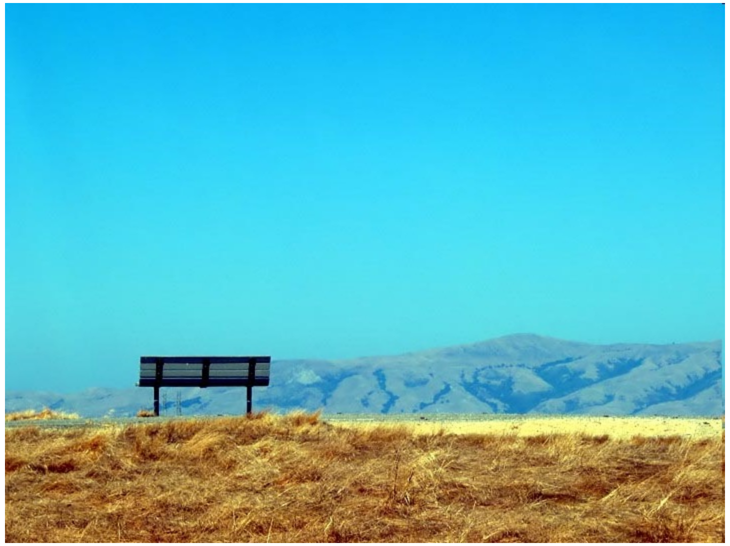
Rules of Thirds





http://www.photo96.com/blog/?p=371

Other examples



Don’t center, especially for motion

Slide credit: Fredo Durand

Don't center, especially for motion

Slide credit: Fredo Durand

... or do center

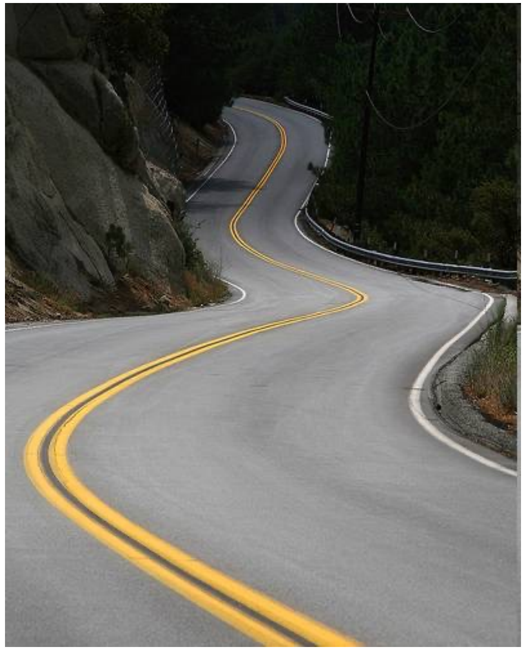
Slide credit: Fredo Durand

Leading Lines

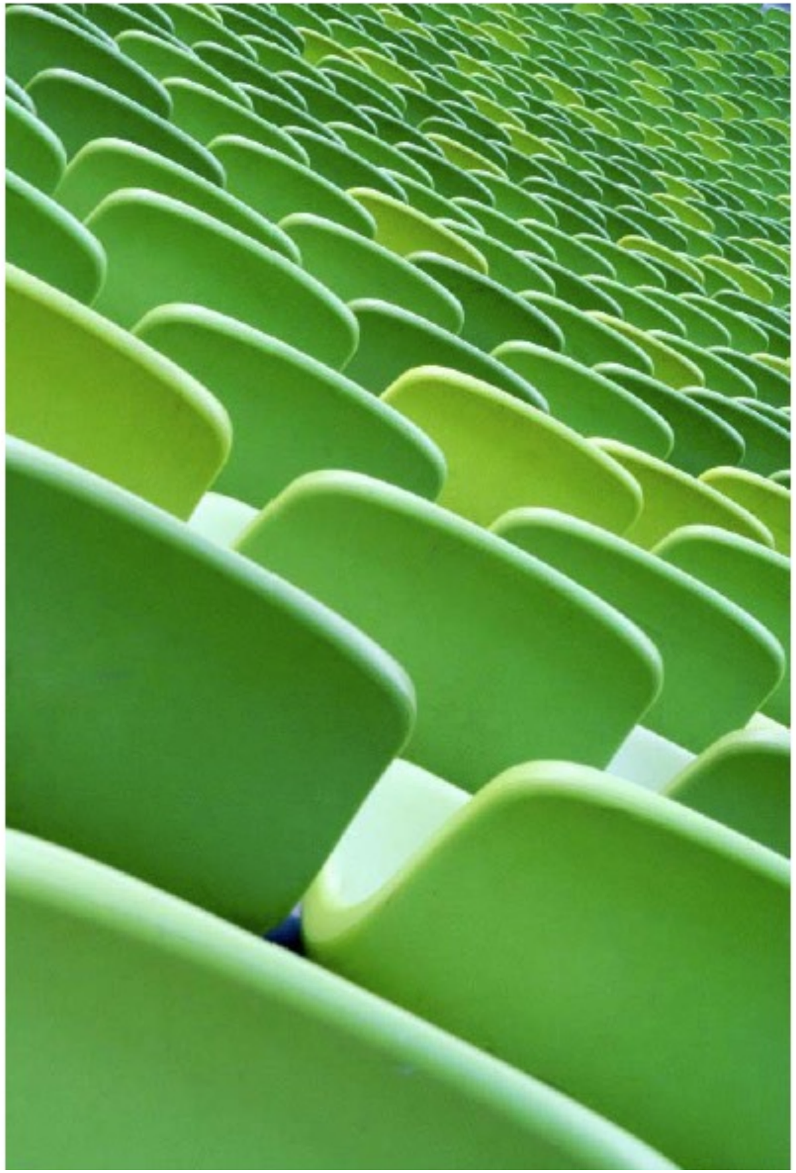
Leading Lines

Slide credit: Fredo Durand

More examples



Textures and Patterns

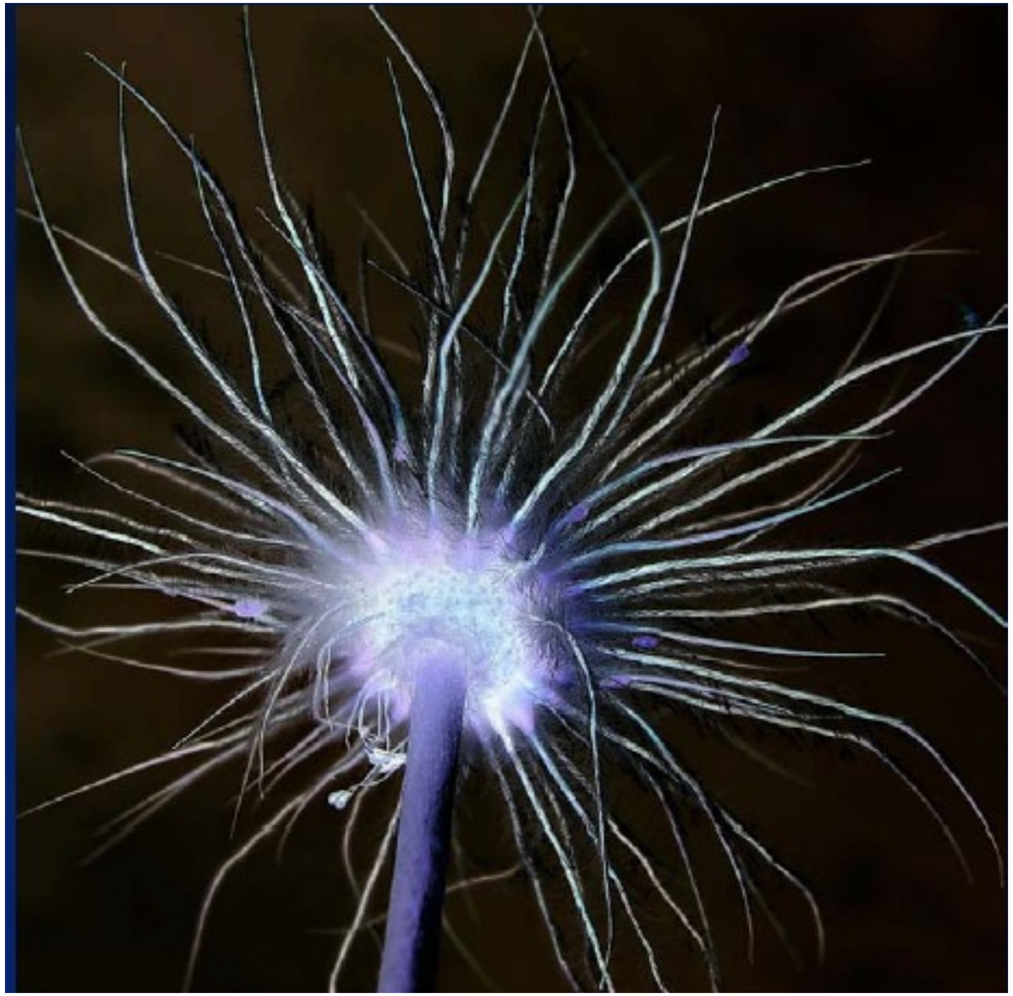


Simplicity

"Look Into" by Josh Brown @ Flickr

Prof - Obvious what one should be looking at, i.e. easy to separate subject from the background. Snap – unstructured, busy, filled with clutter.

Simplicity



"alien flower" by Josef F. Stuefer @ Flickr

Simplicity

“Waiting in line!” by Imapix @ Flickr

B&W for Simplicity

Photo by A. A. Efros

B&W for Simplicity

Photo by A. A. Efros

B&W for Simplicity

Photo by A. A. Efros

B&W for Simplicity

Photo by A. A. Efros

...but not always

Photo by A. A. Efros

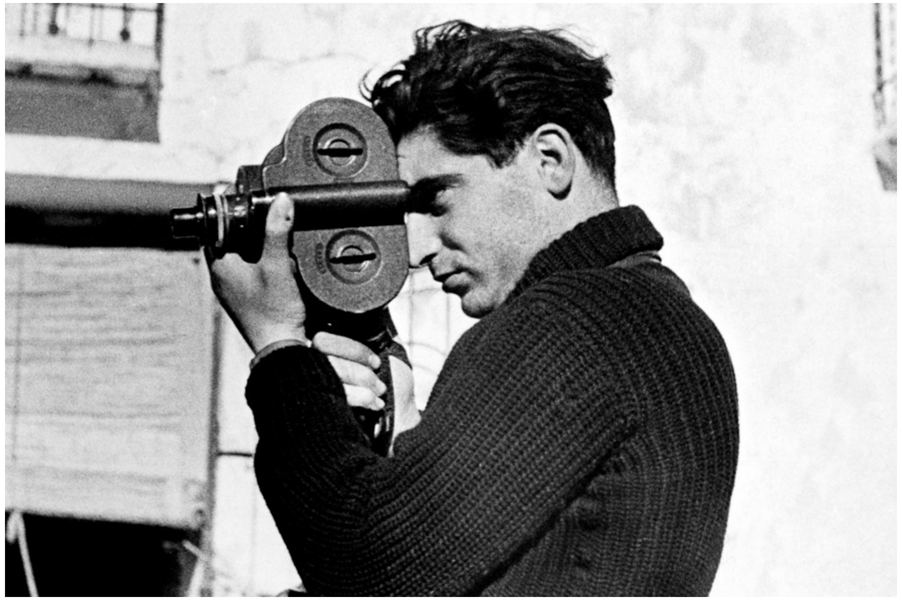
...but not always

Photo by A. A. Efros

Crop for Simplicity

Slide credit: Fredo Durand

Crop for Simplicity



If your pictures aren't good enough, you're not close enough"
— Robert Capa

Slide credit: Fredo Durand

Clean Backgrounds

Slide credit: Fredo Durand

Simplicity for Portraits

https://vimeo.com/29722267

And now, all together...

Photo by A. A. Efros

And now, all together...

Photo by A. A. Efros

Slide credit: Fredo Durand

Get low

Try to be at eye level

Bad

Better

Slide credit: Fredo Durand

Get low

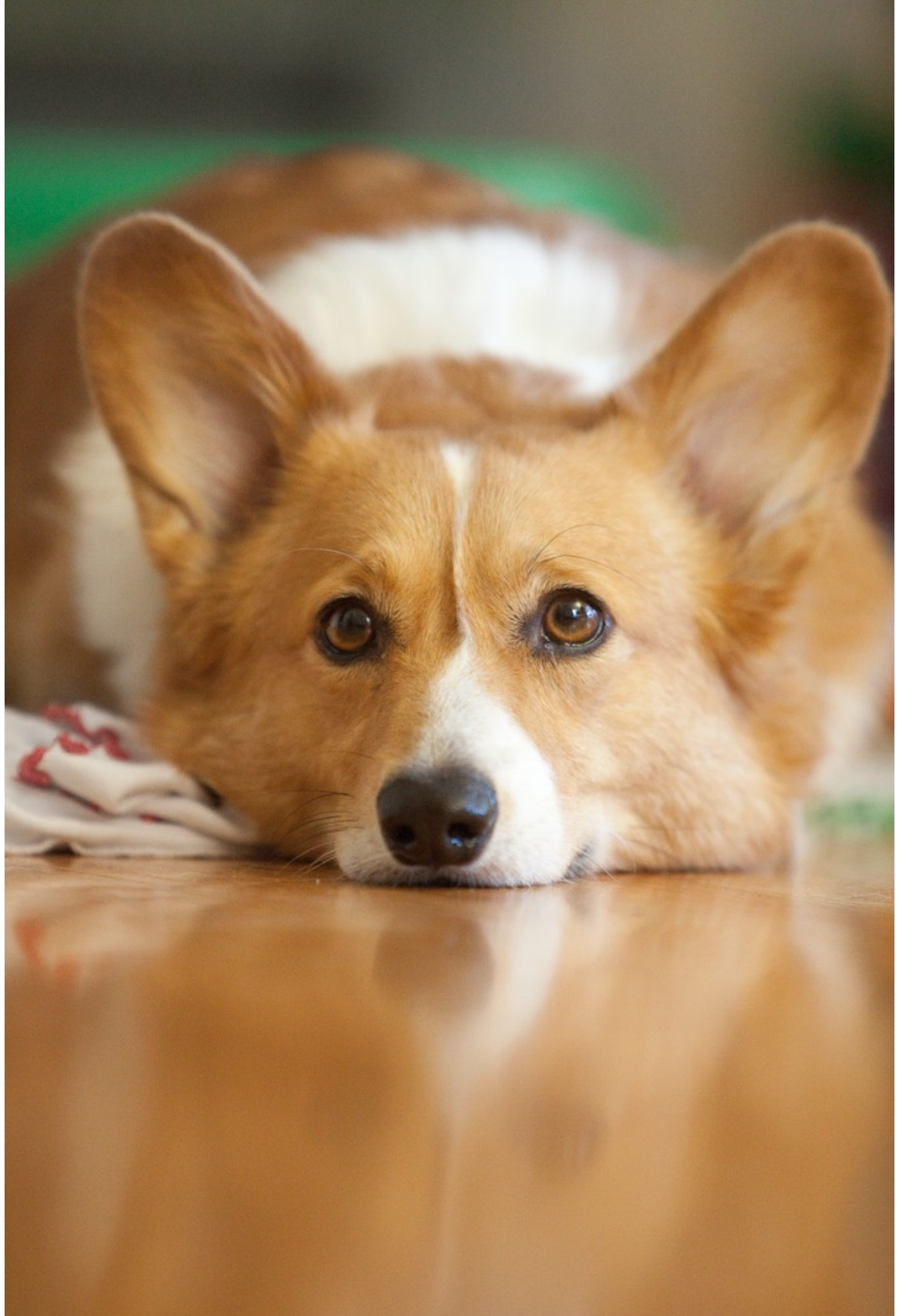
Slide credit: Fredo Durand

Bad angles

https://www.youtube.com/watch?v=8EmRZO9fwvk&feature=youtu.be

Slide credit: Fredo Durand

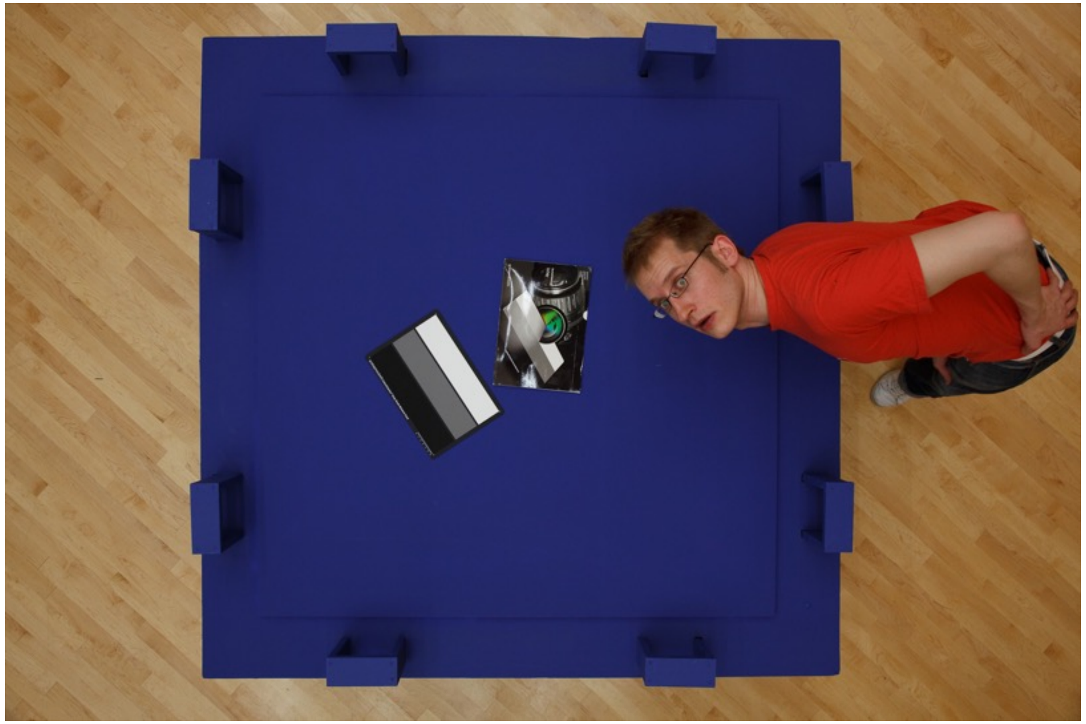
Eye level



Slide credit: Fredo Durand

Or really get high

As usual, follow a rule
or really break it.

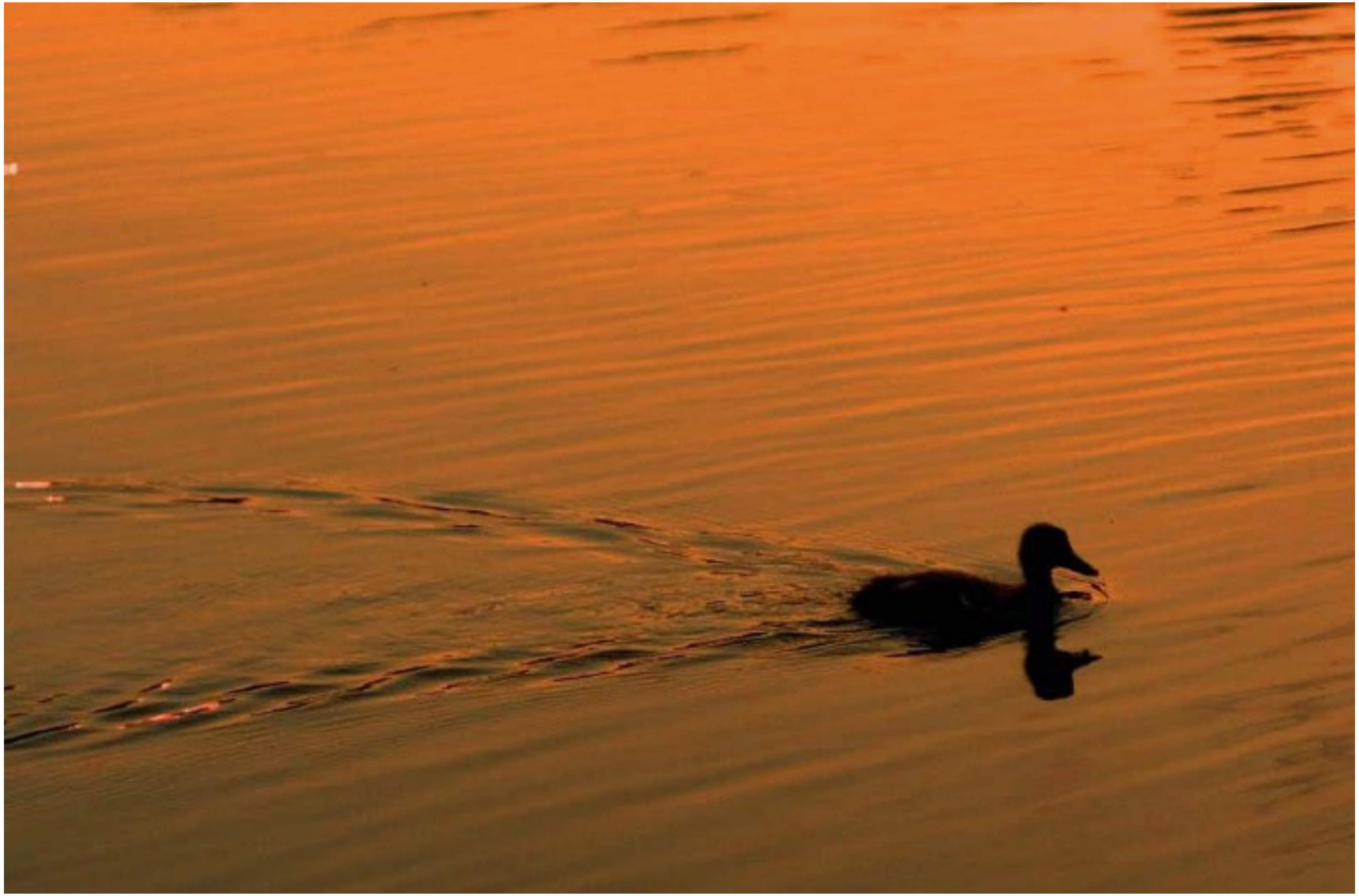


Slide credit: Fredo Durand

Front Lighting

Side Lighting

Back Lighting



Color Coordination

Complementary colors (of opposite hue on color wheel)

Go in the shade

Light is more diffuse

Bad

Better

Slide credit: Fredo Durand

Overcast days are the best

Just don’t put the sky in the frame

The weather conditions

The pictures

Other overcast-day pictures

Slide credit: Fredo Durand

Bottom line

Don't get married on a sunny day!

Slide credit: Fredo Durand

Cloudy day

Slide credit: Fredo Durand

Best time of day: sunset & sunrise

+/- 1 hour "Golden hours"

Night photography: always near sunset/sunrise

- because of nice diffuse light

Mid day:
often not great

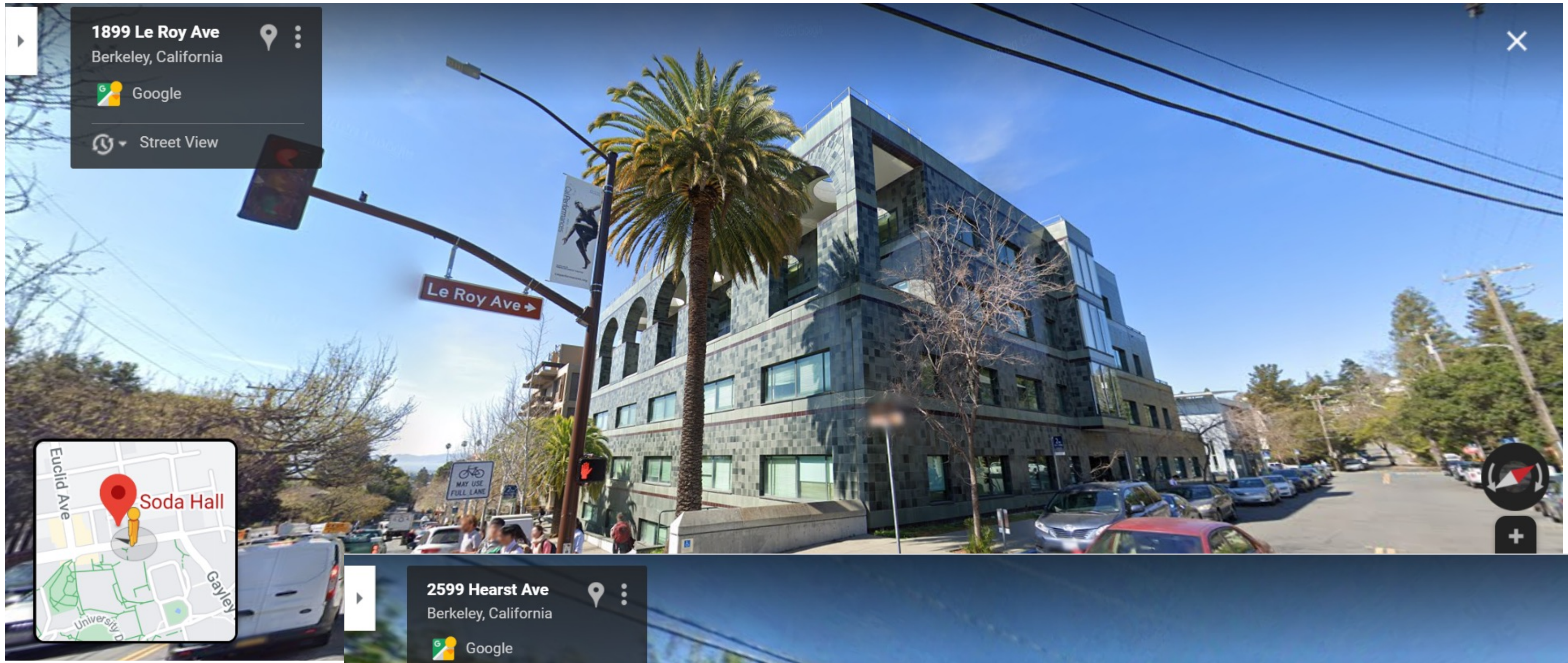
less than 1 hour
after sunrise/
before sunset

During sunset or
sunrise

After sunset

Slide credit: Fredo Durand

"Golden Hour"



1899 Le Roy Ave
Berkeley, California
Google
Street View
Le Roy Ave
Euclid Ave
Soda Hall
Gayley

2599 Hearst Ave
Berkeley, California
Google
Street View
Euclid Ave
Soda Hall
Gayley
Google
Image capture: Apr 2008
© 2020 Google
United States
Terms
Report a problem

less than 1 hour after sunrise

During sunset/sunrise

After sunset

Slide credit: Fredo Durand

After sunset: blue hour

Slide credit: Fredo Durand

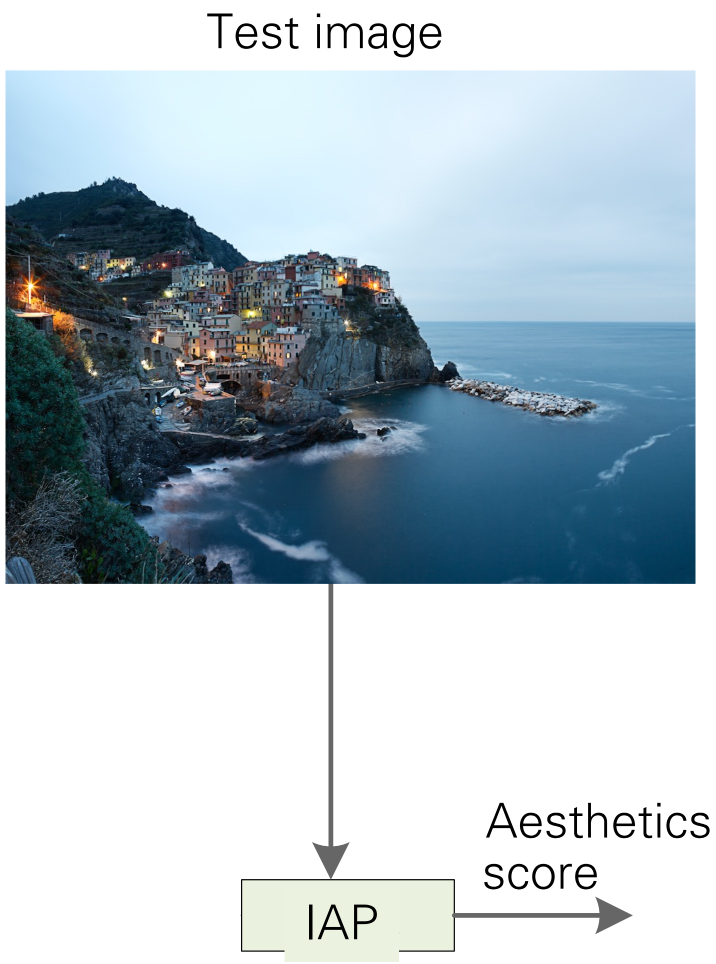
Blue Hour (Russian River)

Photo by A. A. Efros

Image Aesthetics Prediction

- Goal: Build computational models that accurately predict human perception of image aesthetics
- No-reference models in nature.

Image Aesthetics Prediction

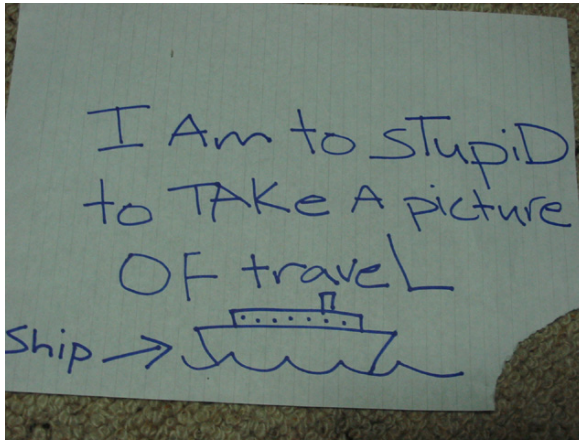


AVA Dataset

- Images: 255000 (rated based on aesthetic qualities by amateur photographer).
- Each photo is scored by an average of 200 people in response to photography contests.

(a) 6.36 (±1.04)

(b) 7.84 (±2.08)

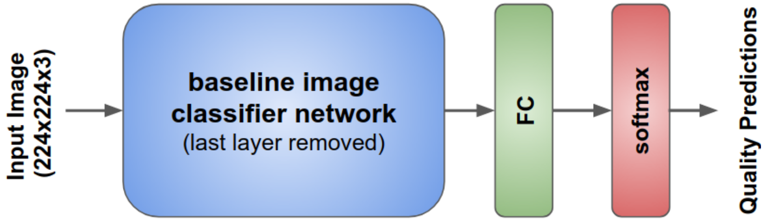


(c) 2.62 (±2.15)

(d) 3.12 (±1.28)

N. Murray, L. Marchesotti, and F. Perronnin, AVA: A large-scale database for aesthetic visual analysis, CVPR 2012

NIMA: Neural Image Assessment [Talebi and Milanfar, 2018]



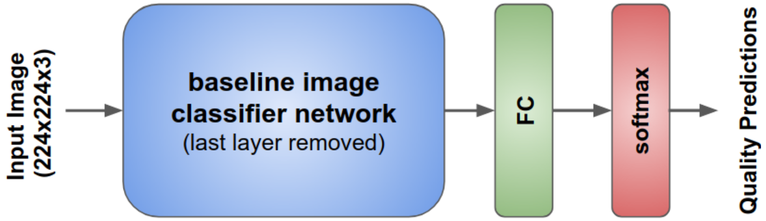
- Instead of predicting the mean opinion score, it predicts the distribution of human opinion scores using a CNN

(a) 6.38 (7.16) (b) 6.24 (6.79) (c) 6.22 (6.64) (d) 6.16 (6.93) (e) 5.92 (6.23)

(f) 5.71 (5.78) (g) 5.61 (5.54) (h) 5.28 (5.32) (i) 5.11 (5.23) (j) 5.03 (5.35)

(k) 4.90 (4.91) (l) 4.83 (4.89) (m) 4.77 (4.55) (n) 4.48 (3.95) (o) 3.55 (3.53)

NIMA: Neural Image Assessment [Talebi and Milanfar, 2018]



- It can be used for automatic parameter tuning to enhance the quality of the outputs

Next Lecture:
Advanced Topics