Machine Learning for Quality Assessment
Adopt – Adapt – Improve

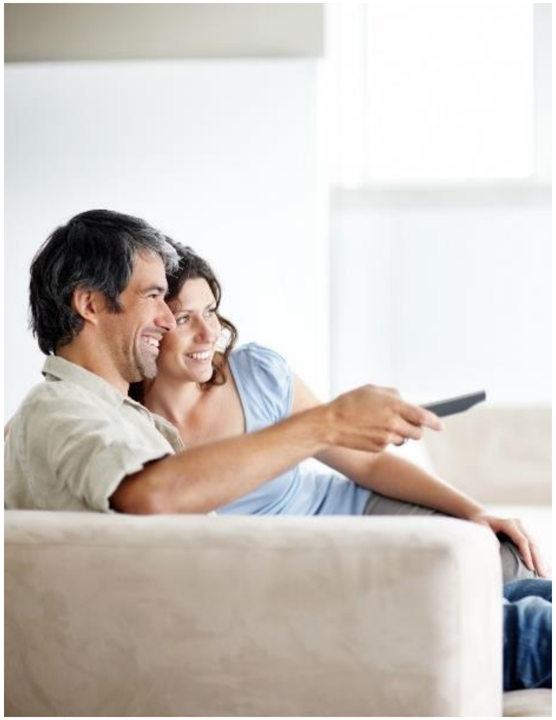
Adriaan Barri
abarri@etro.vub.ac.be

Quality of Experience can make or break your product.
Do you really want to rely on machines to monitor it?

The user expectations on quality are application-dependent

The user expectations on quality are application-dependent

Audience



The user expectations on quality are application-dependent

Device

The user expectations on quality are application-dependent

Content

For more reliable quality estimations, optimize your quality measures

Application A

Quality Measure 1

For more reliable quality estimations, optimize your quality measures

~~Quality Measure 1~~

Quality Measure 2

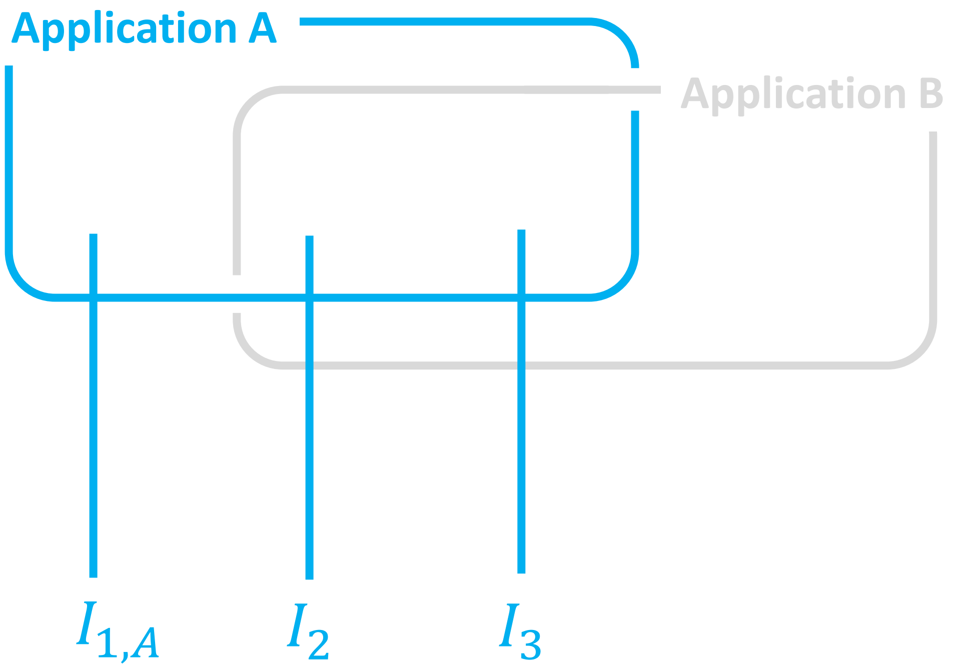
Application B

+

+

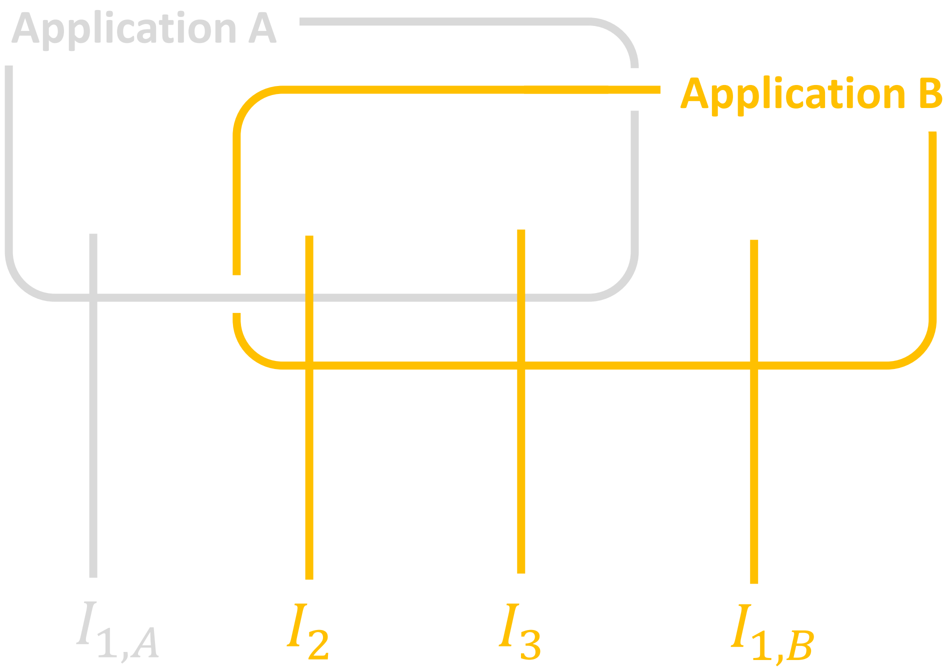
For more reliable quality estimations, optimize your quality measures

- Carefully **select** your quality indicators



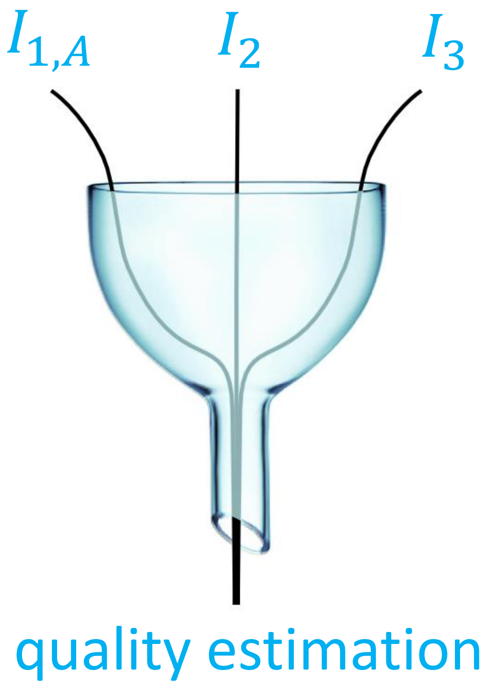
For more reliable quality estimations, optimize your quality measures

- Carefully **select** your quality indicators



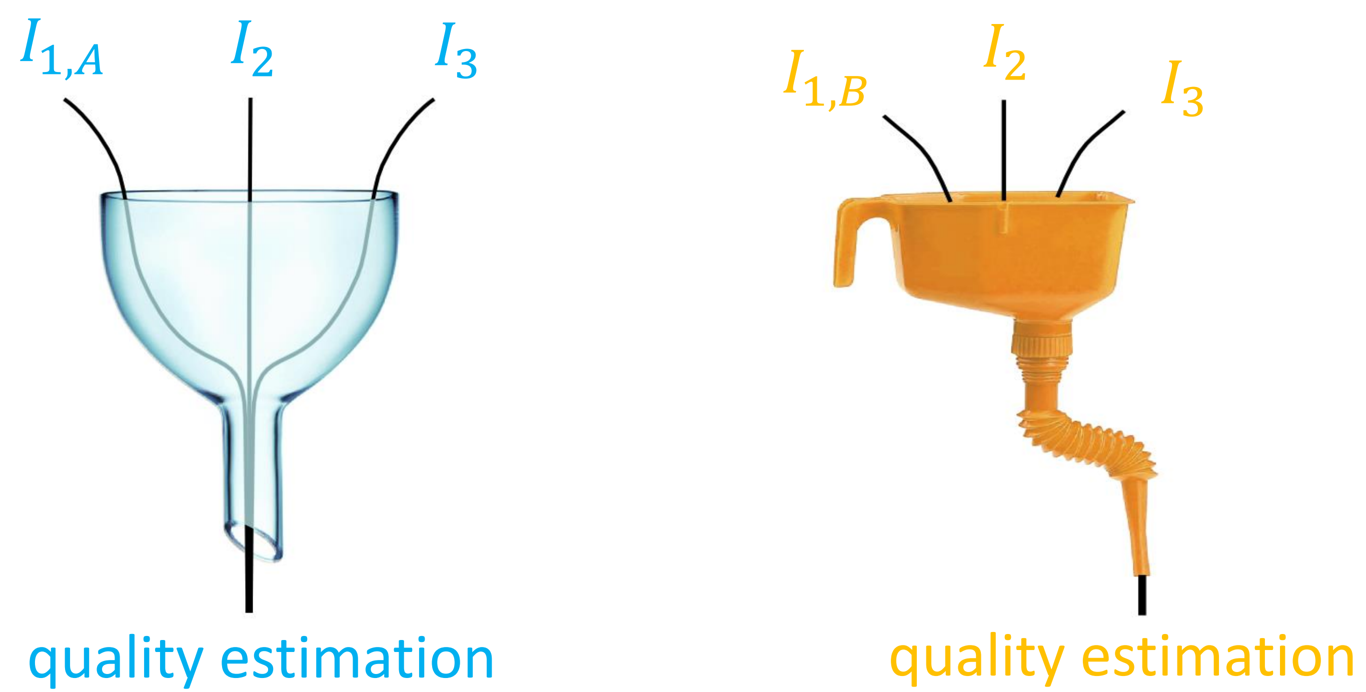
For more reliable quality estimations, optimize your quality measures

- Carefully **select** your quality indicators
- Carefully **combine** your quality indicators



For more reliable quality estimations, optimize your quality measures

- Carefully **select** your quality indicators
- Carefully **combine** your quality indicators

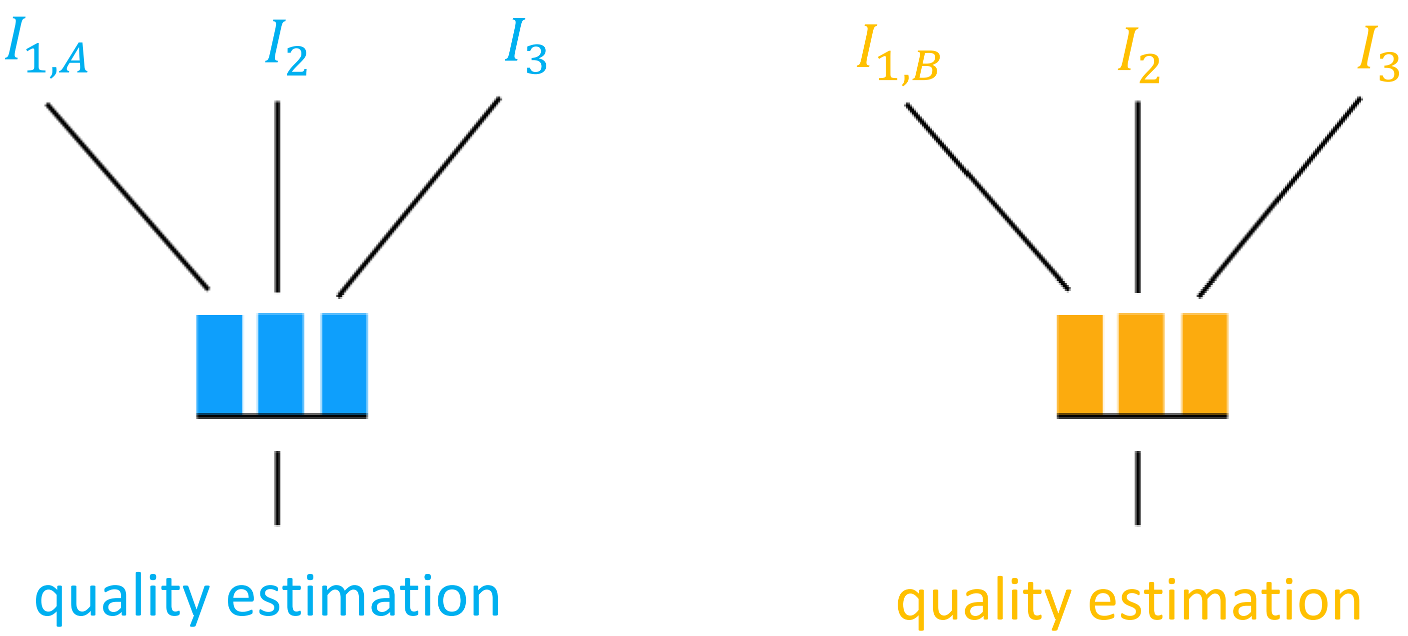


For more reliable quality estimations,
optimize your quality measures

- Carefully **select** your quality indicators
- Carefully **combine** your quality indicators

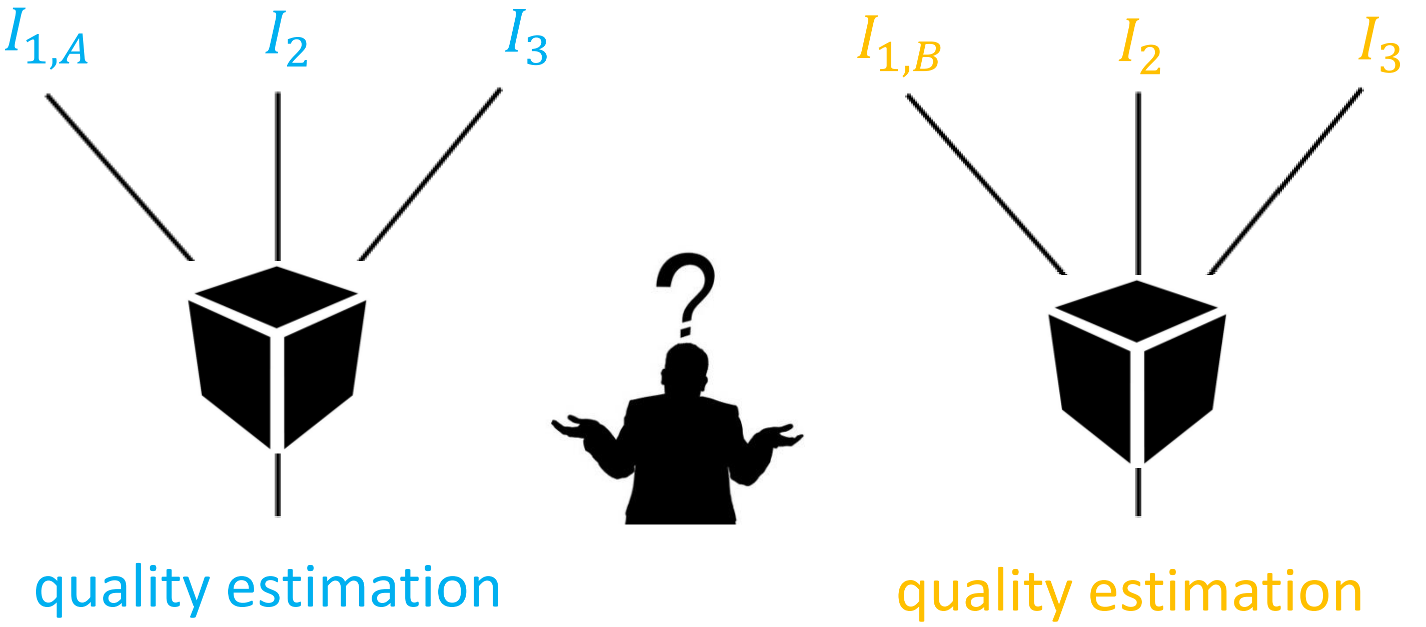
Do not forget to also optimize your combination method!

Current combination methods are not optimized for quality assessment



Heuristic combination methods do not adapt to the application

Current combination methods are not optimized for quality assessment



Traditional machine learning is often difficult to interpret

Machine learning for quality assessment in three steps: adopt, adapt and improve

Adopt machine learning in quality assessment

Adapt machine learning to the quality framework

Improve the reliability of the quality predictions

Machine learning for quality assessment in three steps: adopt, adapt and improve

Adopt machine learning in quality assessment

Adapt machine learning to the quality framework

Improve the reliability of the quality predictions

LOCALLY ADAPTIVE FUSION

For the purpose of quality assessment machine learning is incorporated as follows

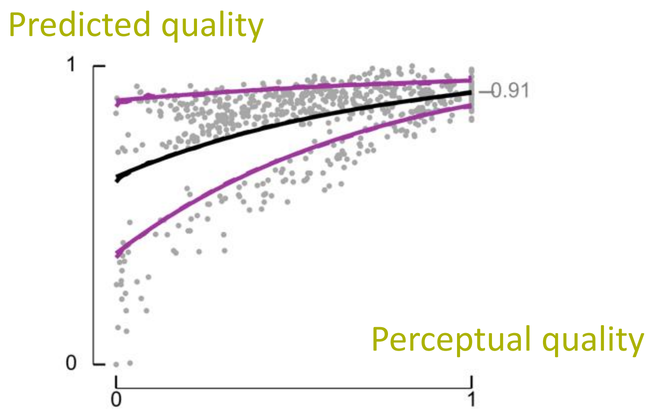
1. **Select input quality indicators**
   simple, fast formulas for quality predictions
   in specific content and distortion classes

ADOPT

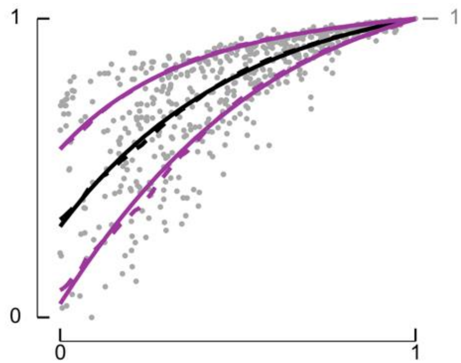
2. **Train the ML system on a quality assessment database**
   on which the quality indicators are evaluated and
   combined to maximize the prediction accuracy

3. **Apply the ML system to newly received signals**
   using the weights obtained during training

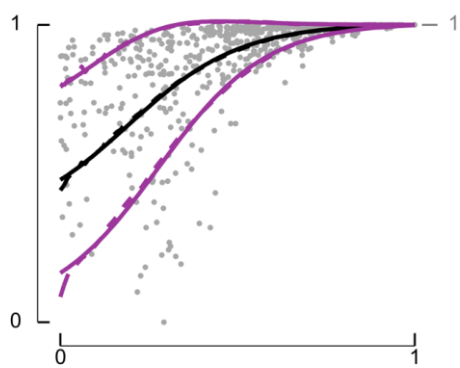
EXAMPLE on the LIVE image database



Indicator 1
Blockiness



Indicator 2
Information Loss



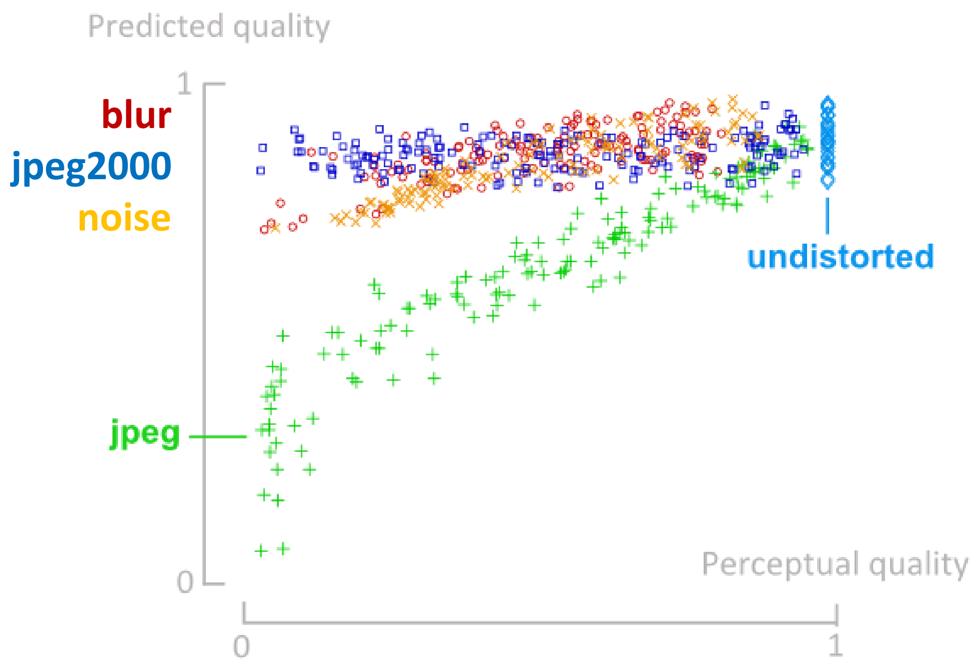
Indicator 3
Contrast Similarity

ADOPT

1. Input selection
2. Training
3. Application

Indicator 1

Blockiness (LIVE)

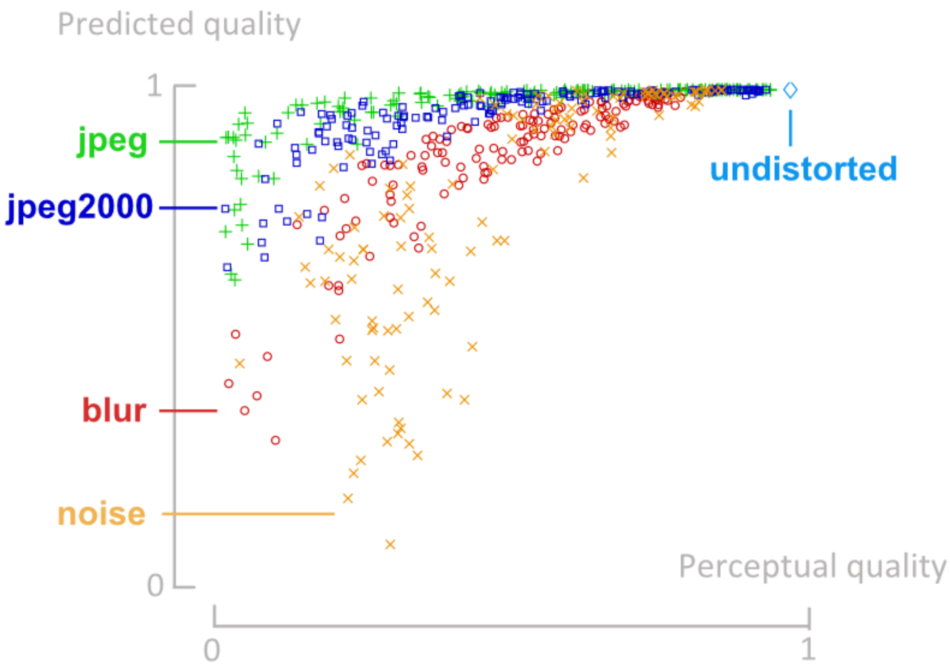


- **jpeg**
- only for high distortion rates
- no-reference

ADOPT

1. Input selection
2. Training
3. Application

Indicator 2
Contrast (LIVE)

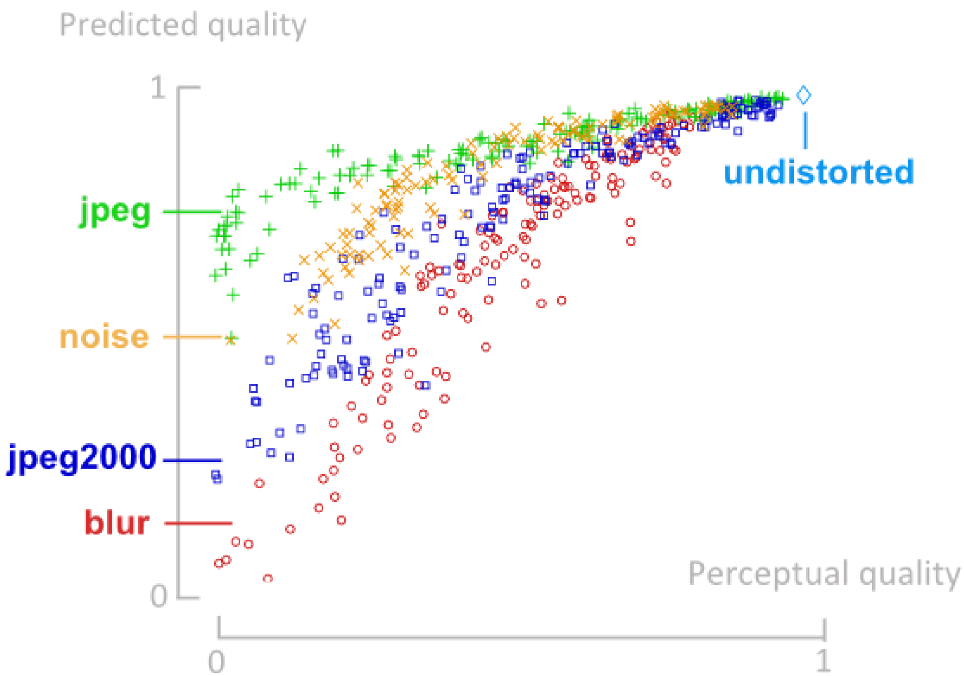


ADOPT

1. Input selection
2. Training
3. Application

- noise and blur
- low and high distortion rates
- reduced-reference

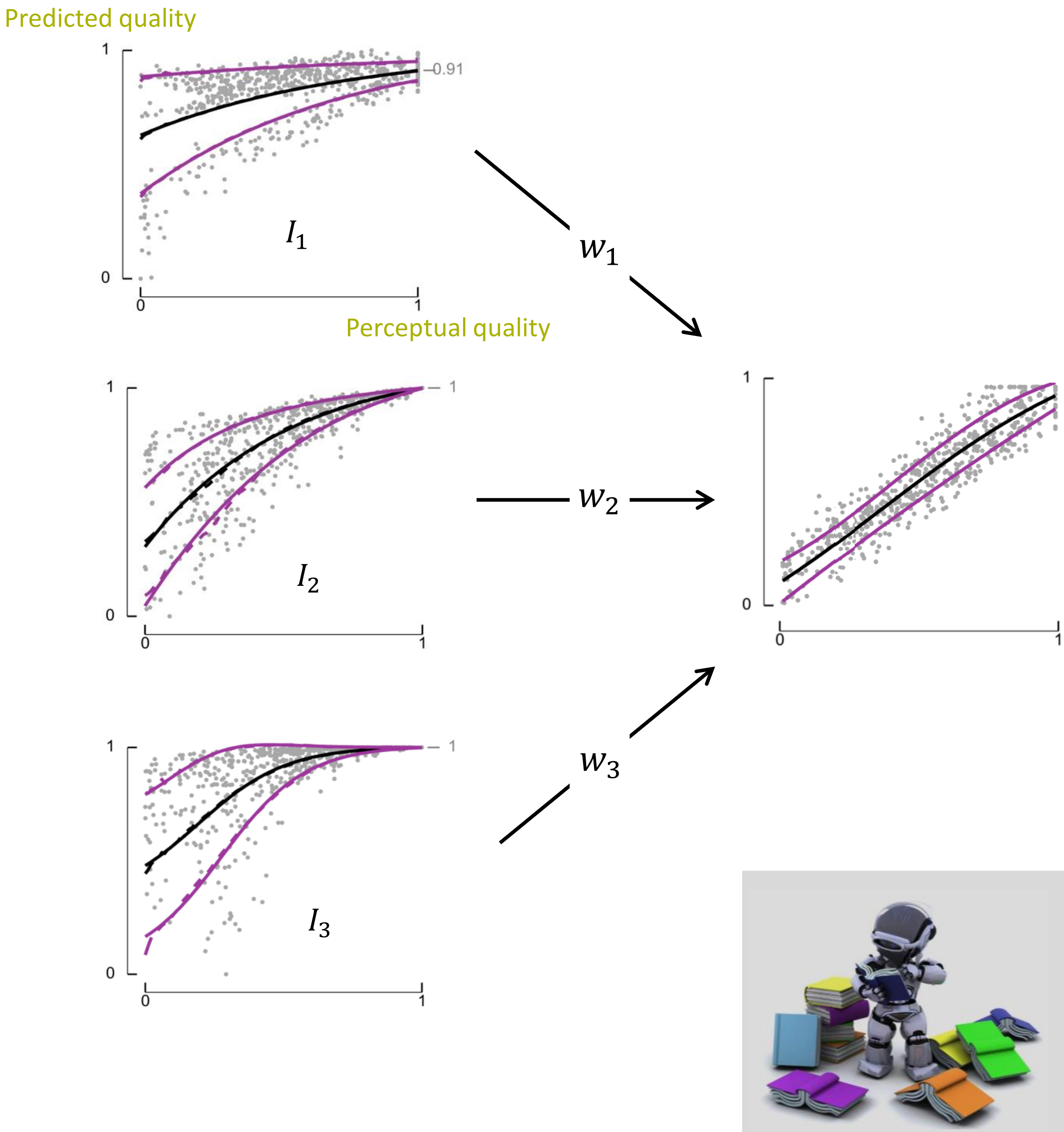
Indicator 3
Information loss (NTIA)



ADOPT

1. Input selection
2. Training
3. Application

- **blur** and **jpeg2000**
- low and high distortion rates
- reduced-reference



ADOPT

1. Input selection
2. **Training**
3. Application

$$w_1 = 0.12$$
$$w_2 = 0.57$$
$$w_3 = 0.31$$

x
newly received signal

$I_1(x)$

$I_2(x)$

$I_3(x)$

ML-based quality prediction
$\sum_i w_i I_i(x)$

ADOPT

1. Input selection
2. Training
3. **Application**

$w_1 = 0.12$
$w_2 = 0.57$
$w_3 = 0.31$

x

newly received signal

$I_1(x)$

$I_2(x)$

$I_3(x)$

ML-based quality prediction

$\sum_i w_i I_i(x)$

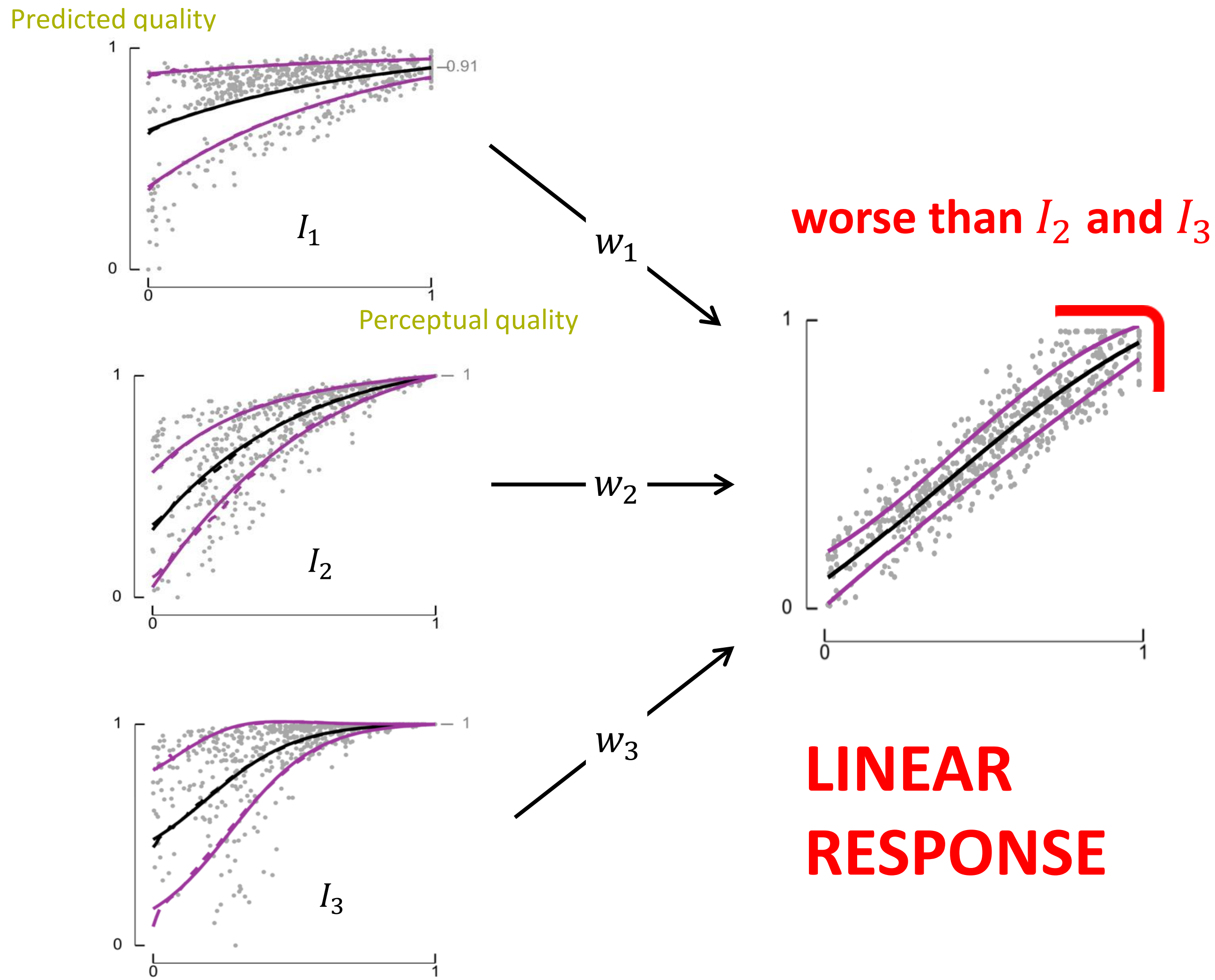
ADOPT

1. Input selection
2. Training
3. Application

Simple linear combinations do not provide enough flexibility

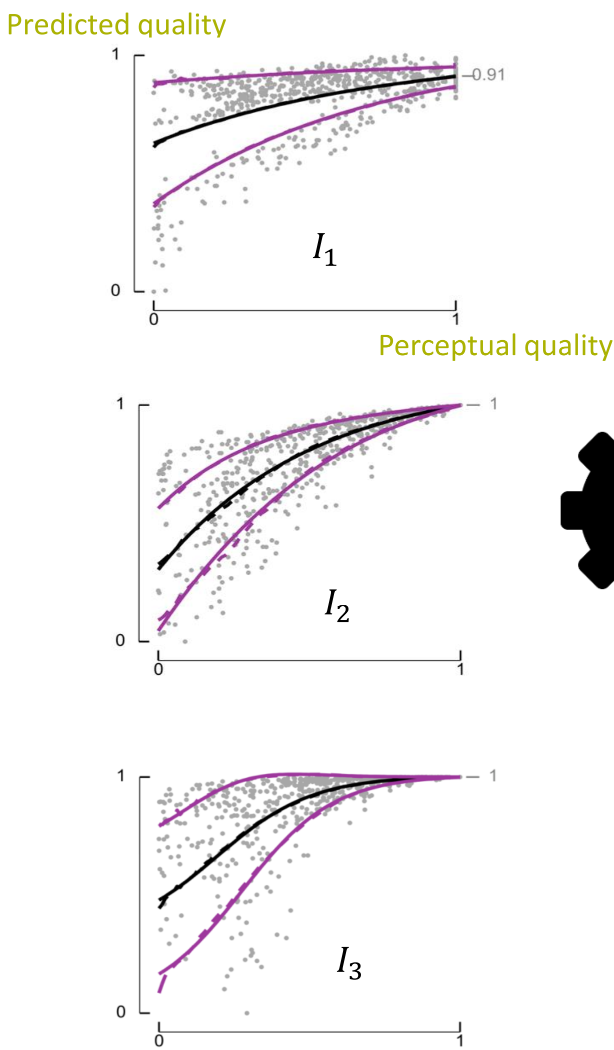
$$w_1 = 0.12$$
$$w_2 = 0.57$$
$$w_3 = 0.31$$

To better adapt to the perceptual mechanisms the ML response is preferably nonlinear

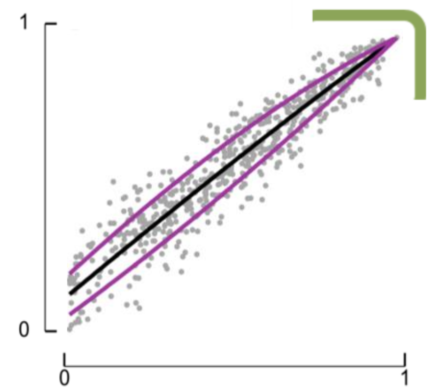


ADAPT

To better adapt to the perceptual mechanisms the ML response is preferably nonlinear



More flexibility allows a higher prediction accuracy

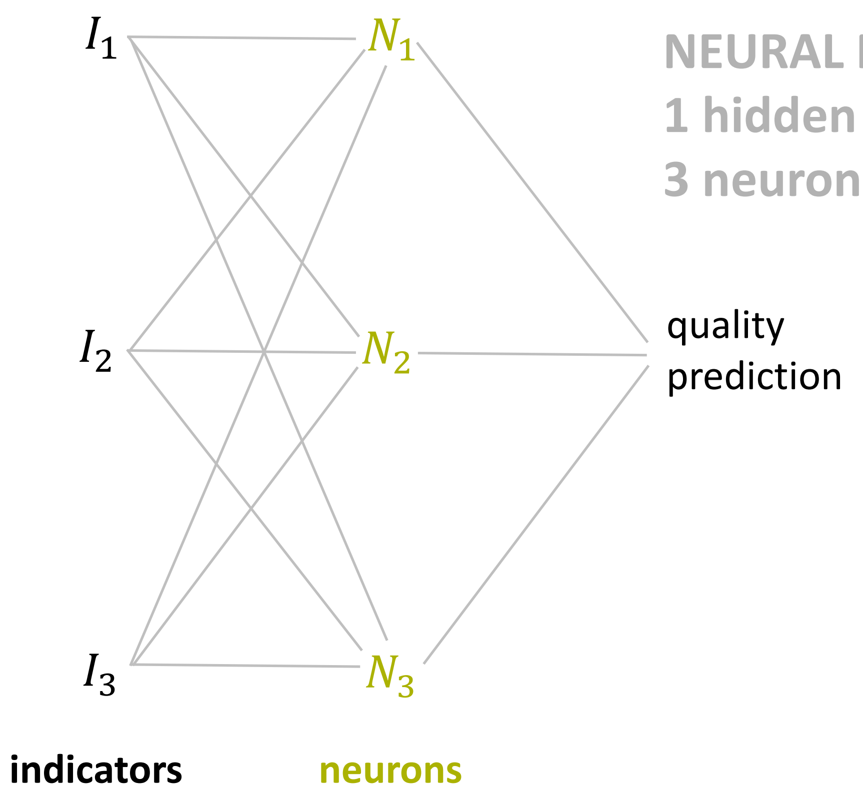


ADAPT

NONLINEAR RESPONSE

For nonlinear ML systems, the weights are typically very difficult to interpret

The hidden neurons of neural networks do not necessarily have a meaning



NEURAL NETWORK
1 hidden layer
3 neurons

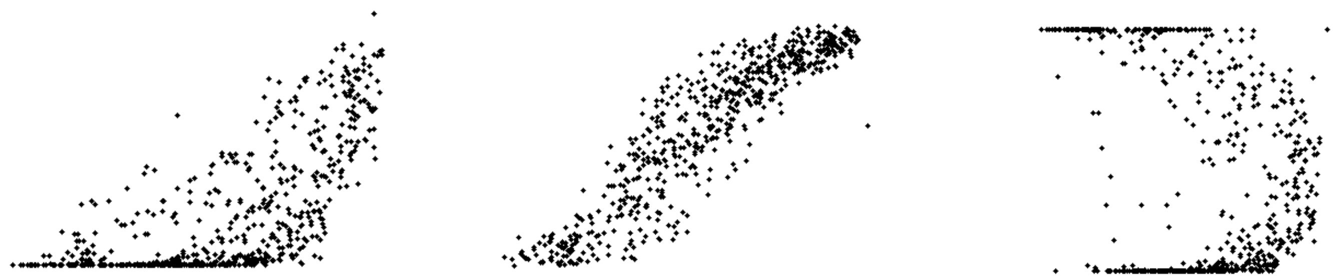
ADAPT

For nonlinear ML systems, the weights are typically very difficult to interpret

The hidden neurons of neural networks
do not necessarily have a meaning

EXAMPLE on the LIVE image database

N_1 N_2 N_3



ADAPT

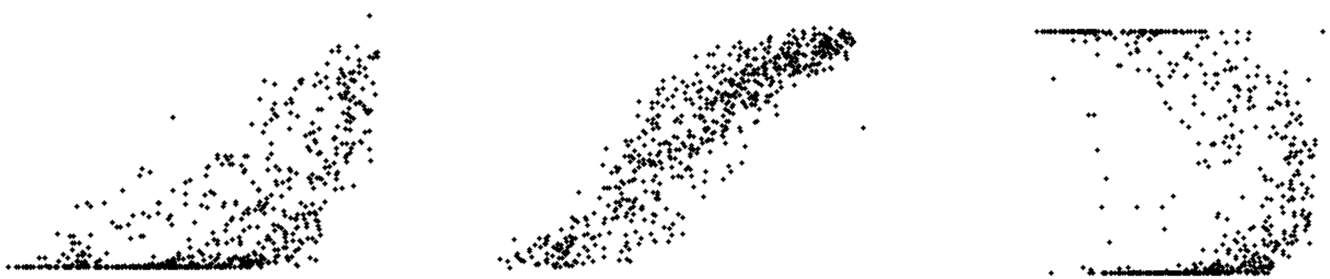
For nonlinear ML systems, the weights are typically very difficult to interpret

ADAPT

The hidden neurons of neural networks
do not necessarily have a meaning

EXAMPLE on the LIVE image database

N_1 N_2 N_3

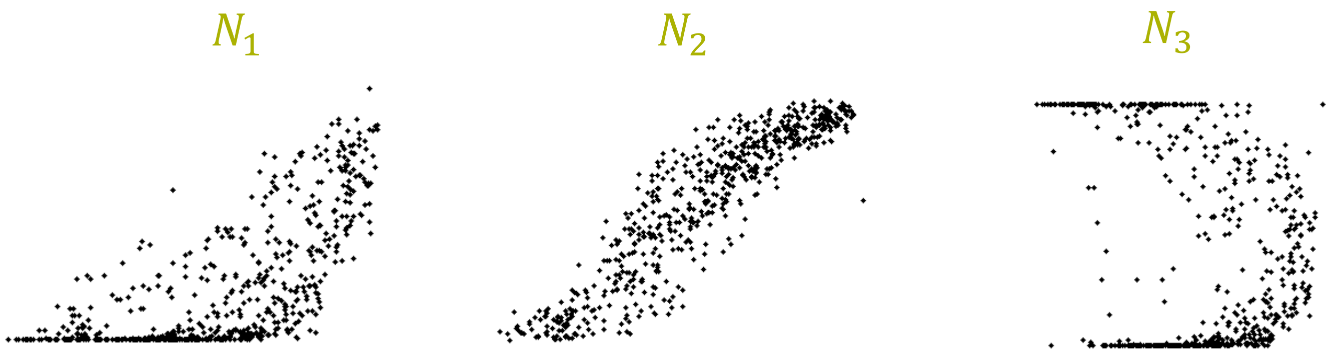


Do not look for patterns
after receiving the data!

For nonlinear ML systems, the weights are typically very difficult to interpret

The hidden neurons of neural networks
do not necessarily have a meaning

EXAMPLE on the LIVE image database



ADAPT

The harder it is to interpret the ML behavior
The easier it will be to disguise the vulnerabilities

There are plenty of vulnerabilities of traditional machine learning

Many vulnerabilities are found using a
large unannotated **stress test database**
Idea of F. Ciaramello and A. Reibman

ADAPT

We performed **three stress test** on a database
of 650 reference images from Wikimedia Commons
and 26000 distorted images

Input quality indicators: Blockiness, Contrast, Information Loss
Training database: LIVE image database

Stress test 1

Machine learning inconsistencies

All quality indicators prefer signal x_1 → x_1

x_2 ← Traditional machine learning may prefer signal x_2

ADAPT

Stress test 1

Machine learning inconsistencies

All quality indicators prefer signal x_1 → x_1

x_2 ← Traditional machine learning may prefer signal x_2

Linear Regression
Principal Component Regression (PCR)

Parametric ML
Feed Forward Neural Network (FFNN)

Kernel-based ML
General Regression Neural Network (GRNN)

ADAPT

Stress test 1
Machine learning inconsistencies

All quality indicators prefer signal x_1 → x_1

x_2 ← Traditional machine learning may prefer signal x_2

Linear Regression
Principal Component Regression (PCR)

Parametric ML
Feed Forward Neural Network (FFNN)

Kernel-based ML
General Regression Neural Network (GRNN)

ADAPT

Linear regression will never cause inconsistencies (when properly trained)

Stress test 1
Machine learning inconsistencies

All quality indicators prefer signal x_1 → x_1

x_2 ← Traditional machine learning may prefer signal x_2

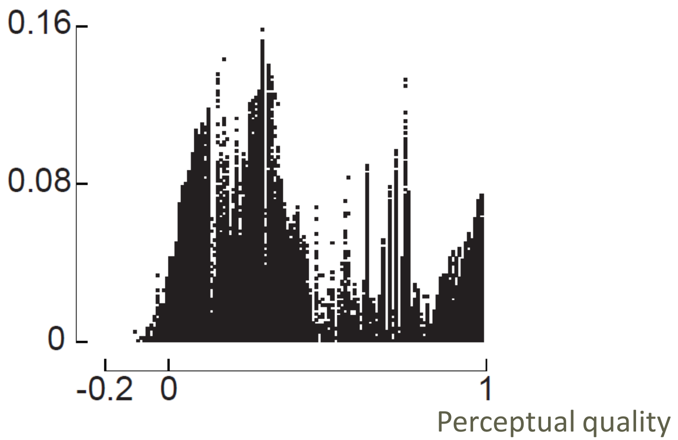
Linear Regression
Principal Component Regression (PCR)

Parametric ML
Feed Forward Neural Network (FFNN)

Kernel-based ML
General Regression Neural Network (GRNN)

ADAPT

Error magnitude up to 16%



more than 100,000 inconsistencies on the stress test database

Stress test 1

Machine learning inconsistencies

All quality indicators prefer signal x_1 → x_1

x_2 ← Traditional machine learning may prefer signal x_2

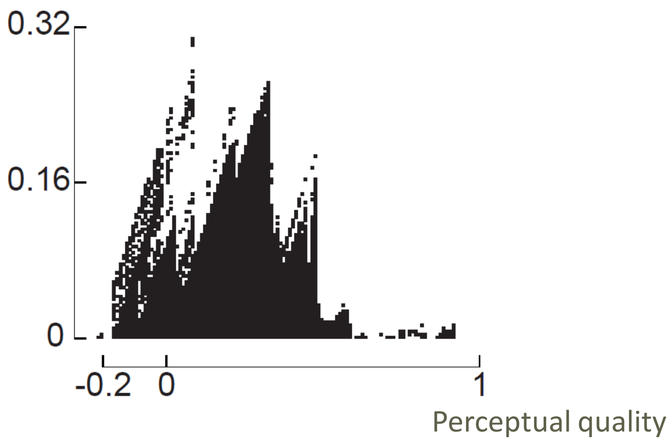
Linear Regression
Principal Component Regression (PCR)

Parametric ML
Feed Forward Neural Network (FFNN)

Kernel-based ML
General Regression Neural Network (GRNN)

ADAPT

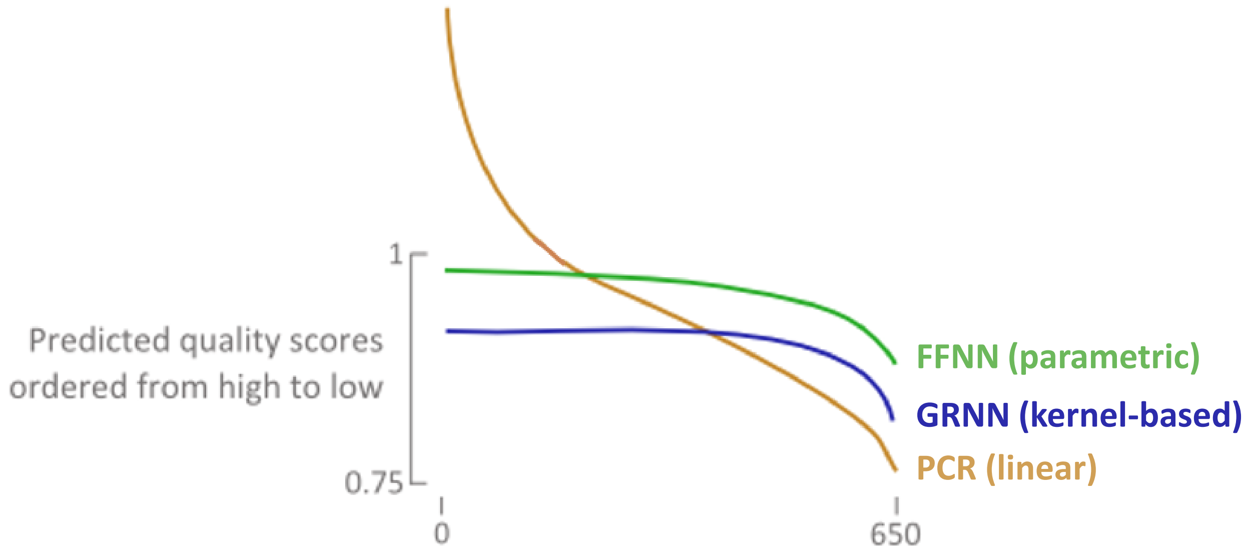
Error magnitude up to 32%



more than 1,000,000 inconsistencies on the stress test database

Stress test 2
Quality estimations of the reference images

- **Behavior of linear, parametric, and kernel-based ML when applied to 650 high quality reference images**



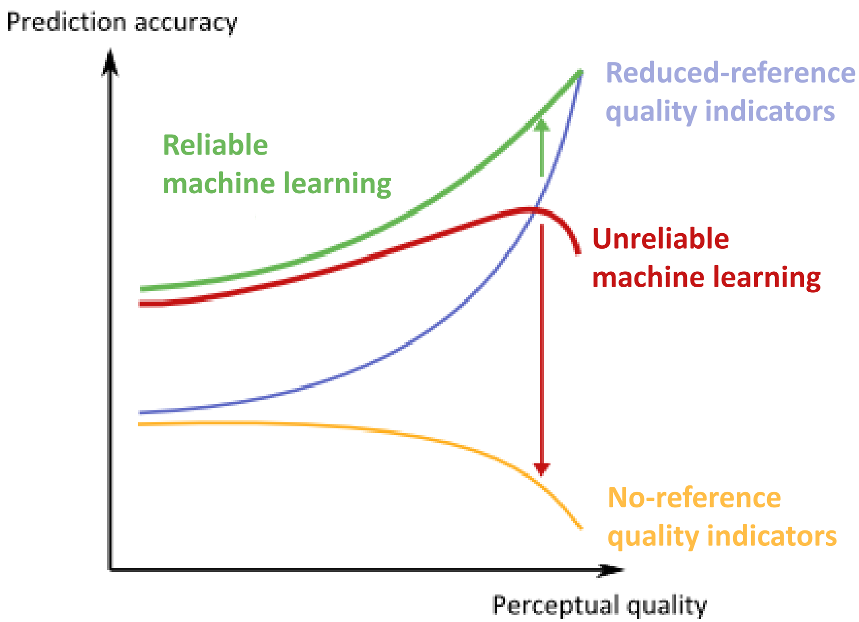
ADAPT

Stress test 2
Quality estimations of the reference images

- Behavior of linear, parametric, and kernel-based ML when applied to 650 high quality reference images

- **Explanation of the unreliable quality scores**

ADAPT



Stress test 3

False orderings

The quality predictions should tend to decrease when the distortion rate is increased.

ADAPT

Stress test 3
False orderings

The quality predictions should tend to decrease when the distortion rate is increased.

Nonlinear machine learning may cause severe false orderings

ADAPT

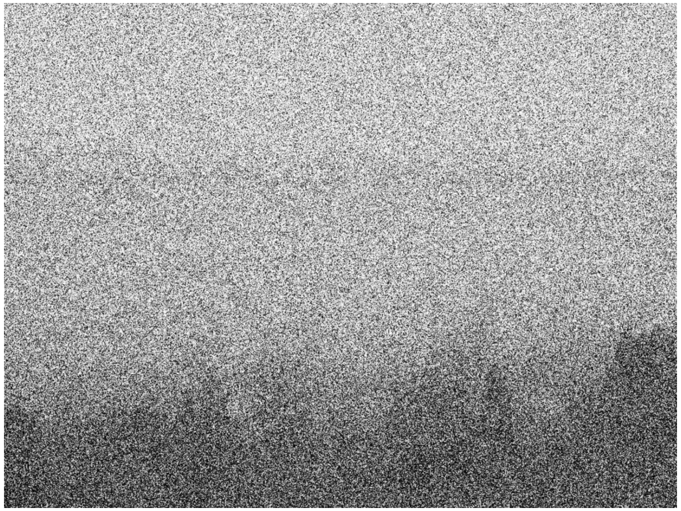
Stress test 3
False orderings

The quality predictions should tend to decrease when the distortion rate is increased.

Nonlinear machine learning may cause severe false orderings

ADAPT

Preferred by FFNN (parametric ML)



Stress test 3
False orderings

The quality predictions should tend to decrease when the distortion rate is increased.

Nonlinear machine learning may cause severe false orderings

ADAPT

Preferred by GRNN (kernel-based ML)

To improve the reliability, we need new forms of machine learning

The machine learning system
should be **flexible**, and also **reliable**

IMPROVE

To improve the reliability, we need new forms of machine learning

The machine learning system
should be **flexible**, and also **reliable**

The ML weights should be **interpretable**

IMPROVE

To improve the reliability, we need new forms of machine learning

The machine learning system
should be **flexible**, and also **reliable**

The ML weights should be **interpretable**

The training process should be **reproducible**

Random initializations during training can be abused

IMPROVE

To improve the reliability, we need new forms of machine learning

The machine learning system
should be **flexible**, and also **reliable**

The ML weights should be **interpretable**

The training process should be **reproducible**

Random initializations during training can be abused

Machine learning should be **consistent**

Ignoring the quality indicators to better
fit the training data reduces the reliability

IMPROVE

To improve the reliability, we need new forms of machine learning

The machine learning system
should be **flexible**, and also **reliable**

The ML weights should be **interpretable**

The training process should be **reproducible**

Random initializations during training can be abused

Machine learning should be **consistent**

Ignoring the quality indicators to better
fit the training data reduces the reliability

The combined measure should be
optimized on the entire quality range

IMPROVE

Combining quality indicators using the Locally Adaptive Fusion (LAF)

The LAF system involves a training and an application phase

1. **Training on an annotated quality database**
   The quality indicators are transformed
   into locally optimized fusion units

2. **Application on a newly received signal**
   by combining the fusion unit values
   using a set of adaptive weights

IMPROVE

The LAF training phase consists of two steps

1. Determine a set of target values r_i of the perceptual quality
2. Associate a fusion unit U_i with each target value r_i

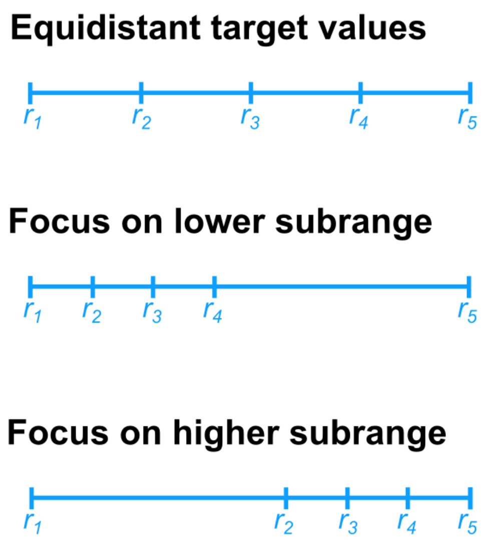
IMPROVE

The LAF training phase consists of two steps

1. **Determine a set of target values r_i of the perceptual quality**
2. Associate a fusion unit U_i with each target value r_i

The more target values, the better the covering of the perceptual quality.

Non-equidistant target values focus on subranges of the perceptual quality



IMPROVE

1. LAF training
2. LAF application

The LAF training phase consists of two steps

1. Determine a set of target values r_i of the perceptual quality
2. **Associate a fusion unit U_i with each target value r_i**

Each fusion unit is a weighted sum of the quality indicators. The used weights $w_{i,j}$ are fixed.

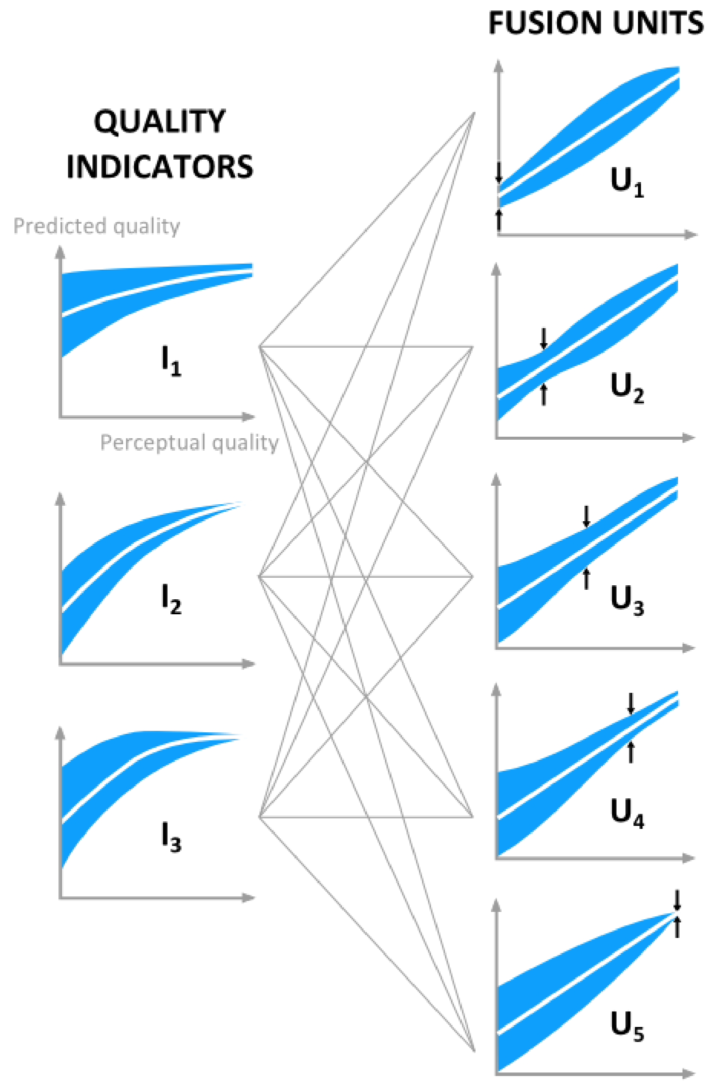
Each fusion unit U_i is optimized for quality predictions near the target value r_i.

	I_1	I_2	I_3
U_1	$w_{1,1}$	$w_{1,2}$	$w_{1,3}$
U_2	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$
U_3	$w_{3,1}$	$w_{3,2}$	$w_{3,3}$
U_4	$w_{4,1}$	$w_{4,2}$	$w_{4,3}$
U_5	$w_{5,1}$	$w_{5,2}$	$w_{5,3}$

During training, LAF transforms the quality indicators into fusion units

ILLUSTRATION

Three quality indicators
Five equidistant target values
Five fusion units



IMPROVE

1. LAF training
2. LAF application

The weights of the fusion units are optimized using the separation ratio

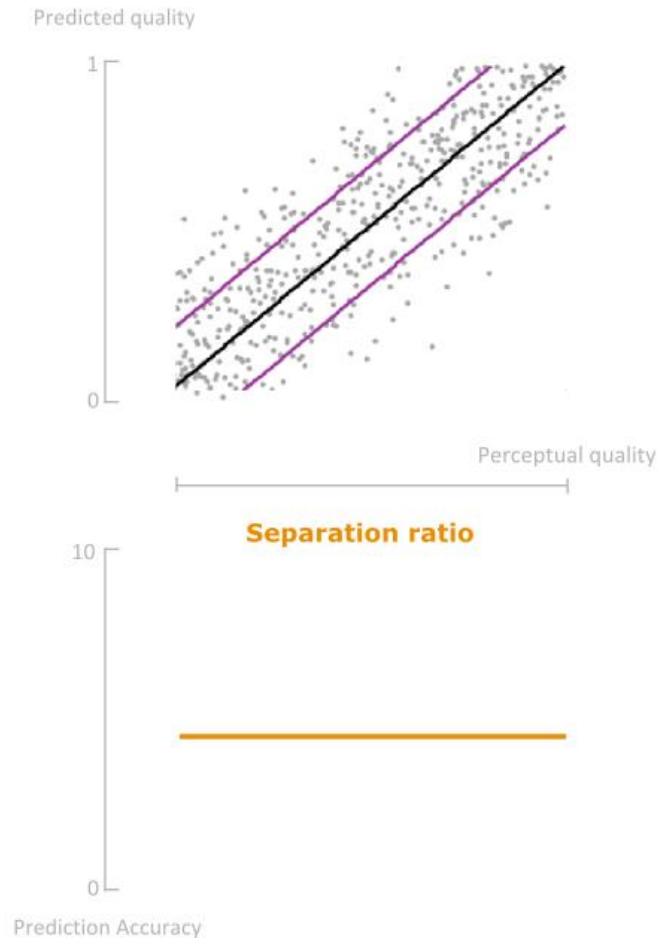
- **Definition of the separation ratio**

The separation ratio of a measure M shows how the local prediction accuracy of M changes in funtion of the perceptual quality

Steepness regression line

$$\text{sep}_{\mathcal{M}}[r] = \frac{\frac{d}{dq}\mathbb{E}[\mathcal{M}(X_q)]\Big|_{q=r}}{\text{std}[\mathcal{M}(X_r)]}$$

Prediction noise



IMPROVE

1. **LAF training**
2. **LAF application**

The weights of the fusion units are optimized using the separation ratio

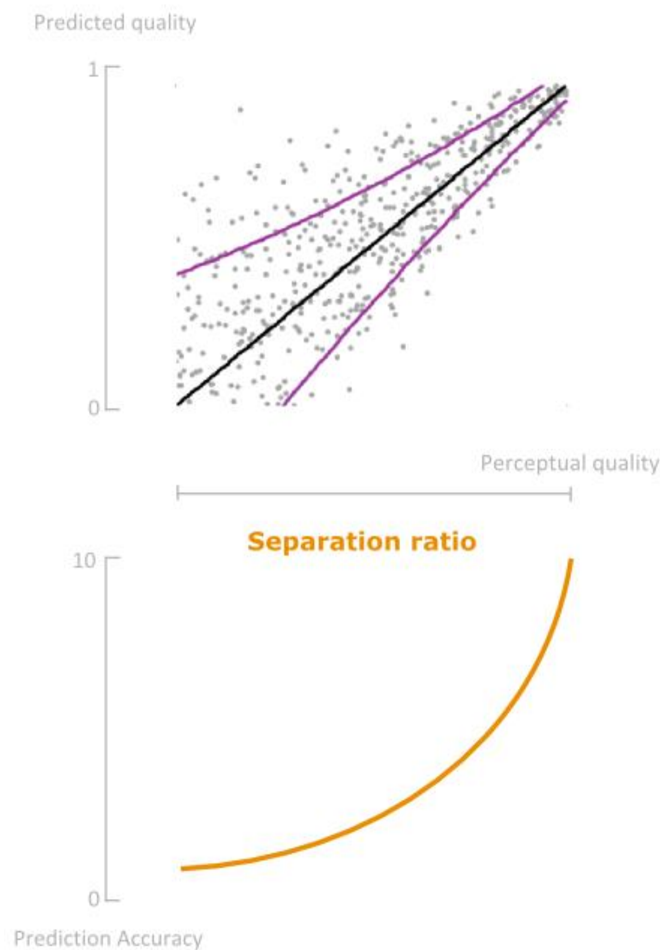
- **Definition of the separation ratio**

The separation ratio of a measure M shows how the local prediction accuracy of M changes in funtion of the perceptual quality

Steepness regression line

$$\mathrm{sep}_{\mathcal{M}}[r] = \frac{\frac{d}{dq}\mathbb{E}[\mathcal{M}(X_q)]\Big|_{q=r}}{\mathrm{std}[\mathcal{M}(X_r)]}$$

Prediction noise



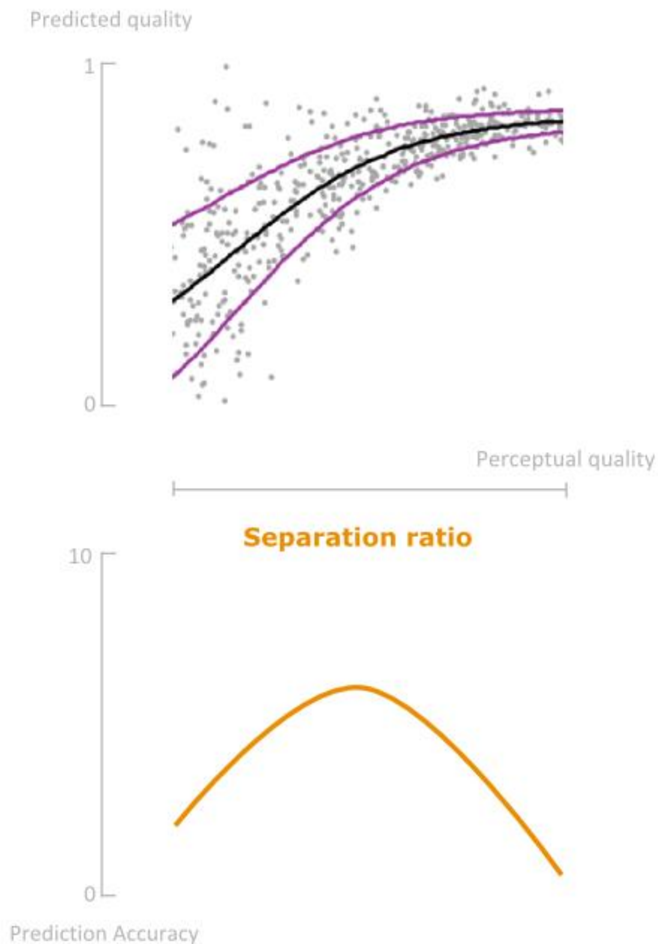
IMPROVE

1. **LAF training**
2. LAF application

The weights of the fusion units are optimized using the separation ratio

- **Definition of the separation ratio**

The separation ratio of a measure M shows how the local prediction accuracy of M changes in funtion of the perceptual quality



Steepness regression line

$$\mathrm{sep}_{\mathcal{M}}[r] = \frac{\frac{d}{dq}\mathbb{E}[\mathcal{M}(X_q)]\Big|_{q=r}}{\mathrm{std}[\mathcal{M}(X_r)]}$$

Prediction noise

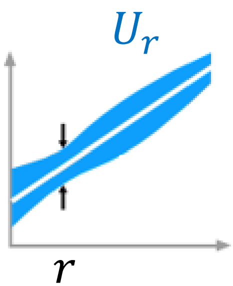
IMPROVE

1. LAF training

2. LAF application

The weights of the fusion units are optimized using the separation ratio

- Definition of the separation ratio
- **Integration of the separation ratio**



Denote $\mathbf{I}$ for the vector of all quality indicators I_j
Then a fusion unit is of the form $U_r = \boldsymbol{w}_r^T \boldsymbol{I}$

The weight vector $\boldsymbol{w}_r$ maximizes
the separation ratio in r

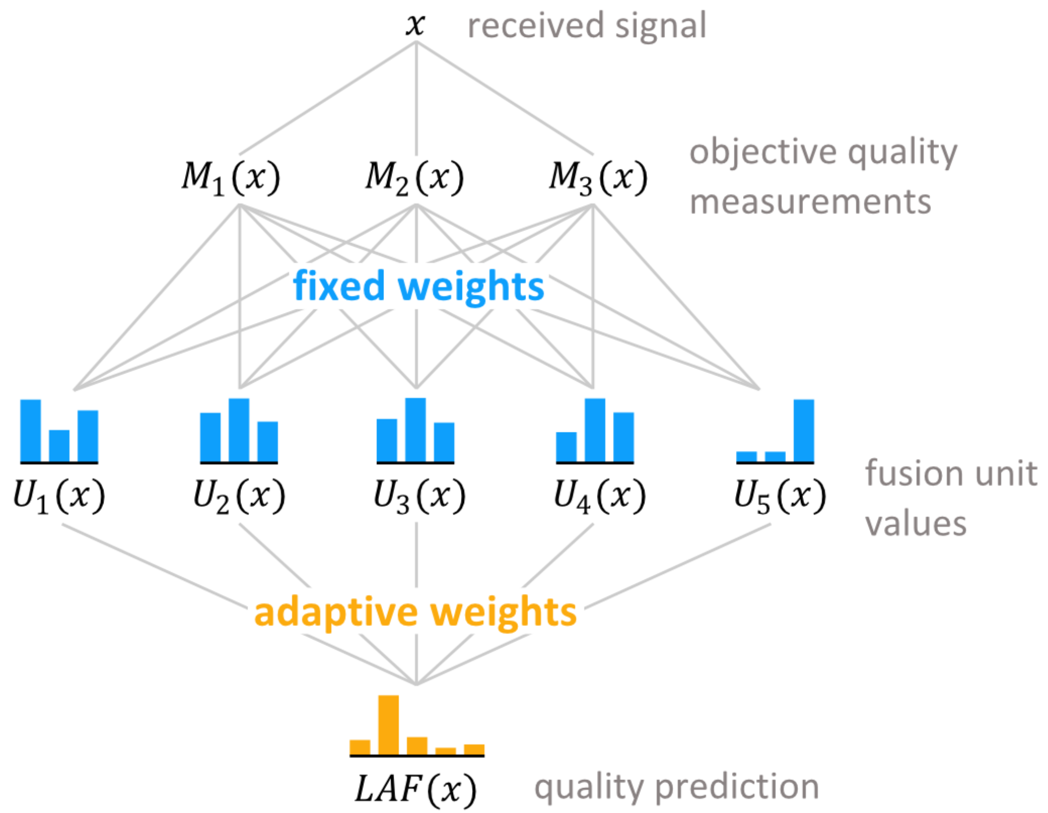
$$\boldsymbol{w}_r = \arg\max_{\boldsymbol{w} \geq 0} \ \mathrm{sep}_{\boldsymbol{w}^T \boldsymbol{I}}[r]$$

This is a convex quadratic programming problem
which can be solved efficiently and accurately

IMPROVE

1. LAF training
2. LAF application

Once the fixed weights are trained LAF can be applied on new signals

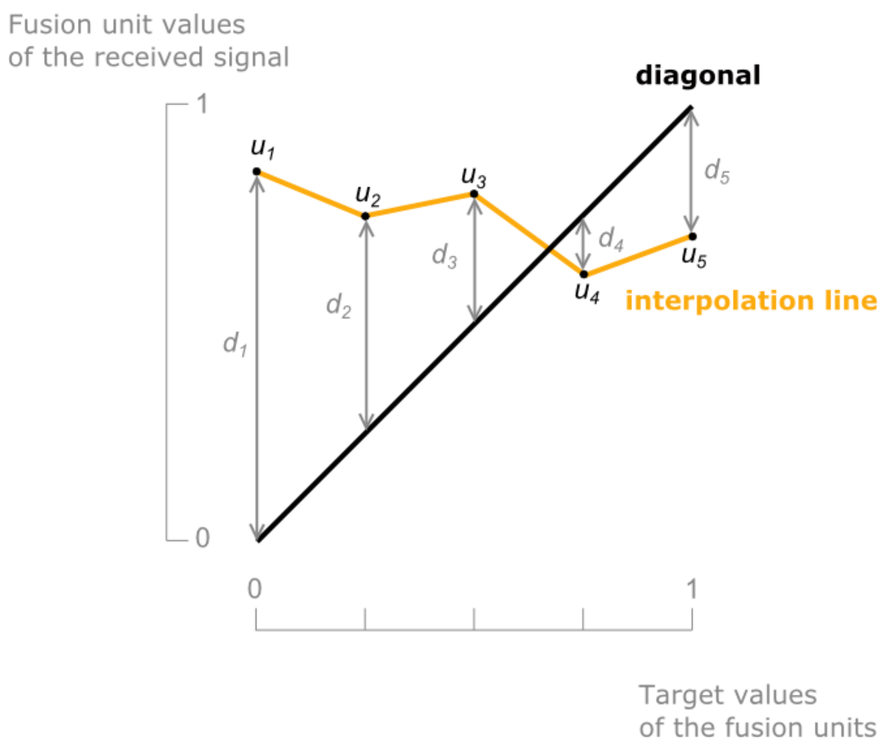


IMPROVE

1. LAF training
2. LAF application

The highest weights are assigned to the most accurate fusion unit values

- Fusion unit values that were not calculated are interpolated



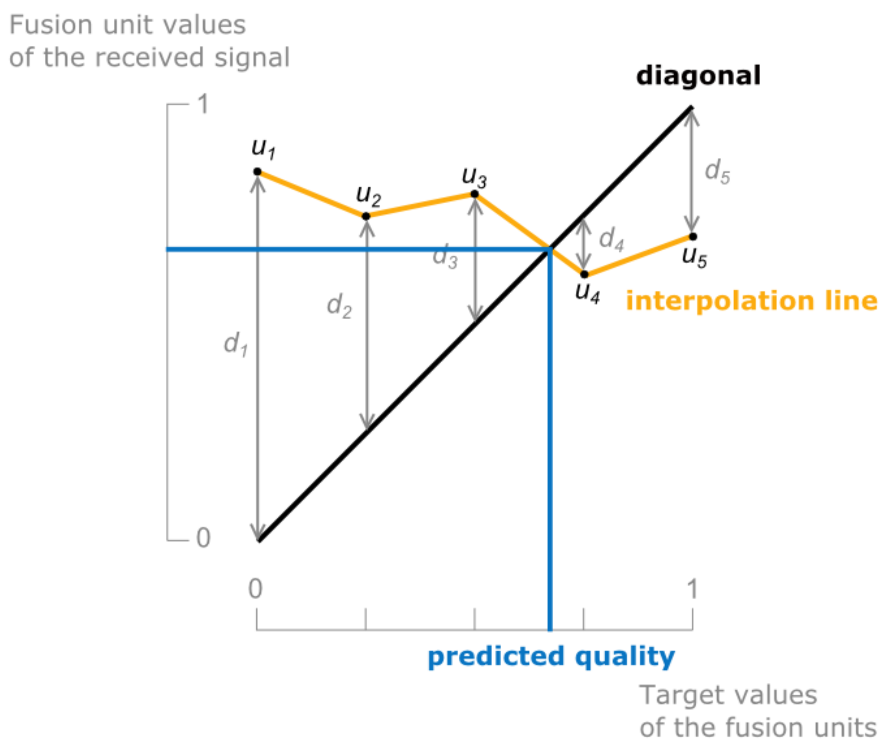
IMPROVE

1. LAF training
2. LAF application

The highest weights are assigned to the most accurate fusion unit values

- Fusion unit values that were not calculated are interpolated

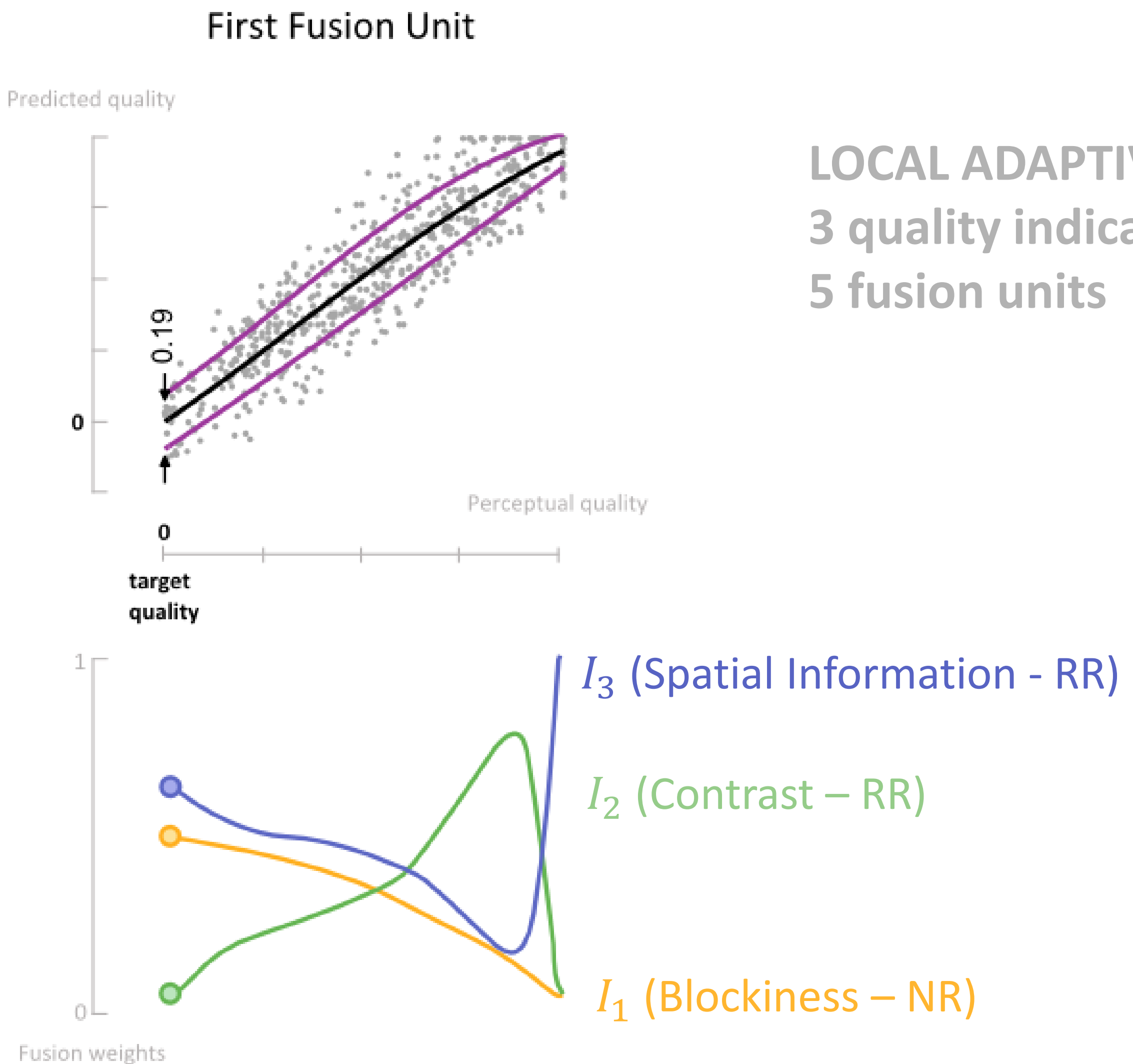
- Best quality prediction is the fixed-point, where fusion unit value = target quality score



IMPROVE

1. LAF training
2. LAF application

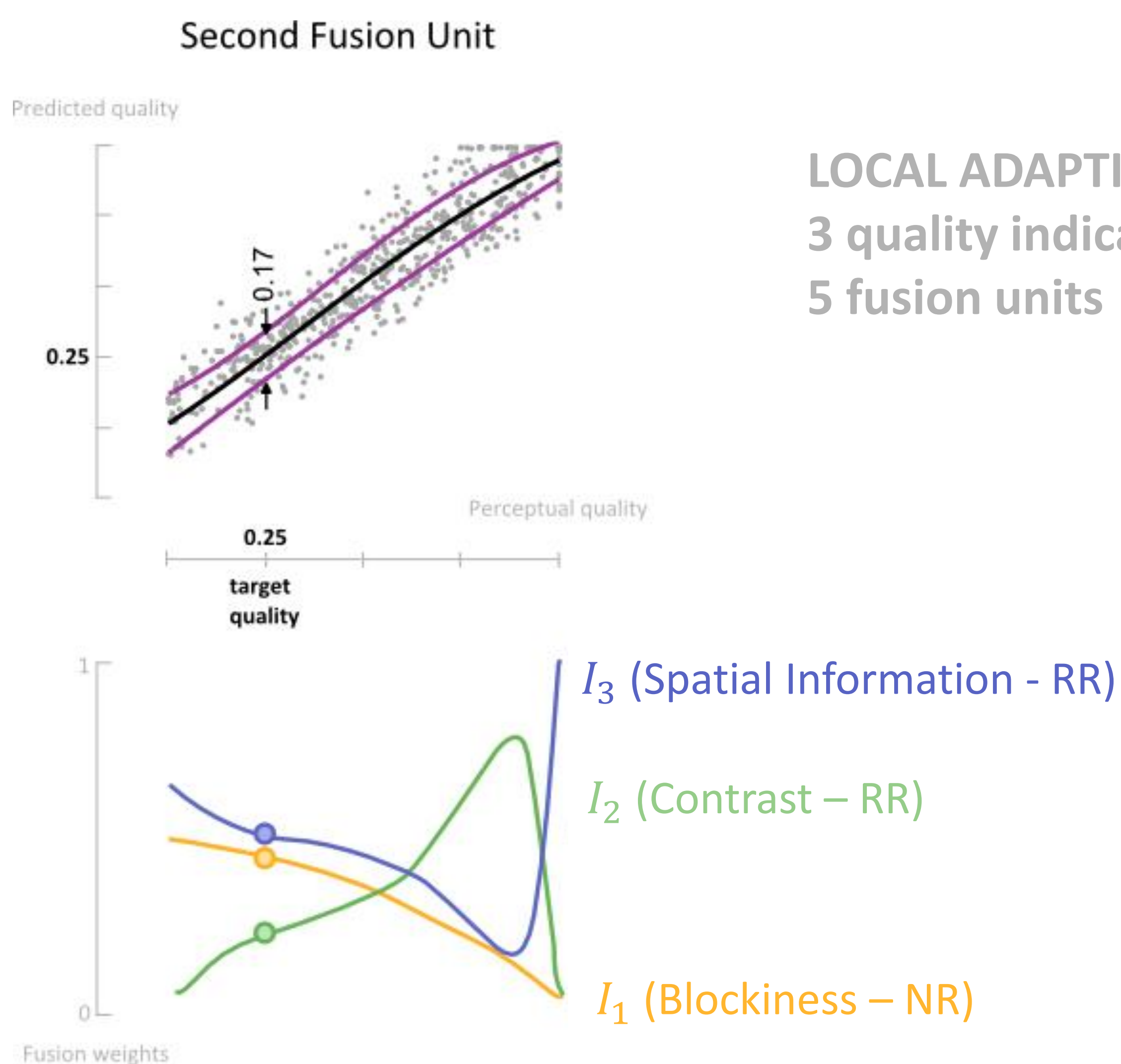
Example   The fusion units are locally optimized



LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

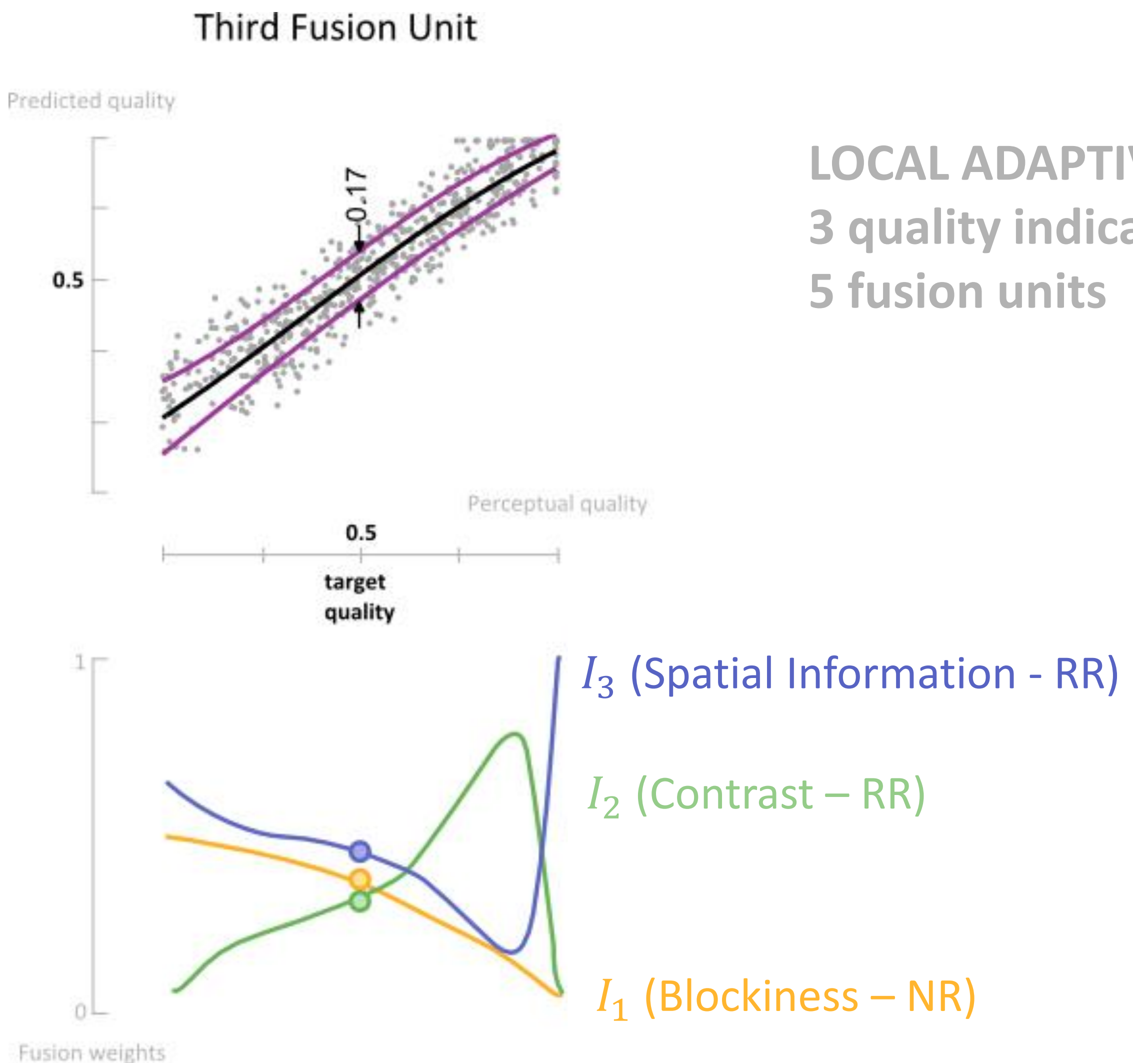
Example The fusion units are locally optimized



LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

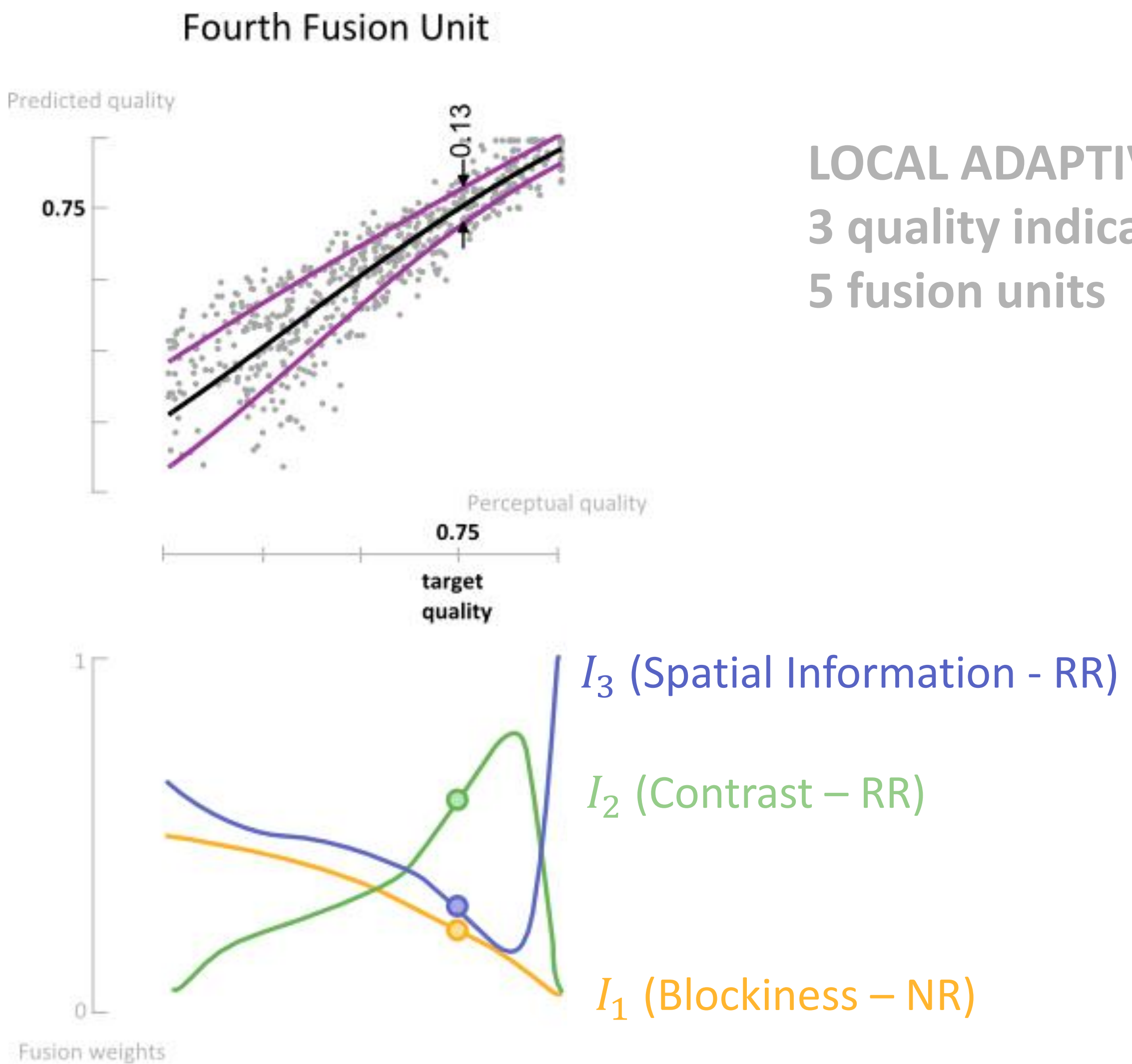
Example The fusion units are locally optimized



LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

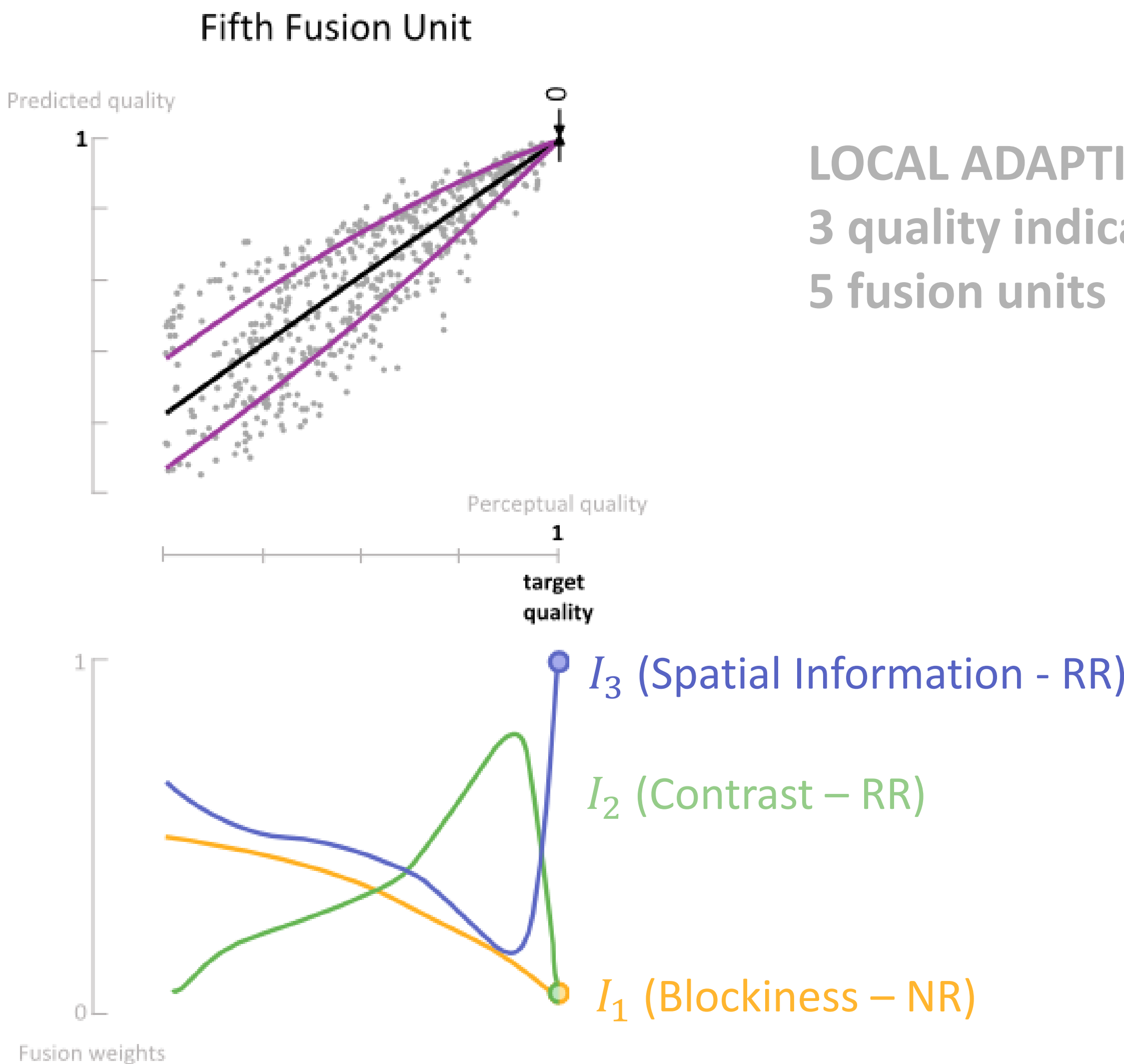
Example The fusion units are locally optimized



LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

Example   The fusion units are locally optimized

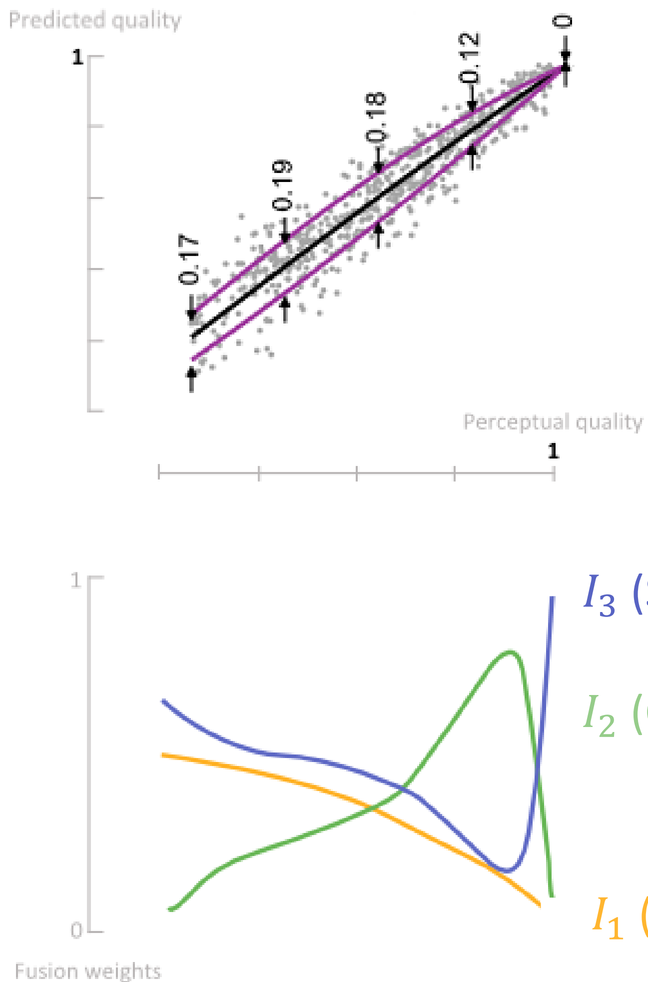


LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

Example

The LAF system is optimized on the entire quality range



LOCAL ADAPTIVE FUSION
3 quality indicators
5 fusion units

IMPROVE

I_3 (Spatial Information - RR)

I_2 (Contrast – RR)

I_1 (Blockiness – NR)

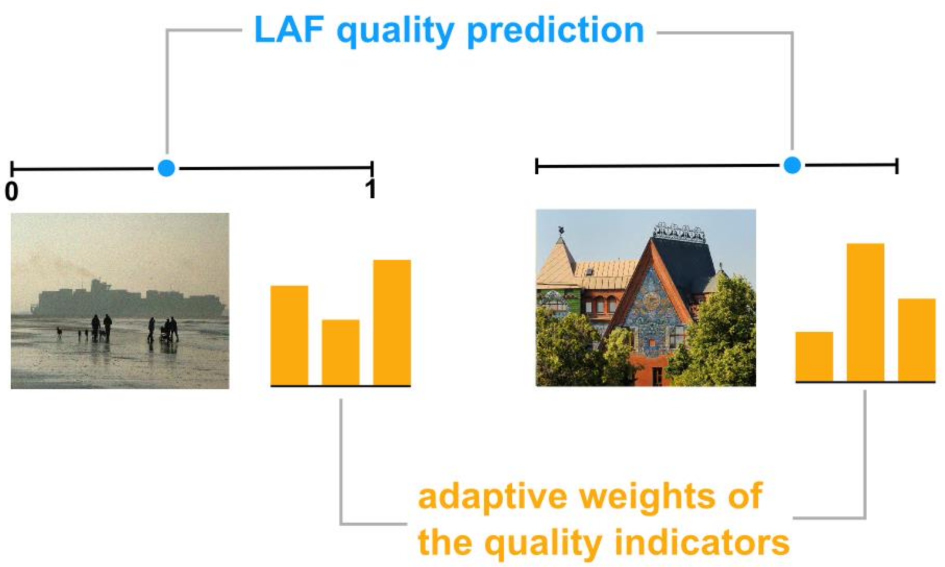
Key features of Locally Adaptive Fusion (LAF)

1. Flexible and interpretable
2. Reproducible
3. Consistent
4. Optimized on the entire quality range

IMPROVE

Key features of Locally Adaptive Fusion (LAF)

1. **Flexible and interpretable**
2. Reproducible
3. Consistent
4. Optimized on the entire quality range

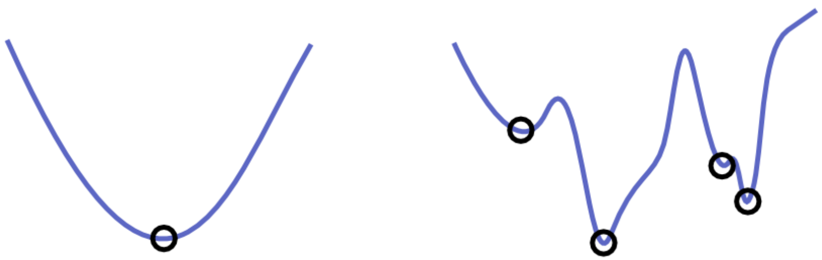


IMPROVE

Key features of Locally Adaptive Fusion (LAF)

1. Flexible and interpretable
2. **Reproducible**
3. Consistent
4. Optimized on the entire quality range

Optimization function
of the training process



IMPROVE

Key features of Locally Adaptive Fusion (LAF)

1. Flexible and interpretable
2. Reproducible
3. **Consistent**
4. Optimized on the entire quality range

Traditional machine learning systems
may violate the consistency rule

IMPROVE

Input quality measures
prefer signal x_1

Locally adaptive fusion
also prefers signal x_1

x_1

x_2

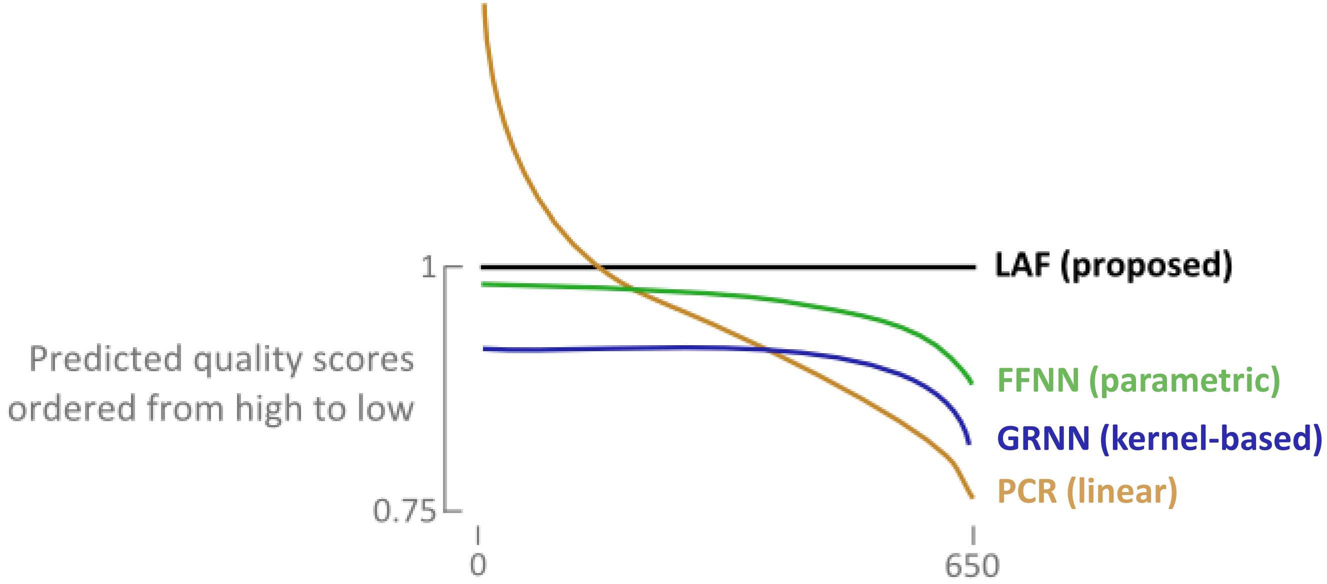
Traditional machine learning
may prefer signal x_2

Key features of Locally Adaptive Fusion (LAF)

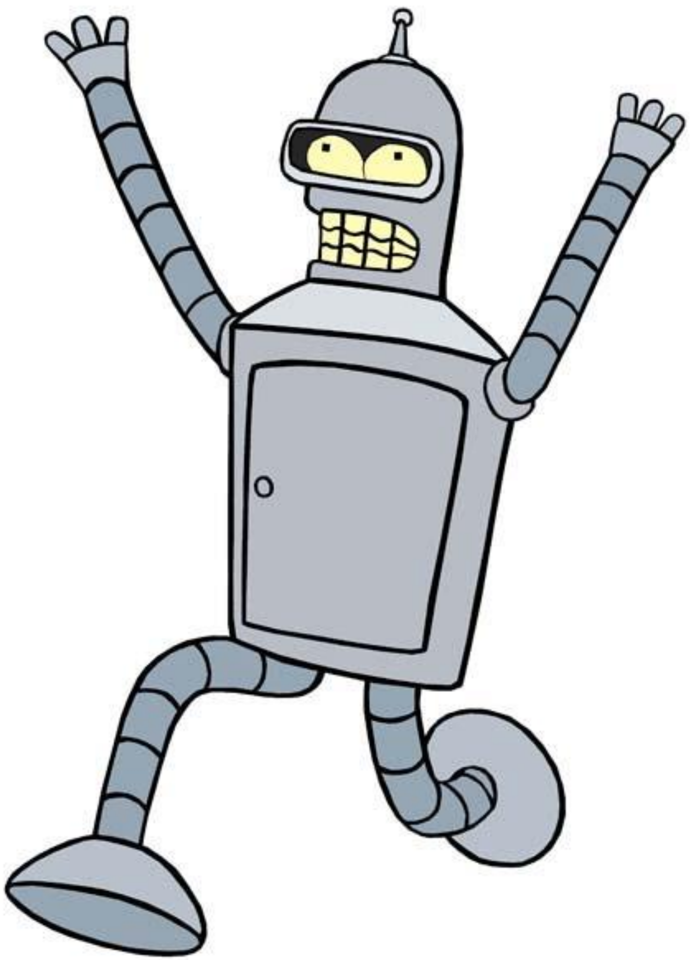
1. Flexible and interpretable
2. Reproducible
3. Consistent
4. **Optimized on the entire quality range**

The behavior of the machine learning systems applied to 650 high quality reference images

IMPROVE



Locally Adaptive Fusions yield
more reliable quality predictions
by imposing strict regulations
on the machine learning behavior



More information:
www.locally-adaptive-fusion.com

A locally adaptive system for the fusion
of objective quality measures
IEEE Transactions on Image Processing

Adriaan Barri
abarri@etro.vub.ac.be

Questions?

Adriaan Barri
abarri@etro.vub.ac.be