ACM multimedia

Chengdu, China OCT 20-24 2021

# Part II: Objective IQA Models

## From Full-Reference to No-Reference Approaches

**Kede Ma**

CityU Department of Computer Science

香港城市大學

City University of Hong Kong

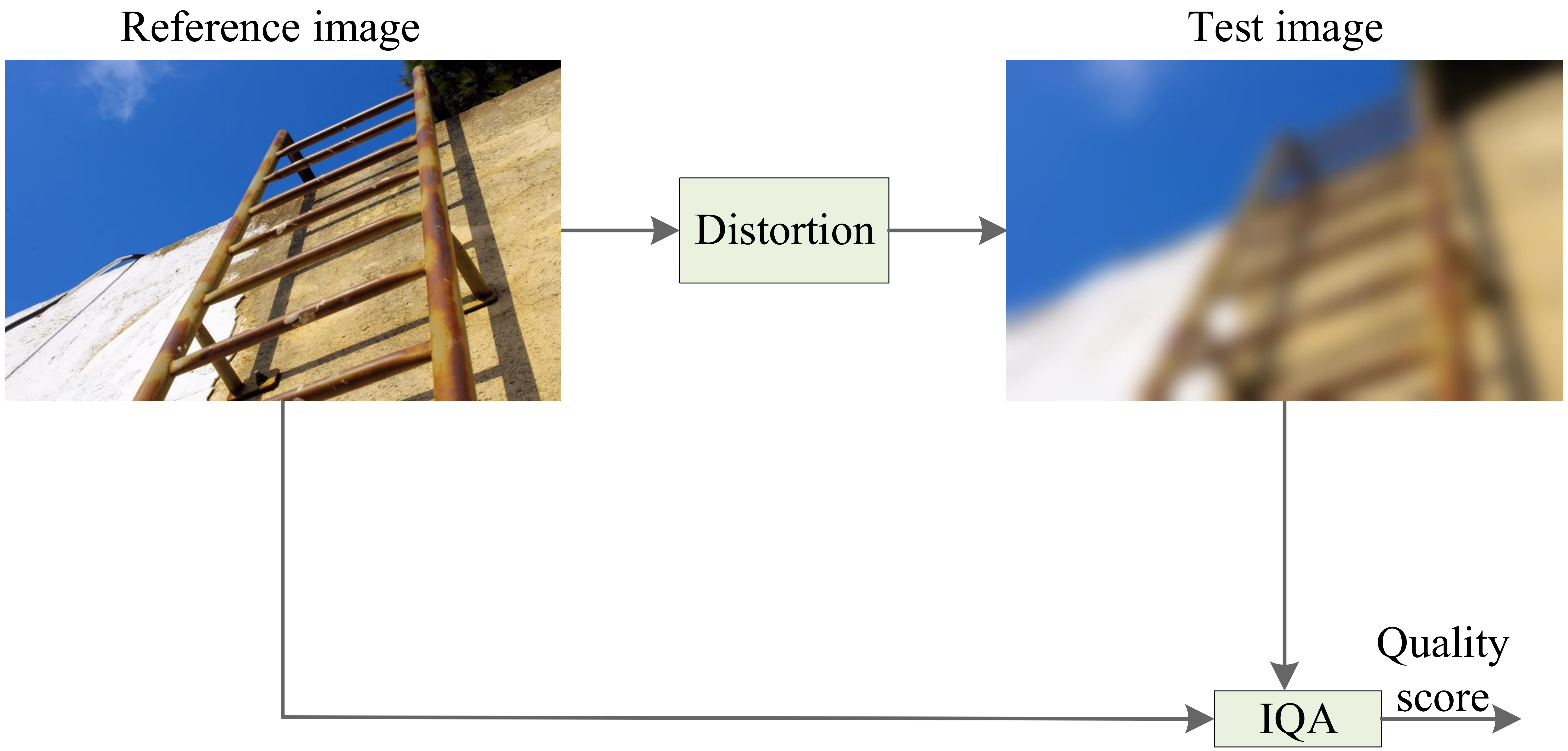
# Outline

- Full-Reference IQA Models
  - Mean Squared Error (MSE) and Structural Similarity (SSIM)
  - Other Representative Methods
- No-Reference IQA Models
  - Natural Scene Statistics (NSS)
  - (Deep) Learning based Approaches
- Discussion

# Goal of Objective IQA

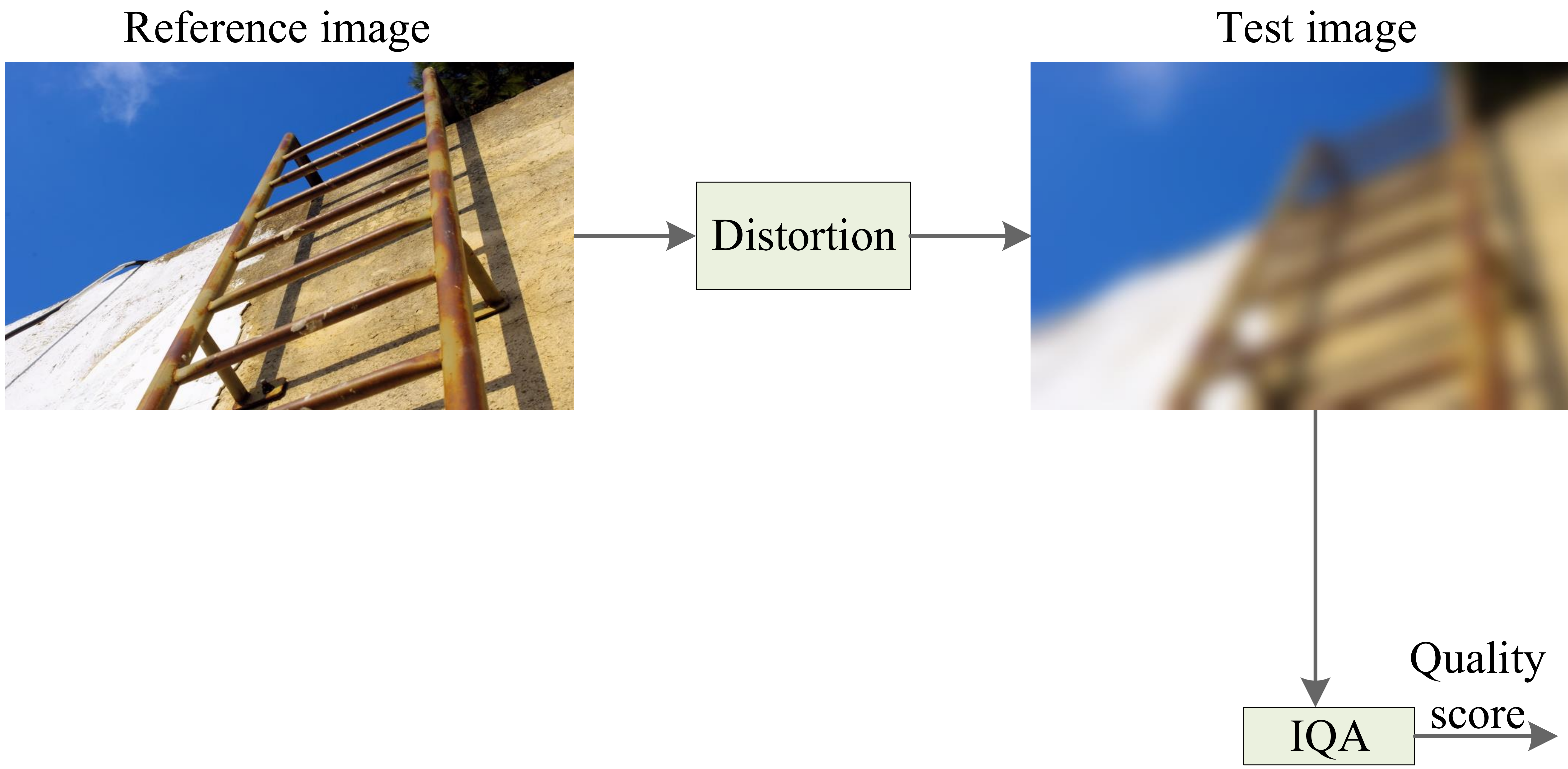
Build computational models that accurately predict human perception of image quality

# Full-Reference IQA



# No-Reference IQA

## Blind IQA (BIQA)



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

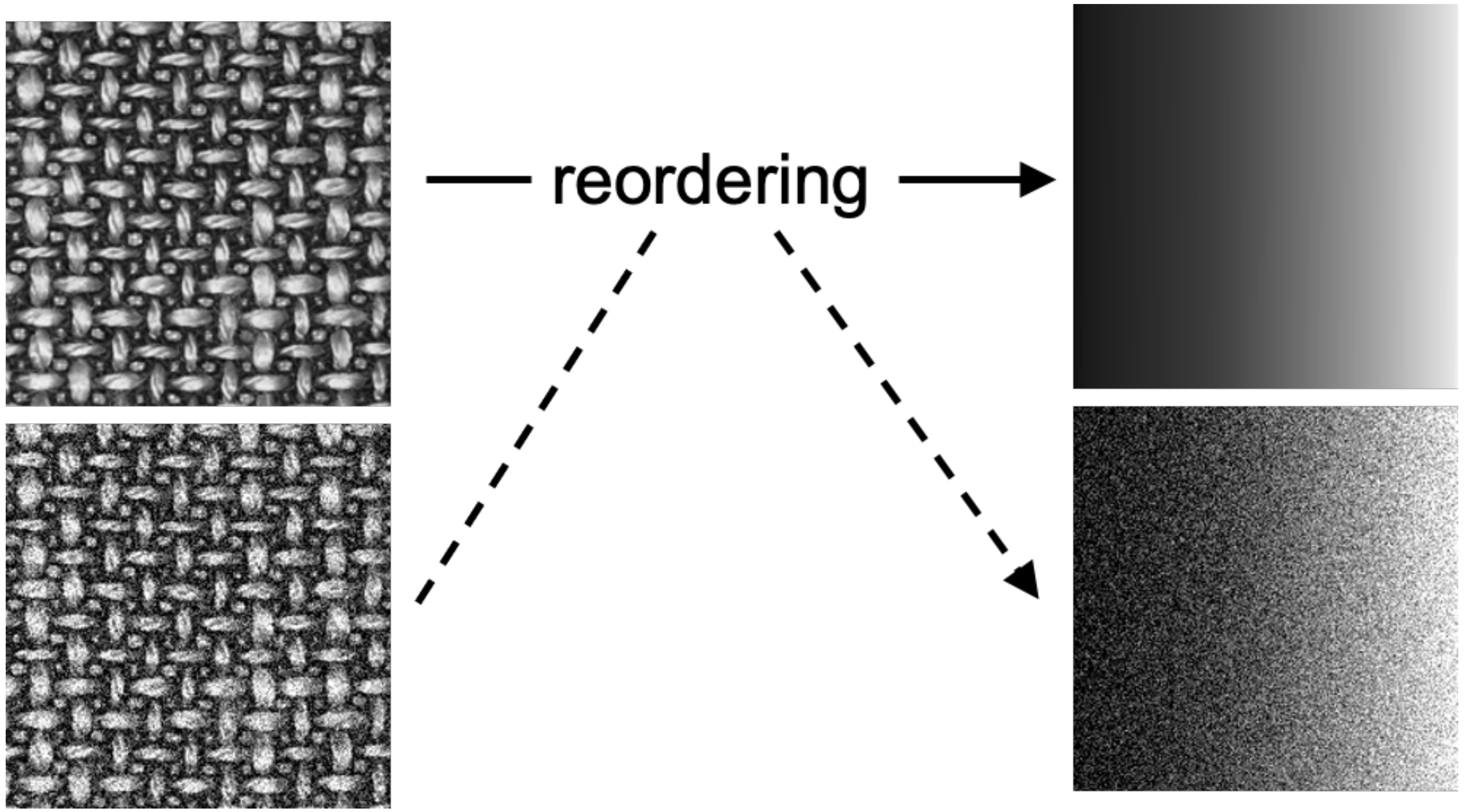
# What is Wrong with MSE?

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

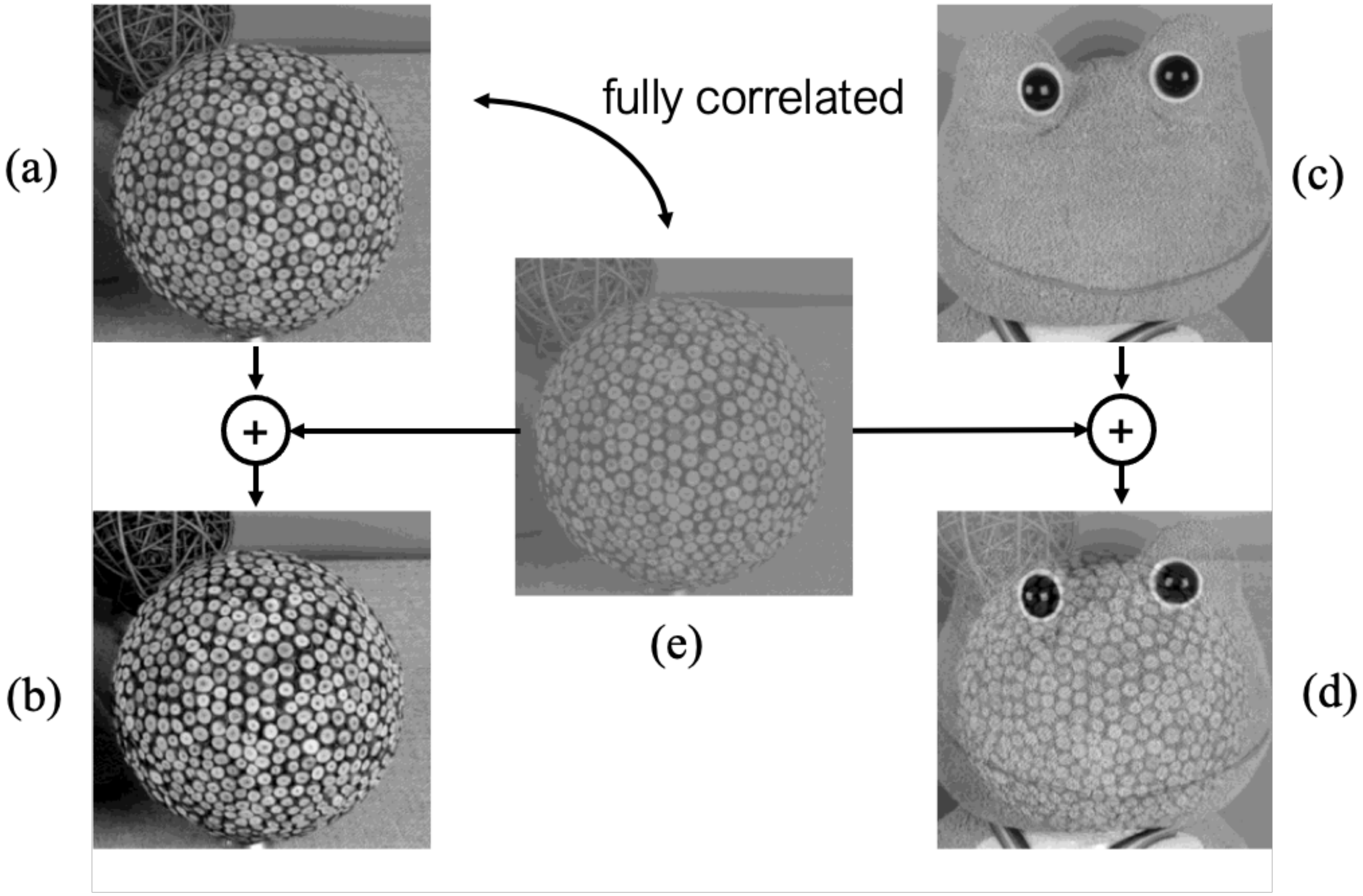



Image Credit: Wang

# 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

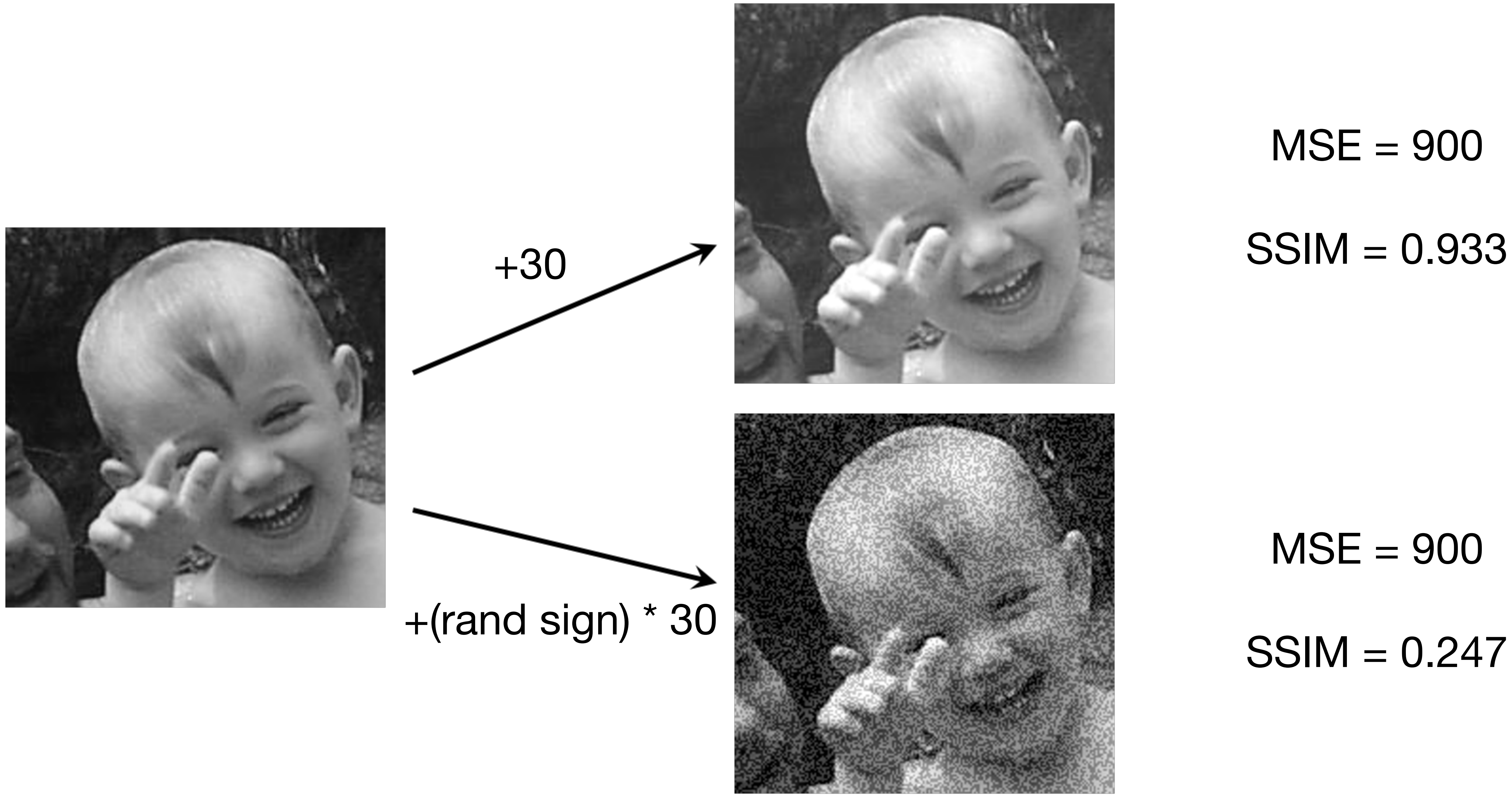



Image Credit: Wang

# 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

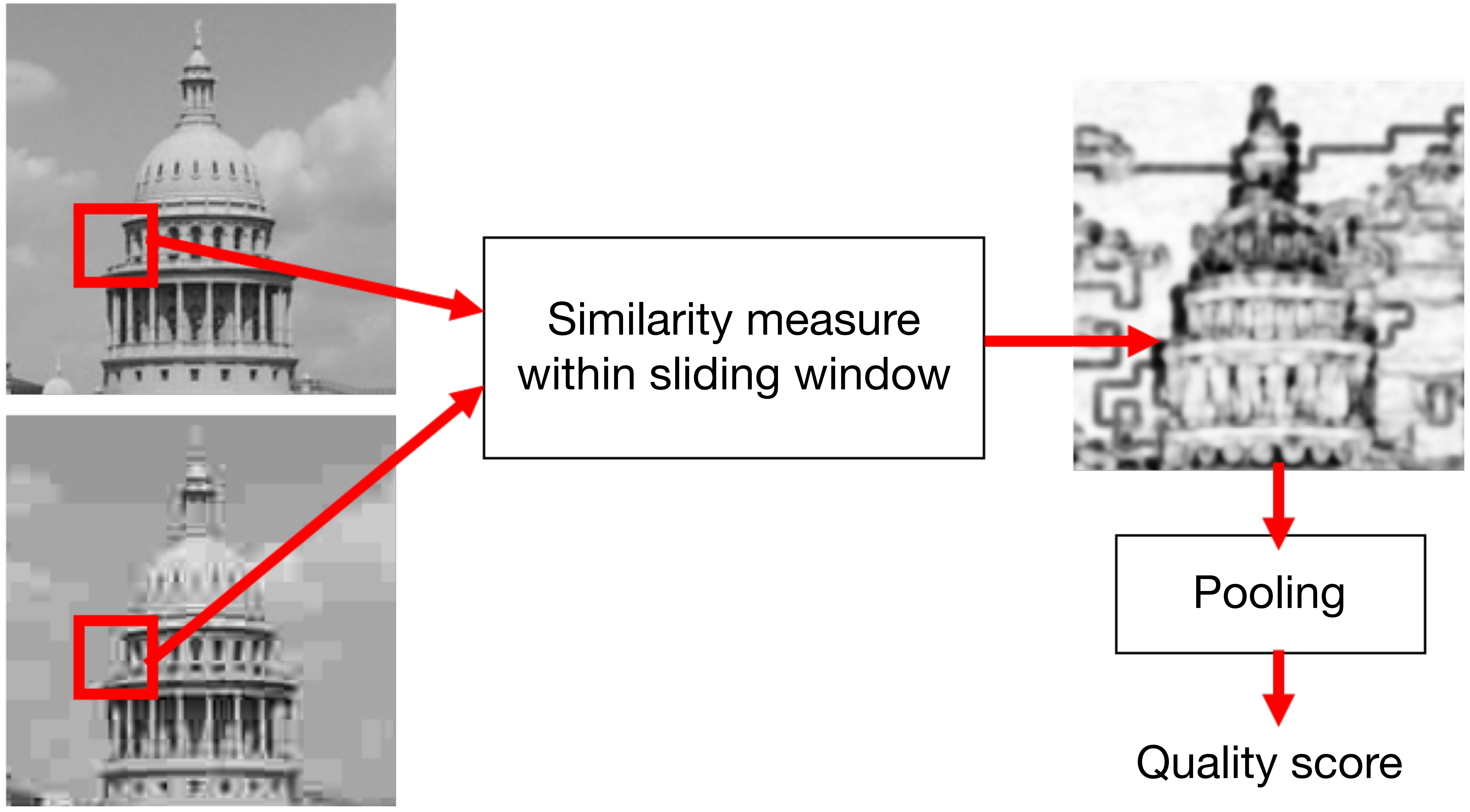


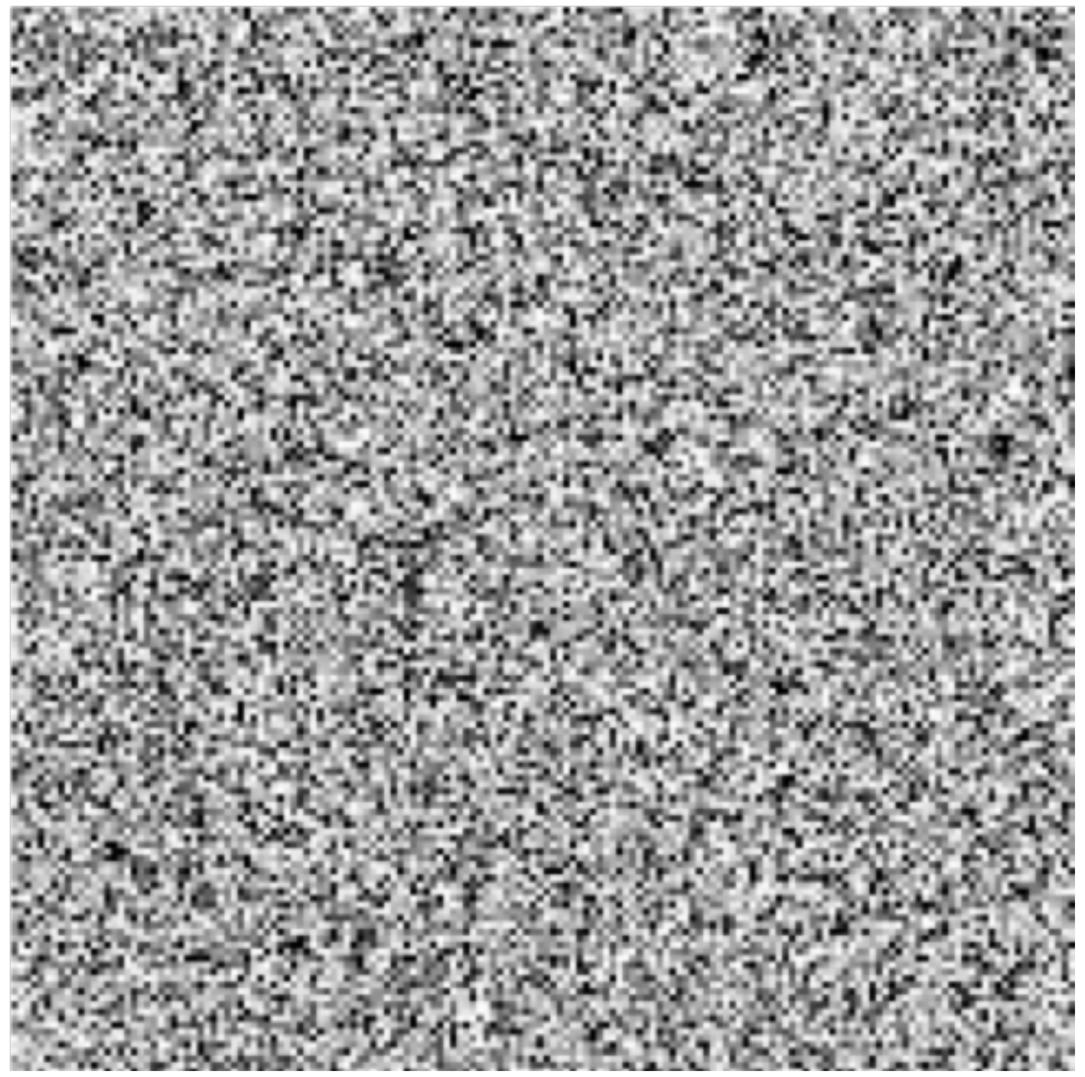

Image Credit: Wang

# Quality Map

Gaussian noise
corrupted image

Original image

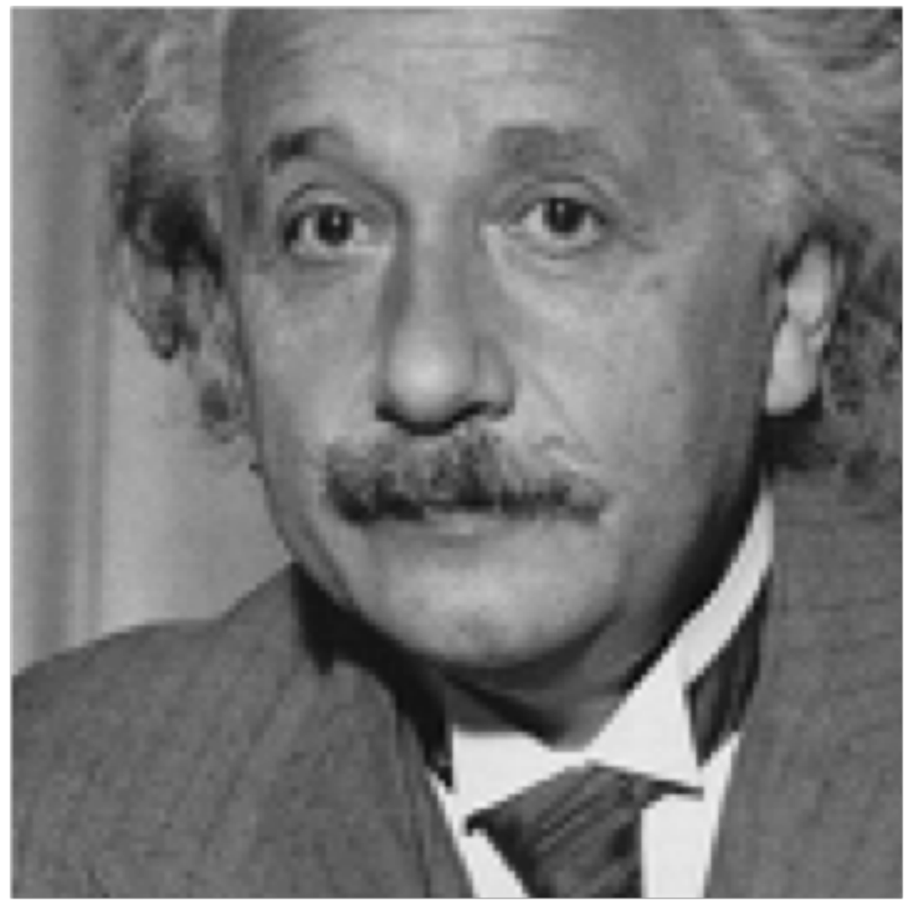
SSIM map



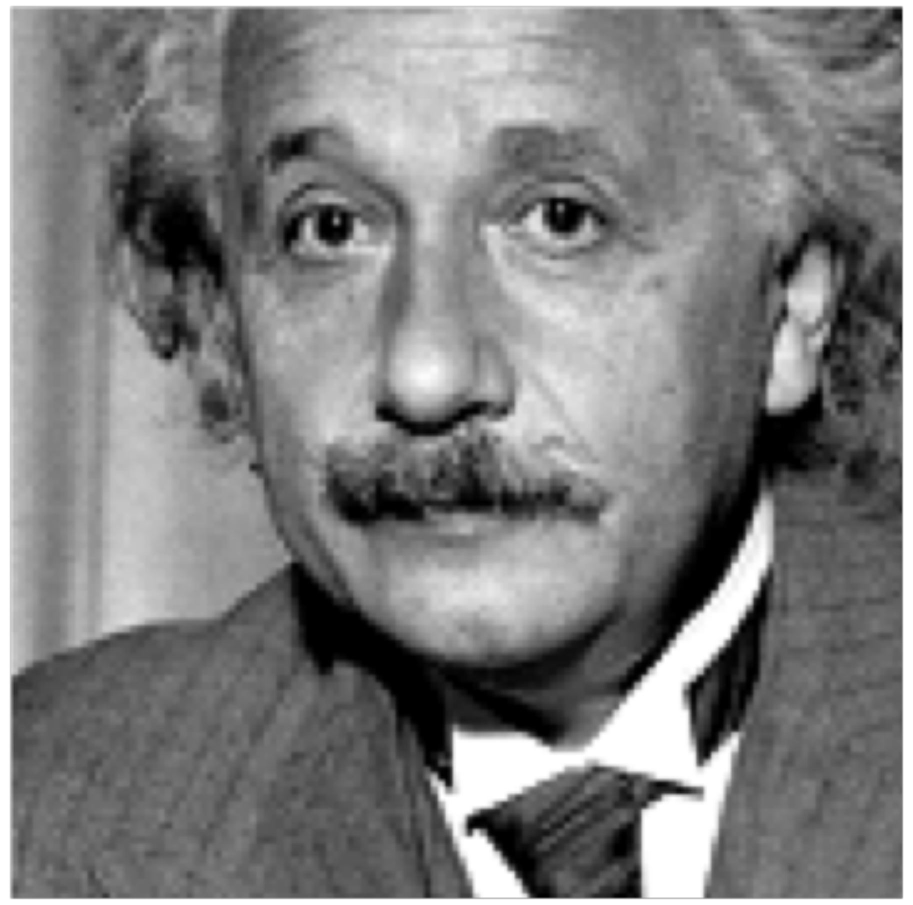
Absolute error map

Image Credit: Wang

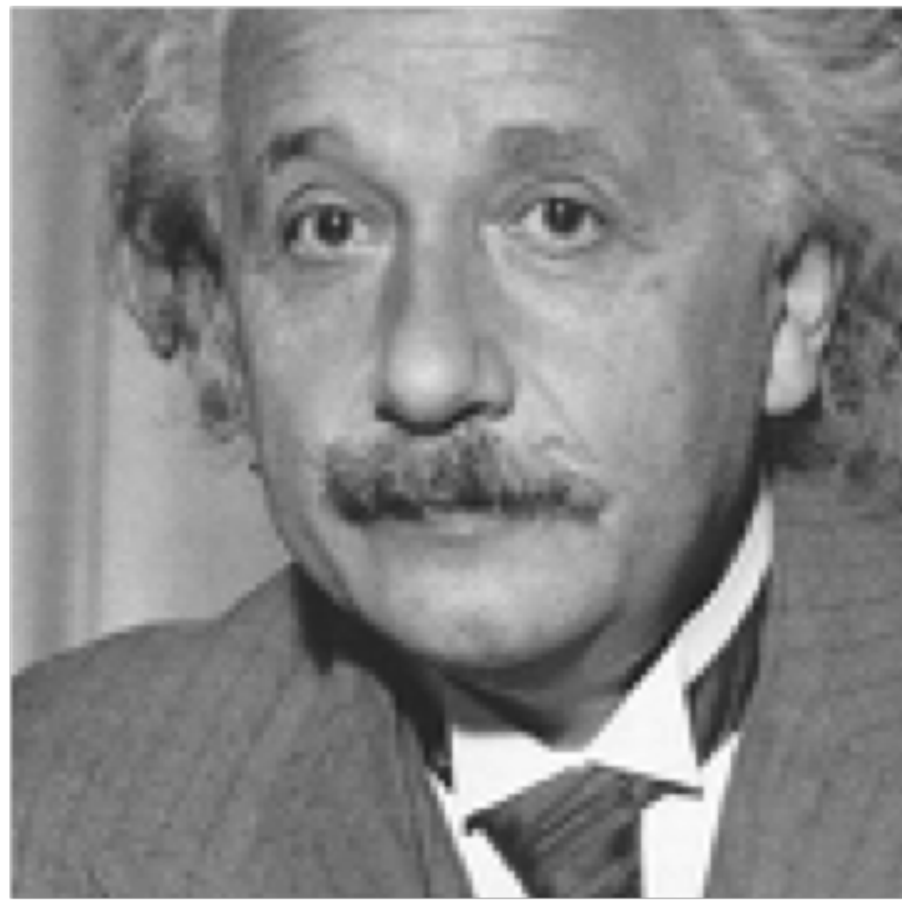
# SSIM vs MSE



MSE = 0, SSIM = 1

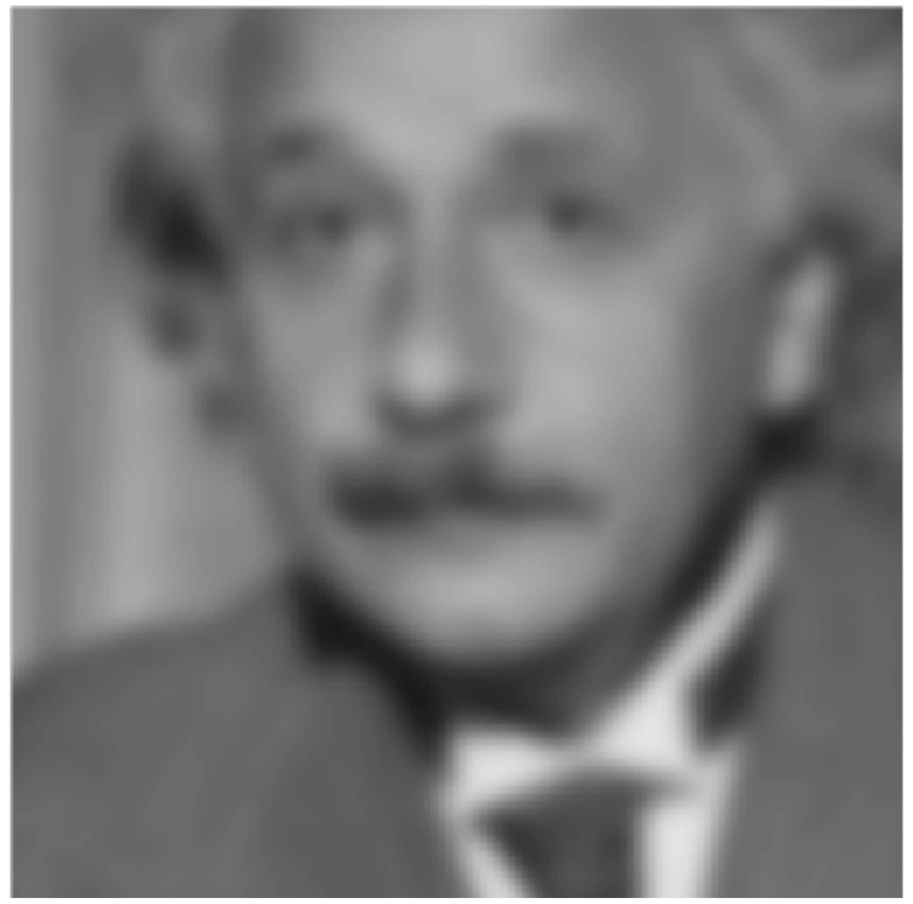


MSE = 309, SSIM = 0.93

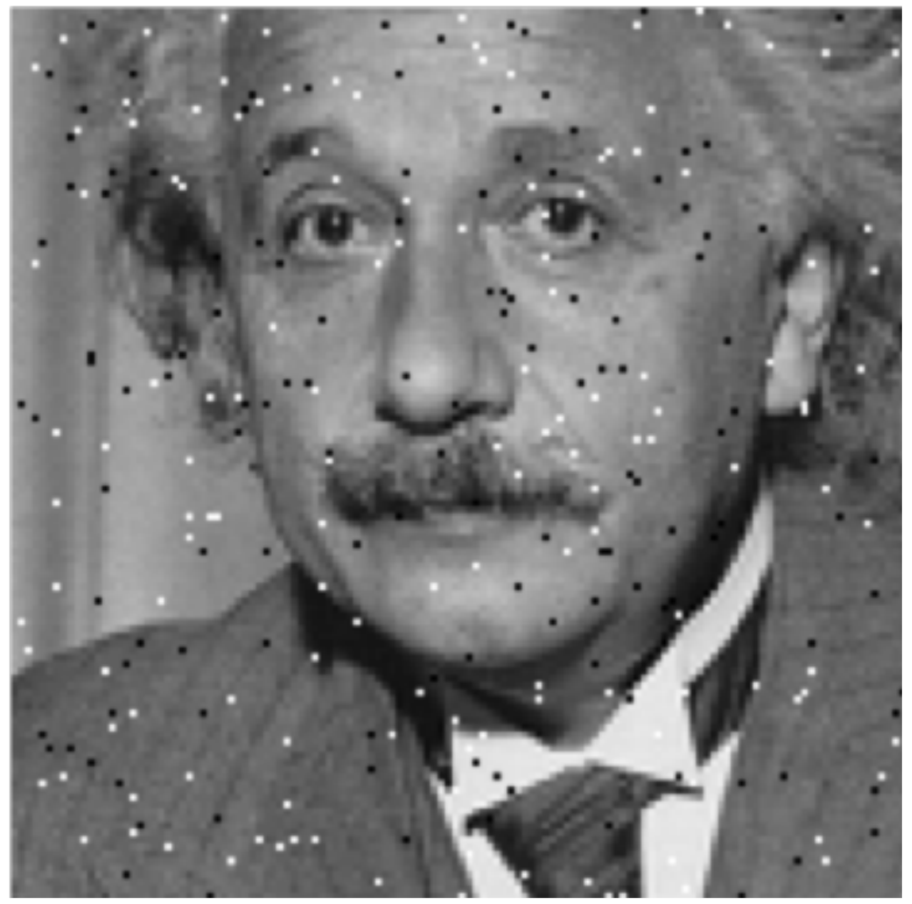


MSE = 309, SSIM = 0.99

MSE = 309, SSIM = 0.58



MSE = 308, SSIM = 0.64



MSE = 309, SSIM = 0.73

Image Credit: Wang

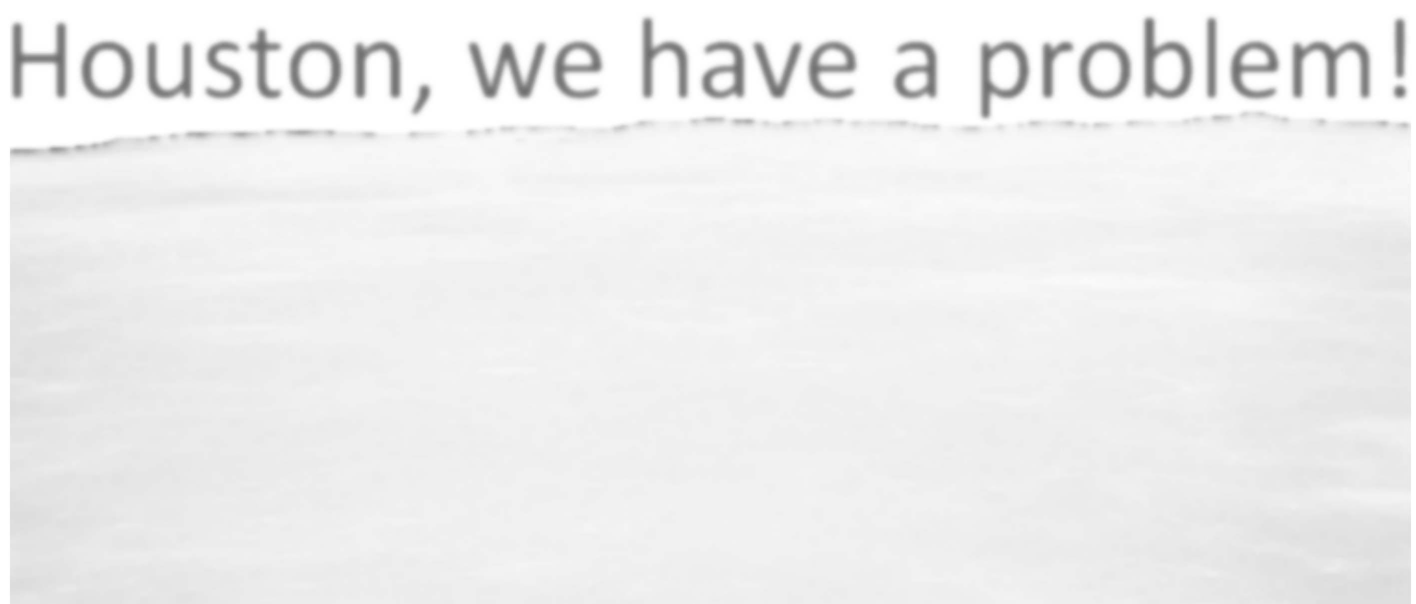
# What is Wrong with SSIM?

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

Normalization is sensitive to low intensities

Original image

Distorted image




SSIM map

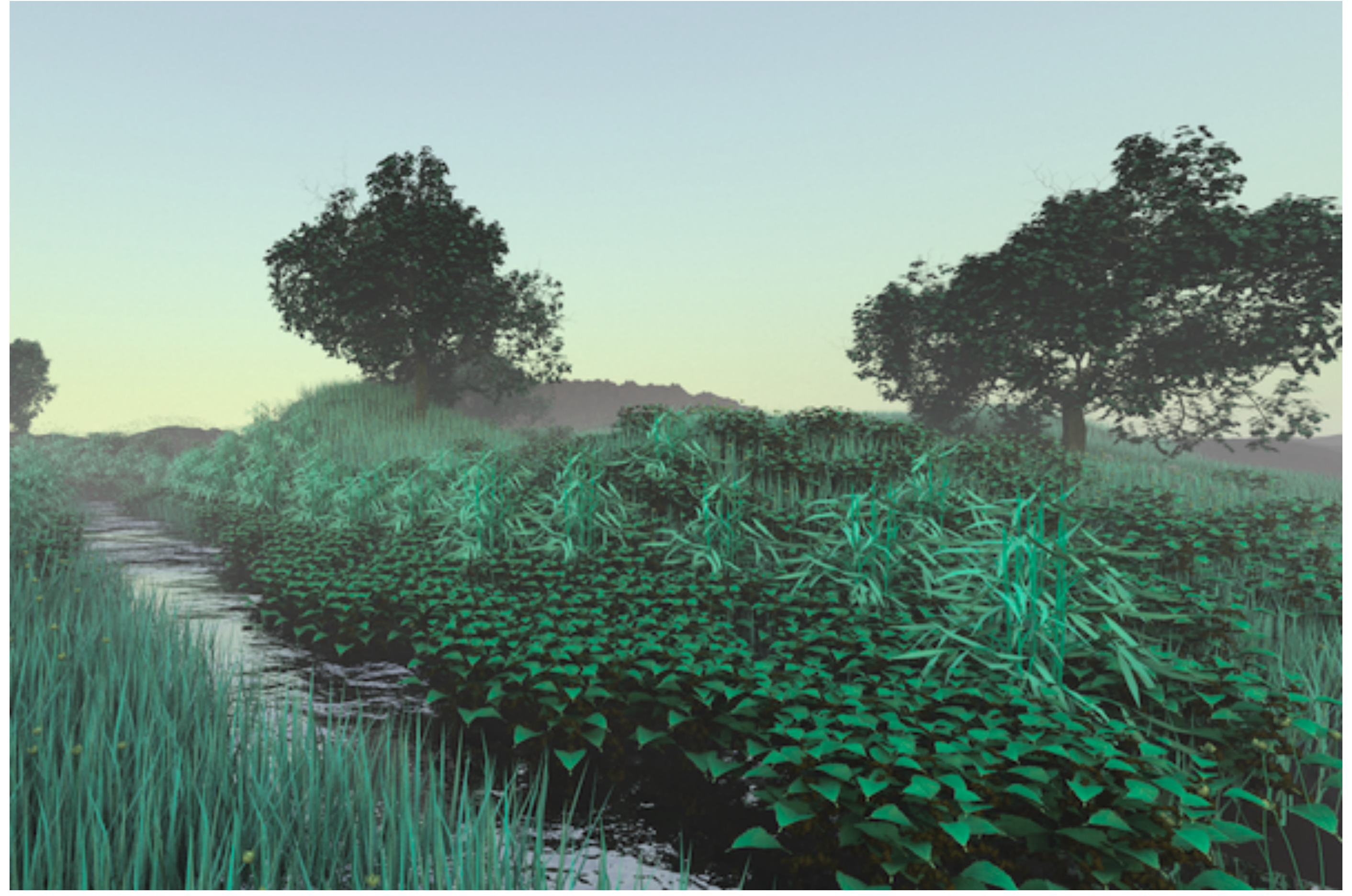
Image Credit: Nilsson and Akenine-Möller

# 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

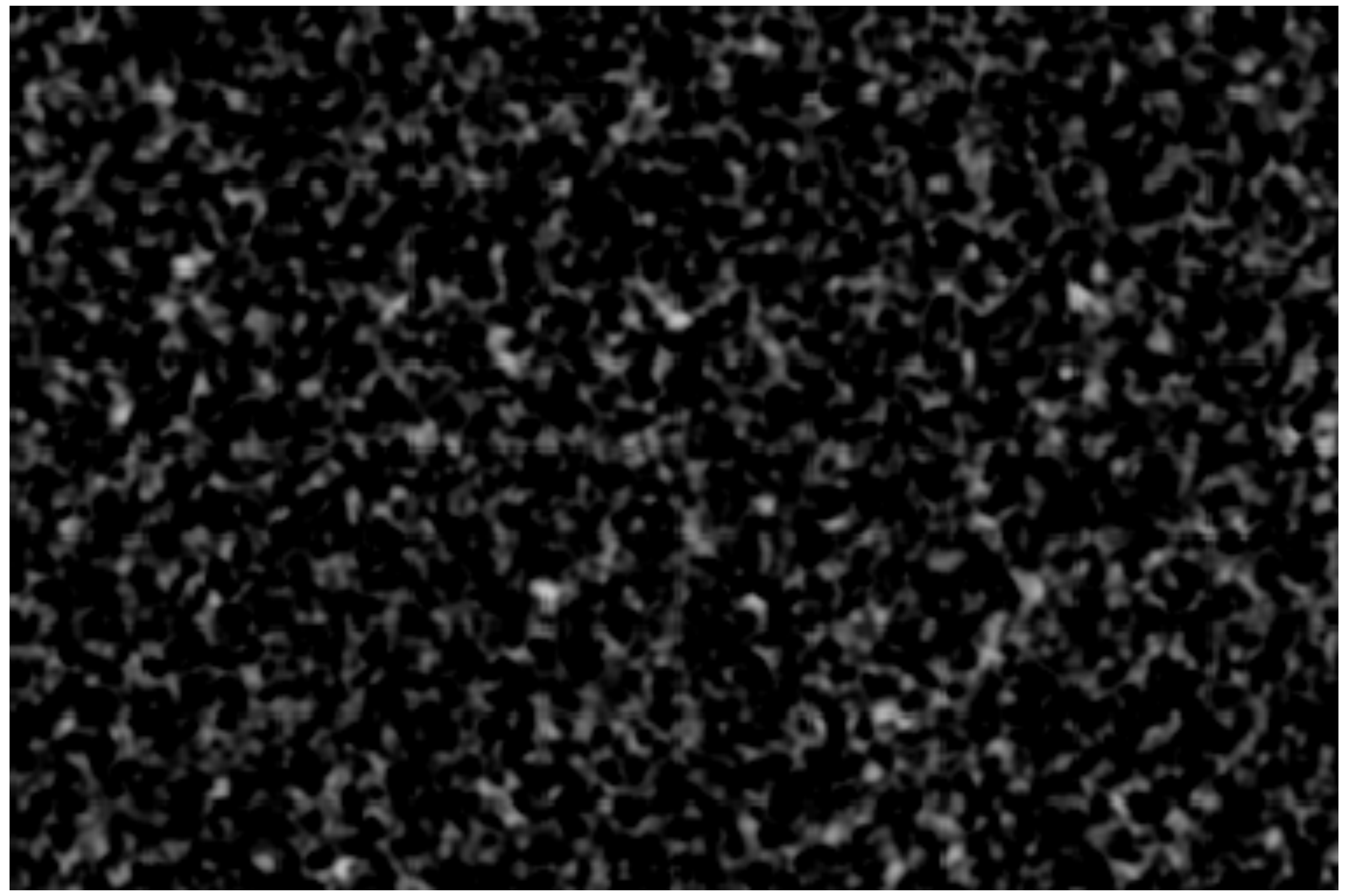
# What is Wrong with SSIM?

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

Rely on point-by-point comparison

Original image

Distorted image



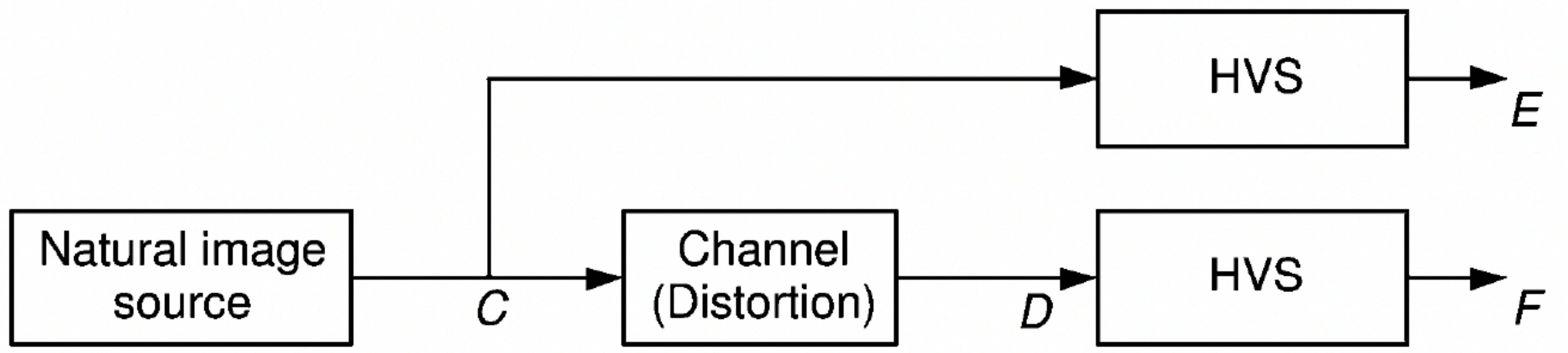
SSIM map

# 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

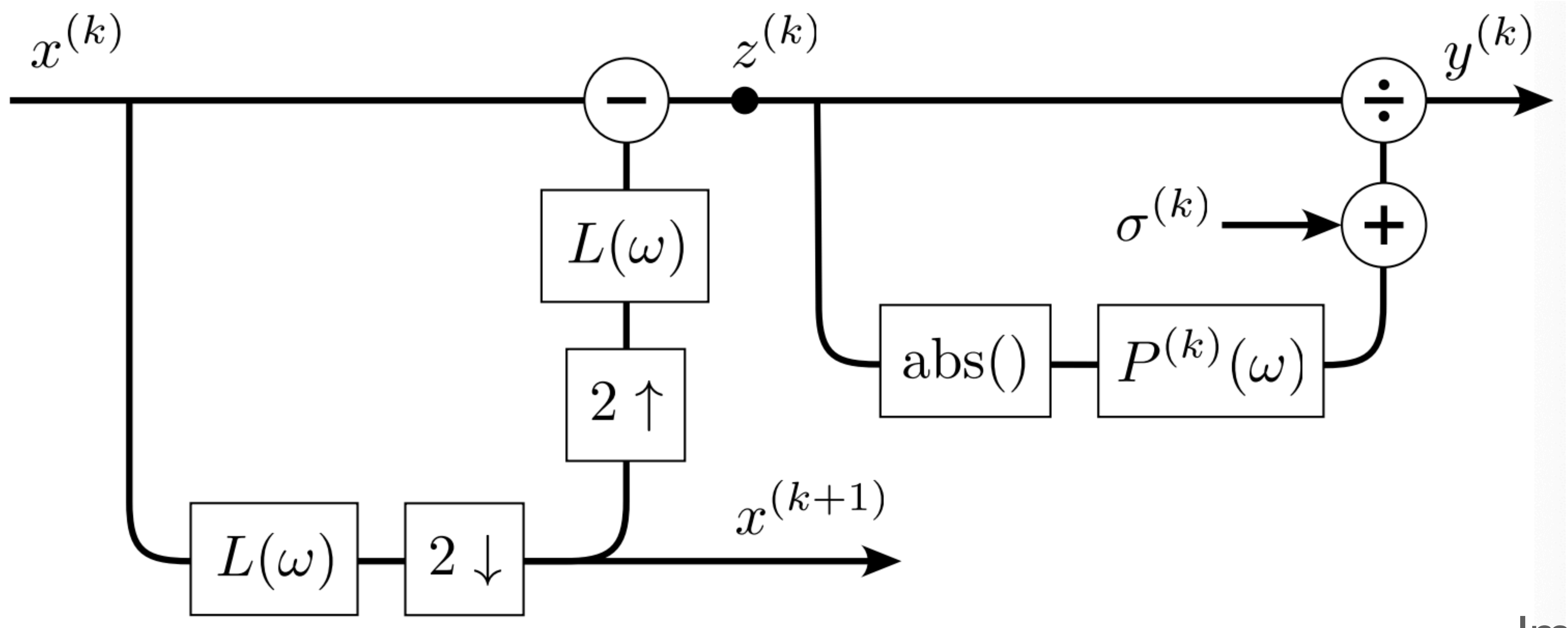
# Most Apparent Distortion (MAD)
## [Larson and Chandler, 2010]

- 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]

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

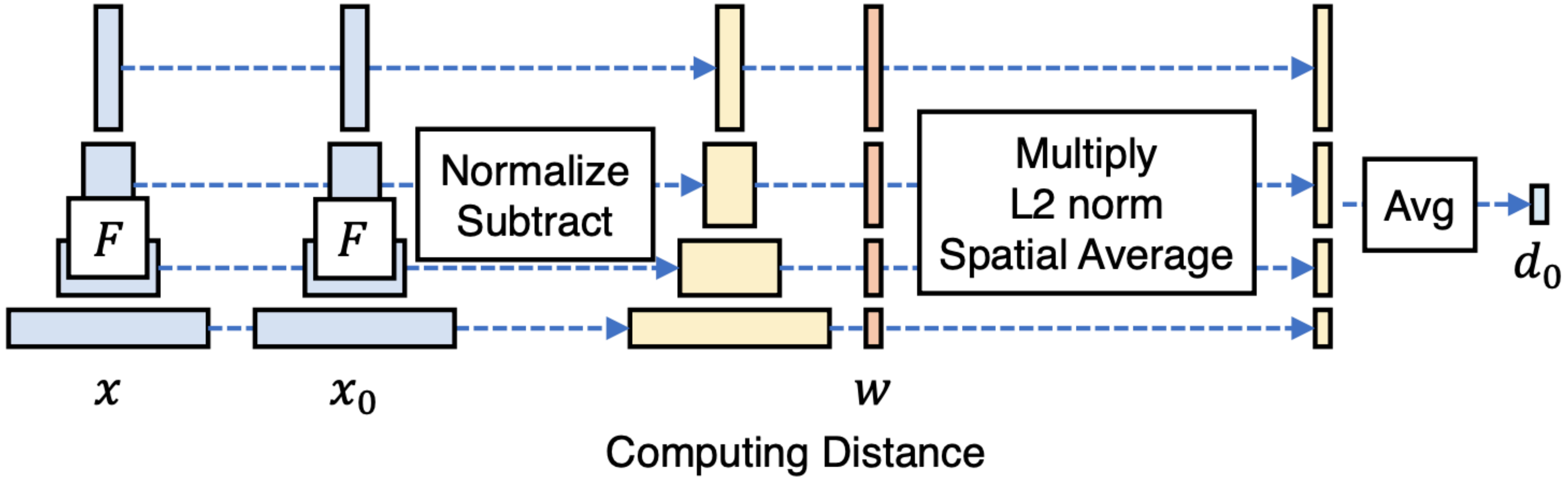
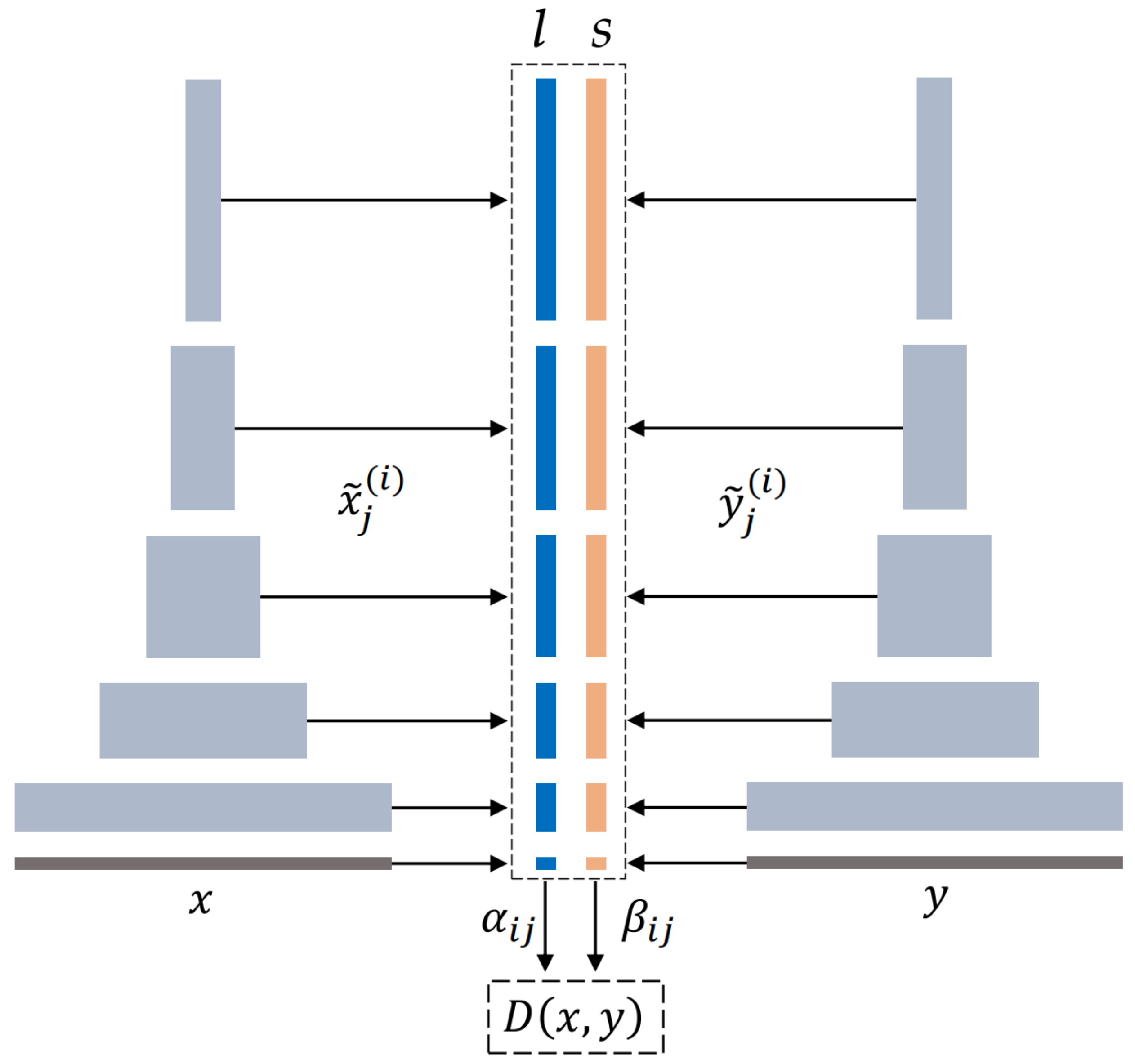



Image Credit: Zhang

# 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

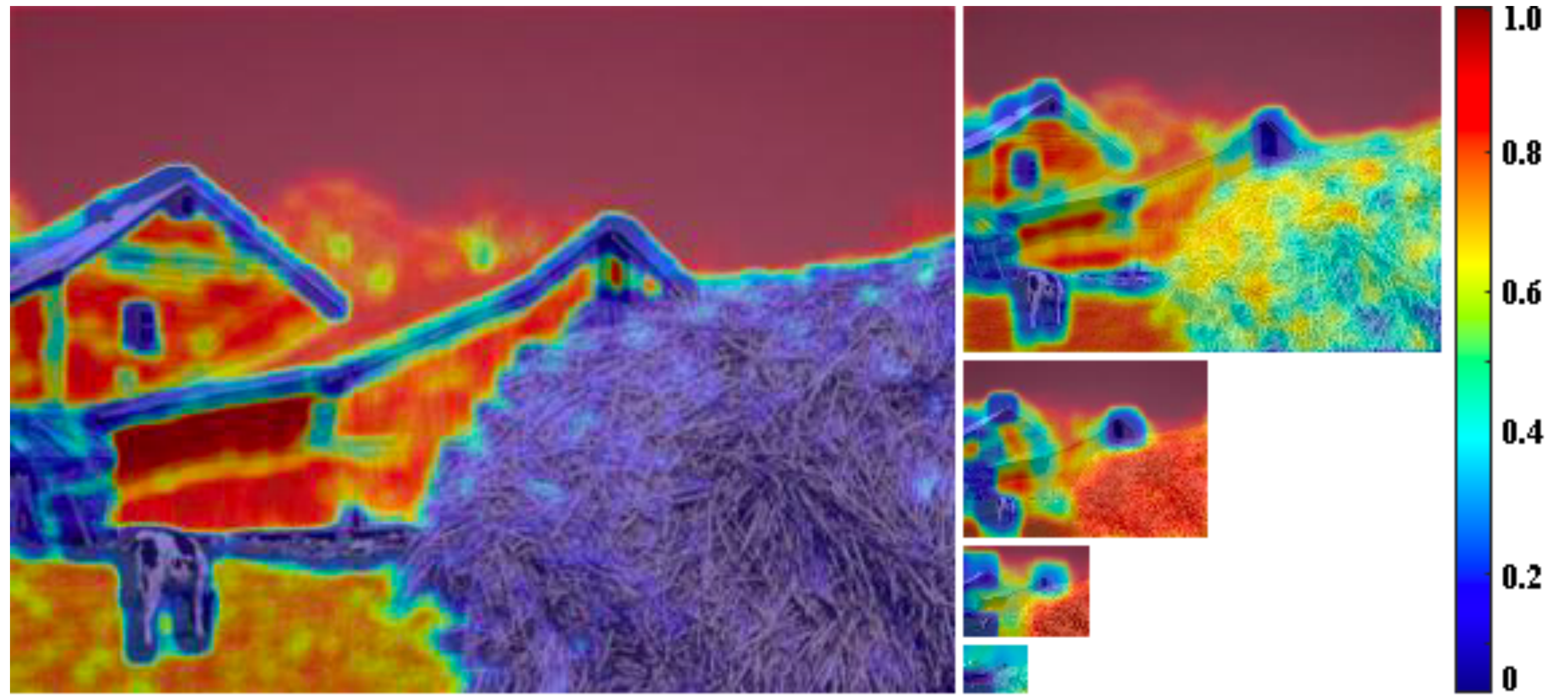



$$\mathrm{DISTS}(x, y) = 1 - \sum_{i=0}^{m} \sum_{j=1}^{n_i} \left( \alpha_{ij} l(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) + \beta_{ij} s(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) \right)$$

# Locally Adaptive DISTS

## [Ding et al., 2021]

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



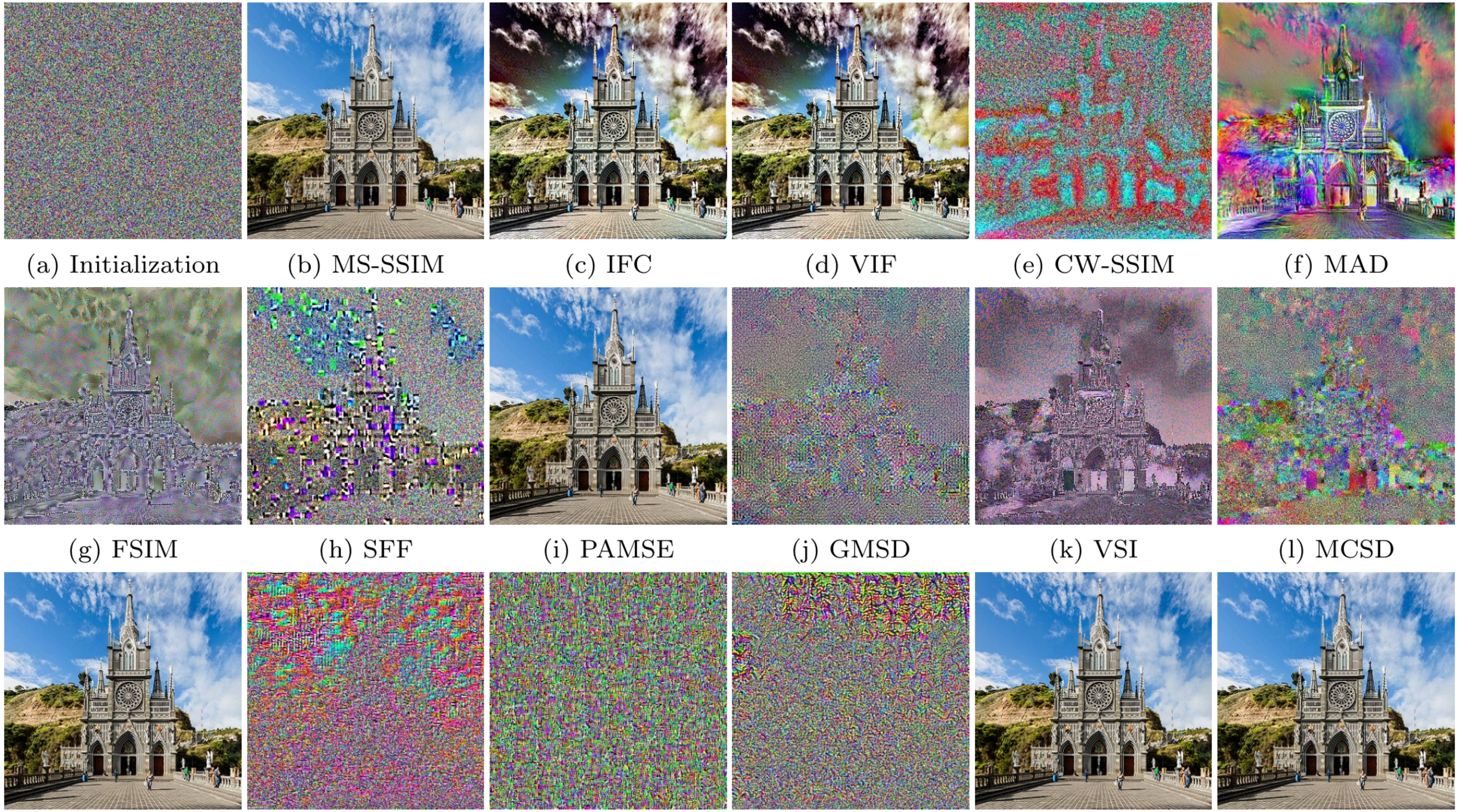

$$\text{A-DISTS}(X, Y) = 1 - \frac{1}{N}\sum_{i=0}^{M}\sum_{j=1}^{N_i} S\left(\tilde{X}_j^{(i)}, \tilde{Y}_j^{(i)}\right)$$

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

# Full-Reference IQA: An Embarrassing Fact

## Reference Image Recovery

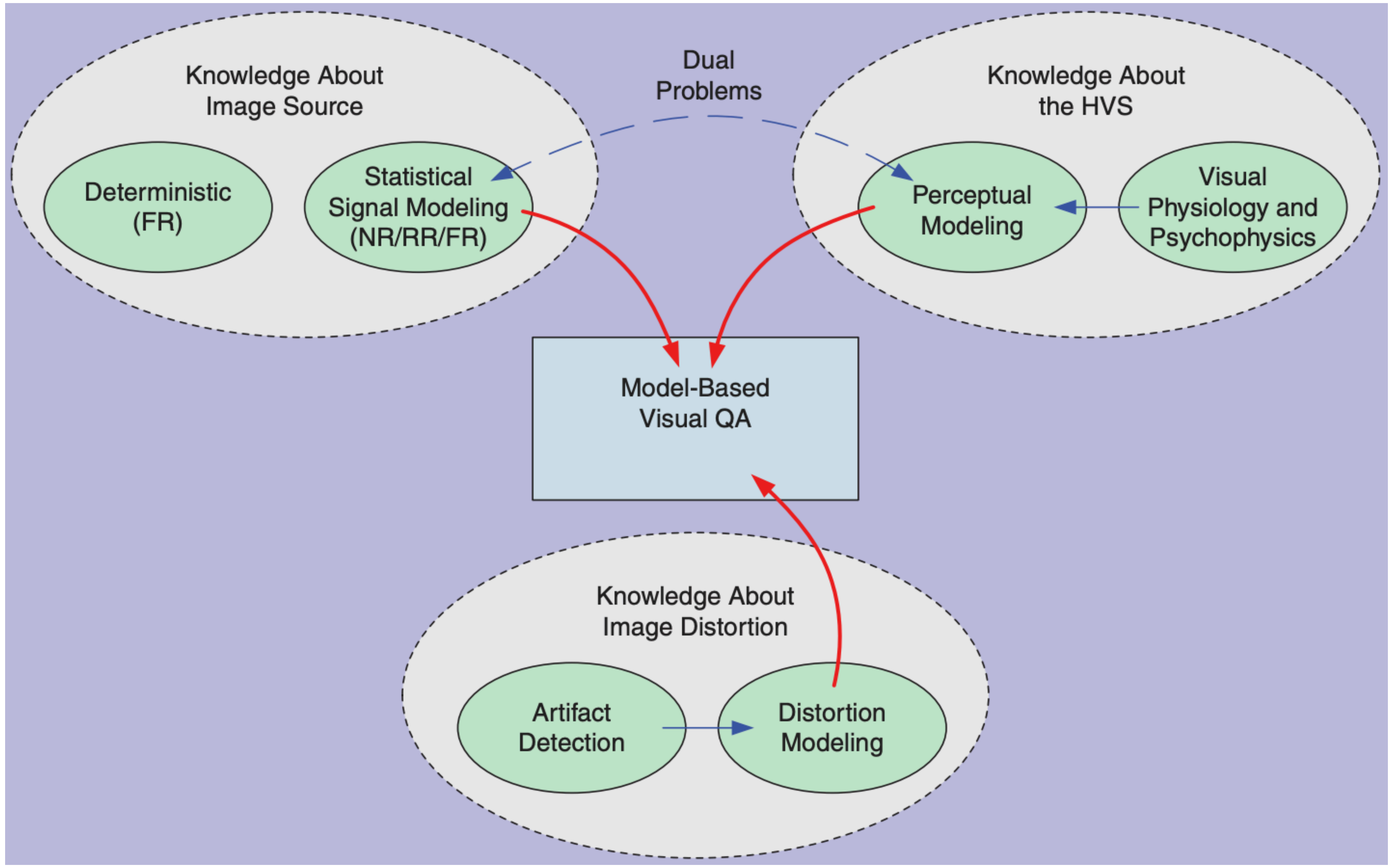
$$y^{\star} = \arg\min_{y} D(x, y)$$



(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

- Spatial domain
  - Edge intensity/spread, sample entropy, BRISQUE, NIQE, IL-NIQE, …
- 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

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

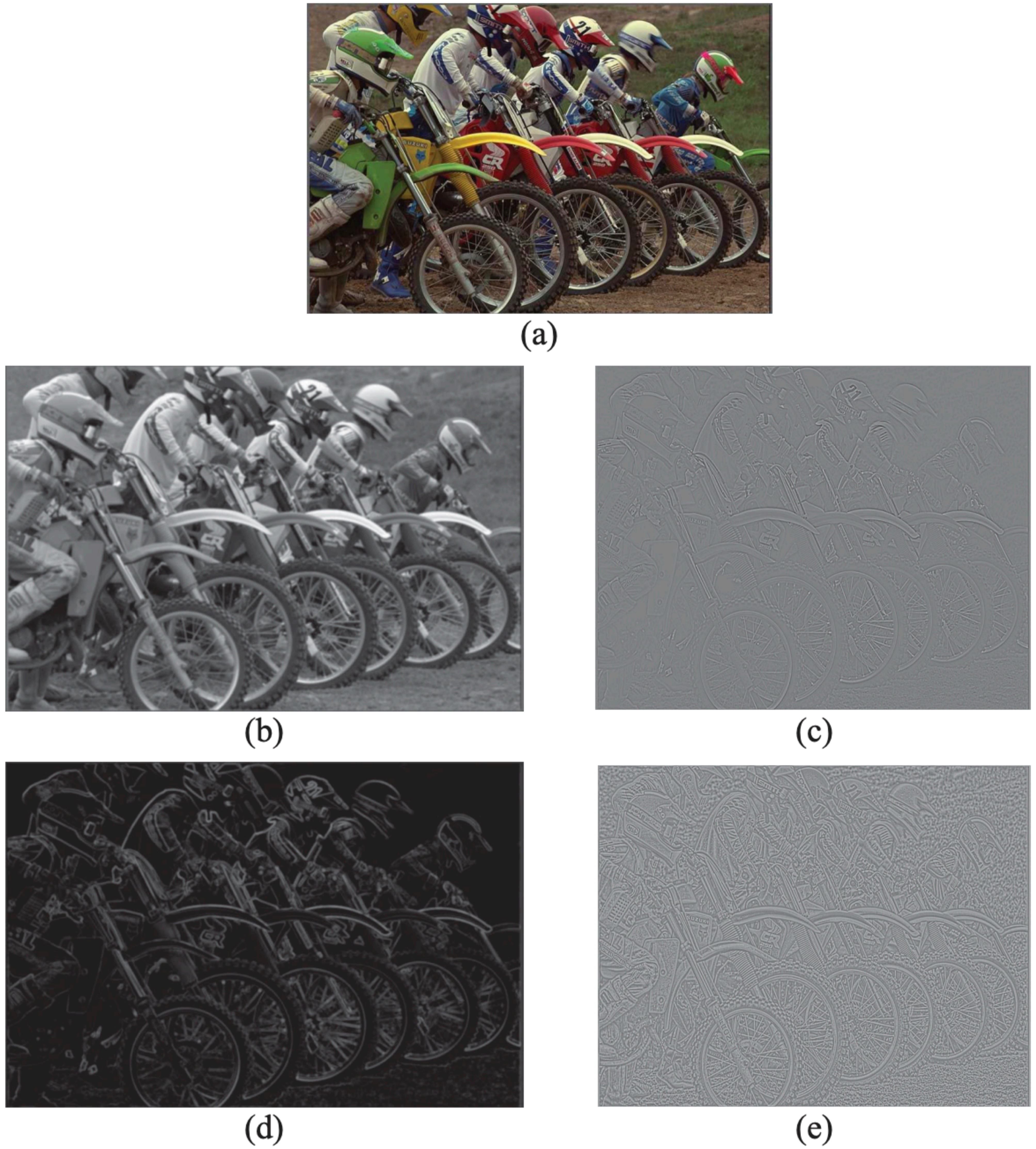



Image Credit: Mittal

# (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

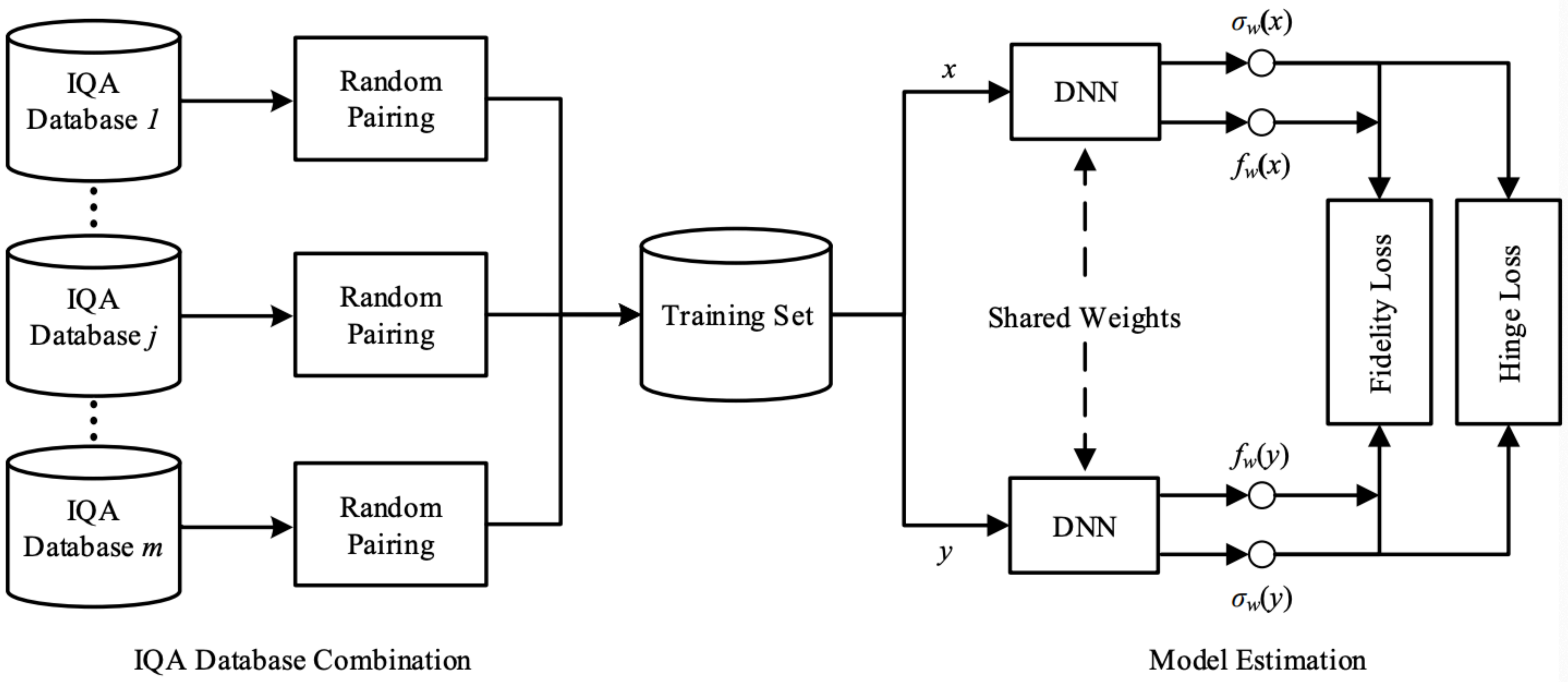
# (Deep) Learning based Approaches

- 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

# New Learning Paradigm

## Unified Learning for No-Reference IQA [Zhang et al., 2021]

- Goal: Learn a unified no-reference IQA model from multiple IQA datasets

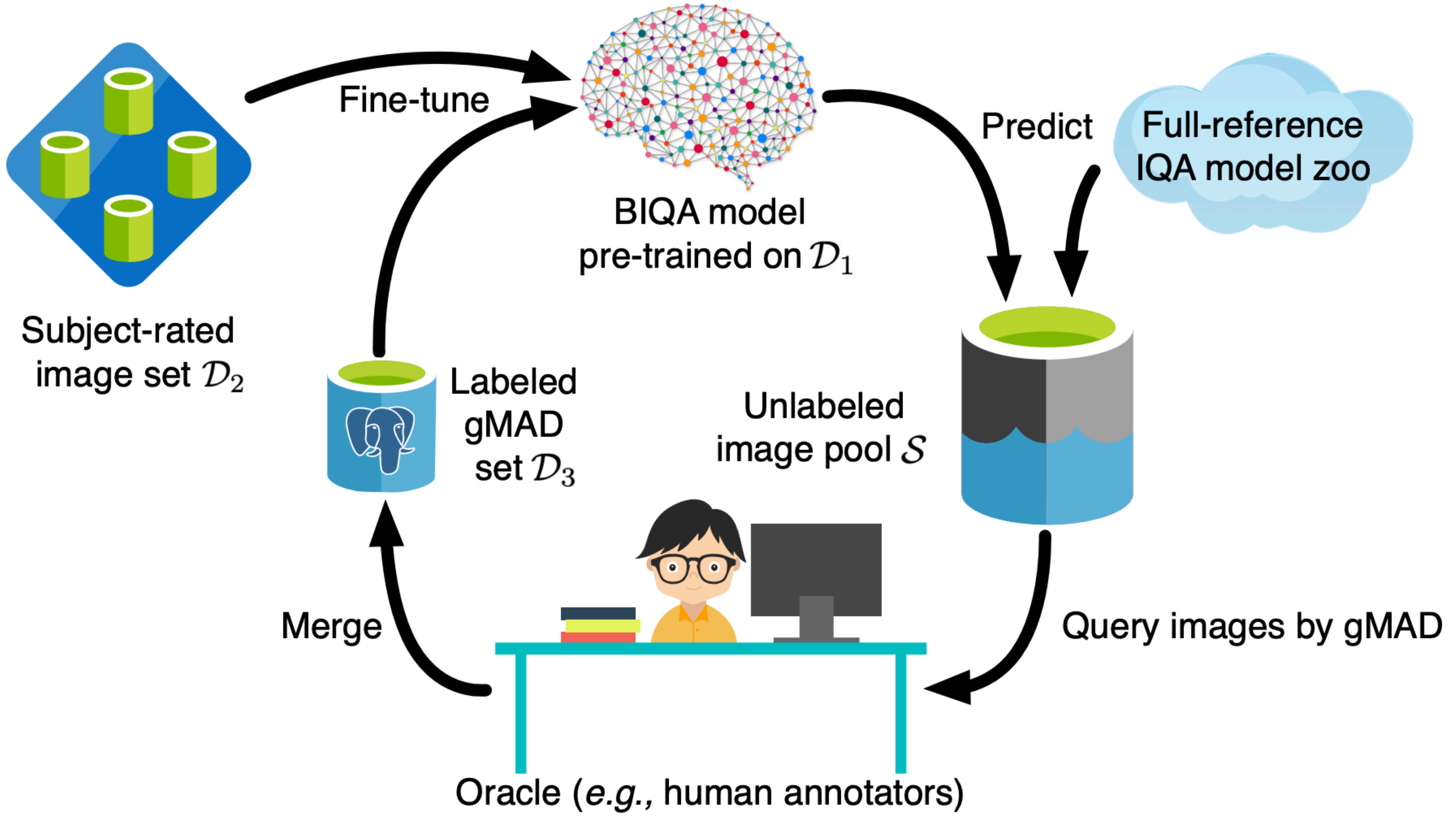



- Question: How to incorporate viewing conditions into model design to make this paradigm more sensible?

# New Learning Paradigm

## Active Learning for No-Reference IQA [Wang et al., 2020, 2021]

- Goal: Identify and learn from the failures of “top-performing” models

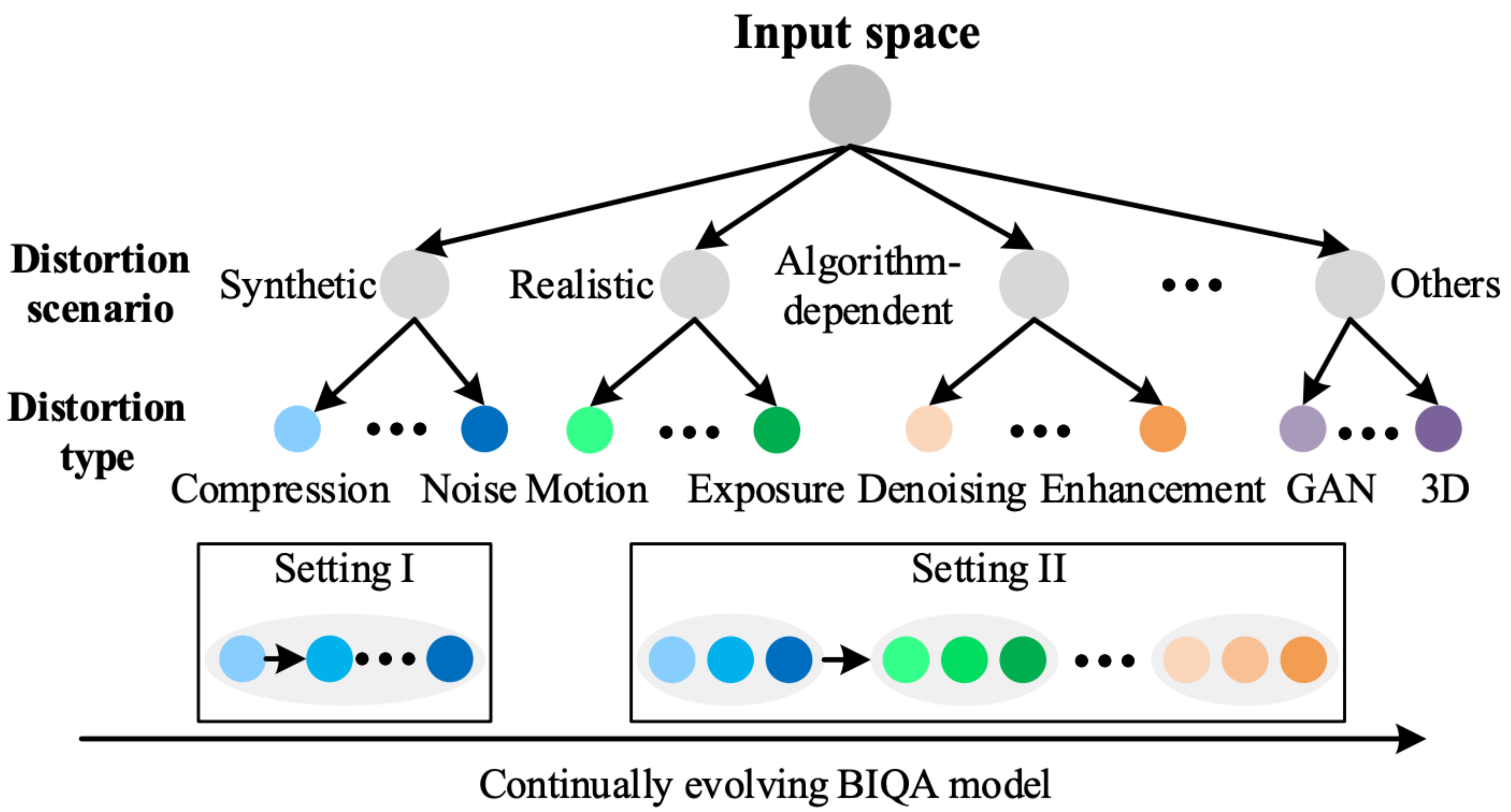



- Question: Are there other effective methods for failure identification?

# New Learning Paradigm

## Continual Learning for No-Reference IQA [Zhang et al., 2021]

- Goal: Learn continually from a stream of IQA datasets, building on what was learned from previously seen data




- Question: What are the desiderata in continual learning for no-reference IQA?

# Discussion

# Discussion

- Transformer-based IQA?
- Absolute vs relative quality
- Generative IQA
    - Model $p(x \mid q)$ (or equivalently $p(x, q)$) rather than $p(q \mid x)$