Maximum Differentiation Competition:
Direct Comparison of Discriminability Models

Zhou Wang & Eero P. Simoncelli

Howard Hughes Medical Institute,
Center for Neural Science, and
Courant Institute for Mathematical Sciences
New York University
Which model best accounts for perceived image quality?

Image Quality Assessment

Wang & Simoncelli, VSS-2005
Which model best accounts for perceived image quality?
Image Quality Assessment

Which model best accounts for perceived image quality?

Wang & Simoncelli, VSS-2005
Example Models

MSE: Mean Squared Error
\[ E(X, Y) = \frac{1}{N} \sum_i (x_i - y_i)^2 \]

SSIM: Structural Similarity
[Wang, et. al. ’04]

- local cross-correlation measure:
\[ s(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \]

- pooling
\[ S(X, Y) = \frac{\sum_i w(x_i, y_i) s(x_i, y_i)}{\sum_i w(x_i, y_i)} \]

where
\[ w(x, y) = \log_2(1 + \sigma_x^2 / C) + \log_2(1 + \sigma_y^2 / C) \]

Wang & Simoncelli, VSS-2005
Conventional Method

• Procedure

1. Choose set of reference and distorted images
2. Perform subjective tests
3. Compare model prediction with subjective responses

• Difficulties
  – Subjective experiments expensive
  – “Curse of dimensionality”: impossible to cover image space
Conventional Method: MSE vs. SSIM

"LIVE" image database, UT Austin
Conventional Method: MSE vs. SSIM

<table>
<thead>
<tr>
<th>Distortion:</th>
<th>JP2(1)</th>
<th>JP2(2)</th>
<th>JPG(1)</th>
<th>JPG(2)</th>
<th>Noise</th>
<th>Blur</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td># images:</td>
<td>87</td>
<td>82</td>
<td>87</td>
<td>88</td>
<td>145</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>MSE</td>
<td>0.934</td>
<td>0.895</td>
<td>0.902</td>
<td>0.914</td>
<td><strong>0.987</strong></td>
<td>0.774</td>
<td>0.881</td>
</tr>
<tr>
<td>SSIM</td>
<td><strong>0.968</strong></td>
<td><strong>0.967</strong></td>
<td><strong>0.965</strong></td>
<td><strong>0.986</strong></td>
<td>0.971</td>
<td><strong>0.936</strong></td>
<td><strong>0.944</strong></td>
</tr>
</tbody>
</table>

Wang & Simoncelli, VSS-2005
Proposed Method:

MMaximum Differentiation (MAD) Competition
Proposed Method:
MAximum Differentiation (MAD) Competition

• Let two models compete
Proposed Method:
MAximum Differentiation (MAD) Competition

• Let two models compete

• ... by synthesizing optimal stimuli
Proposed Method: MAximum Differentiation (MAD) Competition

- Let two models compete
- ... by synthesizing optimal stimuli
- ... that maximally differentiate the models
Geometric Description in Image Space

reference image

initial image
Geometric Description in Image Space

all images with same MSE
Geometric Description in Image Space

all images with same SSIM
Geometric Description in Image Space

worst MSE

reference image

initial image
Geometric Description in Image Space

- Best MSE
- Worst MSE
- Reference Image
- Initial Image

Wang & Simoncelli, VSS-2005
Geometric Description in Image Space
Geometric Description in Image Space

worst SSIM

reference image

initial image
Geometric Description in Image Space

reference image

initial image
MAD Competition: MSE vs. SSIM

reference → add noise
MAD Competition: MSE vs. SSIM

reference

best MSE

worst MSE
MAD Competition: MSE vs. SSIM

Reference

Best MSE

Best SSIM

Worst MSE

Worst SSIM
2AFC Experiment

distortion level (MSE)

2^2  2^3  2^4  2^5  2^6  2^7  2^8

initial image

best SSIM

worst SSIM
• Subjects: 5 (4 naïve, 1 author)
• Images: 10 reference, viewed at 16 pixels/degree
• Trials: 20 per distortion-level per subject
Psychometric Functions

![Graph showing psychometric functions with initial distortion level (MSE) on the x-axis and % correct on the y-axis. The graph includes data points for best/worst SSIM and best/worst MSE.]
Psychometric Functions

all 5 subjects chose top

1 chose top twice
2 chose bottom twice
2 gave 1-1 tie

Wang & Simoncelli, VSS-2005
Summary

• MAximum Differentiation (MAD) Competition
  – Let two models compete
  – ... by synthesizing optimal stimuli
  – ... that maximally differentiate the models

• Advantages
  – Optimized images maximize opportunity for model failure
  – Efficient (minimal # of 2-alternative comparisons)
  – Images reveal model weaknesses => potential improvements

• To Do
  – Full experiment, with more reference images
  – Application to other discriminable quantities
  – Physiology