# *Image Quality Assessment and Perceptual Optimization*

## No-reference IQA via Non-local Modeling

**Shuyue Jia**
M.Phil. Student
April 2023

innovation you

# Learning Objectives

- **What is** Objective Image Quality Assessment (IQA)?
- **Why do we need** IQA with Perceptual Optimization?
- Current **Related Work**
- A **Proposed Non-local Modeling method** for IQA

Note: Subjective IQA: Image quality collected from human subjects (observers)

# What is Image Quality Assessment (IQA)? Synthetic Distortion

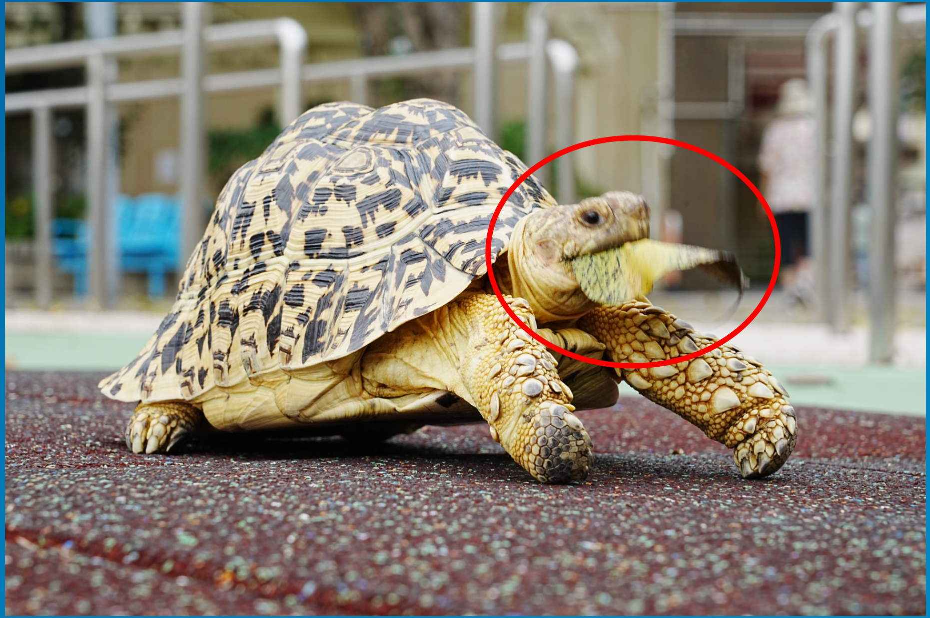
Reference/Pristine Image

Distorted Image
by Gaussian Noise

Image Credit: Shuyue Jia and Ka-Po Chan

# What is Image Quality Assessment (IQA)? Authentic Distortion

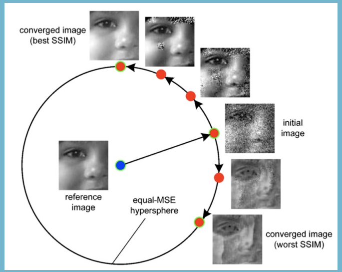
Reference/Pristine Image



Motion Blur
due to low shutter speed

Image Credit: Shuyue Jia and Ka-Po Chan

# Problem Definition




## *Definitions*

- *Natural Image*: images captured by optical cameras
- *Fidelity:* keep the content of the distorted images (semantic information) unchanged
- *Image Quality (Fidelity) Assessment*: measure the input image's visual (perceptual) quality
- *Visual Quality and Perceptual Optimization*: people's overall subjective visual experience when viewing images
- *Synthetic Distortion*: synthetic distortions added to the whole area of image (mainly global uniform distortions)
- *Authentic Distortion*: images captured in the wild include varies contents and diverse types of distortions (global uniform distortions + non-uniform distortions in local areas)

## *Image Quality Assessment Category*

- *Full-Reference IQA*: with Reference/Pristine Image
- *Reduced-Reference IQA*: with partial information from Reference Image, *e.g.*, a subset of features
- *No-Reference (Blind) IQA*: *without any information from* Reference Image

## *Measurements*

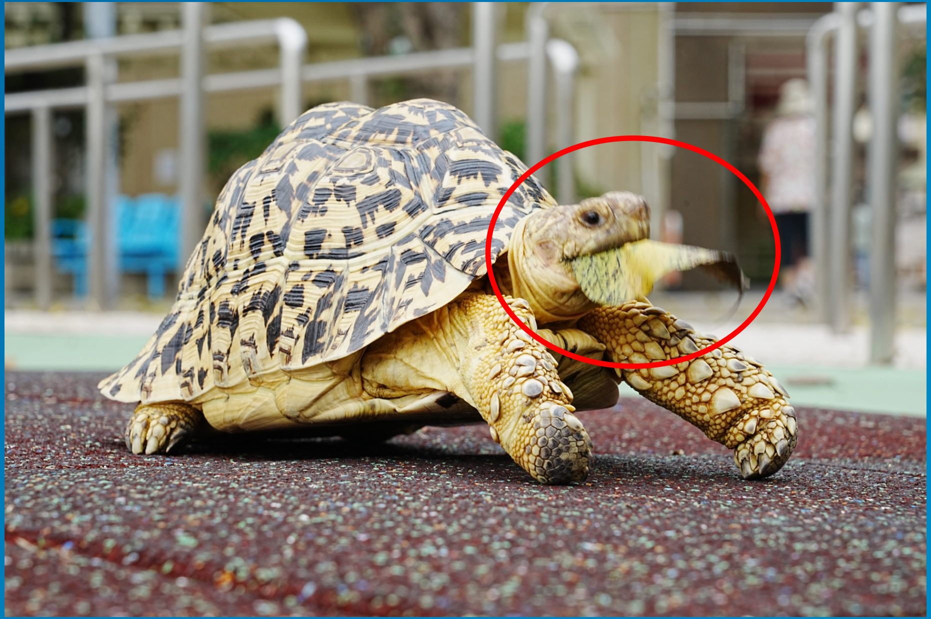
- *Label:* Mean Opinion Score (MOS) *vs. Model Output:* one scalar score
- *Pearson Linear Correlation Coefficient (PLCC)*: prediction accuracy
- *Spearman Rank-order Correlation Coefficient (SRCC)*: prediction monotonicity

Image Credit: Prof. Zhou Wang

# Why do we need Image Quality Assessment?

*"If you can't measure it, you can't improve it."* (Peter Drucker)

Reference/Pristine Image



Motion Blur
due to low shutter speed

Image Credit: Shuyue Jia and Ka-Po Chan

# Why do we need Image Quality Assessment?

## Automatic Image Quality Assessment

Reference Image

Distorted Image

IQA Model

Image Quality

Image Credit: TID2013 Database

# Why do we need Image Quality Assessment?

## Performance Evaluation of Image Processing Systems

## by a Full-Reference IQA Method

Distorted Image

Model Input

Image Restoration System

De-blurred Image

Model Output

Evaluation

MSE
PSNR
SSIM
LPIPS
etc.

Reference Image

Label

FR-IQA Method

Image Credit: TID2013 Database

The picture can't be displayed.

Tools | Predictive Coding Based Multiscale Network with EncoderDecoder LSTM for Video Prediction

130% Annotate Edit Scan & OCR Export

Conv-TT-LSTM 

Fig. 6. Visualization examples on the KTH datasets. We use 10 frames as input to predict next 30 frames. The other results are obtained from [49]. Zoom in for a better view.

### B. *Experimental results*

**KTH** The KTH is one of the most commonly used datasets for task of video prediction. It is very popular due to its moderate complexity of scene and event (the dataset contains only 6 action categories on a simple backgrounds). Similar to previous works, we use person 1-16 for training and person 17-25 for testing, and use 10 frames as input to predict the next 30 frames. The quantitative evaluation results are shown in Table II. The results of previous works are excerpted from [25], [48] and [49]. As shown in the table, our method achieve comparable or better performance compared with the state-of-the-art methods. In addition, our model converges fast, which only takes about 30 epochs to achieve good performance. Figure 6 shows the visualization examples, our method can also achieve good visual results, while the Conv-TT-LSTM [50], which obtains the highest SSIM score, shows poor performance on the qualitative evaluation. The mismatch between quantitative and qualitative assessments remains an unsolved problem for video prediction tasks.

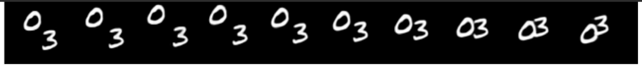


Fig. 7. Visualization examples on the MNIST datasets. We use 10 frames as input to predict next 10 frames. In each group, the first row indicates the ground truth frames and the second row indicates predicted frames. Zoom in for a better view.

**Moving MNIST** The Moving MNIST is an early popular synthetic dataset for video representation learning. Its scenarios and events are very simple, each sequence is 20 in length and shows how 2 digits move at a constant speed and bounce within a $64 \times 64$ box. Similarly, we use 10 frames as input to predict the next 10 frames. Table III shows the quantitative evaluation results on SSIM and MSE, the results of previous works are excerpted from [57] and [48]. Figure 7 shows the visualization examples, we have tried our best to search the demonstrated examples of previous works, but unfortunately none have been found, so we only present our results in the figure. Since the scenes and events are simple, we did not use adversarial training to sharpen the generated images to save training overhead. Nevertheless, we can still achieve good performance from the perspective of qualitative evaluation. Actually, most of the background pixels in this dataset are zeros. If we set a threshold and change the predicted pixels smaller than the threshold to zeros, we will get higher accuracy

Credit: Ling *et al.*, Predictive Coding Based Multiscale Network with Encoder-Decoder LSTM for Video Prediction, Under Review

# Why do we need Image Quality Assessment?

## Optimizing Image Processing Systems (Model Parameter Optimization)

## by a Full-Reference IQA Method

Distorted Image

Model Input

Image Restoration System

De-blurred Image

Model Output

Loss

Reference Image

Label

FR-IQA Method

Image Credit: TID2013 Database

# Current Related Work of Full-Reference IQA

- Signal Fidelity Approaches

Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR)

$I_p$: number of pixels in the image; $x_i$ and $y_i$ are the $i^{\text{th}}$ pixels of the ref. and dis.

$$\mathrm{MSE} = \frac{1}{I_p}\sum_{i=1}^{I_p}(x_i - y_i)^2 , \qquad \mathrm{PSNR} = 10\times \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right).$$

- Bottom-Up Approaches (Error Sensitivity Framework)

separately model each basic module of Human Visual System (HVS)

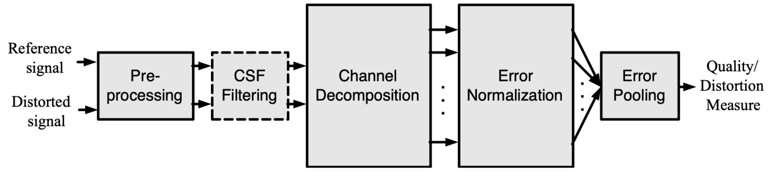



Fig. 1. A prototypical quality assessment system based on error sensitivity. Note that the CSF feature can be implemented either as a separate stage (as shown) or within "Error Normalization."

Image Credit: SSIM Paper (Prof. Zhou Wang)

# Current Related Work of Full-Reference IQA

- Top-Down Approaches

  directly imitate the function of HVS as a single model

  Representative Work:

  (1) Structural Similarity (**SSIM**)

  (2) Visual Information Fidelity (**VIF**)

  (3) Learned Perceptual Image Patch Similarity (**LPIPS**)

# Current Related Work of Full-Reference IQA

- Top-Down Approaches - Structural Similarity (SSIM)

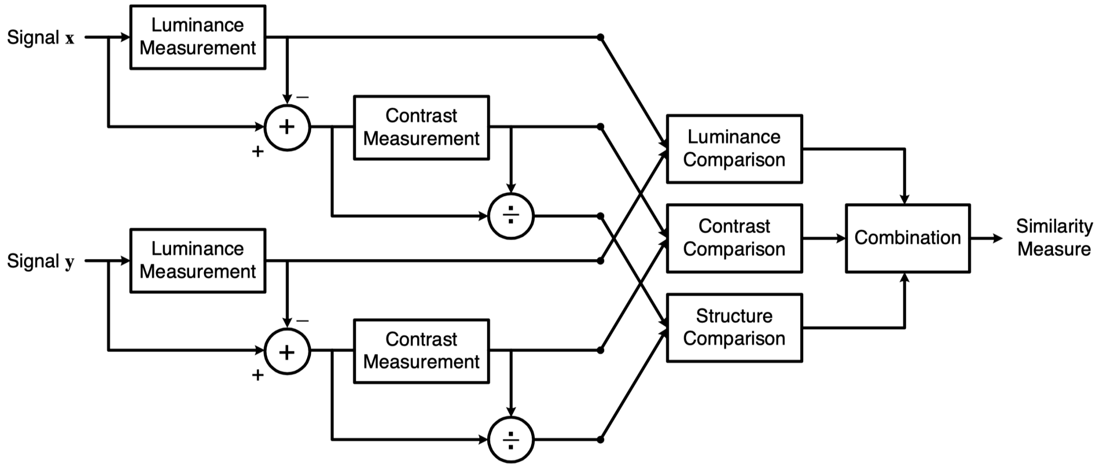



Fig. 3. Diagram of the structural similarity (SSIM) measurement system.

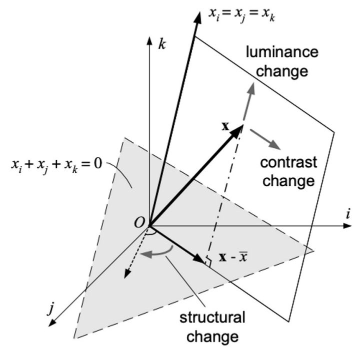



**FIGURE 5** Separation of luminance, contrast and structural changes from a reference image **x** in the image space. This is an illustration in three-dimensional space. In practice, the number of dimensions is equal to the number of image pixels.

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

$\mu$: Mean Intensity → luminance

$\sigma$: Standard Deviation → contrast

$(\mathbf{x} - \mu_x)/\sigma_x$: normalization

Correlation of Normalized Signals → structure

Credit: Wang *et al.*, Image Quality Assessment: From Error Visibility to Structural Similarity, In IEEE T-IP’04

# Current Related Work of Full-Reference IQA

- Top-Down Approaches - Visual Information Fidelity (VIF)

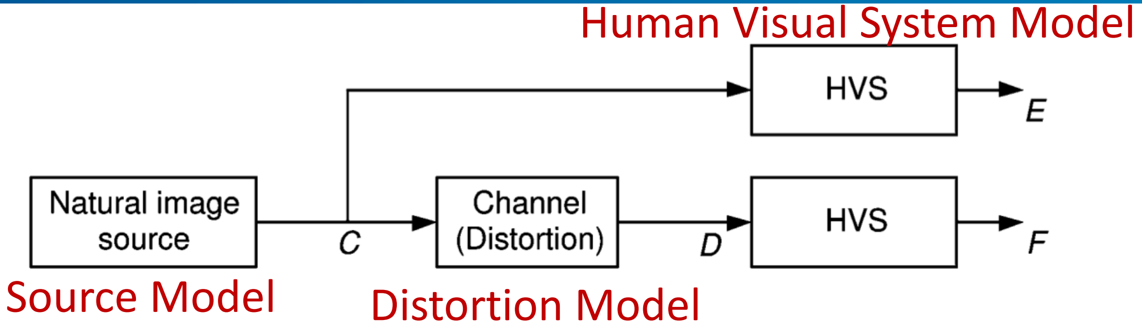



Fig. 1. Mutual information between $\mathcal{C}$ and $\mathcal{E}$ quantifies the information that the brain could ideally extract from the reference image, whereas the mutual information between $\mathcal{C}$ and $\mathcal{F}$ quantifies the corresponding information that could be extracted from the test image.

Credit: Sheikh *et al.*, Image Information and Visual Quality, In IEEE T-IP'06 and Netflix VMAF System (Prof. Jay Kuo, USC)

# Current Related Work of Full-Reference IQA

- **Top-Down Approaches** - Learned Perceptual Image Patch Similarity (**LPIPS**)

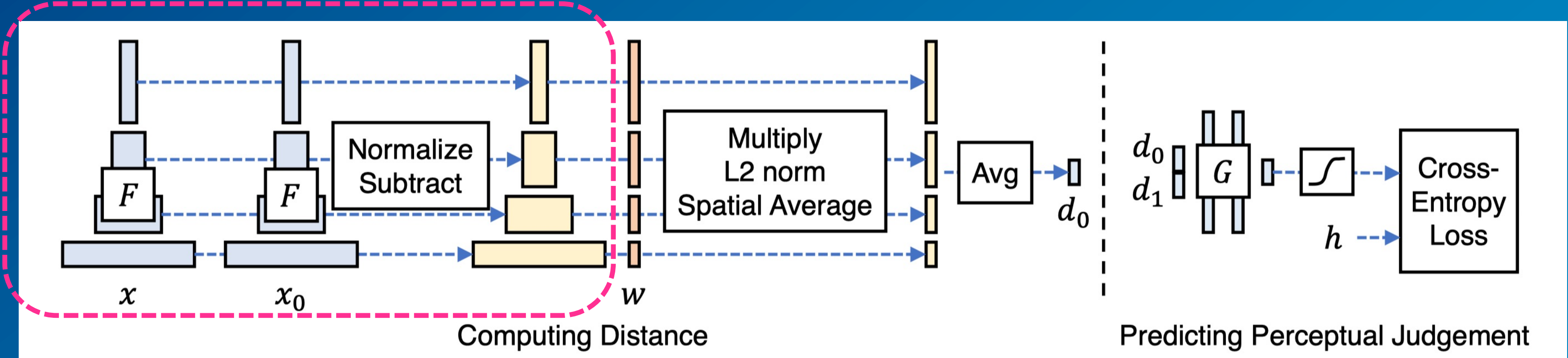



Figure 3: **Computing distance from a network** (Left) To compute a distance $d_0$ between two patches, $x$, $x_0$, given a network $\mathcal{F}$, we first compute deep embeddings, normalize the activations in the channel dimension, scale each channel by vector $w$, and take the $\ell_2$ distance. We then average across spatial dimension and across all layers. (Right) A small network $\mathcal{G}$ is trained to predict perceptual judgment $h$ from distance pair $(d_0, d_1)$.

Based on Deep Features instead of Statistics

Credit: Zhang *et al.*, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, In CVPR'18

Tools [DISTS] Image Quality Assessment- Unifying Structure and Texture Similarity

130% Annotate Edit Scan & OCR Export

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2020.3045810, IEEE Transactions on Pattern Analysis and Machine Intelligence



# Image Quality Assessment: Unifying Structure and Texture Similarity

Keyan Ding, Kede Ma, *Member, IEEE,* Shiqi Wang, *Member, IEEE,* and Eero P. Simoncelli, *Fellow, IEEE*

**Abstract**—Objective measures of image quality generally operate by making local comparisons of pixels of a "degrade" image to those of the original. Relative to human observers, these measures are overly sensitive to resampling of texture regions (*e.g.*, replacing one patch of grass with another). Here we develop the first full-reference image quality model with explicit tolerance to texture resampling. Using a convolutional neural network, we construct an injective and differentiable function that transforms images to multi-scale overcomplete representations. We empirically show that the spatial averages of the feature maps in this representation capture texture appearance, in that they provide a set of sufficient statistical constraints to synthesize a wide variety of texture patterns. We then describe an image quality method that combines correlation of these spatial averages ("texture similarity") with correlation of the feature maps ("structure similarity"). The parameters of the proposed measure are jointly optimized to match human ratings of image quality, while minimizing the reported distances between subimages cropped from the same texture images. Experiments show that the optimized method explains human perceptual scores, both on conventional image quality databases, as well as on texture databases. The measure also offers competitive performance on related tasks such as texture classification and retrieval. Finally, we show that our method is relatively insensitive to geometric transformations (*e.g.*, translation and dilation), without use of any specialized training or data augmentation. Code is available at https://github.com/dingkeyan93/DISTS.

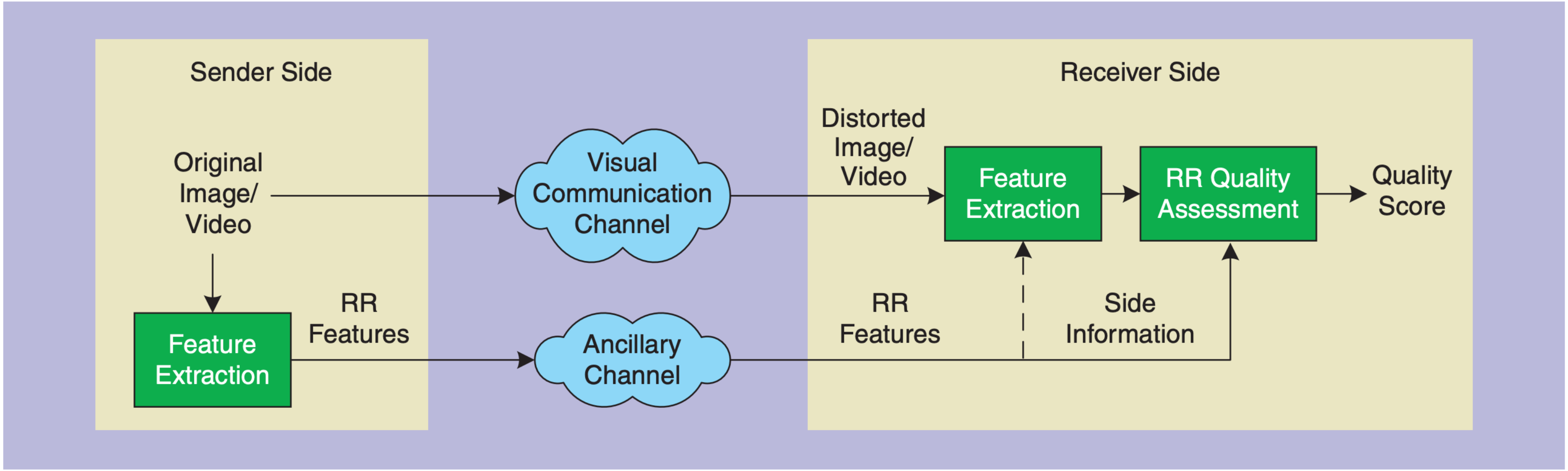
**Index Terms**—Image quality assessment, structure similarity, texture similarity, perceptual optimization.

The picture can't be displayed

Paper Link

Extension: Ding *et al.*, Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems, In IJCV'21

# Framework of Reduced-Reference IQA




[FIG2] **General framework of an RR image or image QA system.**

Credit: Wang *et al.*, Reduced- and No-Reference Image Quality Assessment, In IEEE Signal Processing Magazine'11

# Current Related Work of No-Reference / Blind IQA

- **Distortion-Specific Modeling**

  aware the image distortion types → build distortion-specific models

- **General NR-IQA Modeling**

  **(1) Natural Scene Statistics Modeling**

  Spatial Domain and Transform Domain

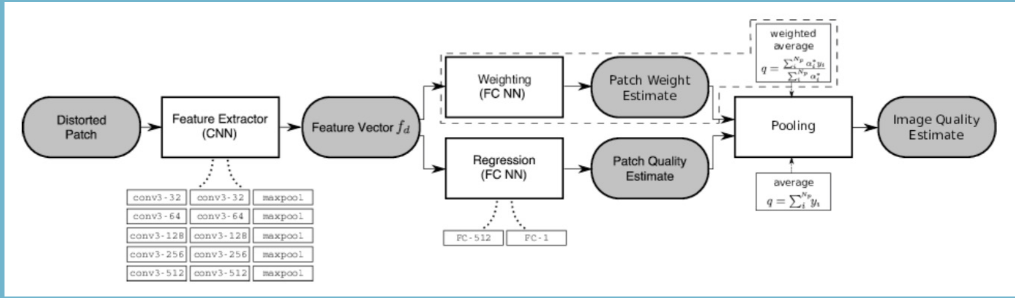
  **(2) Human Visual System Modeling**

  CNN modeling methods, assisted with visual importance information, reference images‘ information during training, ranking-based methods, graph representation learning, etc.

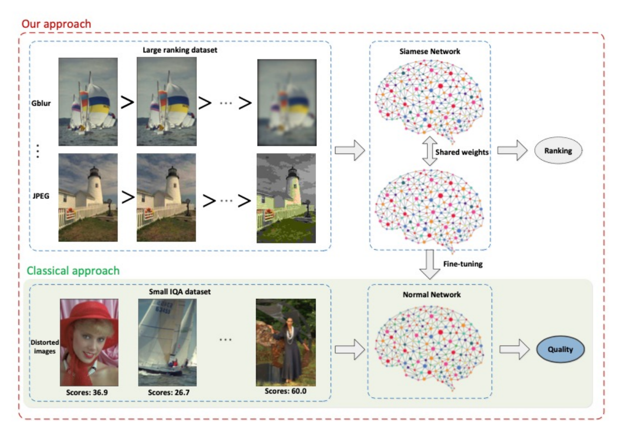
  **(3) Codebook-based Modeling**

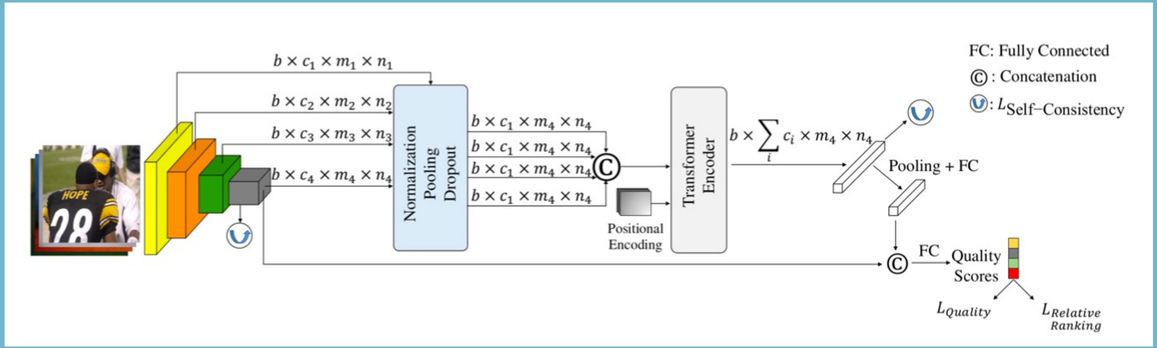
  constructing a codebook

# Selected Recent Progress on No-reference IQA




CNN-based Methods [1]




Ranking-based Methods [2]

Transformer-based Methods [3]

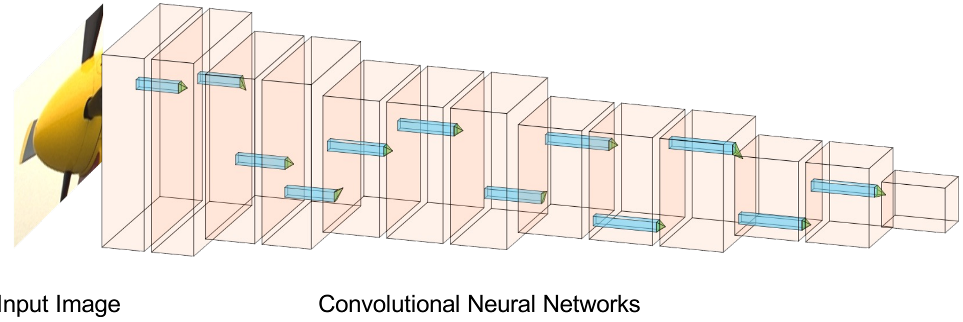
Credit:


[1] Bosse *et al.*, Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment, In IEEE TIP'18

[2] Liu *et al.*, RankIQA: Learning from Rankings for No-reference Image Quality Assessment, In ICCV'17

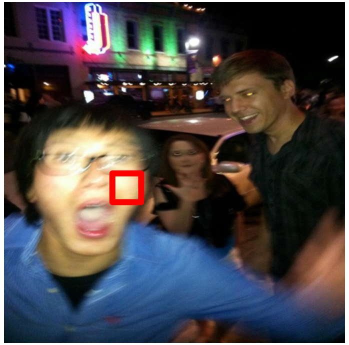
[3] Golestaneh *et al.*, No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency, In WACV'22

# Challenges

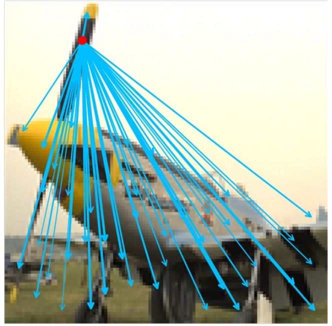



- Local Modeling (Convolutional Neural Networks):
  - ✓ Translation Invariance (Pooling)
  - ✓ Translation Equivalence (Convolution)
  - ✓ Sharable Fewer Parameters (Weight Sharing)
- Limitations:
  - ✓ Small-sized Receptive Field → **Extracted features are too local**
  - ✓ Parameters Fixed across the whole image → **Image content is equally treated**
  - ✓ Lack of Geometric and Relational Modeling → **Missing complex relations and dependencies**

# Motivation



Local Feature
Extraction



Non-local
Dependency

- ✓ HVS is adaptive to the local content

  → ***Local feature extraction*** via a pre-trained CNN

- ✓ HVS perceives image quality with long-range dependency constructed among different regions

  → ***Non-local feature extraction*** for long-range dependency and relational modeling

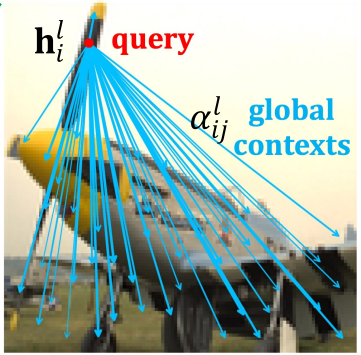


Image Credit: LIVEC and TID2013 Databases

# Definition

**Non-Local:**
**Object-to-Pixel Modeling**


(a)
Local feature extraction is critical




(b)
Non-local dependency learned by the NLNet

Figure 2: Local region feature extraction and non-local dependency feature extraction

**Convolution:**
**Pixel-to-Pixel Modeling**

**Spatial Integration of Information**

$$\mathbf{h}_i^l = \mathrm{ELU}\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^l \mathbf{W}^l \mathbf{h}_j^l\right)$$

**Spatial Weighting Functions**

$$\alpha_{ij}^l = \frac{\exp(\mathrm{a}_{ij}^l)}{\sum_{k\in\mathcal{N}(i)} \mathrm{a}_{ik}^l}$$

$$\mathrm{a}_{ij}^l = \mathrm{LeakyReLU}(\mathrm{FC}([\mathbf{W}^l\mathbf{h}_i^l \parallel \mathbf{W}^l\mathbf{h}_j^l]))$$

- ✓ **Local Modeling**: encodes spatially proximate **Local Neighborhoods**.
- ✓ **Non-local Modeling**: establishes **Spatial Integration of Information** by **Long- and Short-Range Communications** with **different Spatial Weighting Functions**.



Image Credit: TID2013 and LIVEC Databases

# Non-local Behavior

Object-to-Pixel Modeling
Region Feature Extraction

**Non-local Dependency & Relational Modeling**

Semantics and Content Understanding

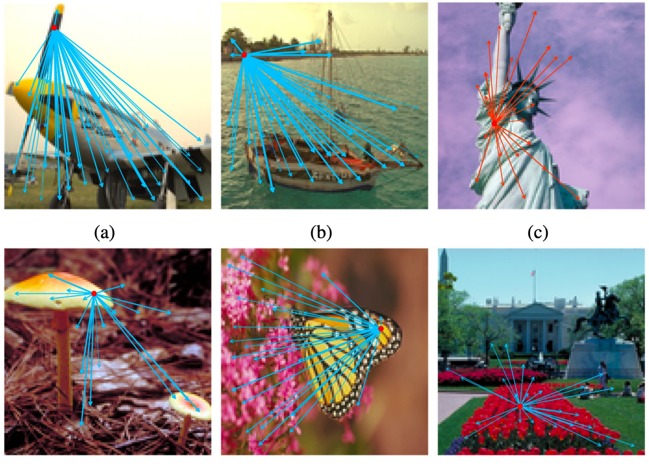



Figure 3.1: The non-local behavior of the long-range dependency and relational modeling. (a) The plane image with a query on wings. (b) The boat image with a query on nearby river bank. (c) The Statue of Liberty image with a query on the lady. (d) The shrooms image with a query on one shroom. (e) The butterfly image with a query on the wing. (f) The Lafayette Square, Washington, D.C. image with a query on flowers.

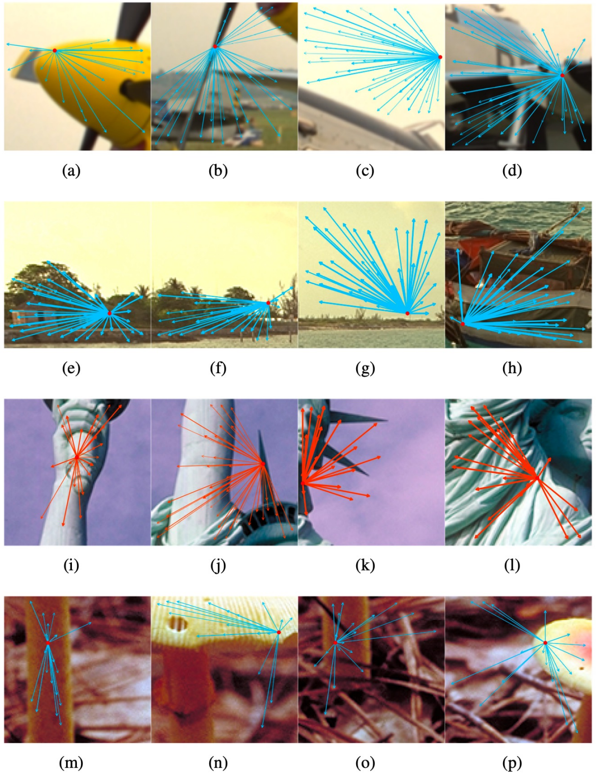



Figure 3.2: Selected demonstrations of the non-local behavior and long-range dependencies with regard to the cropped image patches from the illustrated images. The details of Figure (a) to (p) are described in the thesis.

✓ **Non-local Modeling**: establishes the **Spatial Integration of Information** by **Long- and Short-Range Communications** with **different Spatial Weighting Functions**.



Image Credit: TID2013 and LIVE Databases

# Definition

Local Distortion

Non-Local Recurrence


Figure 4.9: Demonstrations of the global distortions (b/f: GB, c/g: CC, d/h: PN) contaminating the Statue of Liberty and George Rogers Clark Memorial images. Figure (a) and Figure (e) are reference images from the CSIQ database.

Global Distortion

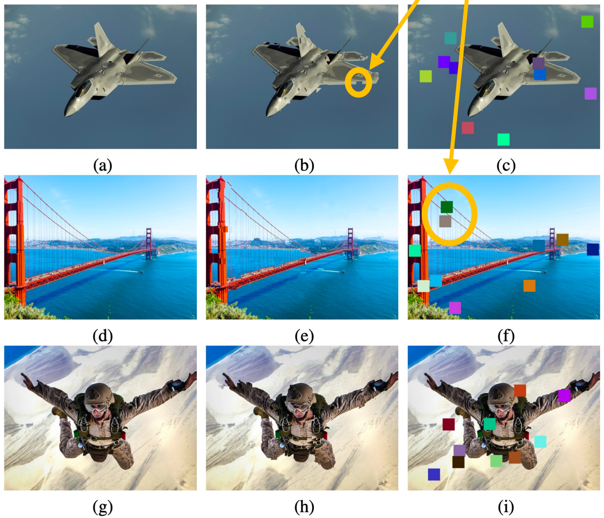



Figure 4.11: Demonstrations of the local distortions (b/e/h: non-eccentricity patch and c/f/i: color block). Figure (a), Figure (d), and Figure (g) are reference images from the KADID-10k database.

Local Distortion

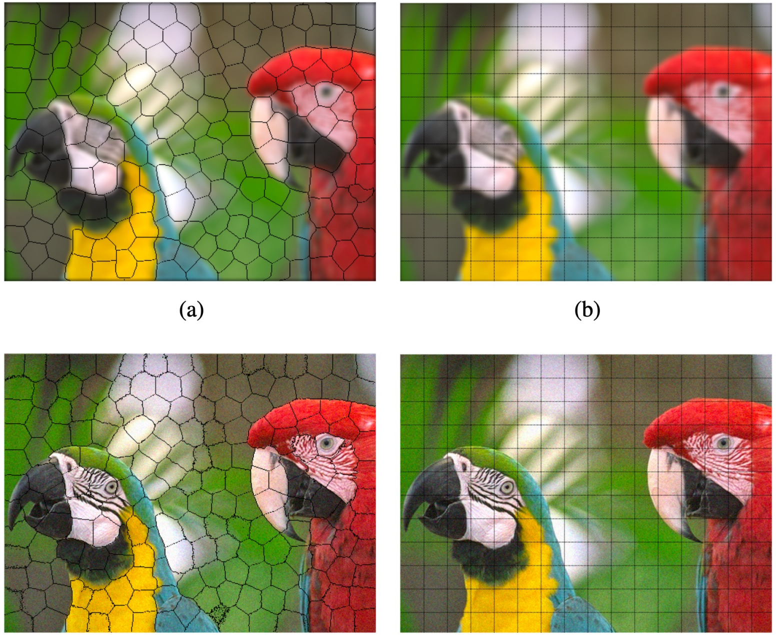
**Global Distortion**: **globally and uniformly distributed** distortions with **non-local recurrences** over the image.

**Local Distortion**: **local nonuniform-distributed** distortions **in a local region**.



Image Credit: CSIQ and KADID-10k Databases

# Superpixel Segmentation



(a) (b) (c) (d)

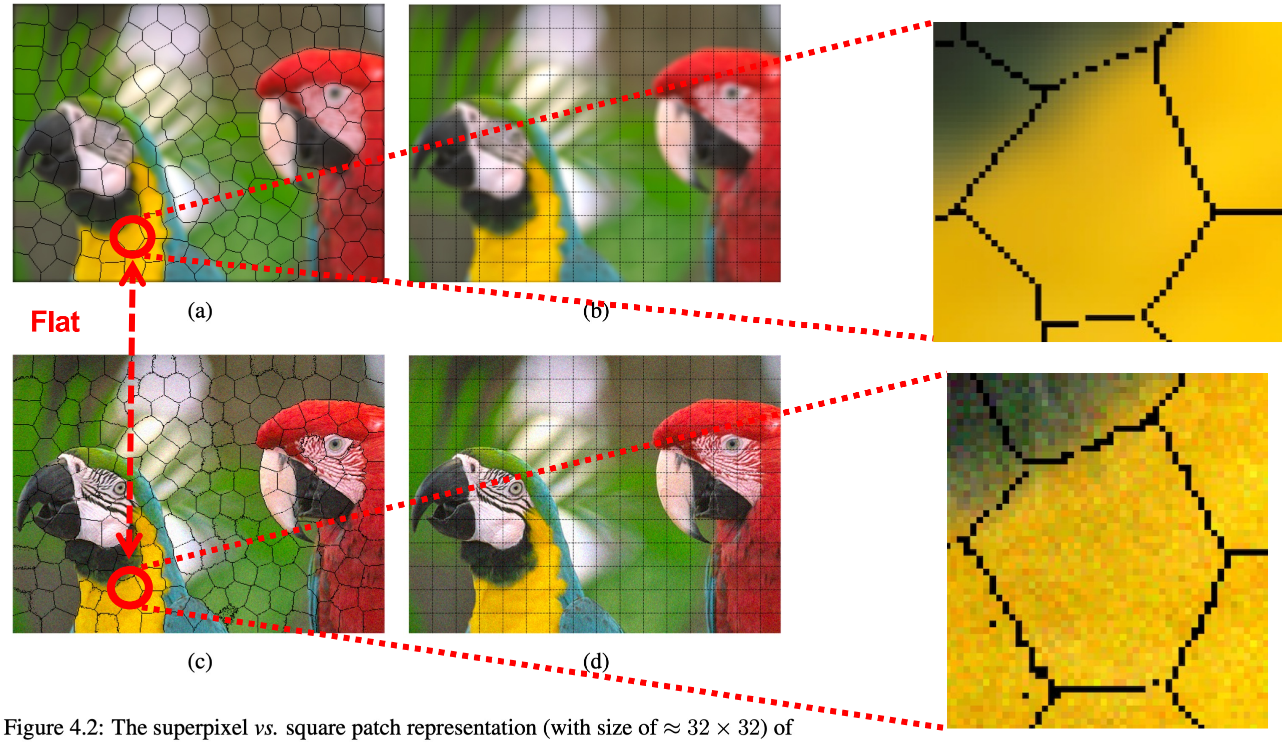
Figure 4.2: The superpixel *vs.* square patch representation (with size of $\approx 32 \times 32$) of the plane image from the TID2013 database.

Superpixel *vs.* Square Patch

- ✓ Adherence to boundaries and visually meaningful
- ✓ Accurate feature extraction

Image Credit: TID2013 Database

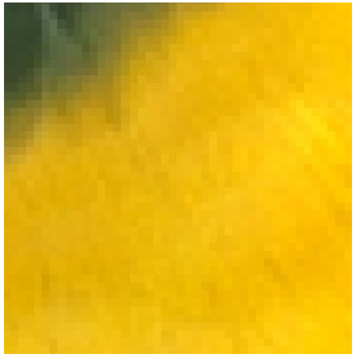
# Superpixel Segmentation




Gaussian Blur

Gaussian Noise

Figure 4.2: The superpixel *vs.* square patch representation (with size of $\approx 32 \times 32$) of the plane image from the TID2013 database.



Reference

Image Credit: TID2013 Database

# Superpixel Segmentation

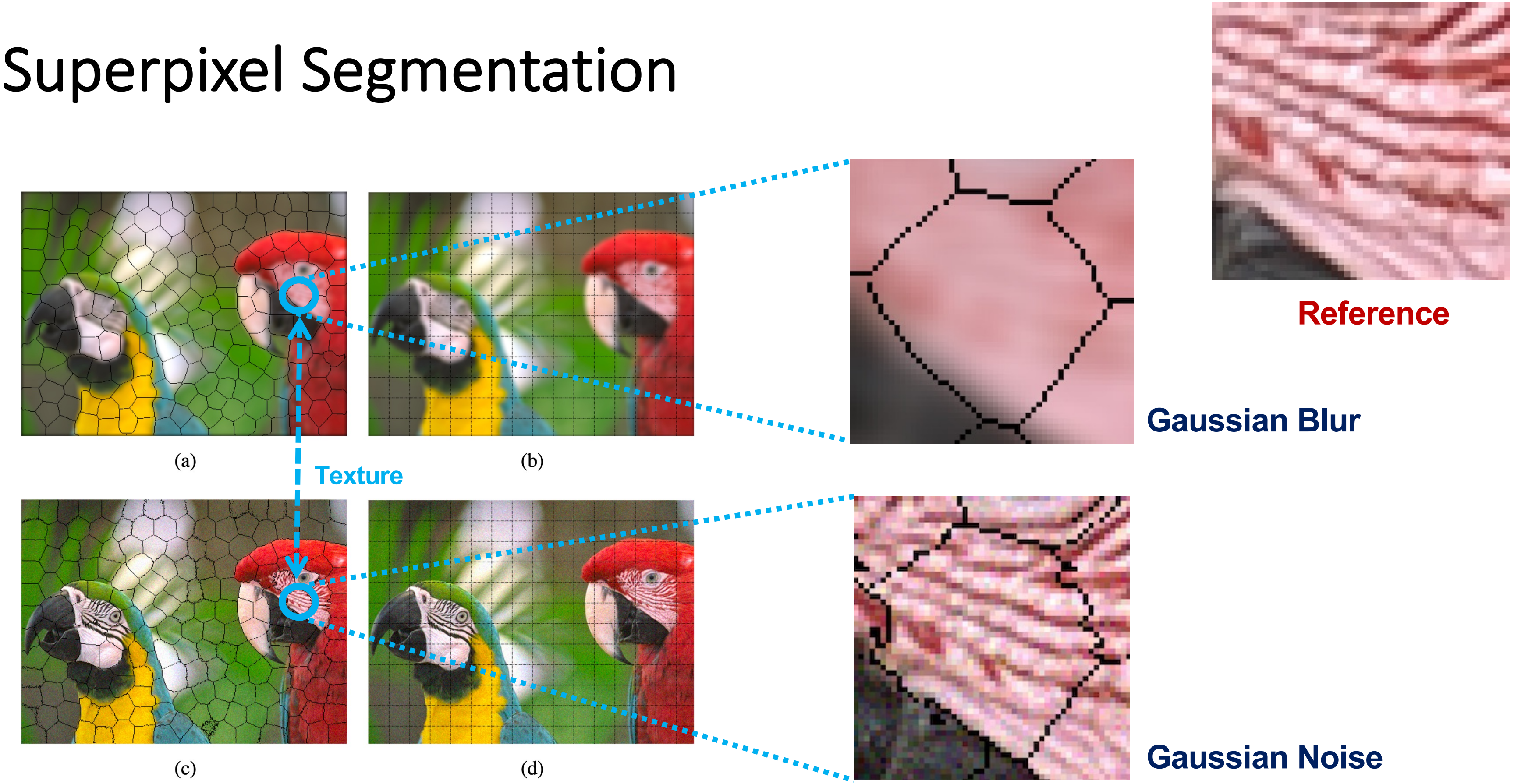



Figure 4.2: The superpixel *vs.* square patch representation (with size of $\approx 32 \times 32$) of the plane image from the TID2013 database.

Image Credit: TID2013 Database

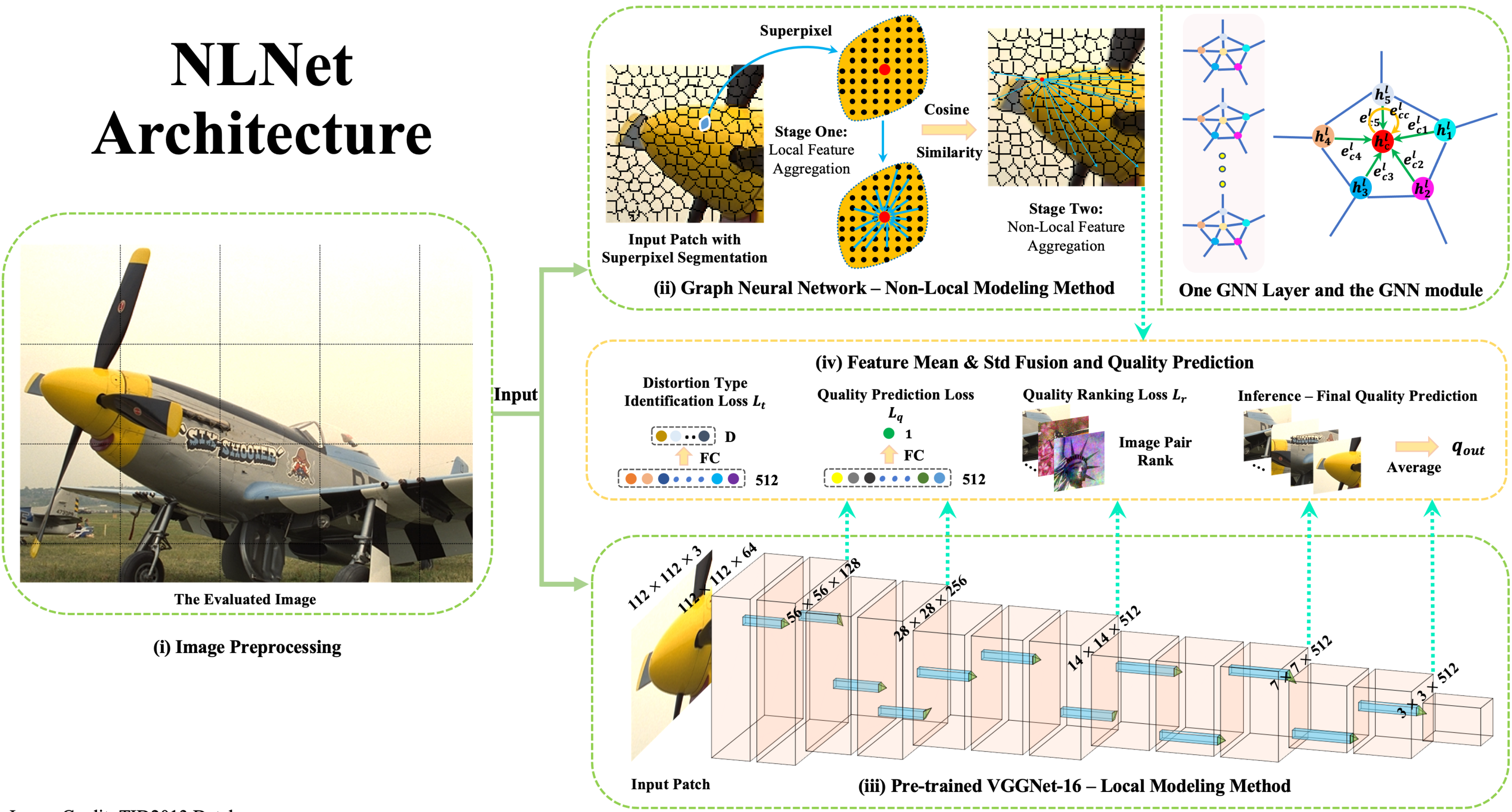



Image Credit: TID2013 Database

# Experimental Setup

**Databases**:

✓ LIVE, CSIQ, TID2013, and KADID-10k

**Experimental Settings**:

✓ Intra-Database Experiments:

→ 60% training, **20% validation**, and 20% testing, with `random` seeds from 1 to 10

→ The median SRCC and PLCC are reported.

✓ Cross-Database Evaluations:

→ One database as the training set, and the other databases as the testing set

→ Report the last epoch's performance

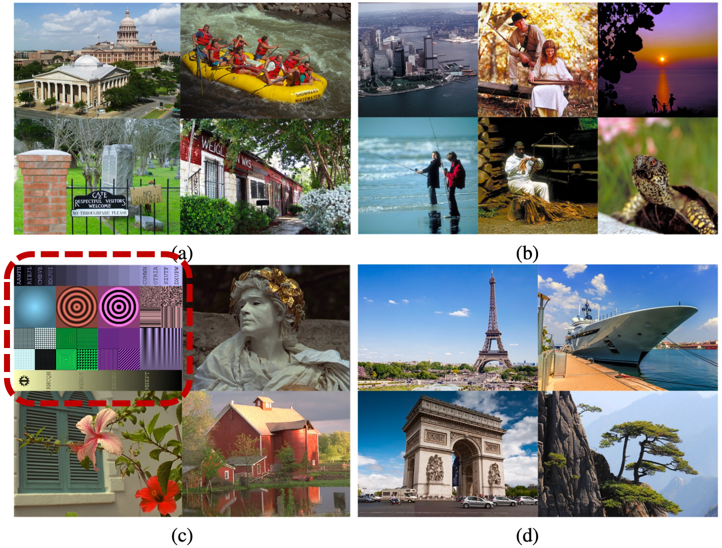



Figure 1.1: Natural images and a screen content image from the constructed databases. (a) LIVE Database [13] (b) CSIQ Database [14] (c) TID2013 Database [15] (d) KADID-10k Database [16].

Table 4.1: Brief summary of the LIVE, CSIQ, TID2013, and KADID-10k databases.

| Database | LIVE [13] | CSIQ [14] | TID2013 [15] | KADID-10k [16] |
|---|---|---|---|---|
| Num. of Reference Images | 29 | 30 | 25 | 81 |
| Num. of Distorted Images | 779 | 866 | 3,000 | 10,125 |
| Num. of Distortion Types | 5 | 6 | 24 | 25 |
| Num. of Distortion Levels | $5 \sim 8$ | $3 \sim 5$ | 5 | 5 |
| Annotation | DMOS | DMOS | MOS | MOS |
| Range | [0, 100] | [0, 1] | [0, 9] | [1, 5] |

# Intra-Database Experiments

Table 4.2: Performance comparisons on the LIVE, CSIQ, and TID2013 databases. Top two results are highlighted in bold.

| Method | LIVE | | CSIQ | | TID2013 | |
|---|---|---|---|---|---|---|
| | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| BRISQUE (2012) [10] | 0.939 | 0.935 | 0.746 | 0.829 | 0.604 | 0.694 |
| CORNIA (2012) [104] | 0.947 | 0.950 | 0.678 | 0.776 | 0.678 | 0.768 |
| M3 (2015) [105] | 0.951 | 0.950 | 0.795 | 0.839 | 0.689 | 0.771 |
| HOSA (2016) [103] | 0.946 | 0.947 | 0.741 | 0.823 | 0.735 | 0.815 |
| FRIQUEE (2017) [90] | 0.940 | 0.944 | 0.835 | 0.874 | 0.68 | 0.753 |
| DIQaM-NR (2018) [35] | 0.960 | **0.972** | - | - | 0.835 | 0.855 |
| DB-CNN (2020) [64] | 0.968 | **0.971** | **0.946** | **0.959** | 0.816 | 0.865 |
| HyperIQA (2020) [65] | 0.962 | 0.966 | 0.923 | 0.942 | 0.729 | 0.775 |
| GraphIQA (2022) [86] | **0.968** | 0.970 | 0.920 | 0.938 | - | - |
| TReS (2022) [87] | **0.969** | 0.968 | 0.922 | 0.942 | **0.863** | **0.883** |
| **NLNet** | 0.962 | 0.963 | **0.941** | **0.958** | **0.856** | **0.880** |

**SOTA Transformer**

**Fewer Training Data**
**↓ 20% Total Data**
**↑ Highly Competitive Performance**

Table 4.3: Performance comparisons on the KADID-10k database. Top two results are highlighted in bold.

| Method | BRISQUE [10] | CORNIA [104] | HOSA [103] | InceptionResNetV2 [16] | DB-CNN [64] | HyperIQA [65] | TReS [87] | **NLNet** |
|---|---|---|---|---|---|---|---|---|
| SRCC | 0.519 | 0.519 | 0.609 | 0.731 | 0.851 | **0.852** | **0.859** | 0.846 |
| PLCC | 0.554 | 0.554 | 0.653 | 0.734 | **0.856** | 0.845 | **0.858** | 0.850 |

# Cross-Database Settings and Evaluations

Table 4.9: Cross-database performance comparisons.

| Training | LIVE | | CSIQ | | TID2013 | |
|---|---|---|---|---|---|---|
| Testing | CSIQ | TID2013 | LIVE | TID2013 | LIVE | CSIQ |
| BRISQUE (2012) [10] | 0.562 | 0.358 | 0.847 | 0.454 | 0.790 | 0.590 |
| CORNIA (2012) [104] | 0.649 | 0.360 | 0.853 | 0.312 | 0.846 | 0.672 |
| M3 (2015) [105] | 0.621 | 0.344 | 0.797 | 0.328 | 0.873 | 0.605 |
| HOSA (2016) [103] | 0.594 | 0.361 | 0.773 | 0.329 | 0.846 | 0.612 |
| FRIQUEE (2017) [90] | 0.722 | 0.461 | 0.879 | 0.463 | 0.755 | 0.635 |
| DIQaM-NR (2018) [35] | 0.681 | 0.392 | - | - | - | 0.717 |
| DB-CNN (2020) [64] | **0.758** | **0.524** | 0.877 | **0.540** | **0.891** | **0.807** |
| HyperIQA (2020) [65] | 0.697 | **0.538** | **0.905** | **0.554** | 0.839 | 0.543 |
| **NLNet** | **0.771** | 0.497 | **0.923** | 0.516 | **0.895** | **0.730** |

**Similar Distortions**

**TID: More Distortion Types & Levels**

# Single Distortion Type Evaluation

Table 4.4: The average SRCC and PLCC results of the individual distortion type on the LIVE database. Top two results are highlighted in bold.

| SRCC | Global Distortion | | | | Local Distortion |
|---|---|---|---|---|---|
| | JPEG | JP2K | WN | GB | FF |
| BRISQUE (2012) [10] | 0.965 | 0.929 | 0.982 | **0.964** | 0.828 |
| CORNIA (2012) [104] | 0.947 | 0.924 | 0.958 | 0.951 | 0.921 |
| M3 (2014) [105] | 0.966 | 0.930 | **0.986** | 0.935 | 0.902 |
| HOSA (2016) [103] | 0.954 | 0.935 | 0.975 | 0.954 | **0.954** |
| FRIQUEE (2017) [90] | 0.947 | 0.919 | 0.983 | 0.937 | 0.884 |
| dipIQ (2017) [82] | 0.969 | **0.956** | 0.975 | 0.940 | - |
| WaDIQaM (2018) [35] | 0.953 | 0.942 | 0.982 | 0.938 | 0.923 |
| DB-CNN (2020) [64] | **0.972** | 0.955 | 0.980 | 0.935 | 0.930 |
| HyperIQA (2020) [65] | 0.961 | 0.949 | 0.982 | 0.926 | 0.934 |
| **NLNet** | **0.979** | **0.958** | **0.990** | **0.964** | **0.941** |

| PLCC | Global Distortion | | | | Local Distortion |
|---|---|---|---|---|---|
| | JPEG | JP2K | WN | GB | FF |
| BRISQUE (2012) [10] | 0.971 | 0.940 | 0.989 | **0.965** | 0.894 |
| CORNIA (2012) [104] | 0.962 | 0.944 | 0.974 | 0.961 | 0.943 |
| M3 (2014) [105] | 0.977 | 0.945 | **0.992** | 0.947 | 0.920 |
| HOSA (2016) [103] | 0.967 | 0.949 | 0.983 | **0.967** | **0.967** |
| FRIQUEE (2017) [90] | 0.955 | 0.935 | 0.991 | 0.949 | 0.936 |
| dipIQ (2017) [82] | 0.980 | **0.964** | 0.983 | 0.948 | - |
| DB-CNN (2020) [64] | **0.986** | **0.967** | 0.988 | 0.956 | **0.961** |
| **NLNet** | **0.986** | 0.961 | **0.993** | 0.964 | 0.951 |

Image Credit: LIVE Database

**Noisy and Compressed Images**

**Global Distortion**

**Non-local Recurrence**

**Local Distortion**

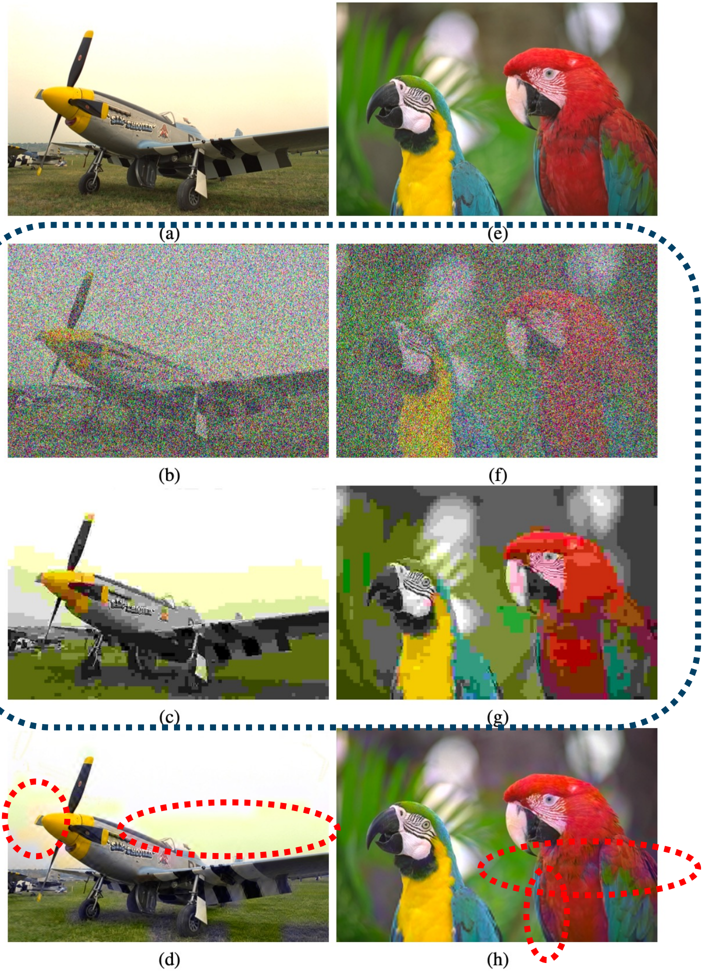



Figure 4.7: Demonstrations of the global distortions (b/f: WN and c/g: JPEG) and local distortions (d/h: FF) contaminating the plane and parrot images. Figure (a) and Figure (e) are reference images from the LIVE database.

# Single Distortion Type Evaluation

Table 4.5: The average SRCC and PLCC results of the individual distortion type on the CSIQ database. Top two results are highlighted in bold.

| SRCC | JPEG | JP2K | WN | GB | PN | CC |
|---|---|---|---|---|---|---|
| BRISQUE (2012) [10] | 0.806 | 0.840 | 0.723 | 0.820 | 0.378 | 0.804 |
| CORNIA (2012) [104] | 0.513 | 0.831 | 0.664 | 0.836 | 0.493 | 0.462 |
| M3 (2014) [105] | 0.740 | 0.911 | 0.741 | 0.868 | 0.663 | 0.770 |
| HOSA (2016) [103] | 0.733 | 0.818 | 0.604 | 0.841 | 0.500 | 0.716 |
| FRIQUEE (2017) [90] | 0.869 | 0.846 | 0.748 | 0.870 | 0.753 | 0.838 |
| dipIQ (2017) [82] | 0.936 | 0.944 | 0.904 | 0.932 | - | - |
| MEON (2018) [71] | **0.948** | 0.898 | **0.951** | 0.918 | - | - |
| WaDIQaM (2018) [35] | 0.853 | 0.947 | 0.974 | 0.979 | 0.882 | **0.923** |
| DB-CNN (2020) [64] | 0.940 | 0.953 | 0.948 | **0.947** | **0.940** | 0.870 |
| HyperIQA (2020) [65] | 0.934 | **0.960** | 0.927 | 0.915 | 0.931 | 0.874 |
| **NLNet** | **0.972** | **0.963** | **0.965** | **0.955** | **0.969** | **0.968** |
| **PLCC** | **JPEG** | **JP2K** | **WN** | **GB** | **PN** | **CC** |
| BRISQUE (2012) [10] | 0.828 | 0.887 | 0.742 | 0.891 | 0.496 | 0.835 |
| CORNIA (2012) [104] | 0.563 | 0.883 | 0.687 | 0.904 | 0.632 | 0.543 |
| M3 (2014) [105] | 0.768 | 0.928 | 0.728 | 0.917 | 0.717 | 0.787 |
| HOSA (2016) [103] | 0.759 | 0.899 | 0.656 | 0.912 | 0.601 | 0.744 |
| FRIQUEE (2017) [90] | 0.885 | 0.883 | 0.778 | 0.905 | 0.769 | 0.864 |
| dipIQ (2017) [82] | 0.975 | **0.959** | 0.927 | 0.958 | - | - |
| MEON (2018) [71] | 0.979 | 0.925 | **0.958** | 0.946 | - | - |
| DB-CNN (2020) [64] | **0.982** | 0.971 | 0.956 | **0.969** | **0.950** | **0.895** |
| **NLNet** | **0.991** | **0.976** | **0.967** | **0.9746** | **0.966** | **0.969** |

Noise-Related Distortions

Global Distortion


Figure 4.9: Demonstrations of the global distortions (b/f: GB, c/g: CC, d/h: PN) contaminating the Statue of Liberty and George Rogers Clark Memorial images. Figure (a) and Figure (e) are reference images from the CSIQ database.



Image Credit: CSIQ Database

# Single Distortion Type Evaluation

Table 4.6: The average SRCC results of the individual distortion type on the TID2013 database. Top two results are highlighted in bold.

| SRCC | Distortion Type | BRISQUE [10] | FRIQUEE [90] | HOSA [103] | MEON [71] | M3 [105] | DB-CNN [64] | CORNIA [104] | **NLNet** |
|---|---|---|---|---|---|---|---|---|---|
| Global Distortion | Additive Gaussian noise | 0.711 | 0.730 | **0.833** ↑8.4% | 0.813 | 0.766 | 0.790 | 0.692 | **0.917** |
| | Lossy compression of noisy images | 0.609 | 0.641 | 0.838 | 0.772 | 0.692 | **0.860** ↑7.5% | 0.712 | **0.935** |
| | Additive noise in color components | 0.432 | 0.573 | 0.551 | **0.722** ↑12.8% | [illegible] | 0.700 | 0.137 | **0.850** |
| | Comfort noise | 0.196 | 0.318 | 0.622 | 0.406 | 0.353 | **0.752** ↑11.8% | 0.617 | **0.870** |
| | Contrast change | -0.001 | **0.585** | 0.362 | 0.252 | 0.155 | 0.548 | 0.254 | **0.793** |
| | Change of color saturation | 0.003 | 0.589 | 0.045 | **0.684** | -0.199 | 0.631 | 0.169 | **0.827** |
| | Spatially correlated noise | 0.746 | 0.866 | 0.842 | **0.926** ↑3.2% | 0.782 | 0.826 | 0.741 | **0.958** |
| | High frequency noise | 0.842 | 0.847 | 0.897 | **0.911** ↑1.0% | 0.900 | 0.879 | 0.815 | **0.921** |
| | Impulse noise | 0.765 | 0.730 | 0.809 | **0.901** ↑1.2% | 0.738 | 0.708 | 0.616 | **0.913** |
| | Quantization noise | 0.662 | 0.764 | 0.815 | **0.888** ↑4.1% | 0.832 | 0.825 | 0.661 | **0.929** |
| | Gaussian blur | 0.871 | 0.881 | 0.883 | 0.887 | **0.896** | 0.859 | 0.850 | **0.912** |
| | Image denoising | 0.612 | 0.839 | 0.854 | 0.797 | 0.709 | **0.865** ↑1.7% | 0.764 | **0.882** |
| | JPEG compression | 0.764 | 0.813 | 0.891 | 0.850 | 0.844 | **0.894** ↑1.1% | 0.797 | **0.905** |
| | JPEG 2000 compression | 0.745 | 0.831 | **0.919** ↑1.1% | 0.891 | 0.885 | 0.916 | 0.846 | **0.930** |
| | Multiplicative Gaussian noise | 0.717 | 0.704 | 0.768 | **0.849** ↑5.5% | 0.738 | 0.711 | 0.593 | **0.904** |
| | Image color quantization with dither | 0.831 | 0.768 | 0.896 | 0.857 | **0.908** | 0.833 | 0.683 | **0.911** |
| | Sparse sampling and reconstruction | 0.807 | 0.891 | **0.909** | 0.855 | 0.893 | 0.902 | 0.865 | **0.940** |
| | Chromatic aberrations | 0.615 | 0.737 | 0.753 | **0.779** | 0.570 | 0.732 | 0.696 | **0.773** |
| | Masked noise | 0.252 | 0.345 | 0.468 | **0.728** | 0.577 | 0.646 | 0.451 | **0.700** |
| | Mean shift (intensity shift) | 0.219 | **0.254** | 0.211 | 0.177 | 0.119 | -0.009 | 0.232 | **0.358** |
| Local Distortion | JPEG transmission errors | 0.301 | 0.498 | 0.730 | 0.746 | 0.375 | **0.772** ↑3.3% | 0.694 | **0.805** |
| | JPEG 2000 transmission errors | 0.748 | 0.660 | 0.710 | 0.716 | 0.718 | **0.773** ↑10.2% | 0.686 | **0.875** |
| | Non eccentricity pattern noise | 0.269 | 0.076 | 0.242 | 0.116 | 0.173 | **0.270** ↑34.6% | 0.200 | **0.616** |
| | Local bock-wise distortions with different intensity | 0.207 | 0.032 | 0.268 | **0.500** | 0.379 | 0.444 | 0.027 | **0.493** |

## Noise and Compression-Related Distortions

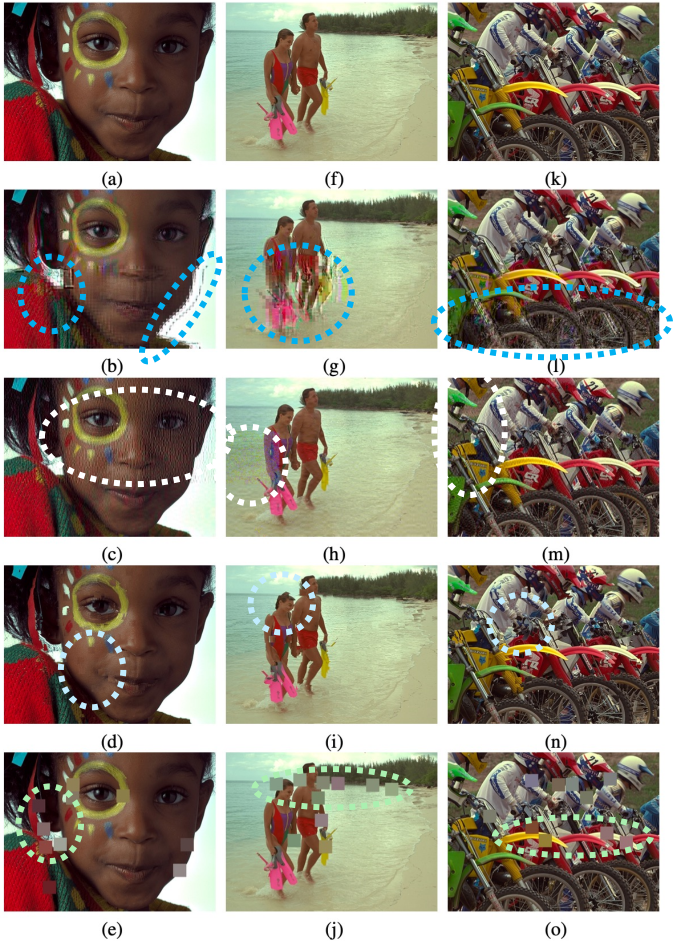



Figure 4.10: Demonstrations of the local distortions (b/g/l: JPEG transmission errors, c/h/m: JPEG2000 transmission errors, d/i/n: non eccentricity pattern noise, e/j/o: local block-wise distortions of different intensity). Figure (a), Figure (f), and Figure (k) are reference images from the TID2013 database.

Image Credit: TID2013 Database

# Single Distortion Type Evaluation

Table 4.7: The average SRCC results of the individual distortion type on the KADID-10k database. The local distortions are highlighted in blue and the top two results are highlighted in bold.

| Distortion Type | | BLIINDS-II [91] | BRISQUE [10] | ILNIQE [102] | CORNIA [104] | HOSA [103] | WaDIQaM [35] | **NLNet** |
|---|---|---|---|---|---|---|---|---|
| Blurs | Lens blur | 0.781 | 0.674 | **0.846** | 0.811 | 0.715 | 0.730 | **0.914** |
| | Gaussian blur | 0.880 | 0.812 | **0.883** | 0.866 | 0.852 | 0.879 | **0.914** |
| | Motion blur | 0.482 | 0.423 | **0.779** | 0.532 | 0.652 | 0.730 | **0.899** |
| Color distortions | Color diffusion | 0.572 | 0.544 | 0.678 | 0.243 | 0.727 | **0.833** | **0.916** |
| | Color saturation 2 | 0.602 | 0.375 | 0.677 | 0.120 | **0.841** | 0.836 | **0.909** |
| | Color quantization | 0.670 | 0.667 | 0.676 | 0.323 | 0.662 | **0.806** | **0.853** |
| | Color shift | -0.139 | -0.182 | 0.090 | -0.002 | 0.050 | **0.421** | **0.777** |
| | Color saturation 1 | 0.091 | 0.071 | 0.027 | -0.019 | **0.216** | 0.148 | **0.604** |
| Compression | JPEG compression | 0.414 | 0.782 | **0.804** ↑6.2% | 0.556 | 0.582 | 0.530 | **0.866** |
| | JPEG 2000 compression | 0.655 | 0.516 | **0.790** ↑6.3% | 0.342 | 0.608 | 0.539 | **0.853** |
| Noise | Denoise | 0.457 | 0.221 | **0.856** ↑9.7% | 0.229 | 0.247 | 0.765 | **0.953** |
| | White noise in color component | 0.757 | 0.718 | 0.841 | 0.418 | 0.745 ↑1.1% | **0.925** | **0.936** |
| | Multiplicative noise | 0.702 | 0.674 | 0.682 | 0.306 | 0.776 ↑5.0% | **0.884** | **0.934** |
| | Impulse noise | 0.547 | -0.543 | 0.808 | 0.219 | 0.254 ↑10.2% | **0.814** | **0.916** |
| | White Gaussian noise | 0.628 | 0.708 | 0.776 | 0.357 | 0.680 ↑1.7% | **0.897** | **0.914** |
| Brightness change | Brighten | 0.458 | 0.575 | 0.301 | 0.227 | **0.753** | 0.685 | **0.822** |
| | Darken | 0.439 | 0.405 | 0.436 | 0.206 | **0.744** | 0.272 | **0.647** |
| | Mean Shift | 0.112 | 0.144 | 0.315 | 0.122 | **0.591** | **0.348** | 0.335 |
| Spatial distortions | Jitter | 0.629 | 0.672 | 0.441 | 0.719 | 0.391 | **0.778** | **0.899** |
| | Pixelate | 0.196 | 0.648 | 0.577 | 0.587 | **0.702** | 0.700 | **0.814** |
| | Quantization | **0.781** | 0.714 | 0.571 | 0.259 | 0.681 | 0.735 | **0.791** |
| | Color block | -0.020 | 0.067 | 0.003 | 0.094 | **0.388** | 0.160 | **0.440** |
| | Non-eccentricity patch | 0.083 | 0.191 | 0.218 | 0.121 | **0.461** | 0.348 | **0.433** |
| Sharpness and contrast | High sharpen | -0.015 | 0.361 | **0.681** | 0.114 | 0.230 | 0.558 | **0.932** |
| | Contrast change | 0.062 | 0.105 | 0.072 | 0.125 | **0.452** | 0.421 | **0.513** |

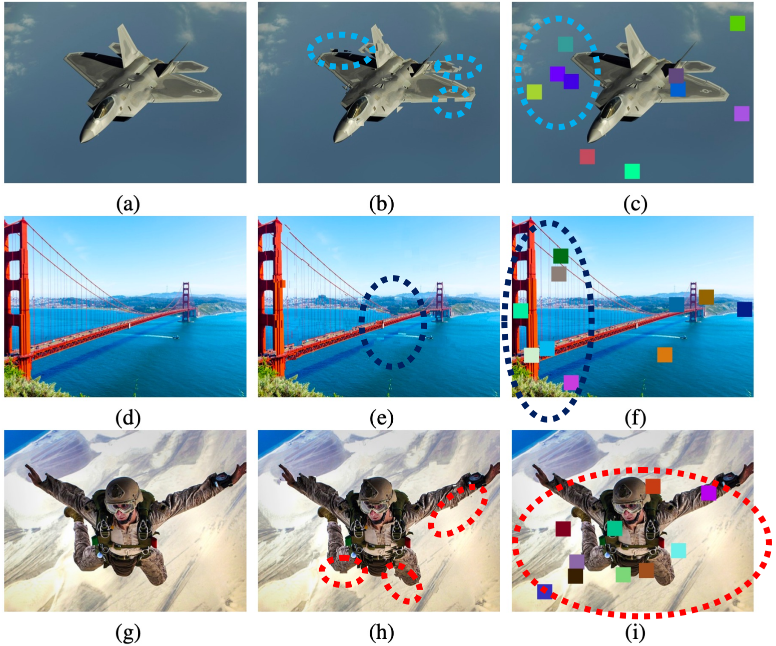



Figure 4.11: Demonstrations of the local distortions (b/e/h: non-eccentricity patch and c/f/i: color block). Figure (a), Figure (d), and Figure (g) are reference images from the KADID-10k database.



Image Credit: KADID-10k Database

# Takeaways and Future Work

✓ Non-local & Local Modeling

(1) The Non-local Modeling is complementary to traditional local methods.

(2) CNN's Local Modeling features are effective and robust.

✓ Global & Local Distortions

(1) Handle a wide variety of Global Distortions: globally and uniformly distributed with non-local recurrences.

(2) Maintain sensitivity to Local Distortions: local nonuniform-distributed distortions in a local region.

(3) Better assess Noisy and Compressed Images quality.

✓ Generalization Capability Cross-Dataset Setting → High Generalization Capability

✓ Future Work Non-local Statistics [1, 2]; PGC → UGC → AIGC: Quality Assessment of AI Generated Content

Credit:
[1] Zontak *et al.*, Internal Statistics of a Single Natural Image, In CVPR 2011
[2] Buades *et al.*, A Non-local Algorithm for Image Denoising, In CVPR 2005