Automatic No-Reference Quality Assessment for Retinal Fundus Images Using Vessel Segmentation

Thomas Köhler, Attila Budai, Martin F. Kraus, Jan Odstrcilik,
Georg Michelson, Joachim Hornegger
20.06.2013
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen Graduate School in Advanced Optical Technologies (SAOT)

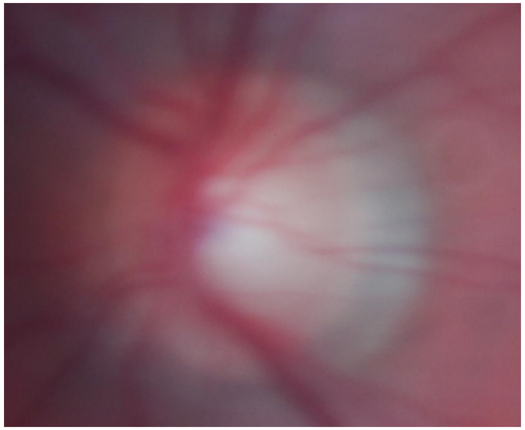
SAOT
erlangen
graduate school in
advanced optical technologies

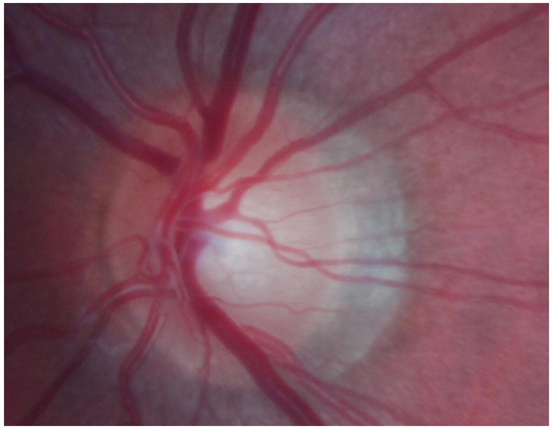
FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Introduction

Motivation

- Image quality assessment is essential for retinal image analysis
- **Example:** Analysis of anatomical structures for manual, computer aided or fully automatic diagnoses





  e. g. segmentation/analysis of optic disk (cup-to-disk ratio) for glaucoma diagnosis: sharp image structures required

- **Goal: objective and automatic quality assessment**

SAOT FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG TECHNISCHE FAKULTÄT

How to assess image quality?

- Qualitative assessment: Ask an expert
  $\Rightarrow$ Subjective, inter- and intra-observer variance

- Quantitative assessment:
  - Ground truth: peak-signal-to-noise ratio (PSNR), structural similarity (SSIM)
    $\Rightarrow$ Not available in practice
  - In the absence of a ground truth: no-reference quality assessment
    $\Rightarrow$ Objective and reproducible

- No-reference quality assessment
  - Classification-based approaches (supervised)

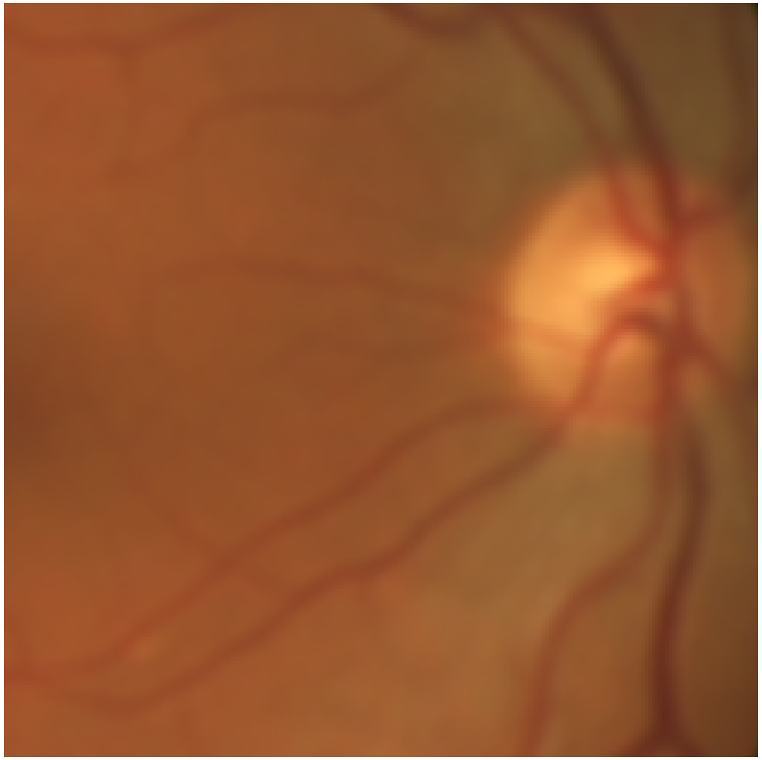
    Niemeijer et al., Med. Image Anal., 2006

    Paulus et al., Int. J. of Computer Assisted Radiology & Surgery, 2010
  - **No-reference quality metrics (unsupervised)**

Objective Image Quality Features

Features for quality assessment:

- **Blur/sharpness**

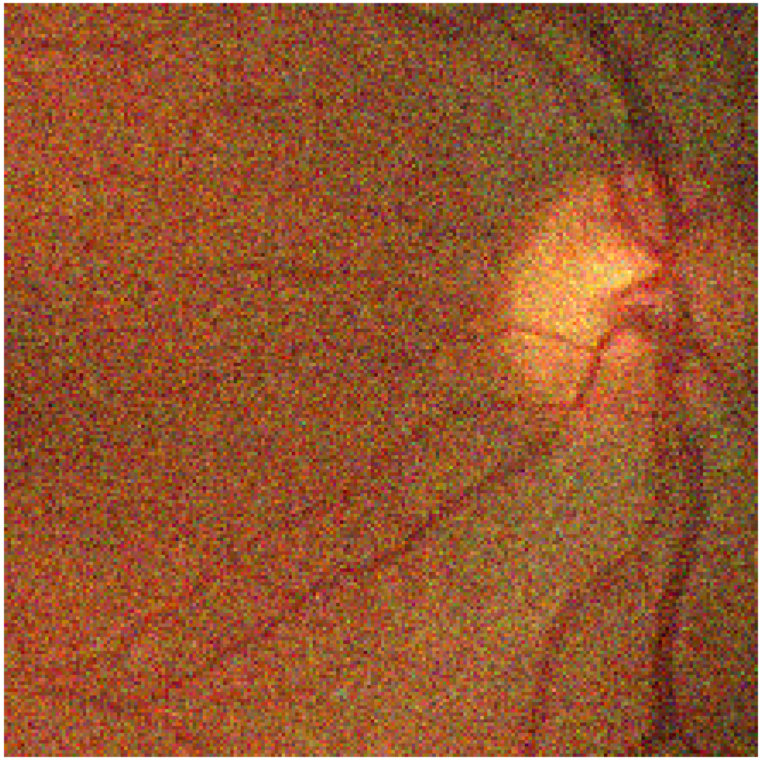


Goal in this work: quantitative assessment of image noise and blur

Objective Image Quality Features

Features for quality assessment:

- **Blur/sharpness**
- **Noise**



Goal in this work: quantitative assessment of image noise and blur

Objective Image Quality Features

Features for quality assessment:

- **Blur/sharpness**
- **Noise**
- Illumination/contrast
- High-level medical features: visibility of blood vessels . . .
- . . .

Goal in this work: quantitative assessment of image noise and blur

No-Reference Quality Metric for Noise and Blur

SAOT
erlangen
graduate school in
advanced optical technologies

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

No-Reference Quality Metric for Noise and Blur[1]

- Decompose image $\boldsymbol{I}$ of size $M \times N$ in distinct patches $\boldsymbol{P}$ of size $n \times n$ (typical parameter: $n = 8$).
- Important quantities:
  - Local gradient matrix

$$\boldsymbol{G} = \begin{pmatrix} P_x(1,1) & P_y(1,1) \\ \vdots & \vdots \\ P_x(n,n) & P_y(n,n) \end{pmatrix} \tag{1}$$

  - Singular value decomposition (SVD) of $\boldsymbol{G}$:

$$\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top = \boldsymbol{U} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \boldsymbol{V}^\top \tag{2}$$

[1] X. Zhu and P. Milanfar, Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content, IEEE Transactions on Image Processing, 2010.

Automatic Quality Assessment – Algorithm

1. Calculate **coherence** for each patch:

$$R = \frac{s_1 - s_2}{s_1 + s_2} \tag{3}$$

Automatic Quality Assessment – Algorithm

1. Calculate **coherence** for each patch:

$$R = \frac{s_1 - s_2}{s_1 + s_2} \tag{3}$$

2. Detect **anisotropic patches** (thresholding: $R > \tau$)

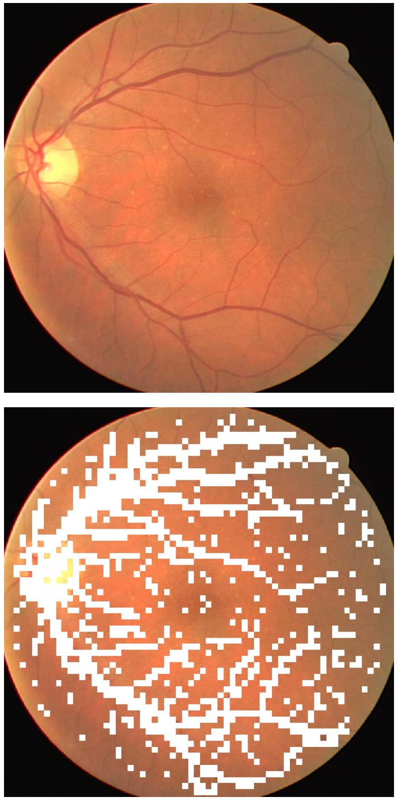
Automatic Quality Assessment – Algorithm

1. Calculate **coherence** for each patch:

$$R = \frac{s_1 - s_2}{s_1 + s_2} \quad (3)$$

2. Detect **anisotropic patches** (thresholding: $R > \tau$)
3. For each anisotropic patch: Calculate **local** score

$$q(\boldsymbol{P}) = s_1 \cdot R \quad (4)$$



Automatic Quality Assessment – Algorithm

1. Calculate **coherence** for each patch:

$$R = \frac{s_1 - s_2}{s_1 + s_2} \quad (3)$$

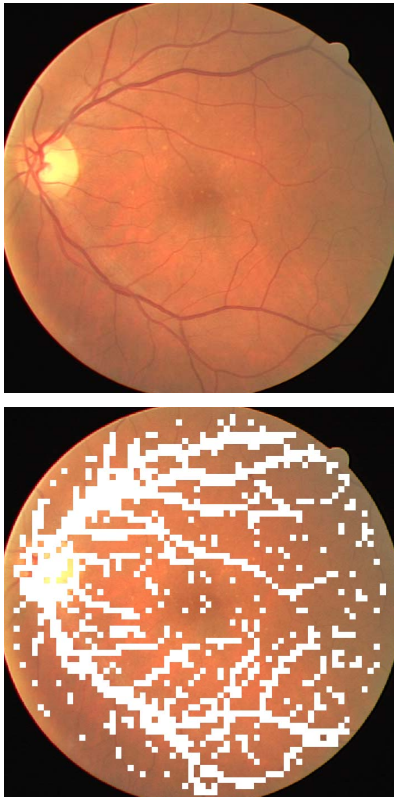
2. Detect **anisotropic patches** (thresholding: $R > \tau$)
3. For each anisotropic patch: Calculate **local** score

$$q(\boldsymbol{P}) = s_1 \cdot R \quad (4)$$

4. Calculate **global** score for noise and blur:

$$Q = \frac{1}{MN} \sum_{i,j:\mathcal{P}(i,j)=1} q(\boldsymbol{P}_{ij}) \quad (5)$$

High $Q \Rightarrow$ better quality (in terms of blur and noise)



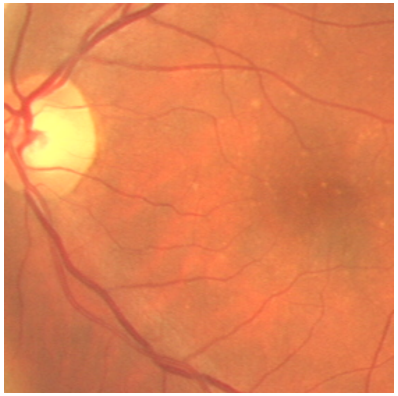
Quality Assessment Using Vessel Segmentation

SAOT erlangen graduate school in advanced optical technologies

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Vessel Segmentation Guidance

- Limitation of metric *Q*: **false-positive** and **false-negative** patch detections



Vessel Segmentation Guidance

- Limitation of metric Q: **false-positive** and **false-negative** patch detections

- **Appropriate guidance:** vessel tree visible in fundus images

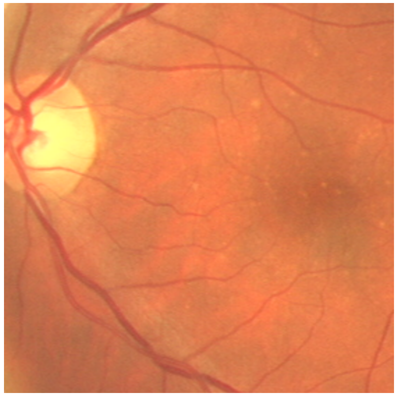
  Vessel tree is detected by **vesselness** measure

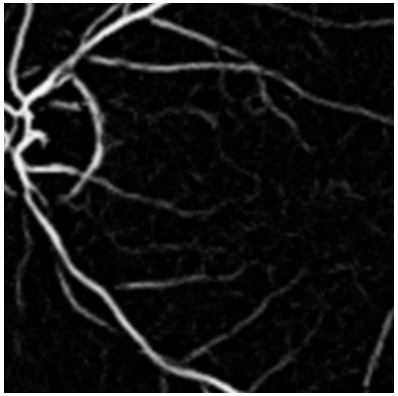
  (Frangi et al., Multiscale vessel enhancement filtering, MICCAI 1998)

$$V = \exp\left(-\frac{\lambda_1^2}{\lambda_2^2}\right)\left(1 - \exp\left(-(\lambda_1^2 + \lambda_2^2)\right)\right) \qquad (6)$$

λ_1, λ_2: Eigenvalues of pixel-wise **Hessian matrix**

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial d^2 \boldsymbol{I}}{\partial x^2} & \frac{\partial d^2 \boldsymbol{I}}{\partial x \partial y} \\ \frac{\partial d^2 \boldsymbol{I}}{\partial x \partial y} & \frac{\partial d^2 \boldsymbol{I}}{\partial y^2} \end{pmatrix} \qquad (7)$$





Weighted Quality Score

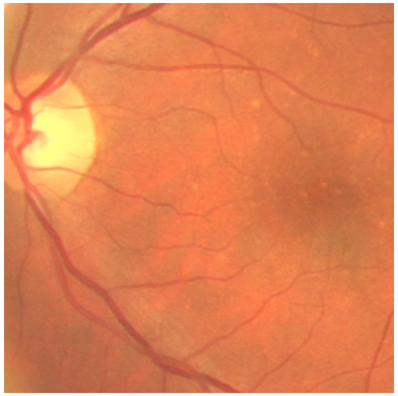
- **Basic idea:** Detected patch on a blood vessel (boundary) is more reliable

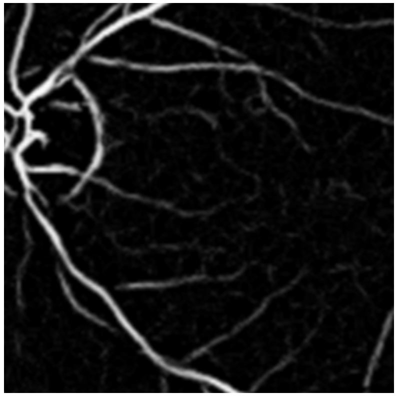
- **Weighted quality score** according to vesselness

$$Q_v = \sum_{i,j:\mathcal{P}(i,j)=1} \tilde{\Sigma}_{ij} \cdot q(\boldsymbol{P}_{ij}) \quad (8)$$

- Weighting factor $\tilde{\Sigma}_{ij}$: **local variance of vesselness** in patch $\boldsymbol{P}_{ij}$

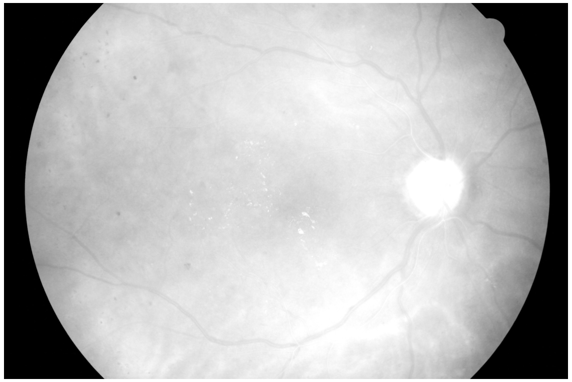
  Blood vessel boundary
  $\Rightarrow$ high $\tilde{\Sigma}_{ij}$ $\Rightarrow$ high reliability of $q(\boldsymbol{P}_{ij})$





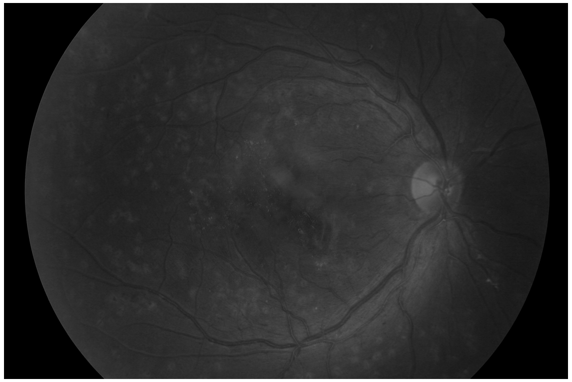
Application to Color Fundus Images

- Quality metric defined for single-channel images
- **Color fundus images:** extraction of green channel for quality assessment
  $\Rightarrow$ good contrast/illumination compared to red and blue channel



Red
(Oversaturated)

Green



Blue
(Underexposed)

Experimental Evaluation

Experimental Evaluation

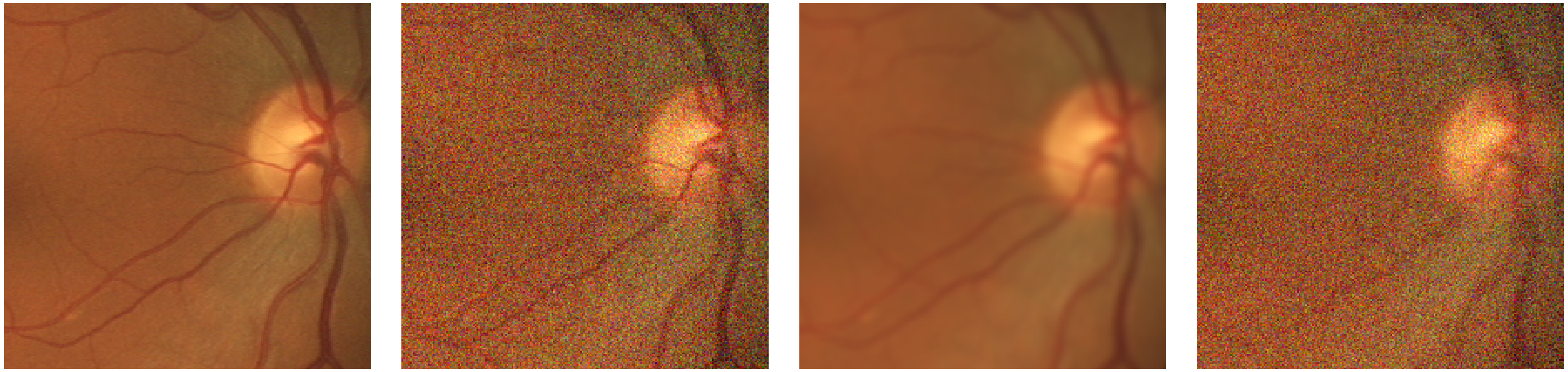
- **Synthetic images:** correlation analysis
  - Ground truth data available
  - How good agrees no-reference quality assessment to established full-reference quality metrics?

- **Real fundus images:**
  - Quality classification
  - Agreement to human camera operator

Synthetic Images – Correlation Analysis

- 40 images out of the DRIVE database: simulation of artificial Gaussian noise and Gaussian blur
- Agreement (correlation) between no-reference and full-reference metrics:
  - Full-ref. metrics: peak-signal-to-noise ratio (PSNR), structural similarity (SSIM)
  - Spearman's rank correlation (Spearman's ρ) to assess agreement

$\Rightarrow$ High corrleation $\Rightarrow$ good agreement to full-reference assessment

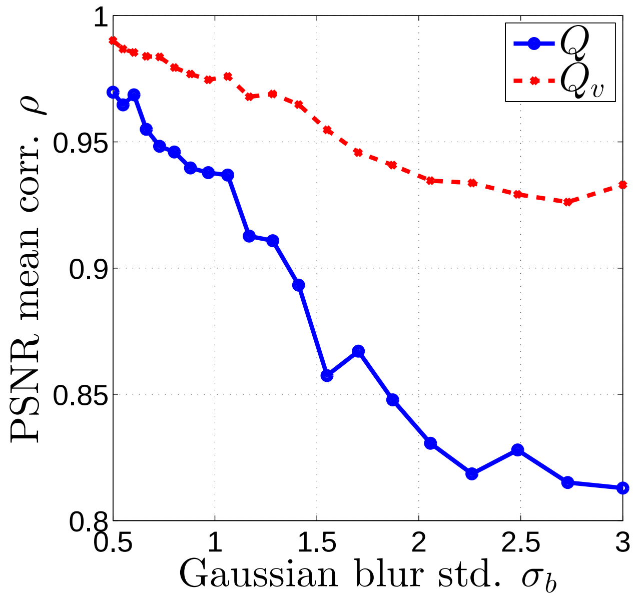


Ground truth | Noisy image | Blurred image | Noise and blur

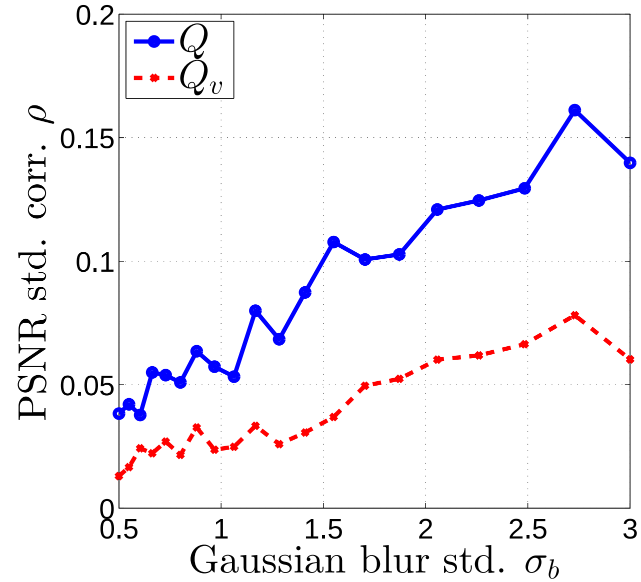
Synthetic Images – Correlation Analysis (cont.)

Spearman's ρ versus amount of **Gaussian blur**:

- Gaussian blur: 7×7 kernel (fixed), $\sigma_b = 0 \ldots 3.0$
- Agreement Q, $Q_v \leftrightarrow$ **PSNR** (averaged over 40 images)



Mean ρ

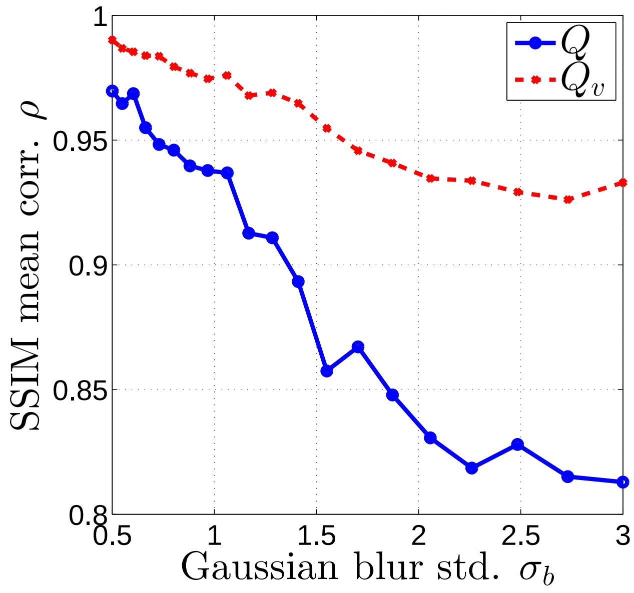


Standard deviation ρ

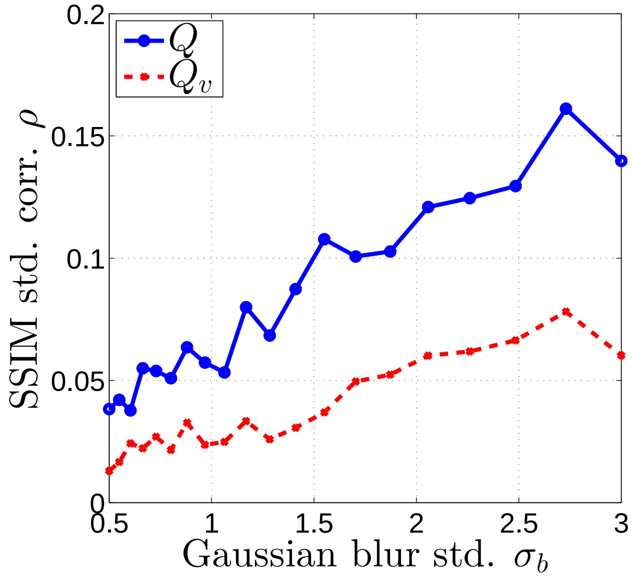
Synthetic Images – Correlation Analysis (cont.)

Spearman's ρ versus amount of **Gaussian blur**:

- Gaussian blur: 7×7 kernel (fixed), $\sigma_b = 0 \ldots 3.0$
- Agreement Q, $Q_v \leftrightarrow$ **SSIM** (averaged over 40 images)



Mean ρ

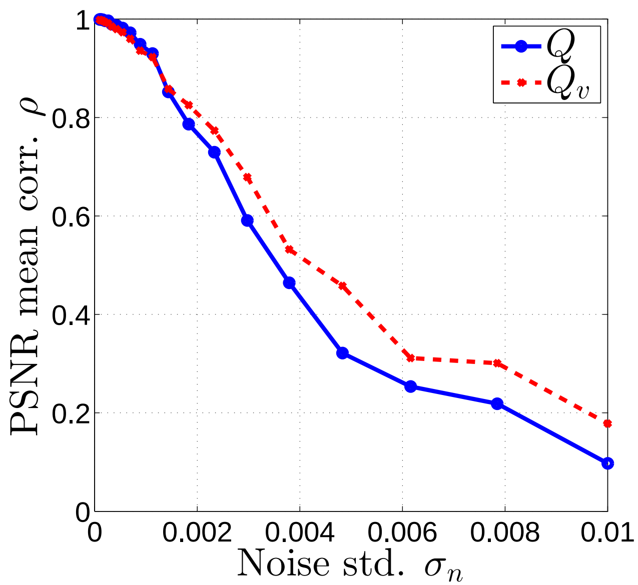


Standard deviation ρ

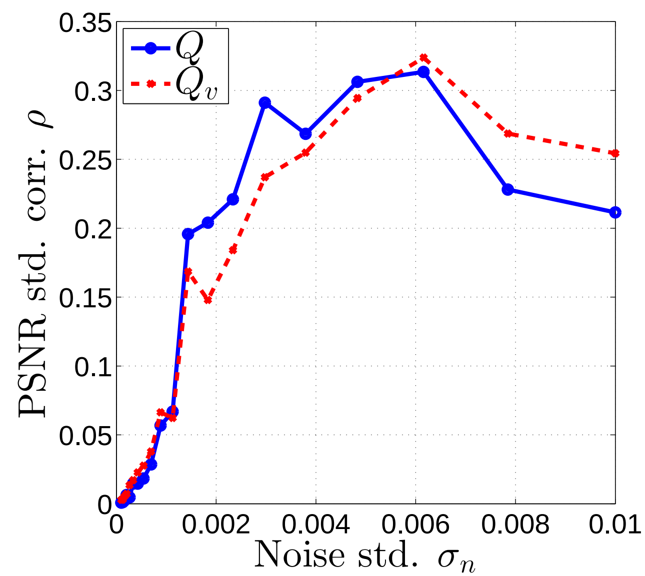
Synthetic Images – Correlation Analysis (cont.)

Spearman's ρ versus amount of **Gaussian noise**:

- Gaussian noise: $\sigma_n = 0 \dots 0.01$ (normalized intensities: [0, 1])
- Agreement Q, $Q_v \leftrightarrow$ **PSNR** (averaged over 40 images)



Mean ρ

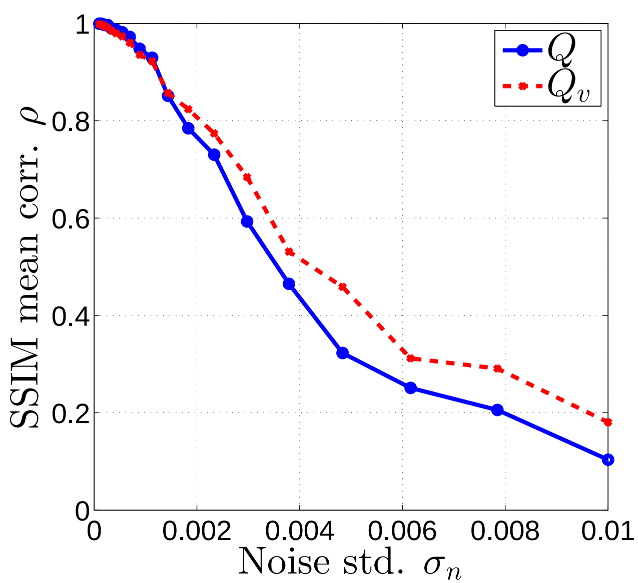


Standard deviation ρ

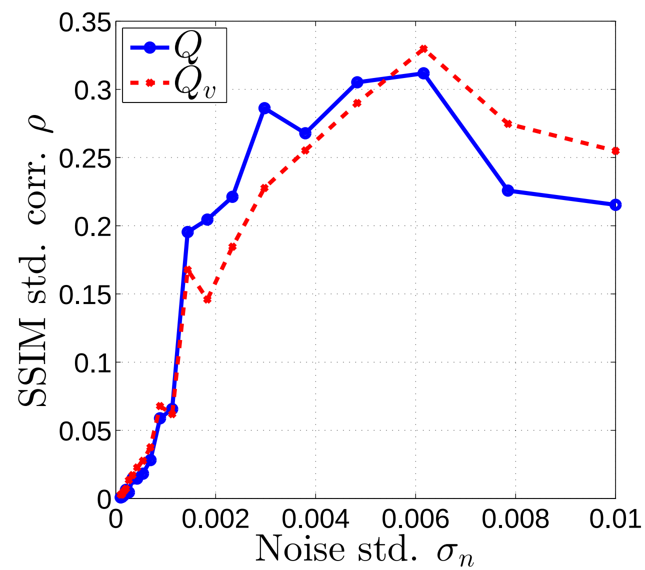
Synthetic Images – Correlation Analysis (cont.)

Spearman's ρ versus amount of **Gaussian noise**:

- Gaussian noise: $\sigma_n = 0 \dots 0.01$ (normalized intensities: [0, 1])
- Agreement Q, $Q_v \leftrightarrow$ **SSIM** (averaged over 40 images)



Mean ρ



Standard deviation ρ

Synthetic Images – Correlation Analysis (cont.)

Overall correlation: simultaneously varying blur and noise

- 40 ground truth images (DRIVE database)
- 20 levels of Gaussian blur: $\sigma_b = 0 \dots 3.0$
- 20 levels of Gaussian noise: $\sigma_n = 0 \dots 0.01$

Spearman's ρ over the whole experiment:

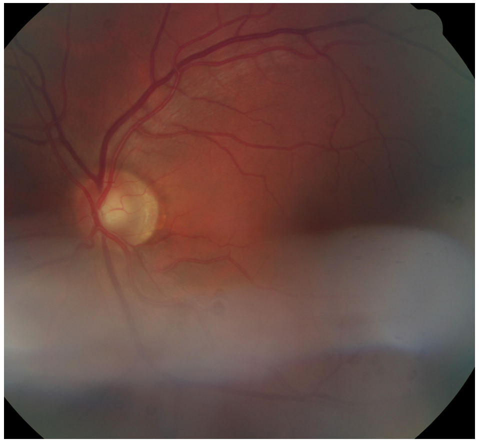
Full-ref. metric	$\rho(Q)$	$\rho(Q_v)$
PSNR	0.8227	**0.8920**
SSIM	0.8412	**0.9076**

$\Rightarrow$ Higher corrleation for proposed Q_v metric

Real Images

High-resolution fundus (HRF) image database
(http://www5.cs.fau.de/research/data/fundus-images/)
Canon CR-1 fundus camera, 45 degree field of view

- 18 pairs of fundus images: good/bad image per pair (36 images)
- Poor quality due to de-focused camera
- In case of poor quality: acquisition was repeated



Experimental evaluation:

- Quality classification
- Agreement to camera operator

Real Images – Quality Classification

Quality classification implemented as thresholding:

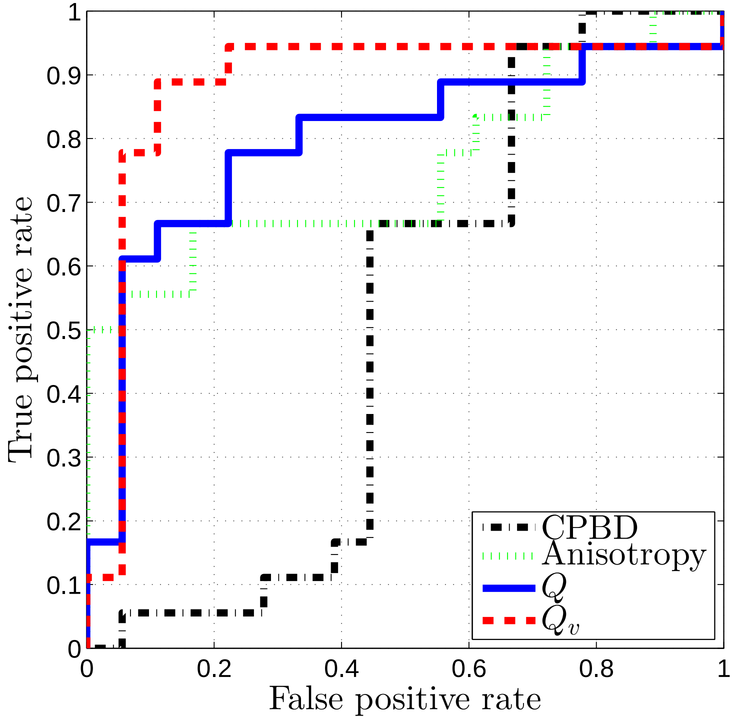
- 2-class problem (class label: y):
  $y = 1$ (good quality) and $y = -1$ (poor quality)
- Decision rule for quality metric x and threshold τ_0:

$$y = \begin{cases} -1 & x < \tau_0 \quad \text{(poor quality)} \\ +1 & x \geq \tau_0 \quad \text{(good quality)} \end{cases} \tag{9}$$

- **Comparison:**
  - Proposed Q_v metric
  - Standard Q metric Zhu and Milanfar, 2010
  - Anisotropy blind quality metric Gabarda and Cristóbal, 2007
  - Cumulative probability of blur detection (CPBD) Narvekar and Karam, 2011

Real Images – Quality Classification (cont.)

- ROC analysis for different classification approaches:

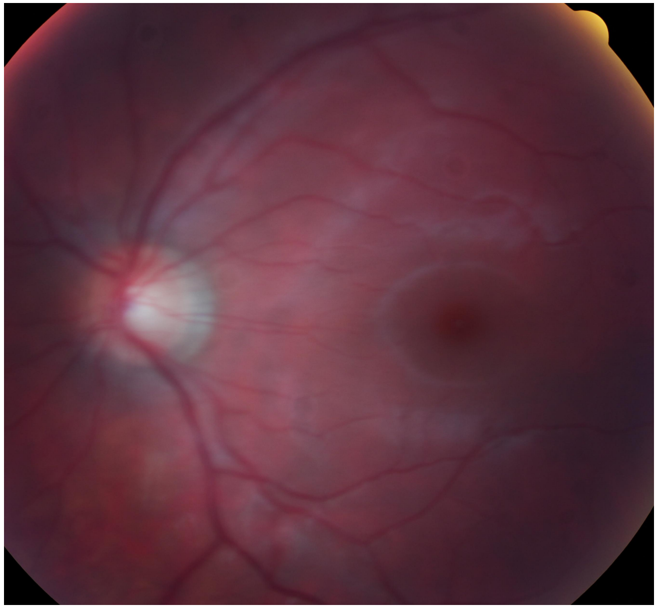


$\Rightarrow$ Good performance of Q_v in terms of **area under ROC curve: 88.3 %**

Agreement with Human Observer

- **Pair-wise agreement** with human observer (good vs. bad image): 16 of 18 pairs **(88.9 %)**

sharp: $Q_v = 0.0240$



defocused: $Q_v = 0.0017$

Agreement with Human Observer (cont.)

- Comparison of pair-wise agreement for different metrics (based on 18 image pairs):

No-ref. Metric	Agreement [%]
CPBD	55.6
Anisotropy	94.4
Q	83.3
Q_v	88.9

$\Rightarrow$ Competitive performance of proposed Q_v metric

Summary and Conclusion

Summary and Conclusion

- No-reference image quality metric to quantify noise and sharpness
  $\Rightarrow$ Unsupervised approach (opposed to classification-based approaches)

- Quality assessment guided by the blood vessel tree
  $\Rightarrow$ Reliable quality score for fundus images
  - High correlation to full-reference quality metrics
  - Quality classification: 88.3% area under ROC curve
  - Agreement to human camera operator: 88.9%

- Applications:
  - Numerical score for image noise/sharpness (e. g. auto-focusing)
  - Feature for quality classification

Acknowledgments

The authors gratefully acknowledge funding of the **Erlangen Graduate School in Advanced Optical Technologies (SAOT)** by the German National Science Foundation (DFG) in the framework of the excellence initiative.

Thank you for your attention!