# DEEP BLIND IMAGE QUALITY ASSESSMENT BY LEARNING SENSITIVITY MAP

Jongyoo Kim, Woojae Kim and Sanghoon Lee

ICASSP 2018

## Deep Learning and Convolutional Neural Networks (CNNs)

### ❖ SOTA in computer vision & image processing

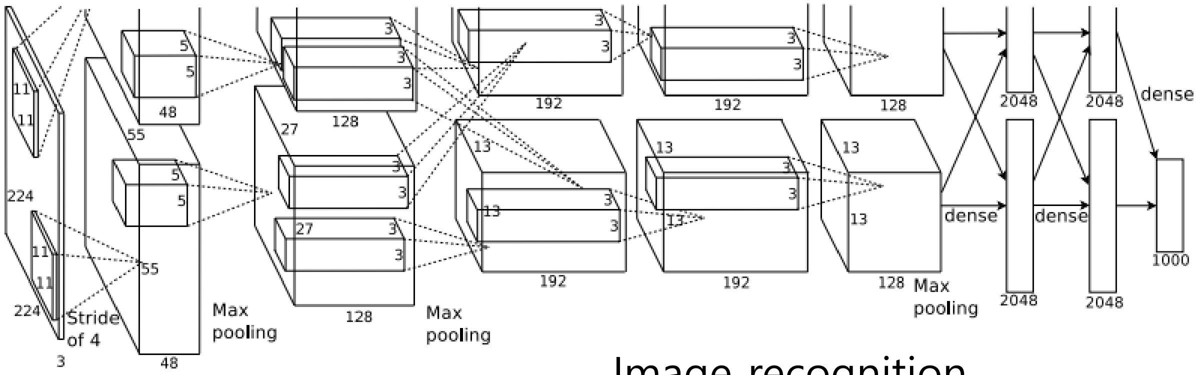
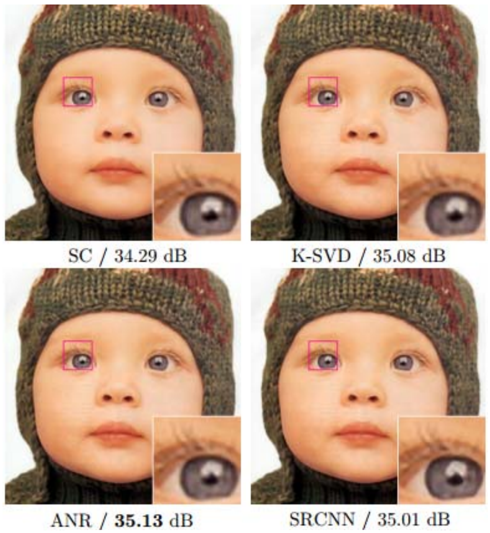
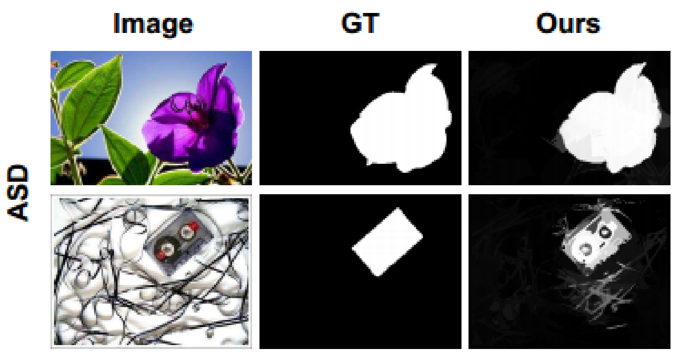




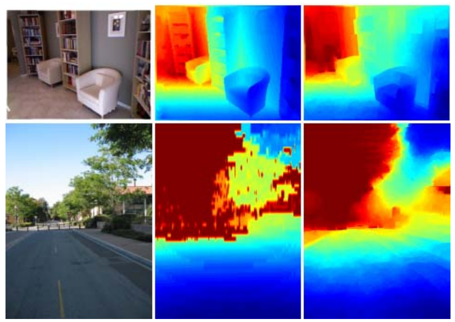
Image recognition




Super-resolution




Saliency detection



Depth map from single image

# Problems of Applying CNN to Image Quality Assessment (IQA)

## ❖ Lack of Training Dataset

| Dataset size | Label (target) | Augmentation |
|---|---|---|
| ***ImageNet database***<br>Millions of labeled images (1.6M) | Semantic object meaning<br>**via Crowdsourcing**<br>Easy decision (short time) | Possible |
| ***LIVE IQA database***<br>**29 reference images,**<br>→ 5 distortion types & 6~7 levels<br>**982 distorted images**<br>***TID2013***<br>**25 reference images**<br>→ 17 distortion types & 4 levels<br>**1,700 distorted images,**<br>Distorted images are highly Correlated with each other | *Subjective score*<br>**via Subjective test**<br>Hard decision (long time)<br>**Controlled environment**<br>Creating a large-scale database is a formidable problem | Only horizontal reflection<br>Any image transformation can change the label |

# An attempt to increase the dataset for IQA

Distorted image

Divided image patches

**Only one score**
DMOS = 74

Local quality score for each patch is different

YONSEI UNIVERSITY



Multidimensional insight

# Transfer Learning (Image Recognition → IQA) ?

**Image recognition**

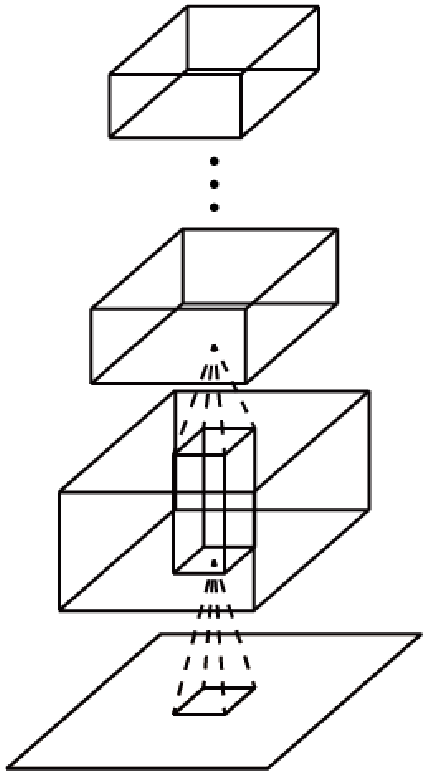
**Synthetic info.**

**Distortion info.**

**Robust to image corruption**

**Synthetic info.**
**Distortion info.**

High-level feature



Distorted image

**IQA**

**Distortion info.**

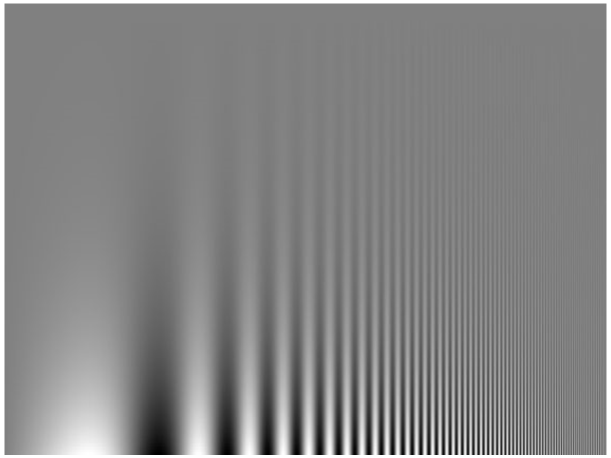
**Synthetic info.**

Various distortion types
- Need low-level texture features
- Need semantic features

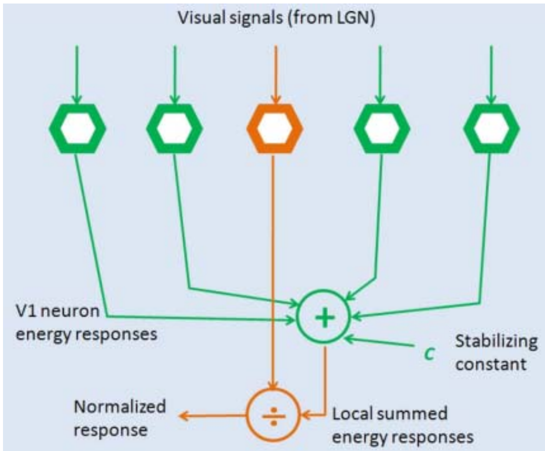
**Synthetic info.**
**Distortion info.**

# Major ISSUE of Applying CNN for IQA

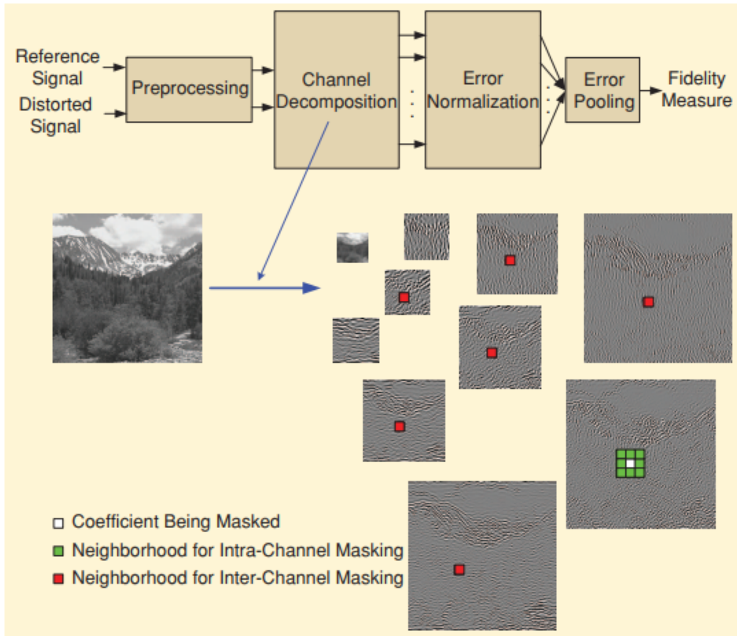
- **Modeling of Accurate Human Visual Perception**
  - Bandpass, multiscale, and directional decompositions
  - Contrast / Texture Masking
  - Luminance Adaption
  - Etc.



Campbell and Robson chart.
Contrast sensitivity.




Divisive normalization

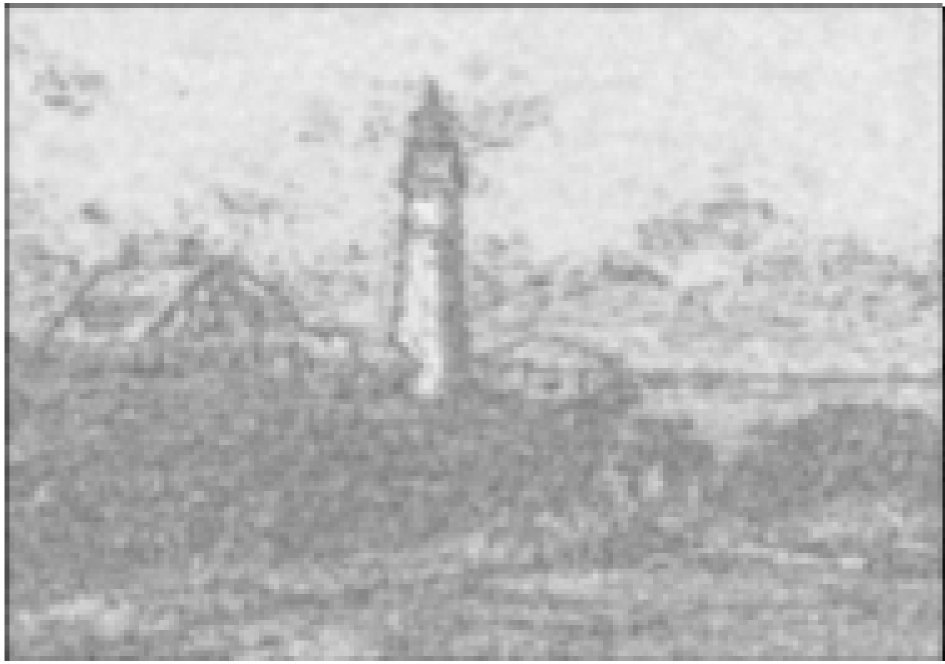


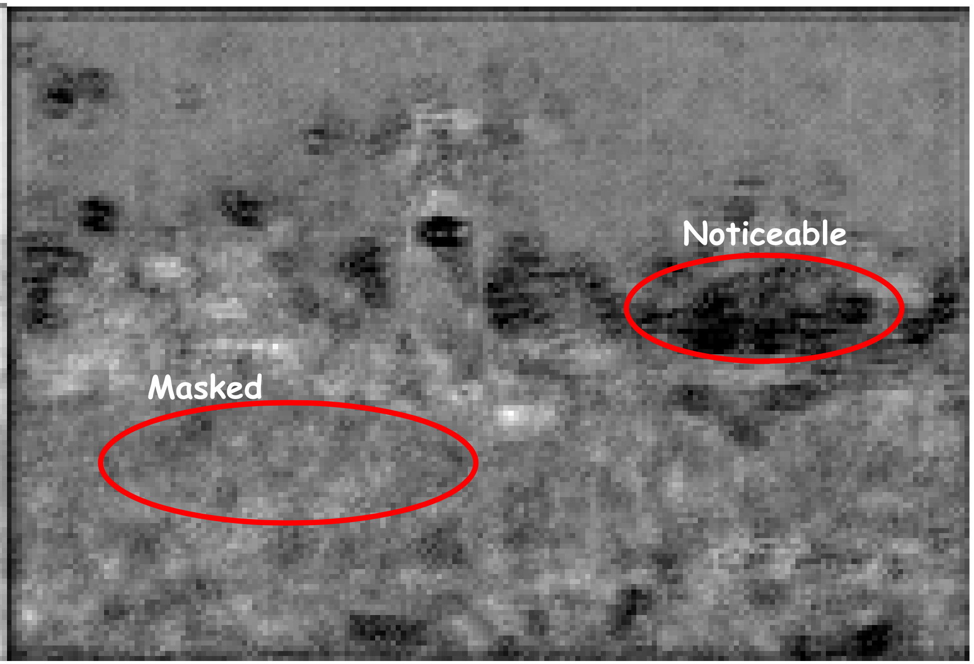
# Motivation – find the Visual Sensitivity in a data-driven way

- A Distorted Image (JPEG 2000)

# Motivation – Deep Learning of Human Visual Sensitivity

- DeepQA learns the visual sensitivity without any prior knowledge
- Using distorted image, objective error map, subjective score






Objective error map
(Derived by simple distance metric)

$$\mathbf{e} = err\left(I_r, I_d\right)$$

Perceptual error map

# Utilizing local objective score as a proxy ground truth

Reference image

Distorted image

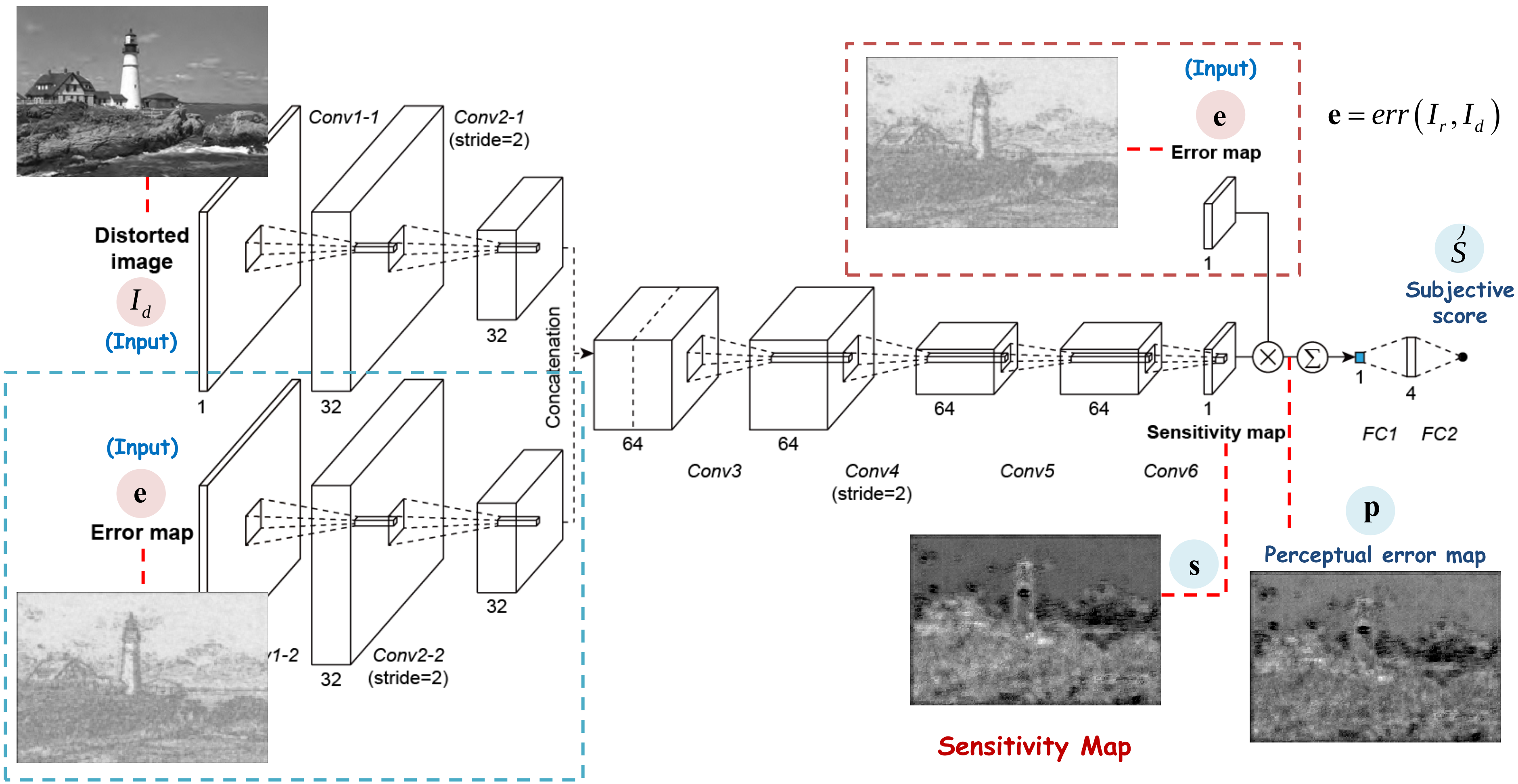
MSE
**Pixel diff** Etc.

Local objective score
(called Error Map in this paper)

# Motivation – Previous Work

J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," CVPR 2017.

## ❖ DeepQA – Full Reference Image Quality Assessment

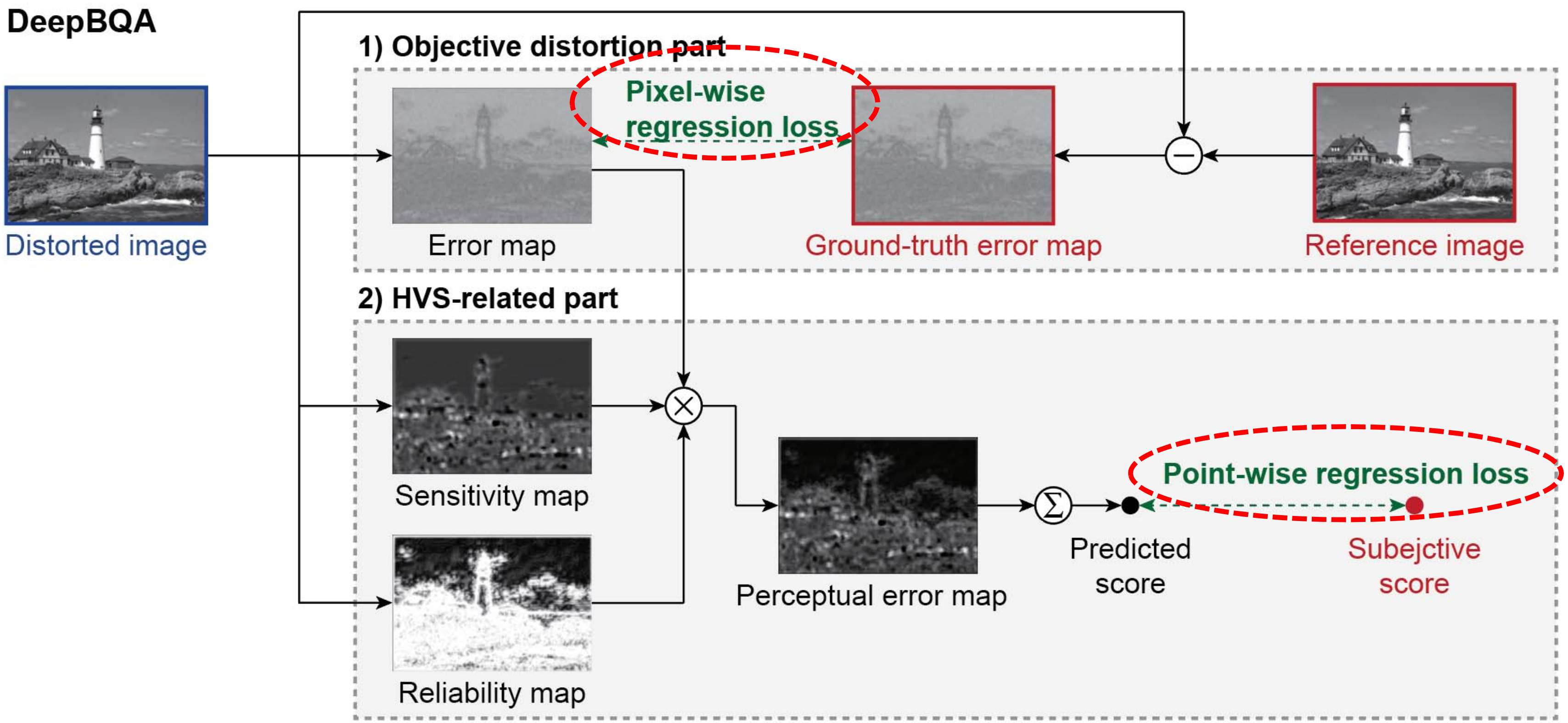





YONSEI UNIVERSITY

Multidimensional Insight

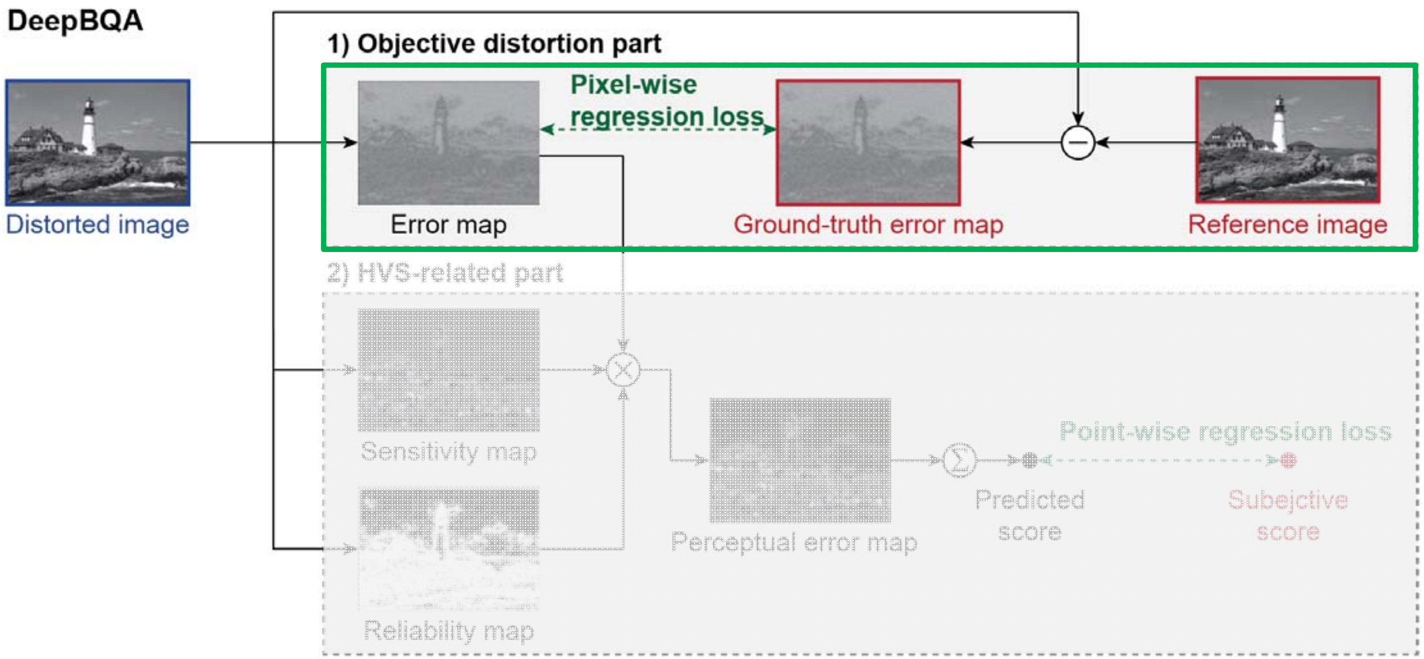
# Proposed Deep Blind Quality Assessment (BQA)




- In the first stage, an objective error map is used as a proxy training target to expand the dataset labels.

# Learning Objective Error Map (1st stage)

## ❖ Obtaining Objective Error Maps

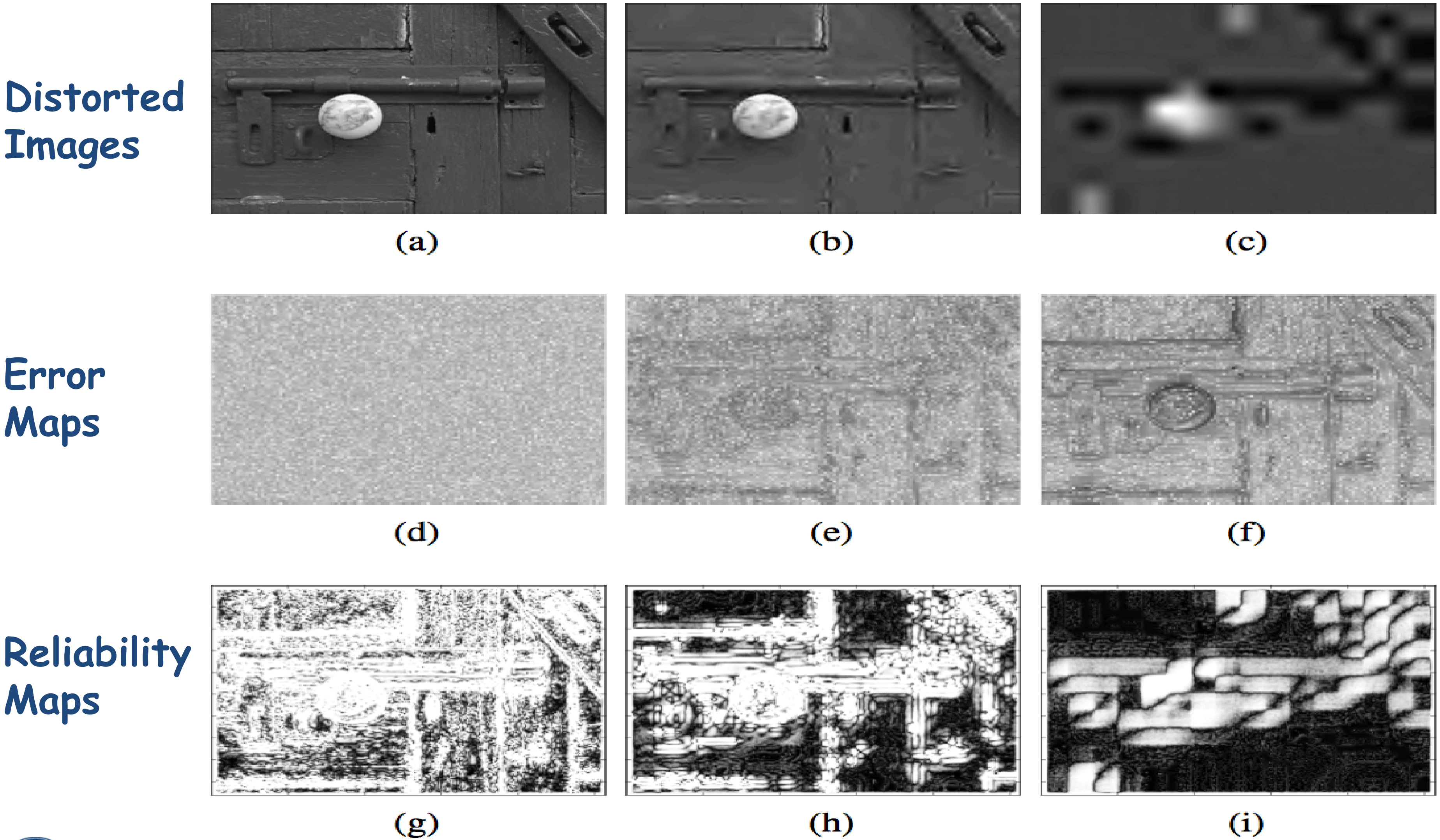



- Error map: $\mathbf{e}_{gt} = |\hat{I}_r - \hat{I}_d|^p$
  - p = 0.2
- Loss Function

$$\mathcal{L}_e(\hat{I}_d, \hat{I}_r; \theta_1) = \left\| (CNN_1(\hat{I}_d; \theta_1) - \mathbf{e}_{gt}) \odot \hat{\mathbf{r}} \right\|_2^2$$

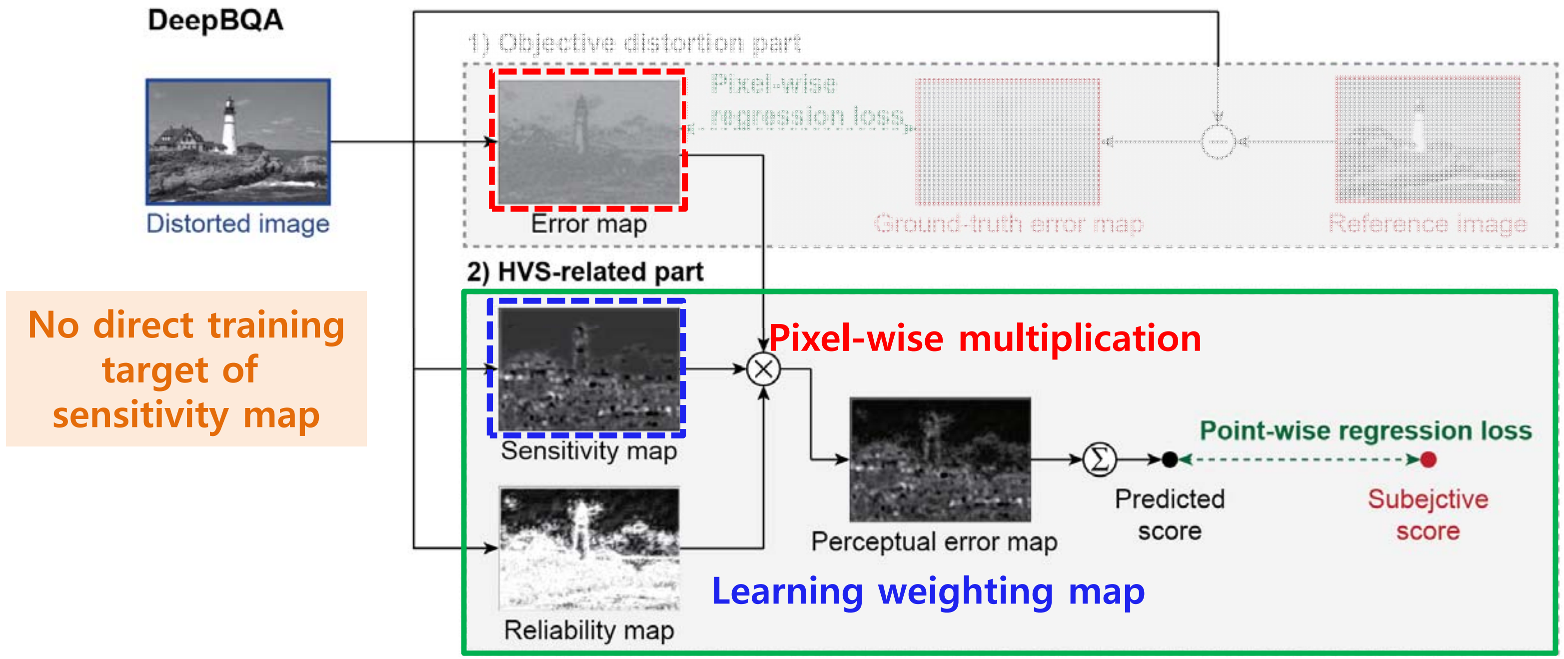
# Reliability Map Prediction

❖ **Blurry Regions have lower reliability than textured regions**



# Learning Sensitivity Map (2nd stage)

## ❖ Loss Function




- Perceptual error map

$$\mathbf{p} = \mathbf{s} \odot \mathbf{e} \odot \hat{\mathbf{r}}.$$

$$\mathbf{s} = CNN_2(\hat{I}_d; \theta_2)$$

- **s**: sensitivity map
- **e**: error map
- **r**: reliability map

- Loss function

$$\mathcal{L}_s(\hat{I}_d; \theta_1, \theta_2) = \|(\underline{f}(\mu_{\mathbf{p}}) - S)\|_2^2$$

MLP regression

# Learning Sensitivity Map (2nd stage)

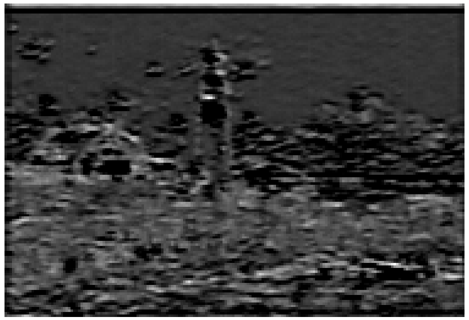
## ❖ Total Variation L2 Norm (Regularization)

- Purpose
  - To smooth the sensitivity map
- Why?
  - Differentiable,
  - Can be added to the loss function for SGD with ease

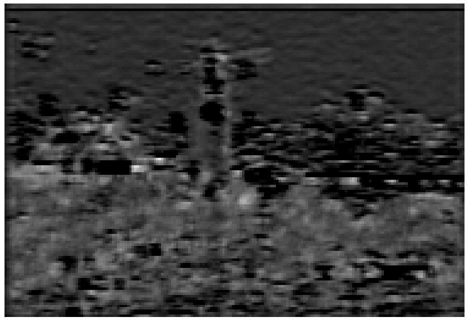
$$TV(\mathbf{s}) = 1/HW \cdot \sum \left( sobel_h(\mathbf{s})^2 + sobel_v(\mathbf{s})^2 \right)^{\beta/2}$$

- Sobel filtered sensitivity map

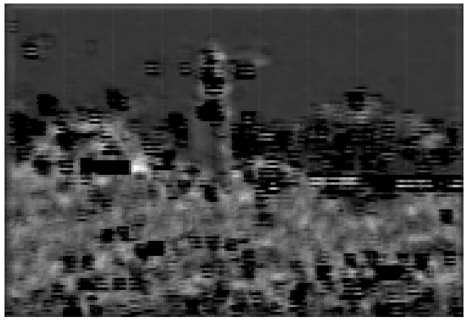
$$L(I_d;\theta) = w_{subj} \left\| f(\mu_{\mathbf{p}}) - S \right\|_F^2 + \underbrace{w_{TV} TV(\mathbf{s})}_{\text{smoothing (Total variation)}}$$



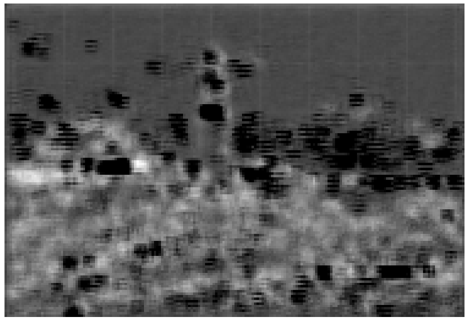
$w_{TV} = 10^{-4}$



$w_{TV} = 10^{-3}$



$w_{TV} = 10^{-2}$



$w_{TV} = 10^{-1}$

# Database

## ❖ Three IQA Databases

| Database | Ref. | Dist. | Dist. Types | Score Type |
|---|---|---|---|---|
| LIVE IQA | 29 | 779 | 5 | DMOS |
| CSIQ | 30 | 866 | 6 | DMOS |
| TID2013 | 25 | 3,000 | 24 | MOS |

- **LIVE IQA**: H. Sheikh, M. Sabir, and A. Bovik, "**A statistical evaluation of recent full reference image quality assessment algorithms**," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006.

- **CSIQ**: E. C. Larson and D. M. Chandler, "**Most apparent distortion: Full-reference image quality assessment and the role of strategy,**" *J. Electron. Imaging*, vol. 19, no. 1, pp. 19 –19 –21, 2010.

- **TID2013**: N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. C. Jay Kuo, "**Image database TID2013: Peculiarities, results and perspectives**," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

# Experiment and Analysis

## ❖ Experimental Setting (Common)

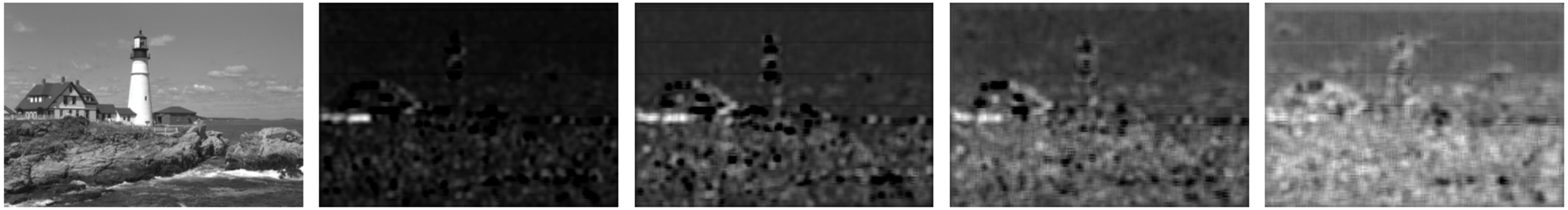
- Evaluation metrics
  - Spearman's rank order correlation coefficient (SRCC)
  - Pearson's linear correlation coefficient (PLCC)

$$SRCC = 1 - \frac{6\sum_i d_i^2}{n\left(n^2 - 1\right)} \qquad PLCC = \frac{\sum_i (\hat{S}_i - \mu_{\hat{S}})(S_i - \mu_S)}{\sqrt{\sum_i (\hat{S}_i - \mu_{\hat{S}})^2}\sqrt{\sum_i (S_i - \mu_S)^2}}$$

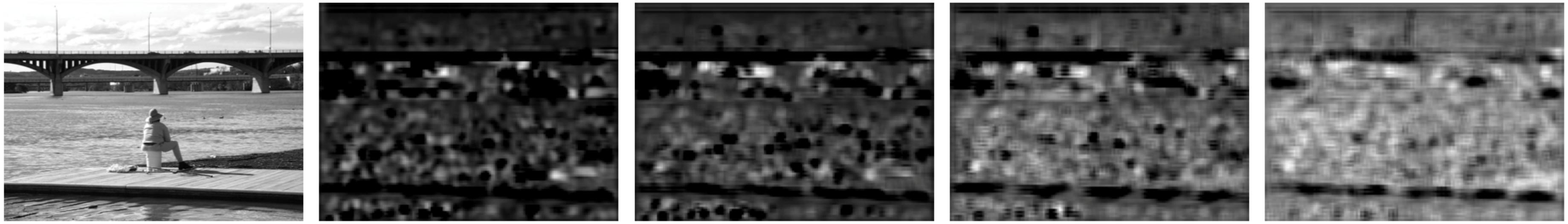
- Training and testing sets
  - For each repetition, reference images were **randomly divided** into two subsets, **80% for training** and **20% for testing**.
  - Then, the corresponding distorted images were divided into the two subsets.
  - Horizontally flipped images were supplemented to the training set.

# Predicted Sensitivity Maps based on Regularization

## ❖ Total Variation L2 Norm Regularization



(a) (b) $w_{TV} = 10^{-4}$ (c) $w_{TV} = 10^{-3}$ (d) $w_{TV} = 10^{-2}$ (e) $w_{TV} = 10^{-1}$



(f) (g) $w_{TV} = 10^{-4}$ (h) $w_{TV} = 10^{-3}$ (i) $w_{TV} = 10^{-2}$ (j) $w_{TV} = 10^{-1}$

| Metric | TV regularization weight ($w_{TV}$ =) | | | | |
|---|---|---|---|---|---|
| | 0 | $10^{-4}$ | $10^{-3}$ | $10^{-2}$ | $10^{-1}$ |
| SRCC | 0.961 | 0.965 | 0.966 | $\boxed{0.971}$ | 0.969 |
| PLCC | 0.963 | 0.966 | 0.967 | $\boxed{0.972}$ | 0.970 |

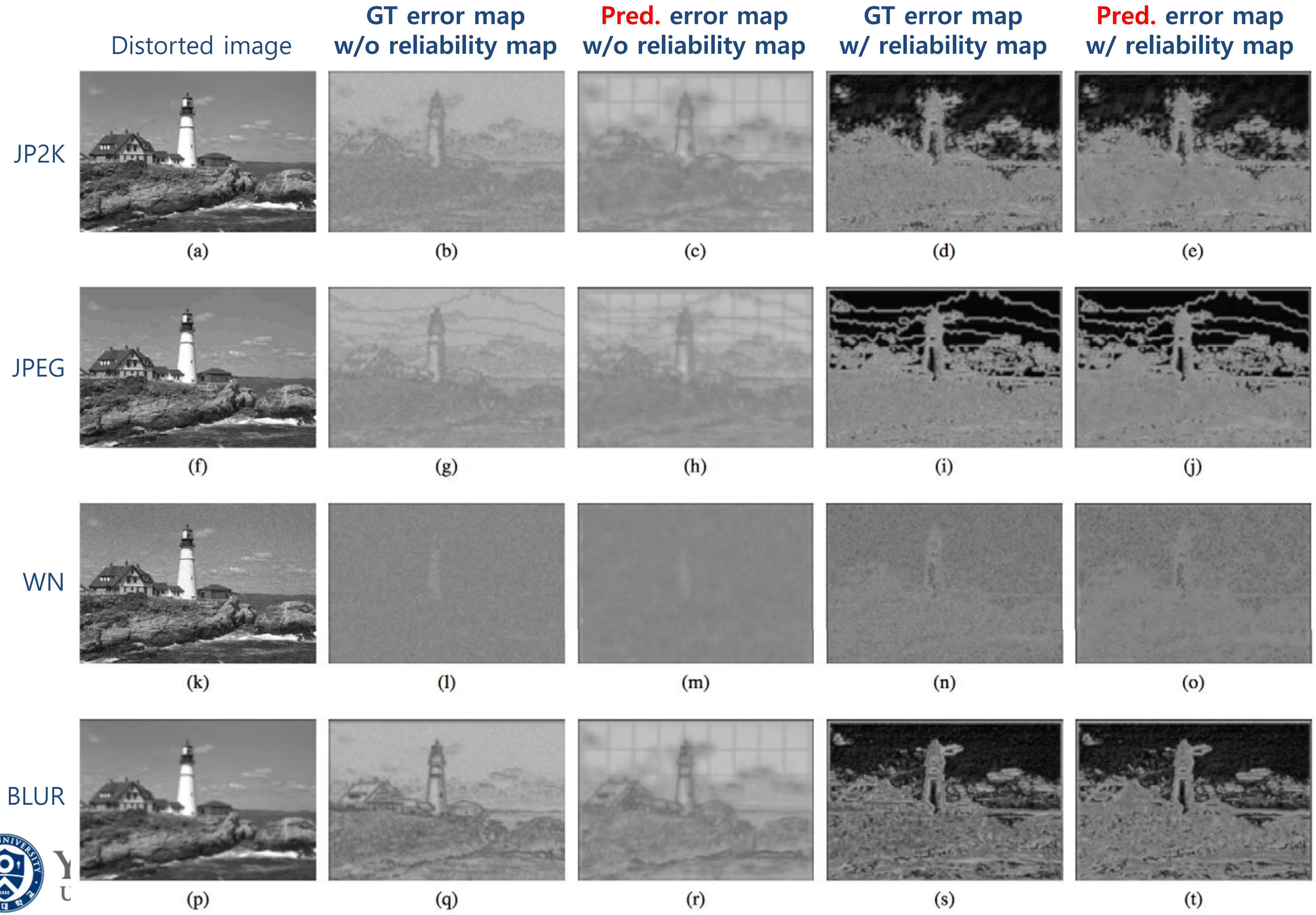
YONSEI UNIVERSITY



Multidimensional Insight

# Error and Perceptual Error Maps Visualization

## ❖ Error Maps Visualization



# Error and Perceptual Error Maps Visualization

## ❖ Perceptual Error Maps Visualization

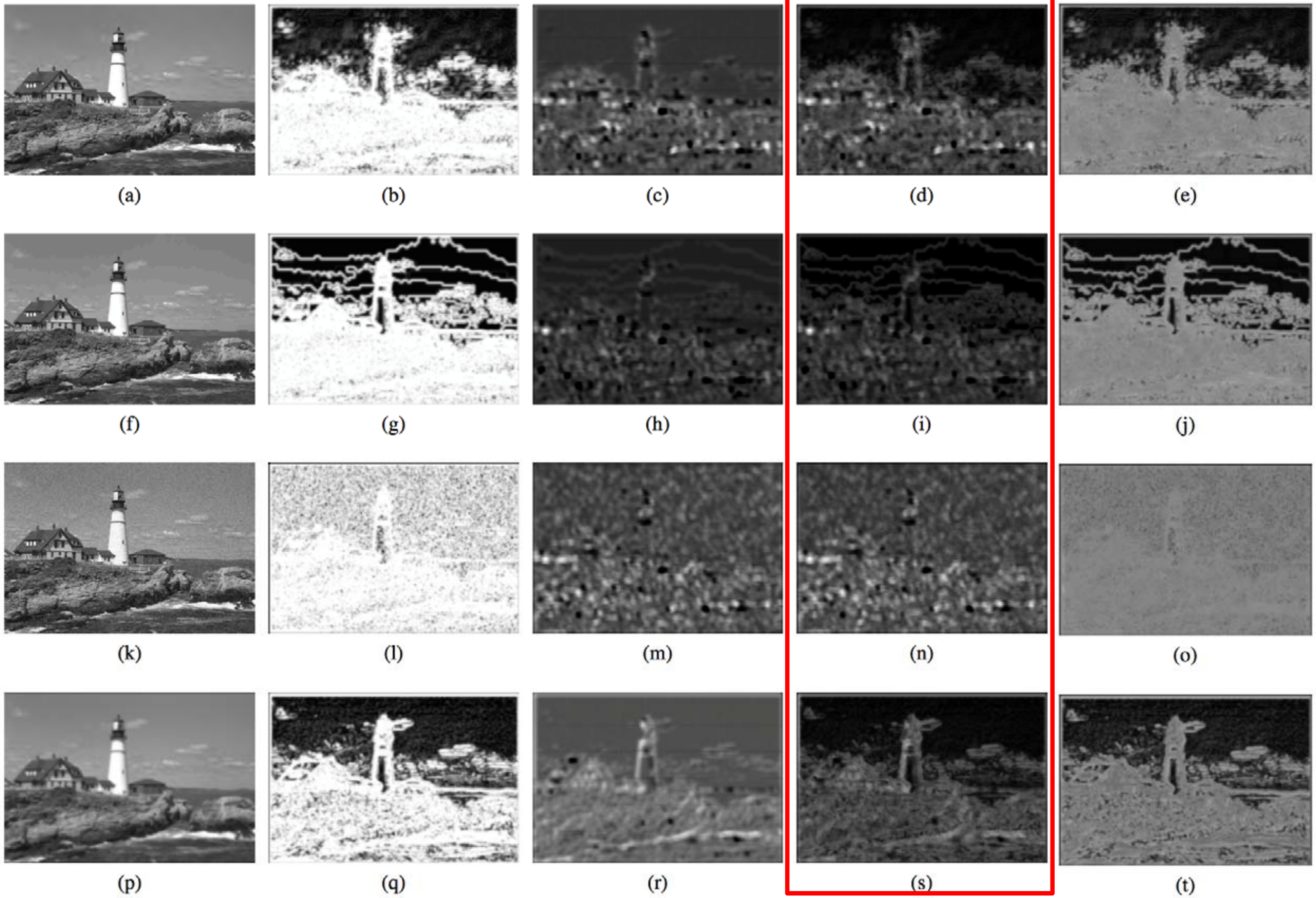
Distorted image | Reliability map | Pred. sensitivity map | Pred. perceptual error map | Pred. error map w/ reliability map

JP2K

JPEG

WN

BLUR



# Benchmark Result

## ❖ SRCC and PLCC Comparison on the 5 Databases

| | | LIVE IQA | | CSIQ | | TID2013 | |
|---|---|---|---|---|---|---|---|
| | | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| FR | PSNR | 0.876 | 0.872 | 0.806 | 0.800 | 0.636 | 0.706 |
| | SSIM | 0.948 | 0.945 | 0.876 | 0.861 | 0.775 | 0.691 |
| | FSIMc | 0.963 | 0.960 | 0.931 | 0.919 | 0.851 | 0.877 |
| | *DeepQA* | 0.981 | 0.982 | 0.961 | 0.956 | 0.939 | 0.947 |
| NR | BLIINDSII | 0.912 | 0.916 | 0.780 | 0.832 | 0.536 | 0.628 |
| | BRISQUE | 0.939 | 0.942 | 0.775 | 0.817 | 0.572 | 0.651 |
| | CORNIA | 0.942 | 0.943 | 0.714 | 0.781 | 0.549 | 0.613 |
| | IL-NIQE | 0.902 | 0.908 | **0.821** | **0.865** | 0.521 | 0.648 |
| | GMLOG | **0.950** | **0.954** | 0.803 | 0.812 | **0.675** | **0.683** |
| | *BIECON* | **0.958** | **0.962** | **0.825** | **0.838** | **0.721** | **0.765** |
| | *DeepBQA* | **0.970** | **0.971** | **0.858** | **0.879** | **0.843** | **0.868** |

- *Italic*: Deep learning

# Conclusion

- Error map – Pixel-level data augmentation
- Reliability map – Getting a more accurate sensitivity map
- Sensitivity map – Analysis of human visual system
- Perceptual error map – Delivery of insight of human perception on distortion
- DeepBQA achieved state of the art performance