CS5489
Lecture 11.2: Neural Networks and Deep Learning Part VI: Low-Level Applications

Kede Ma

City University of Hong Kong (Dongguan)

Slide template by courtesy of Benjamin M. Marlin

Part of slides are borrowed from the Stanford University CS231n course

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

6 Colorization

7 Fusion

What is Low-Level Vision?

- Can you recover the unseen pixels?
- Image inpainting

General Formulation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

- $\mathbf{x}$: Signal to be recovered
- $\mathbf{e}$: Measurement noise
- $\mathbf{A}$: Sensing matrix
- $\mathbf{y}$: Measurements

$$\text{(P0)} \quad \min_{x} \overbrace{\|y - Ax\|_2^2}^{\textbf{Fidelity}}.$$

Beyond Fidelity

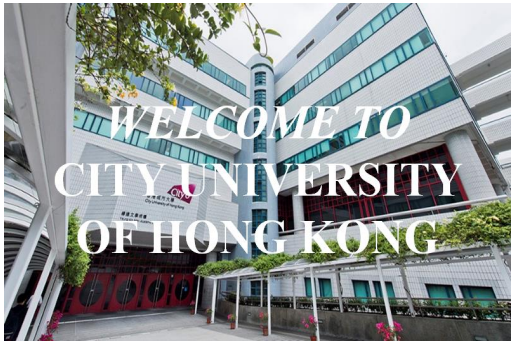
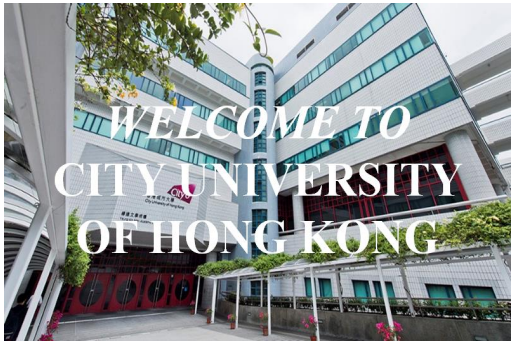


Image Priors Matter



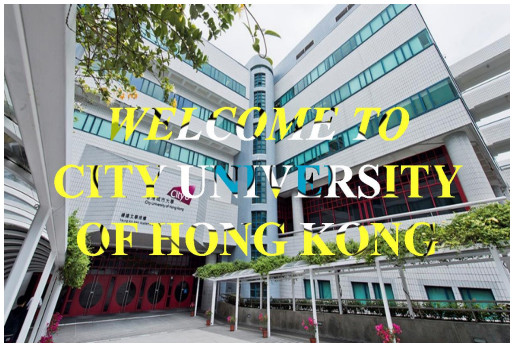


Image Priors Matter

- Prior knowledge is needed

Maximum a Posteriori (MAP) Estimation for Image Enhancement

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

$$\text{(P0)} \quad \min_{x} \overbrace{\|y - Ax\|_2^2}^{\text{Fidelity}} + \overbrace{\lambda \Phi(x)}^{\text{Signal Regularizer}}$$

- $\mathbf{x}$: Signal to be recovered
- $\mathbf{e}$: Measurement errors
- $\mathbf{A}$: Sensing matrix
- $\mathbf{y}$: Measurements
- $\Phi(\mathbf{x})$: Signal regularizer (i.e., image prior)

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

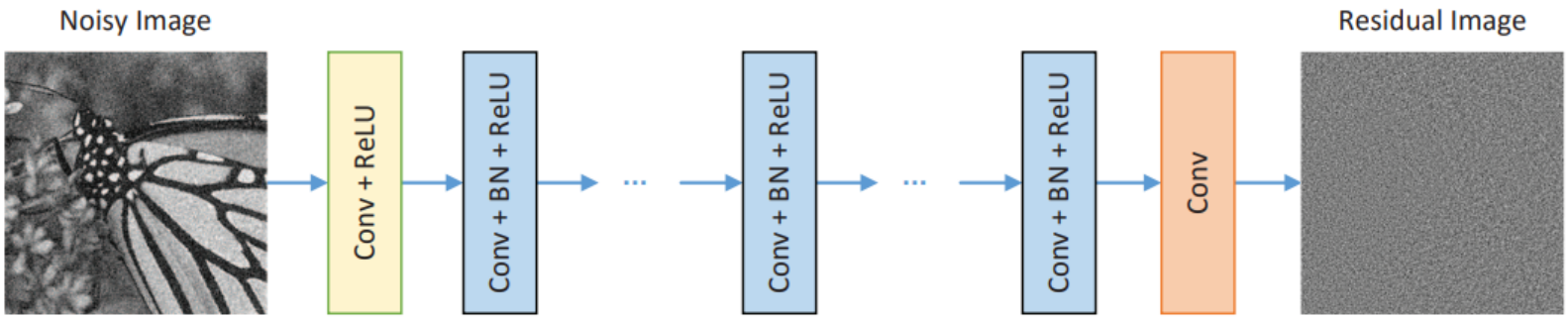
6 Colorization

7 Fusion

Denoising

- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$
- $\mathbf{A} = \mathbf{I}$ is the identity matrix
- Noisy measurement: $\mathbf{y} = \mathbf{x} + \mathbf{e}$
- Simplest and the "most boring" low-level vision problem

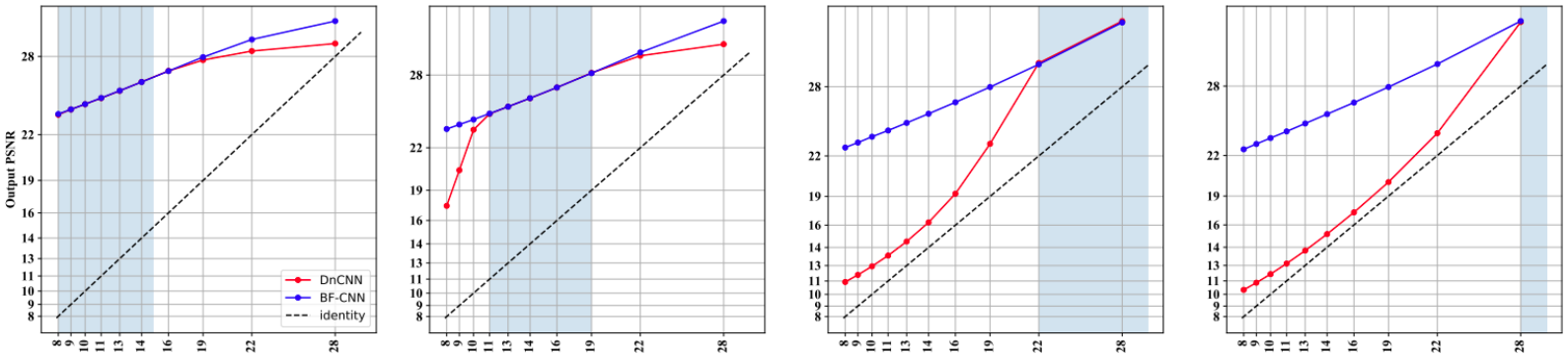
Denoising by Residual Learning (DnCNN)



- Key technical feature: Residual learning
  - MSE as the perfect loss in this setting
  - Much easier to optimize

K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE TIP*, 2017.

Bias-Free CNN for Denoising



- Key technical feature:DnCNN with all bias terms removed
  - Interpretable via linear algebra tools
  - Generalizable to noise levels beyond the training range

S. Mohan, Z. Kadkhodaie, E. P. Simoncelli, and C. Fernandez-Granda, "Robust and interpretable blind image denoising via bias-free convolutional neural networks," *arXiv.org e-prints, Technical Report 1906.05478*, 2019.

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

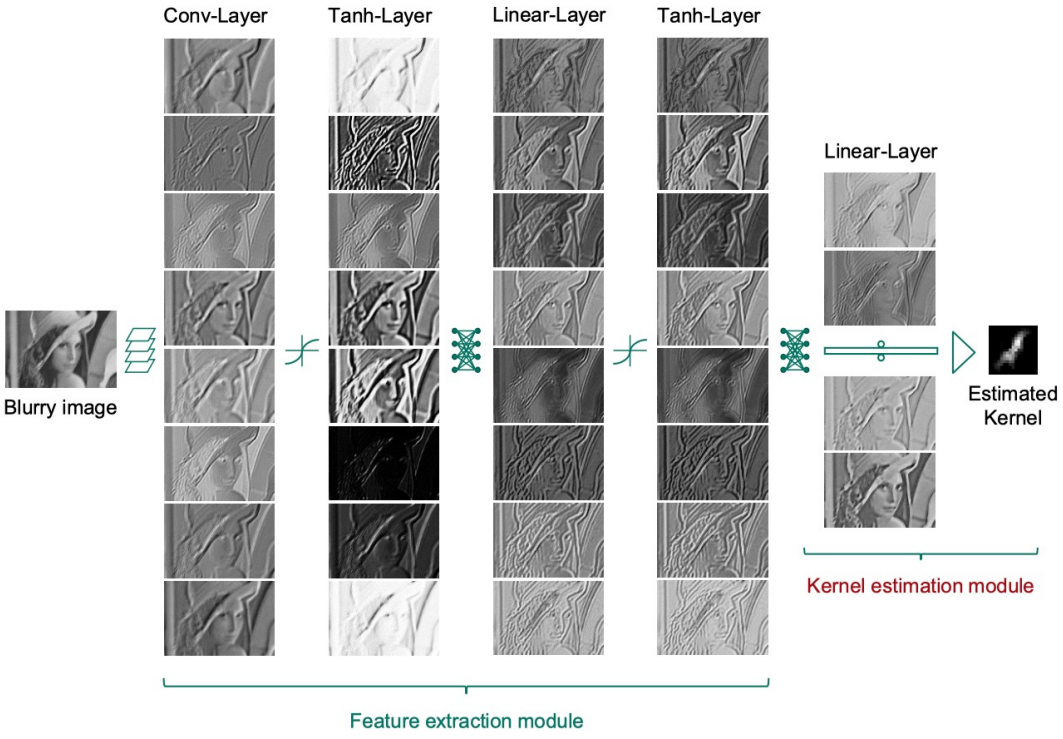
6 Colorization

7 Fusion

Deblurring

- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$
- $\mathbf{A}$ is a blurring operator (i.e., convolution with a blur kernel)

Deblurring by Estimating Blurring Kernel



- Key technical feature: Estimate the blurring kernel
  - MSE as the loss function

C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE TPAMI*, 2015.

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

6 Colorization

7 Fusion

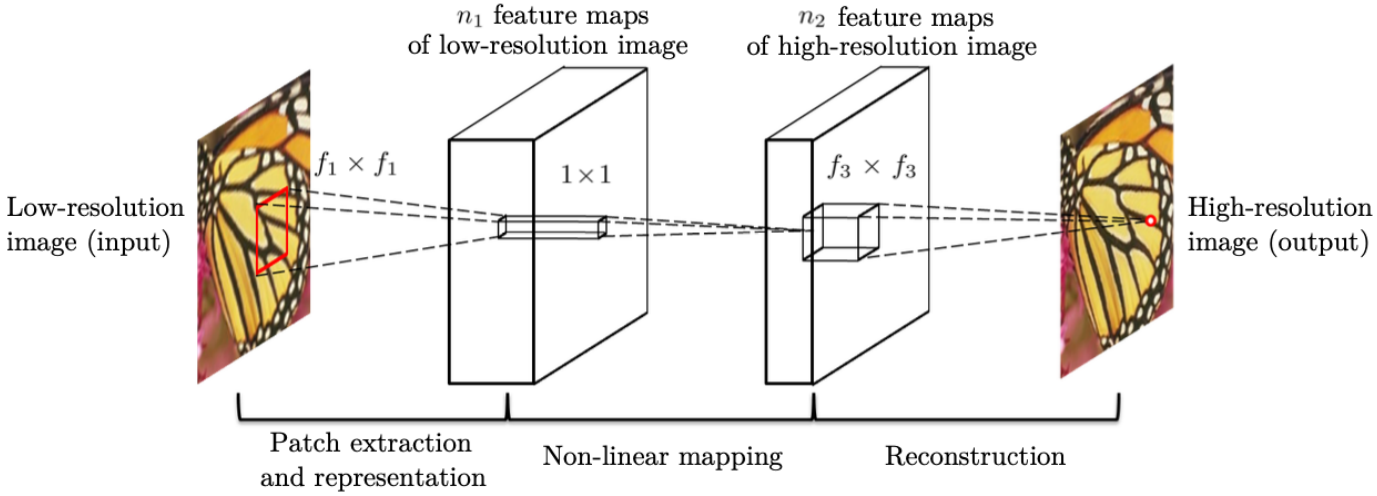
Super-resolution

- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} = \mathbf{D}\mathbf{H}\mathbf{x} + \mathbf{e}$
- $\mathbf{H}$ is a blurring operator and $\mathbf{D}$ is a down-sampling operator

Original high-resolution

Down-sampled

Super-resolution by SRCNN



- Key technical feature: Map low-resolution features nonlinearly to high-resolution representations
  - Good conceptual connections to previous methods
  - MSE as the loss function

C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *ECCV*, 2014.

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

6 Colorization

7 Fusion

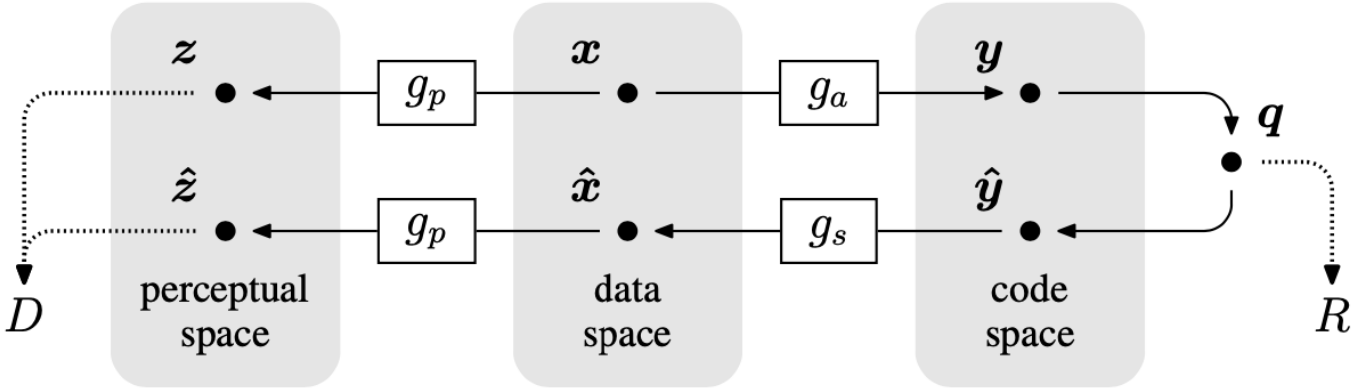
Compression

- Image compression minimizes its size (in bytes) without degrading image quality below an acceptable threshold

JPEG, 4283 bytes (0.121bit/px)

Better Compression, 3986 bytes (0.113bit/px)

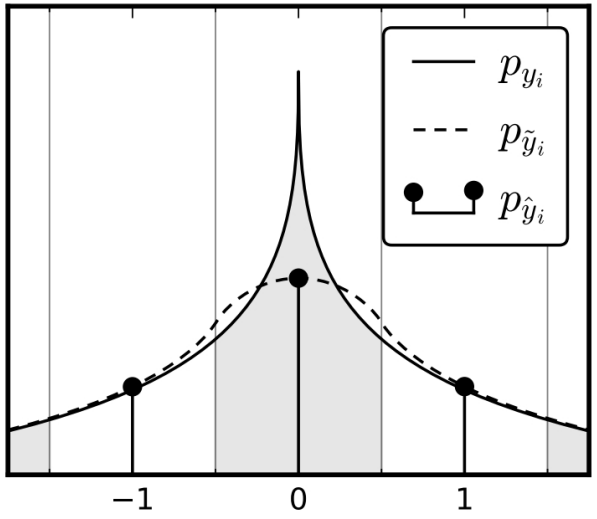
End-to-end Optimized Image Compression



- Key technical feature: Jointly optimize a weighted sum of the rate and distortion
  - Rate (R) as the discrete entropy and distortion (D) as the MSE
  - Loss function: $R + \lambda D$
  - Backpropagate through the non-differentiable quantizer by adding uniform noise

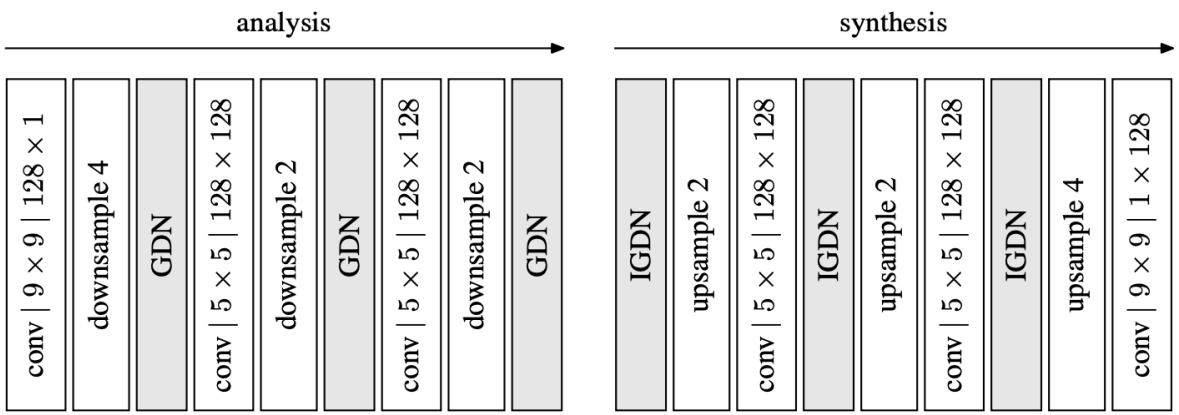
J. Ballé, V. Laparra and E. P. Simoncelli, "End-to-end optimized image compression," *ICLR*, 2017.

Quantizer



- $\hat{y}_i = \text{round}(y_i)$ and $p_{\hat{y}_i}(n) = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} p_{y_i}(t)dt$, for all $n \in \mathbb{Z}$
- $\tilde{y}_i = y_i + \Delta y_i$, for $\Delta y_i \sim \mathcal{U}(-0.5, 0.5)$
  - The density function of $\tilde{y}_i$ is a continuous relaxation of the probability mass function of $\hat{y}_i$, identical at integer values
  - Independent uniform noise is frequently used as a model of quantization error in signal processing

Network Architecture



- Generalized Divisive Normalization (GDN)
  - Let $\boldsymbol{x} \in R^N$ be the input vector to GDN, and the output response $\boldsymbol{z} \in R^N$ can be computed by

$$z_i = \frac{x_i}{\left(\beta_i + \sum_{j=1}^{N} \gamma_{ji} x_j^2\right)^{\frac{1}{2}}},$$

where the weight matrix $\boldsymbol{\gamma} \in R^{N \times N}$ and the bias vector $\boldsymbol{\beta} \in R^N$ are parameters in GDN to be optimized

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

6 Colorization

7 Fusion

Image Colorization

- Grey-level images are obtained by dropping the color representation of color images

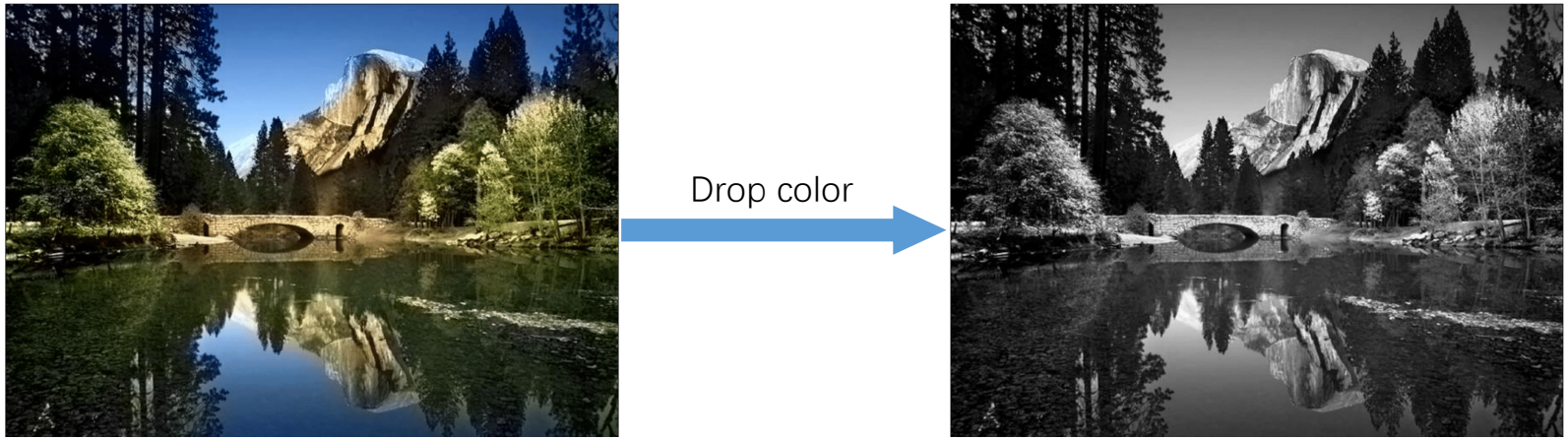
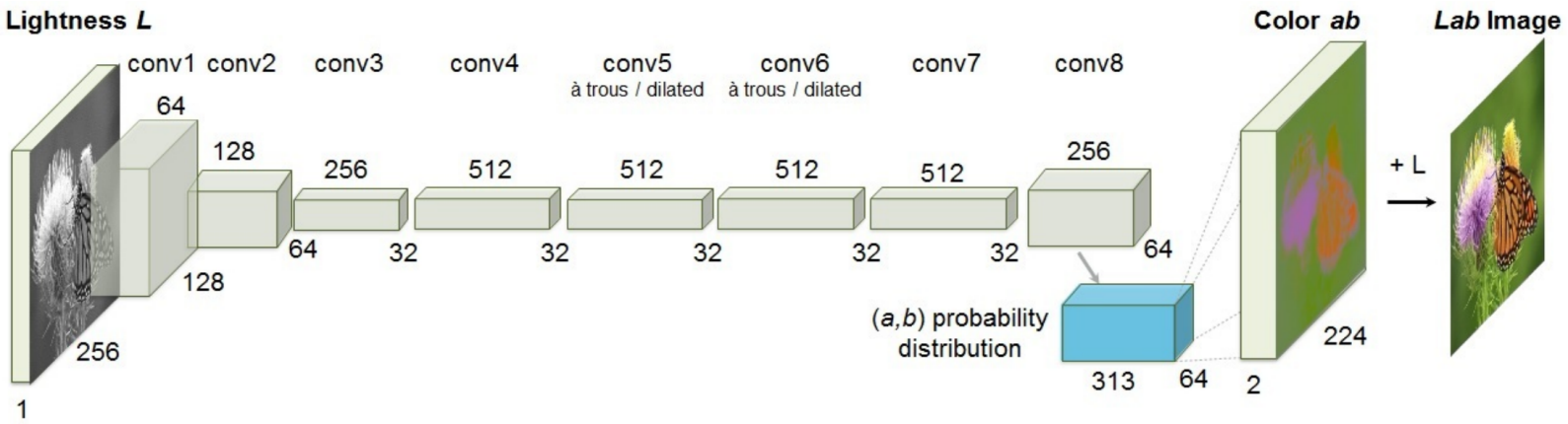


Image Colorization by CNNs



- Key technical feature: Predict a probability distribution of discretized ab values
  - CIE LAB is a more perceptually uniform color space than RGB
  - Cast a regression task into a multiclass classfication task to cope with the inherent ambiguity and multimodal nature of colorization

R. Zhang, P. Isola, and A.A. Efros, "Colorful image colorization," *ECCV*, 2016.

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

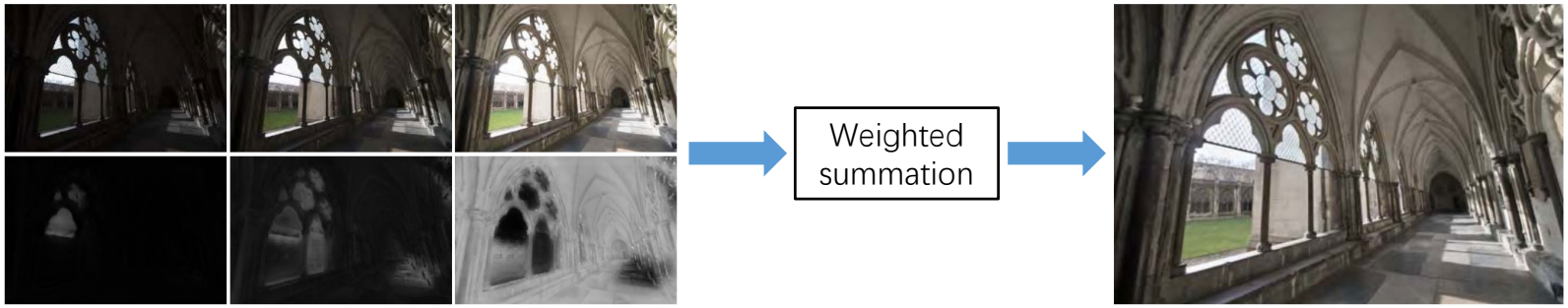
5 Compression

6 Colorization

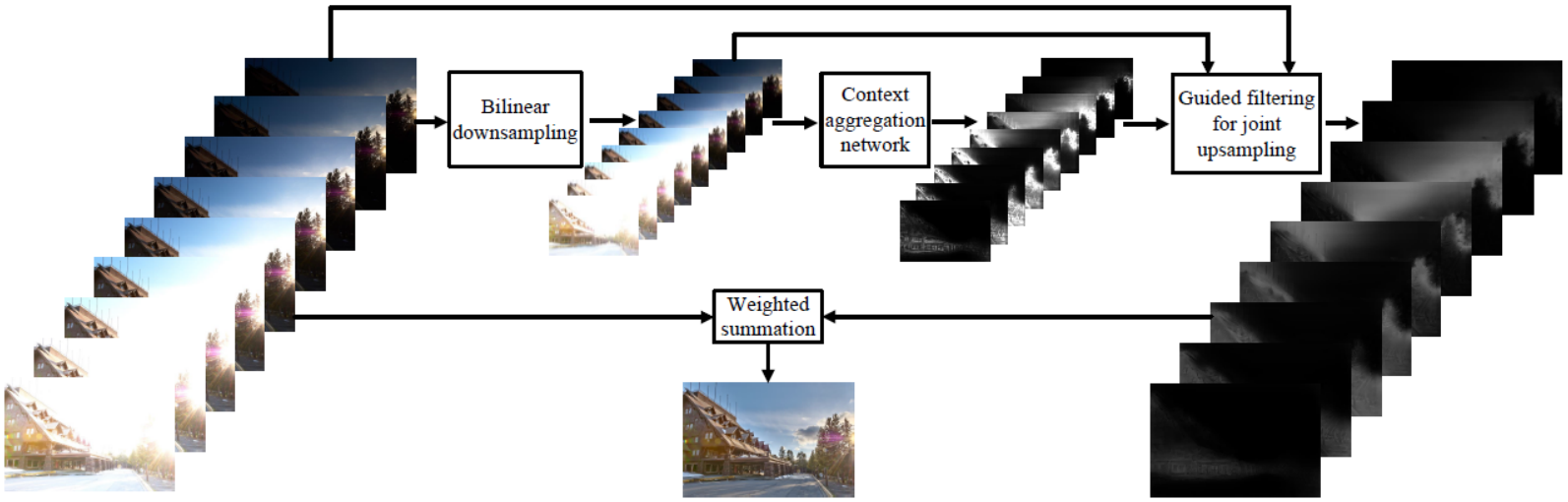
7 Fusion

Multi-Exposure Image Fusion (MEF)

- MEF takes an image sequence with different exposure levels as input and produces a high-quality image with richer details
- This is mainly accomplished by a weighted summation framework $\mathbf{y} = \sum_{k=1}^{K} \mathbf{w}_k \odot \mathbf{x}_k$



Multi-Exposure Image Fusion by MEF-Net



- Key technical feature: Predict downsampled weight maps, followed by guided filtering for upsampling
  - MSE not applicable; MEF structural similarity (MEF-SSIM) index as the objective function
  - Perceptually optimized

K. Ma, Z. Duanmu, H. Zhu, Y. Fang, and Z. Wang, "Deep Guided Learning for Fast Multi-Exposure Image Fusion," *IEEE TIP*, 2020.

Outline

1 Introduction

2 Denoising

3 Deblurring

4 Super-resolution

5 Compression

6 Colorization

7 Fusion

Full-Reference Image Quality Metrics as Loss Functions

- Mean Squared Error (MSE)
- Structural Similarity Index (SSIM)
- Learned Perceptual Image Patch Similarity (LPIPS)
- Deep Image Structure and Texture Similarity (DISTS)
- ...

Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\mathbf{x}}_i)^2$$

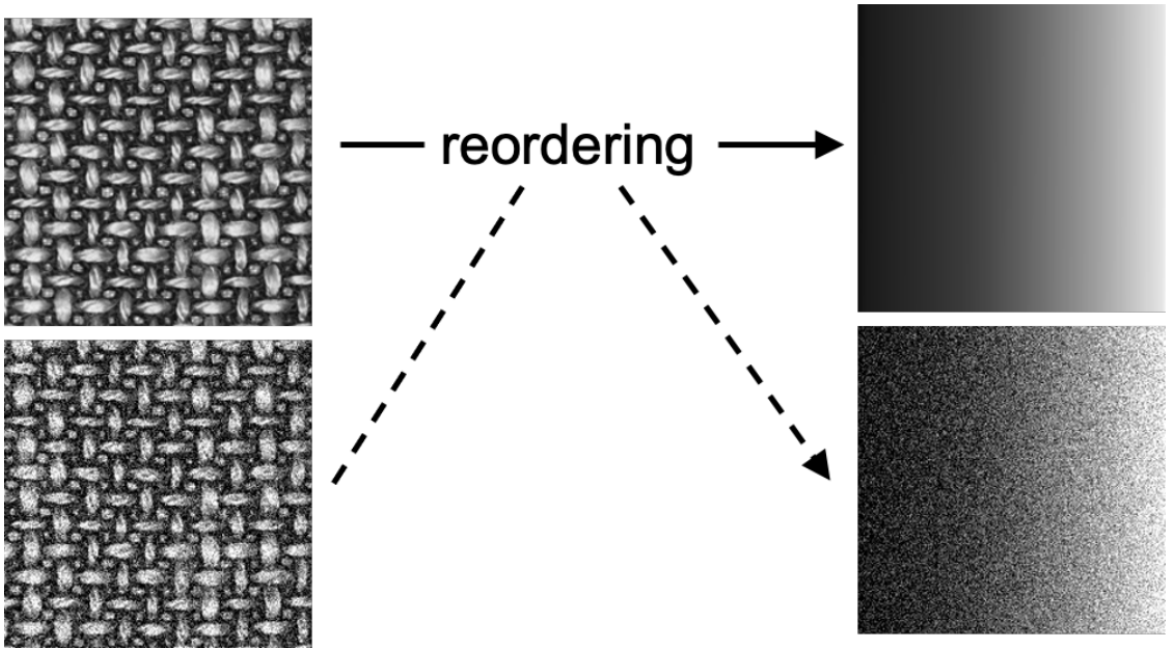
- Target signal: $\mathbf{x} \in \mathbb{R}^N$
- Estimated signal: $\hat{\mathbf{x}} \in \mathbb{R}^N$

Why Do We Love MSE?

- It is simple
  - Parameter-free, cheap to compute, and memory-less
- It is a valid distance metric
  - Satisfying nonnegativity, identity, symmetry, and triangular inequality
- It has a clear physical meaning as energy
- It is an excellent metric in the context of optimization
  - Convex, differentiable, which often admits closed-form analytical solutions

What is Wrong with MSE?

- Do not care about pixel ordering

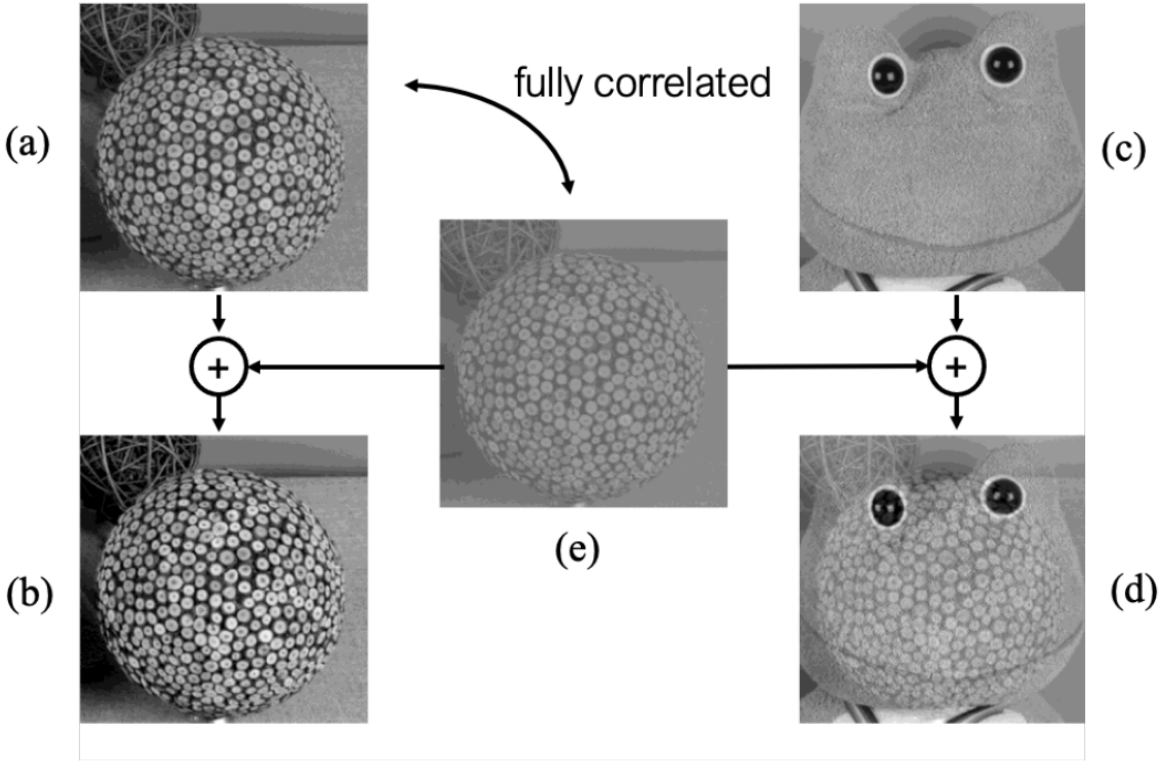


MSE = 1600, SSIM = 0.637

MSE = 1600, SSIM = 0.042

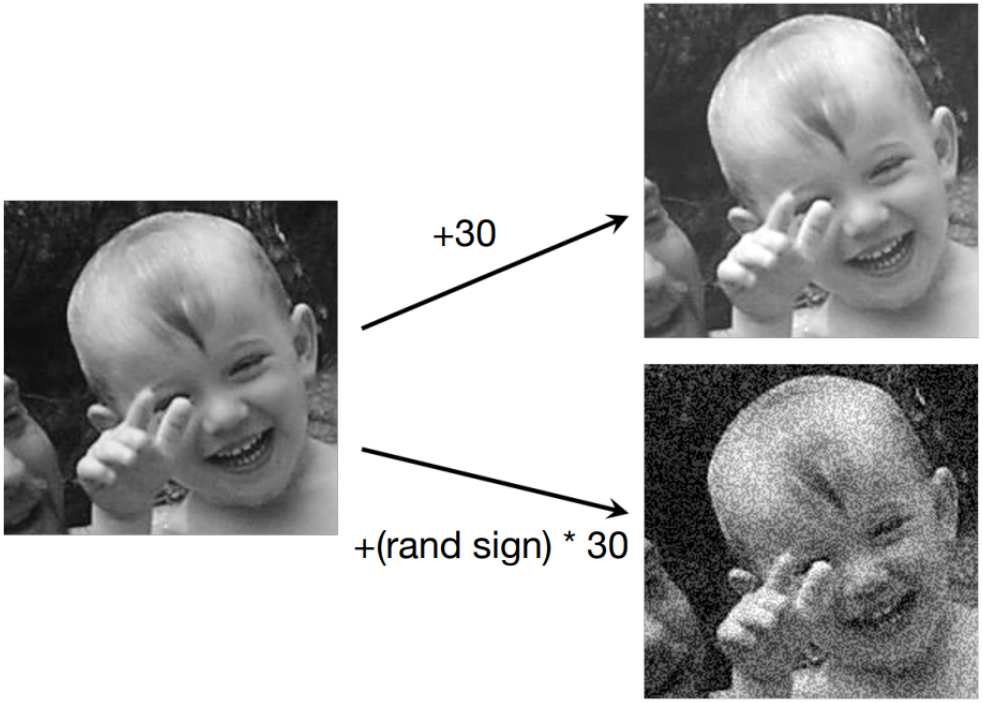
What is Wrong with MSE?

- Care about pixel difference, not the underlying signal



What is Wrong with MSE?

- Do not care about the sign of pixel difference



MSE = 900

SSIM = 0.933

MSE = 900

SSIM = 0.247

What is Wrong with MSE?

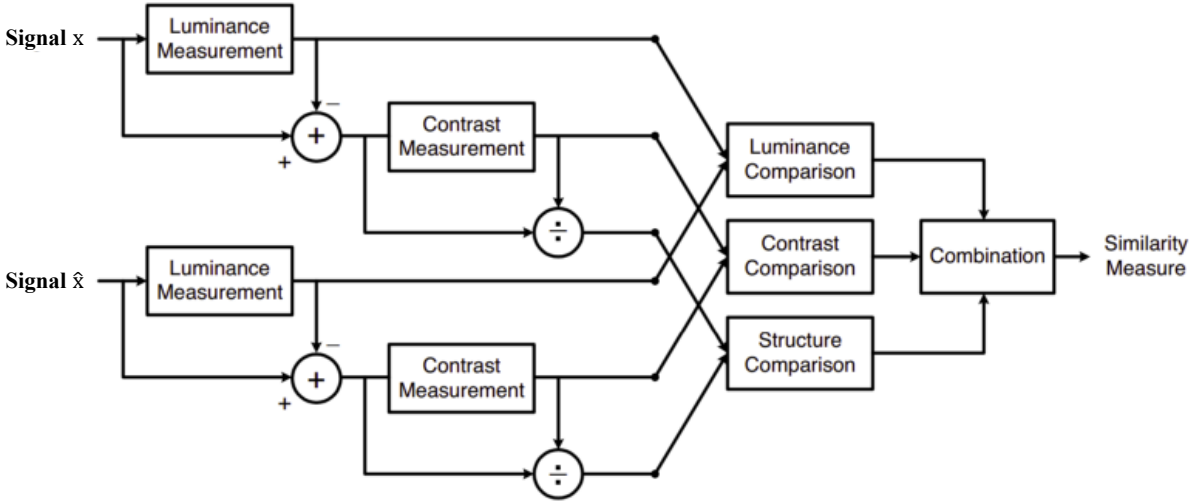
- MSE implicitly assumes that errors are statistically independent
  - True, if spatial dependencies are eliminated prior to computation
  - No easy task as natural images are highly structured (i.e., spatially correlated)
- Possible solution?
  - Learn a "perceptual" transform $\mathbf{f}$:

$$D(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{f}(\mathbf{x})_i - \mathbf{f}(\hat{\mathbf{x}})_i\right)^2$$

- What are the desirable properties of $\mathbf{f}$?

Structural Similarity Index (SSIM)

$$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = l(\mathbf{x}, \hat{\mathbf{x}}) \cdot s(\mathbf{x}, \hat{\mathbf{x}}) = \left(\frac{2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1}{\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + c_1} \right) \cdot \left(\frac{2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + c_2}{\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + c_2} \right)$$

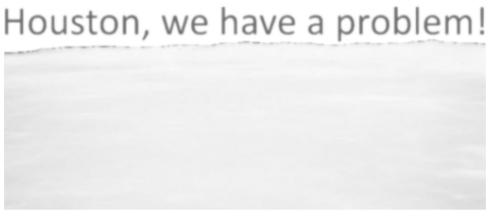


What is Wrong with SSIM?

- Normalization is sensitive to low intensities

Original image

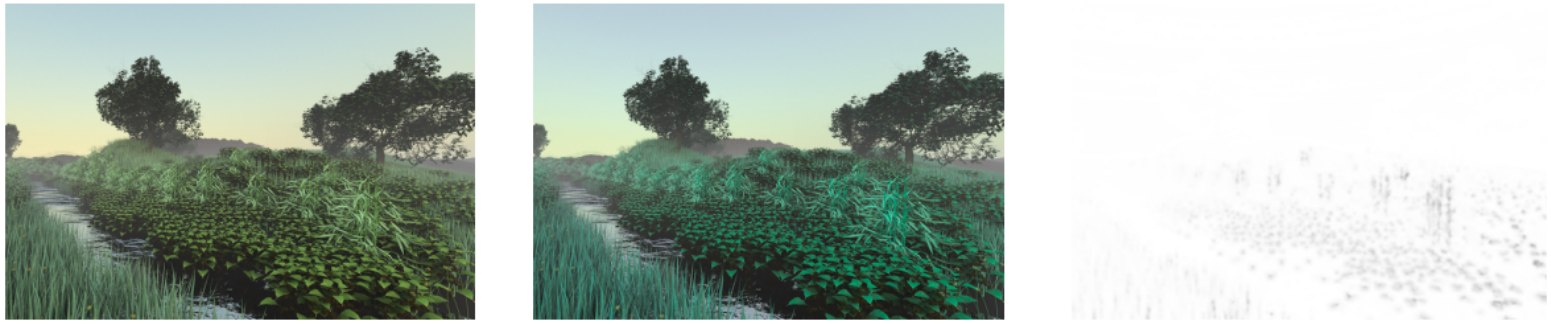
Distorted image



SSIM map

What is Wrong with SSIM?

- Do not consider chrominance



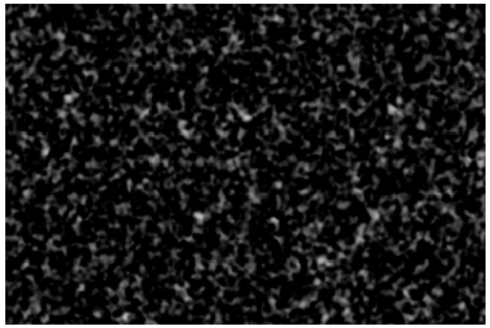
Original image | Distorted image | SSIM map

What is Wrong with SSIM?

- Rely on patch-by-patch comparison

Original image

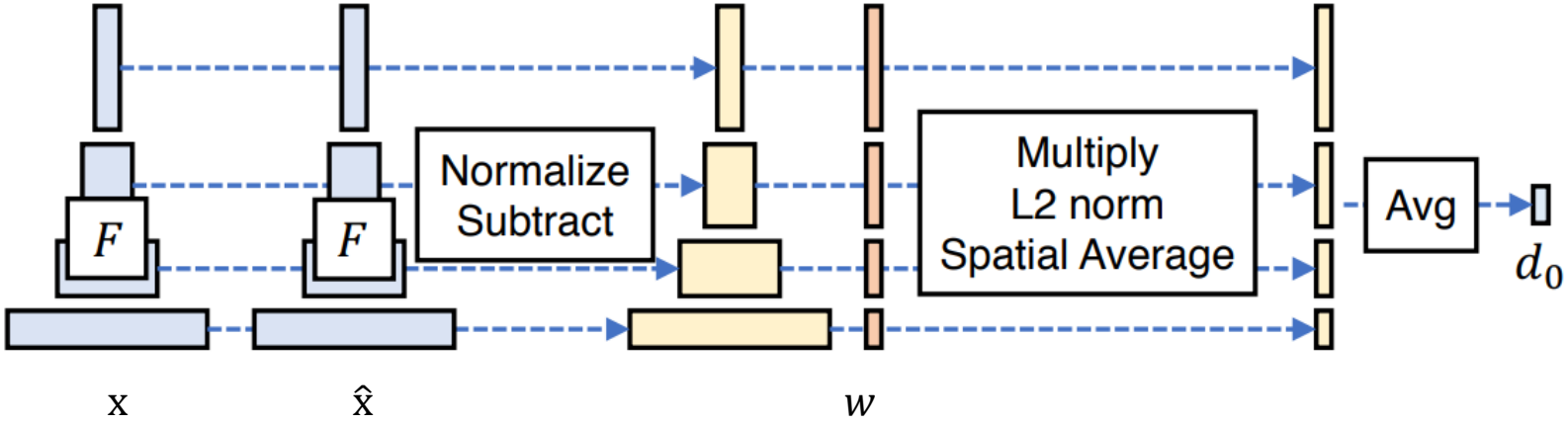
Distorted image



SSIM map

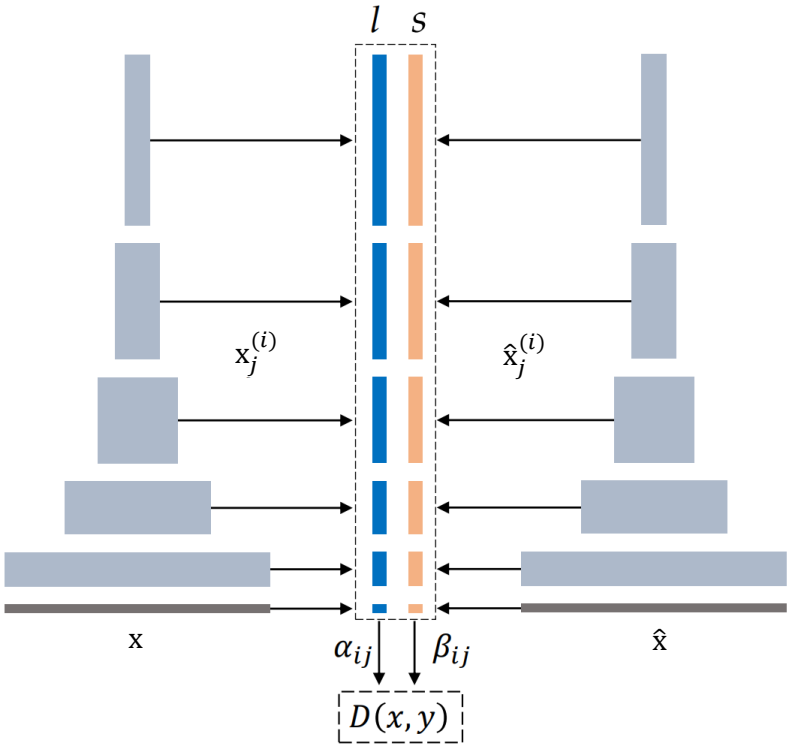
Learned Perceptual Image Patch Similarity (LPIPS)

$$\mathrm{LPIPS}\left(\mathbf{x}, \hat{\mathbf{x}}\right) = \sum_{i=1}^{S} \sum_{j=1}^{N_i} w_{ij} \, \mathrm{MSE}\left(\mathbf{f}(\mathbf{x})_j^{(i)}, \mathbf{f}(\hat{\mathbf{x}})_j^{(i)}\right)$$



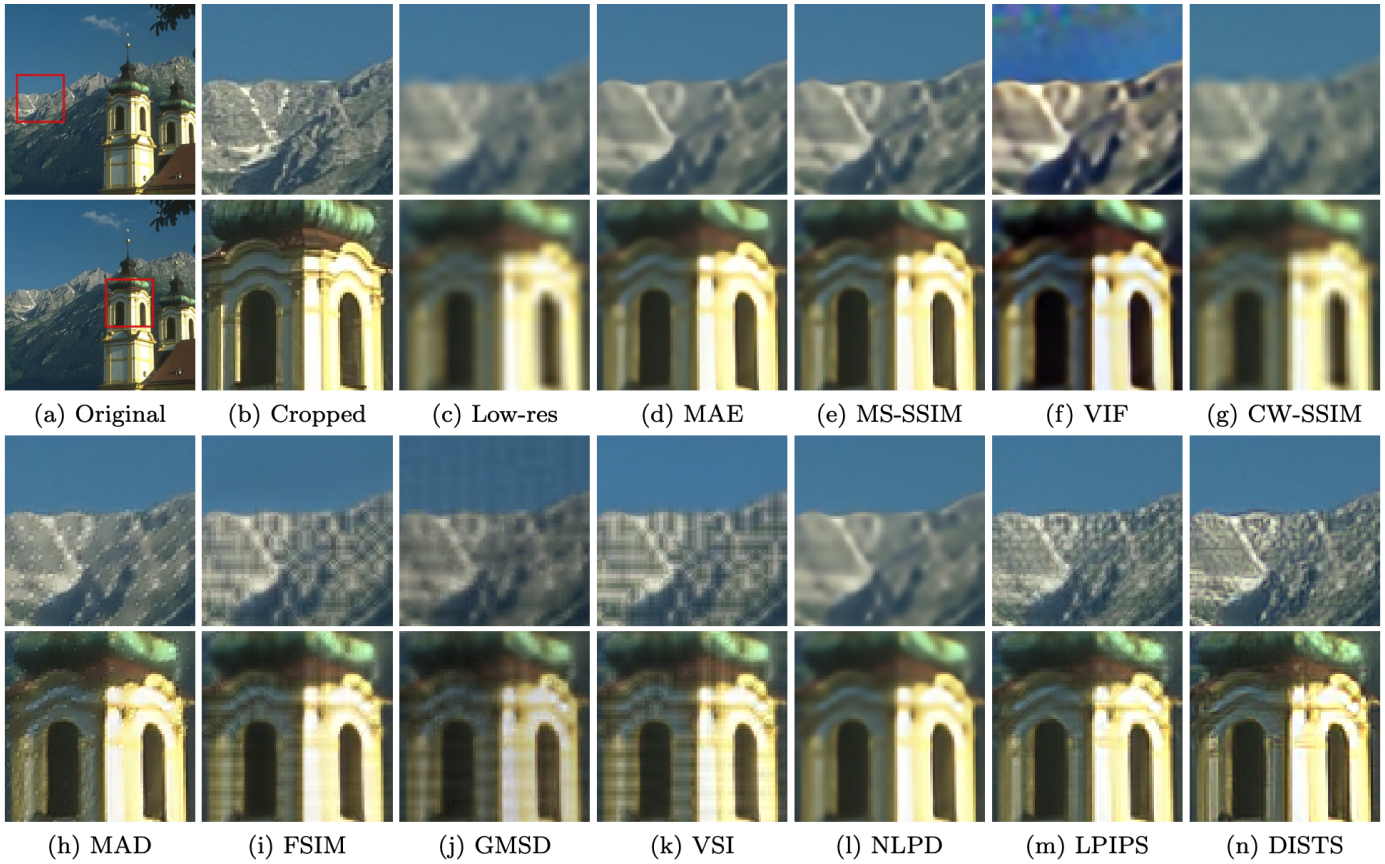
Deep Image Structure and Texture Similarity (DISTS)

$$\mathrm{DISTS}(\mathbf{x}, \hat{\mathbf{x}}) = 1 - \sum_{i=0}^{S} \sum_{j=1}^{N_i} \left(\alpha_{ij} l \left(\mathbf{x}_j^{(i)}, \hat{\mathbf{x}}_j^{(i)} \right) + \beta_{ij} s \left(\mathbf{x}_j^{(i)}, \hat{\mathbf{x}}_j^{(i)} \right) \right)$$



Comparison of Loss Functions using Super-Resolution

- Super-resolution results optimized for different IQA models



K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems," *IJCV*, 2021.

Summary

- Advances of deep learning has been driven by the ImageNet competition
- As depth increases, need to have a smart architecture design to make training more effective
- Deep learning also begins to dominate low-level vision, e.g., image denoising, deburring, super-resolution, compression etc.