Neural Network Nearest-Neighbour Transformation for Segmentation of Retinal Blood Vessels

Bachelor's Thesis Final Talk

Anne-Marie Strauch
Supervisors: Lennart Husvogt, Tino Haderlein, Andreas Maier
28.11.2016
Pattern Recognition Lab (CS 5)

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Outline

- Introduction
- Approach
  - Training
  - Testing
- Experiments
- Discussion
- Conclusion and Outlook

Outline

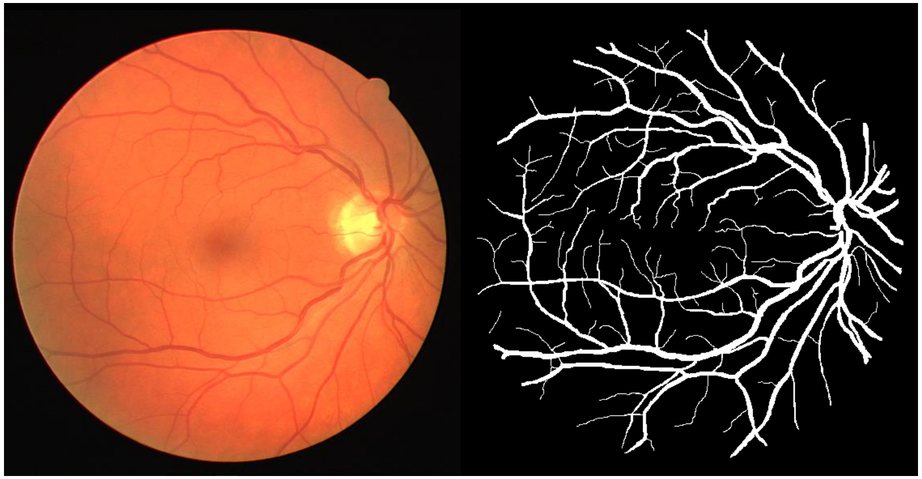
- **Introduction**
- Approach
  - Training
  - Testing
- Experiments
- Discussion
- Conclusion and Outlook

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Fundus Photography

- non-invasive modality for examining eye
- diagnosis of glaucoma, diabetic retinopathy,...
  - diabetic retinopathy leading cause for acquired blindness
- segmentation supports computer-aided diagnosis

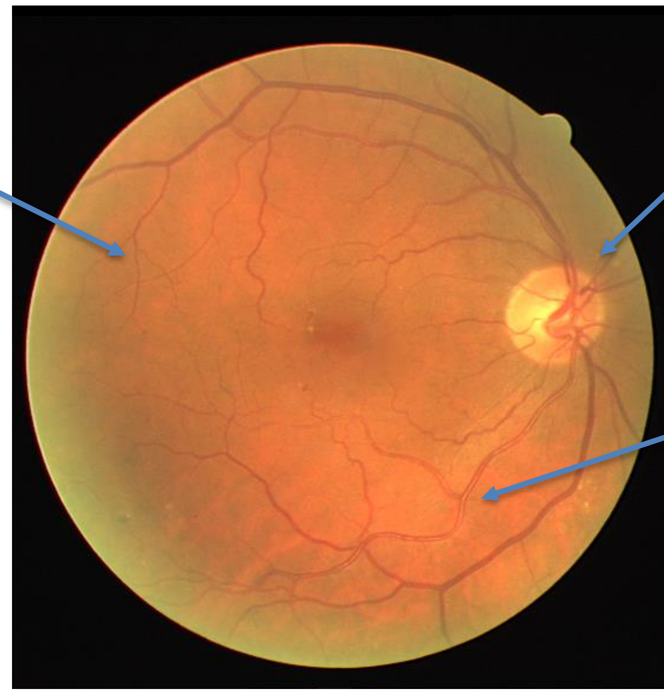


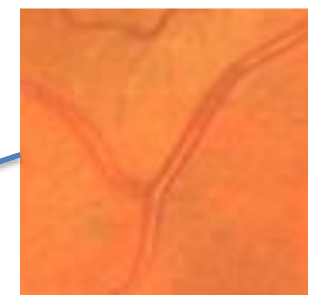
Fundus image and segmentation [DRIVE database]

Some challenges

thin vessels

bright optic disc





central vessel reflex

- bifurcations
- lighting variations
- poor contrast
- pathologies

Introduction

N^4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms

Yaroslav Ganin, Victor Lempitsky
Skolkovo Institute of Science and Technology (Skoltech)

- reimplementation of N^4 Paper
- experiments with varying parameters
- evaluation on DRIVE database

Outline

- Introduction
- **Approach**
  - Training
  - Testing
- Experiments
- Discussion
- Conclusion and Outlook

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

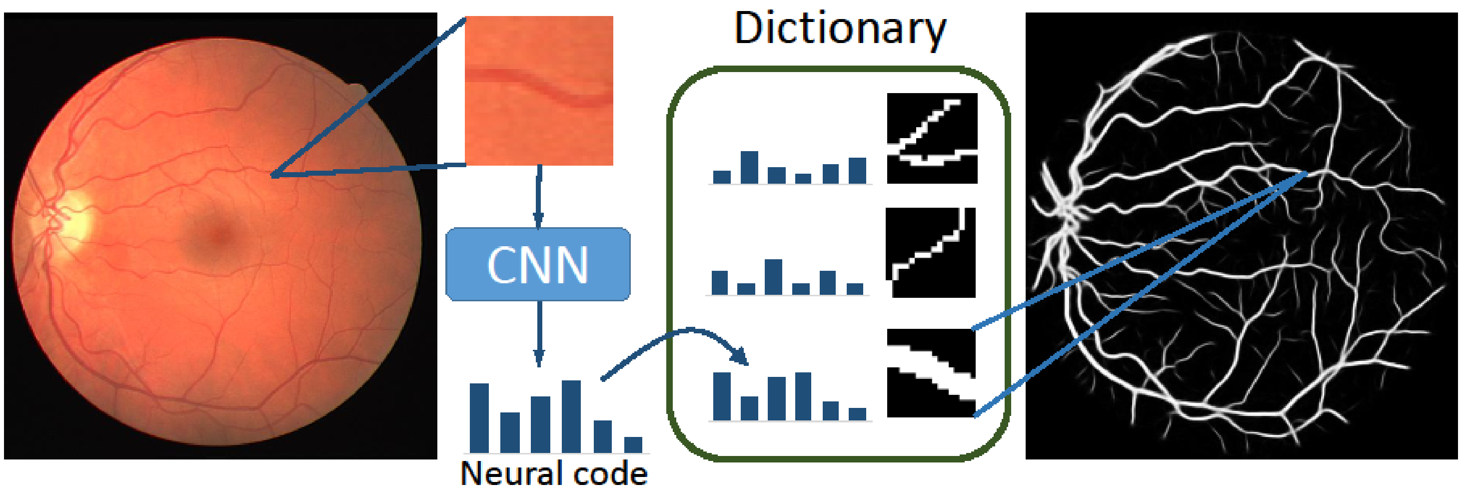
N^4-Fields

N^4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms

Yaroslav Ganin, Victor Lempitsky
Skolkovo Institute of Science and Technology (Skoltech)

- architecture for natural edge detection/thin object segmentation
- neural network and nearest-neighbour search applied sequentially
- process images patch-by-patch

N⁴-Fields



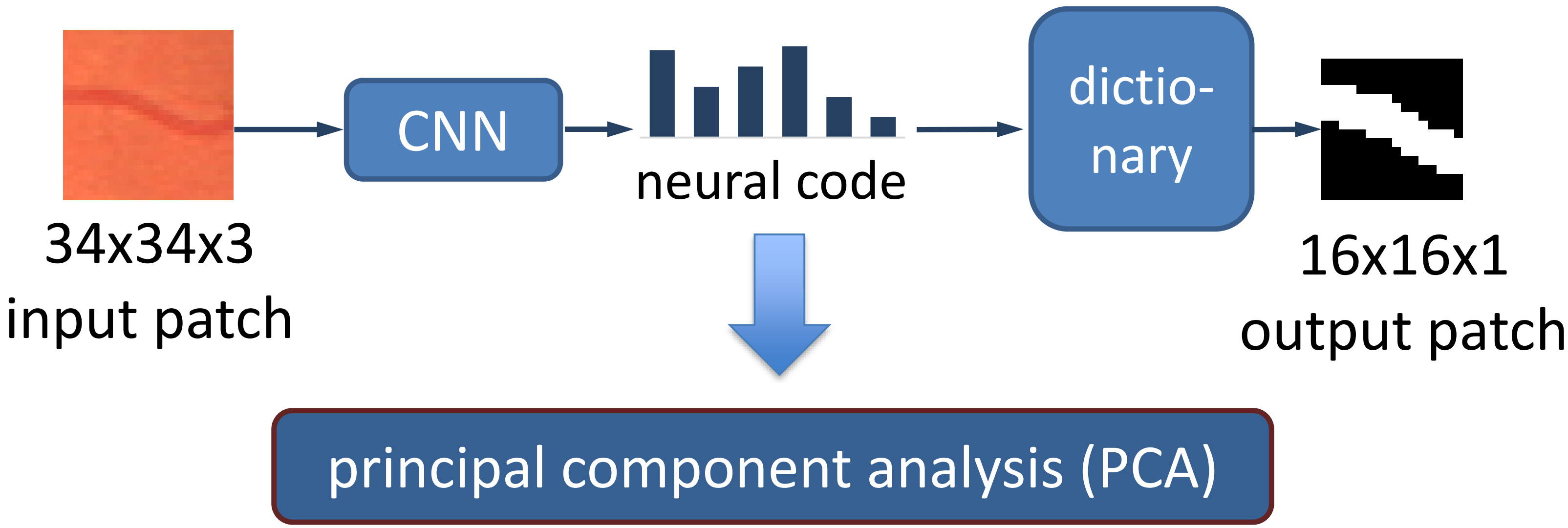
- run patch through convolutional neural network
  → receive neural code
- map to closest dictionary entry using nearest-neighbour classifier
  → retrieve output patch
- average all output patches to segment image

Outline

- Introduction
- Approach
  - **Training**
  - Testing
- Experiments
- Discussion
- Conclusion and Outlook

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

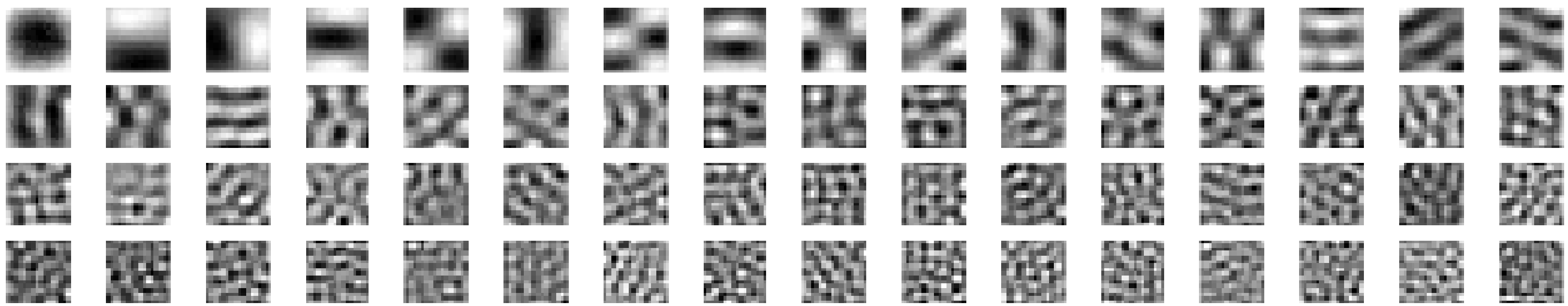
Training



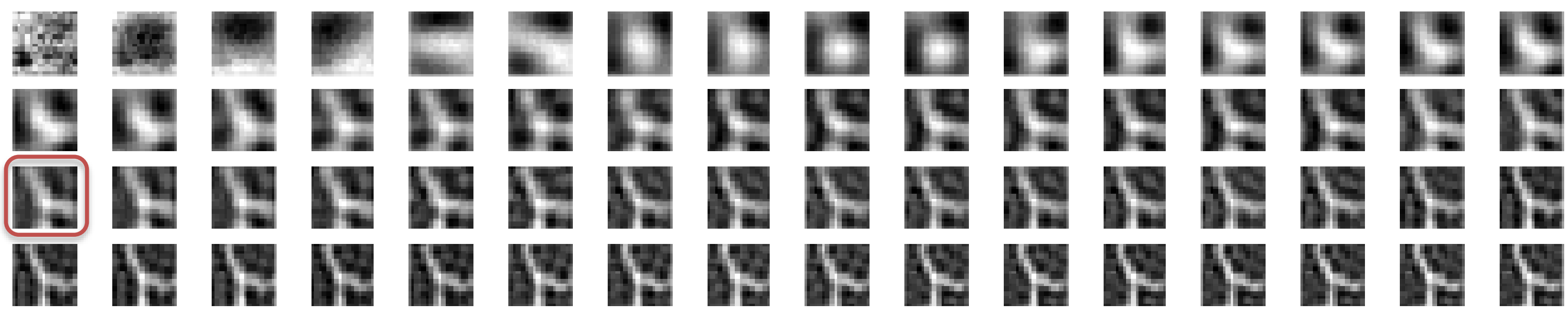
neural code length
→ number of eigenvectors used
in data reduction

$$\boldsymbol{X} = \begin{bmatrix} \text{patch 1} \\ \text{patch 2} \\ \vdots \\ \text{patch N} \end{bmatrix}$$

PCA



first 64 eigenvectors



reconstruction of a patch with a total of 63 eigenvectors

Convolutional Neural Network

#	Layer type	#filters/units	Filter size	Output size
1	convolutional + ReLU +max-pooling	96 -	7x7 2x2	28x28x96 14x14x96
2	convolutional + ReLU + max-pooling	128 -	5x5 2x2	10x10x128 5x5x128
3	convolutional + ReLU	256	3x3	3x3x256
4	fully connected + ReLU	768	-	768
5	fully connected + ReLU	768	-	768
6	fully connected + linear	16	-	16

max-pooling:

6	3
9	12

ReLU: $\varphi(x) = \max(0, x)$

linear: $\varphi(x) = x$

Convolutional Neural Network

- dropout with 50 %
- update rule Adadelta → parameter-wise learning rate

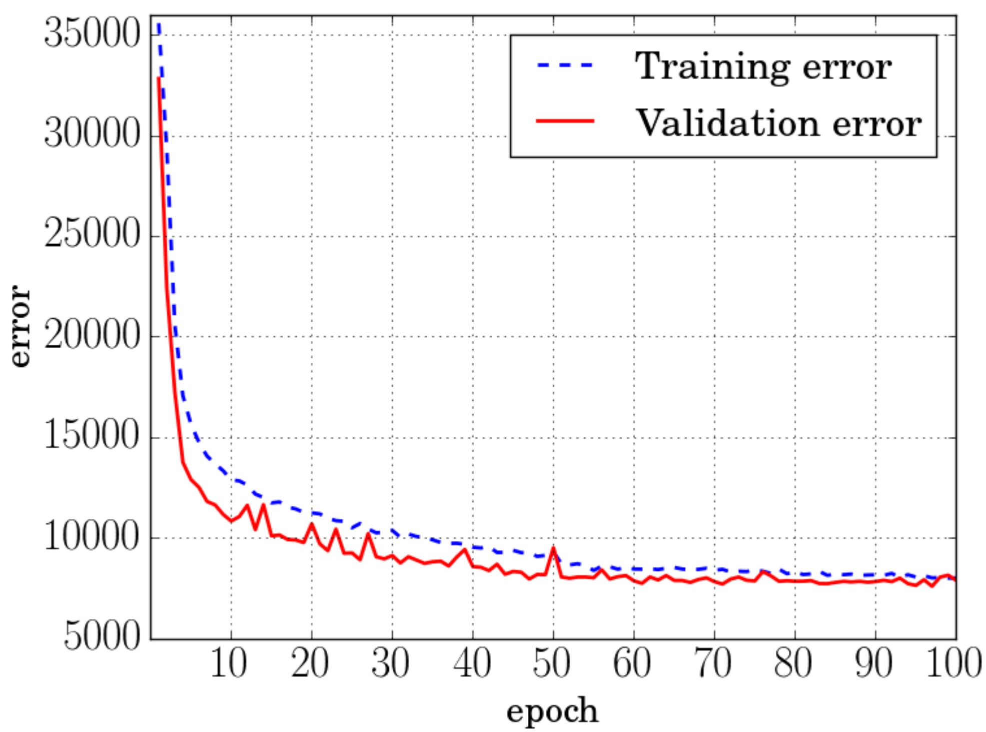
- input patches drawn at random locations with random rotations

- loss function:

$$L = \frac{1}{b \cdot f} \sum_{i=1}^{b} (\mathrm{CNN}(P;W)_i - \mathrm{PCA}(\mathbf{A}(P))_i)^T \cdot (\mathrm{CNN}(P;W)_i - \mathrm{PCA}(\mathbf{A}(P))_i)$$

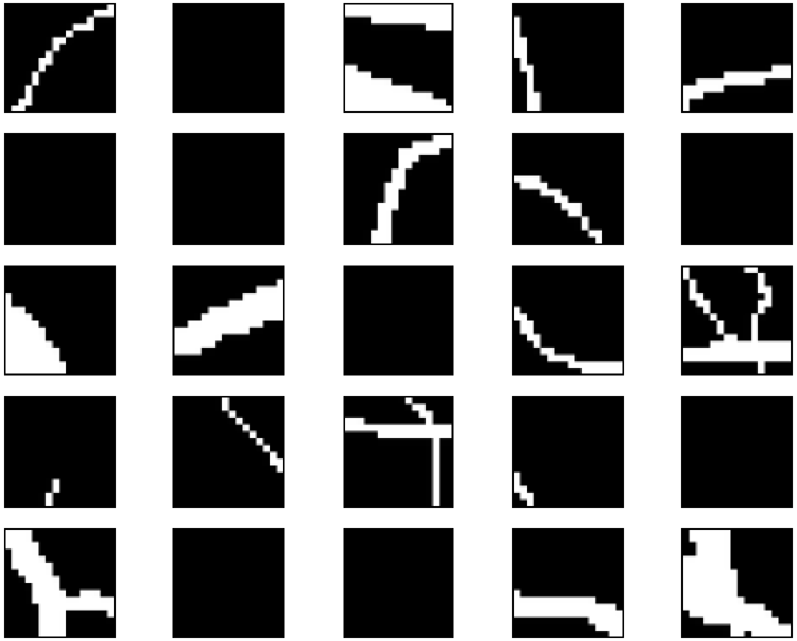
CNN Training Time

- about 80 epochs per model
  → 90 to 120 minutes training

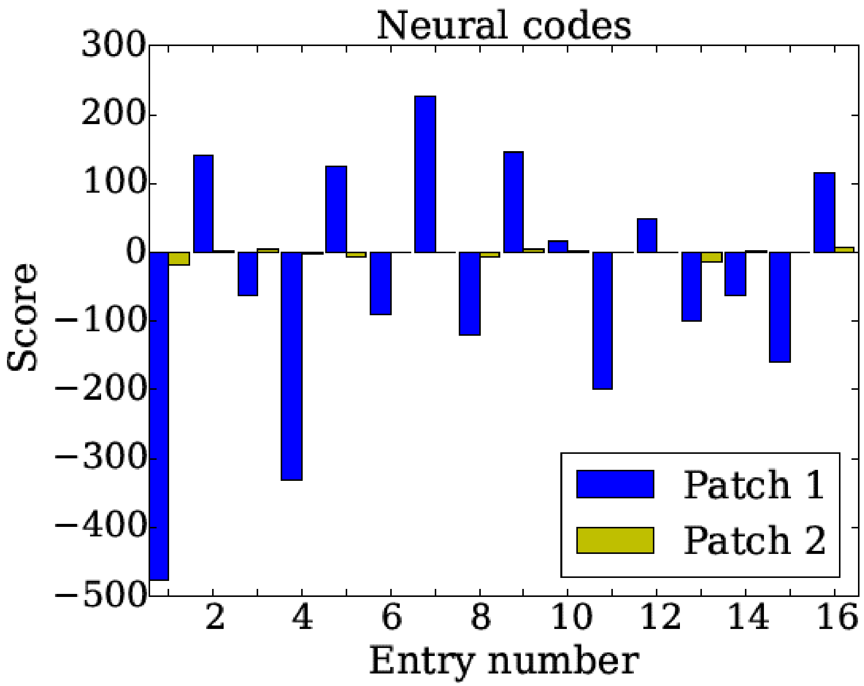


Dictionary

- retrieve neural codes from randomly sampled patches
- set nearest-neighbour classifier



example entries from dictionary



neural codes

Outline

- Introduction
- Approach
  - Training
  - **Testing**
- Experiments
- Discussion
- Conclusion and Outlook

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Testing

- not patches but whole images
- re-implement fully connected layers as convolutional
- average over all output patches

→ output is soft classification

Outline

- Introduction
- Approach
    - Training
    - Testing
- **Experiments**
- Discussion
- Conclusion and Outlook

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

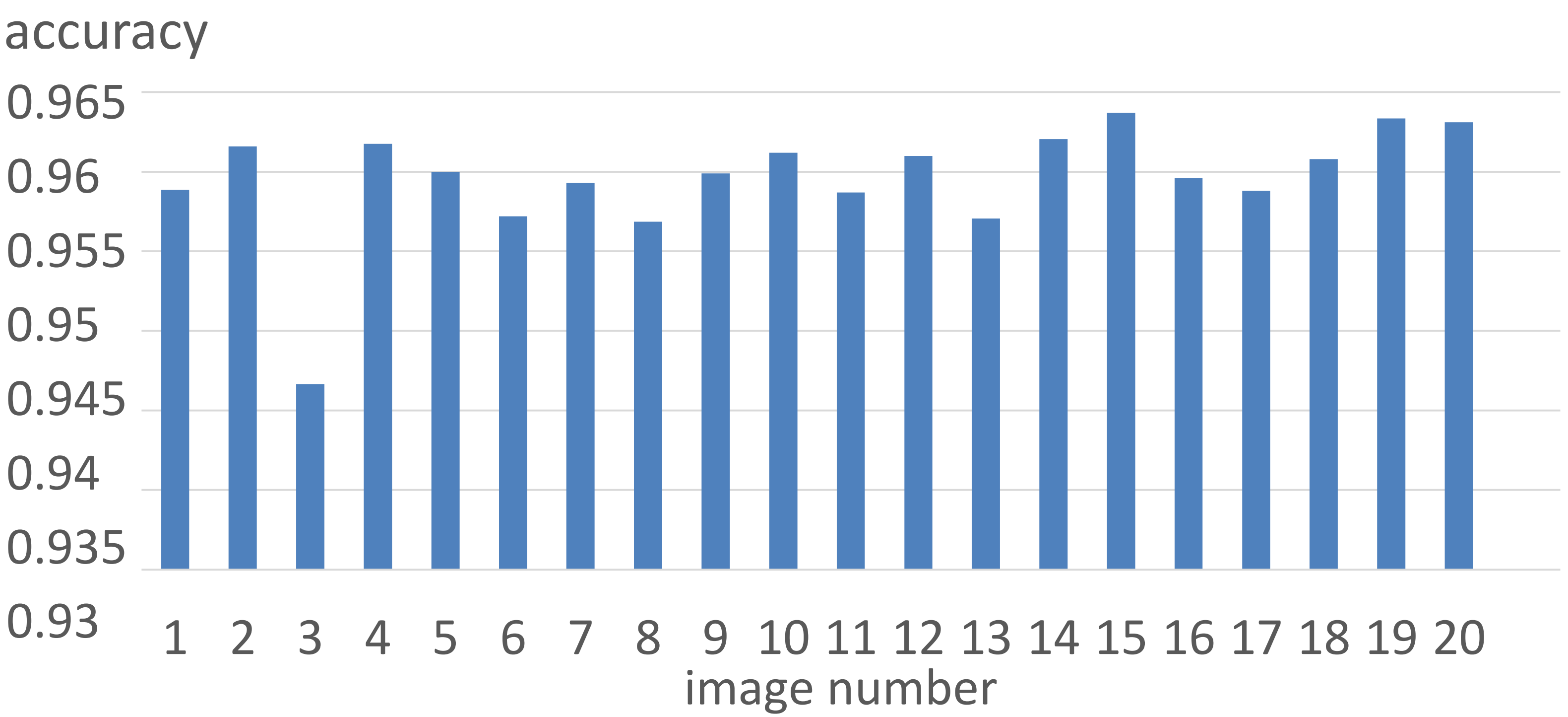
Experiments

- on DRIVE database
- soft classification has to be thresholded
- → simple thresholding that maximizes average accuracy

Trials:

- neural code length
- dictionary size
- multi-resolution
- normalized neural codes
- green colour channel only
- ensemble of multiple models

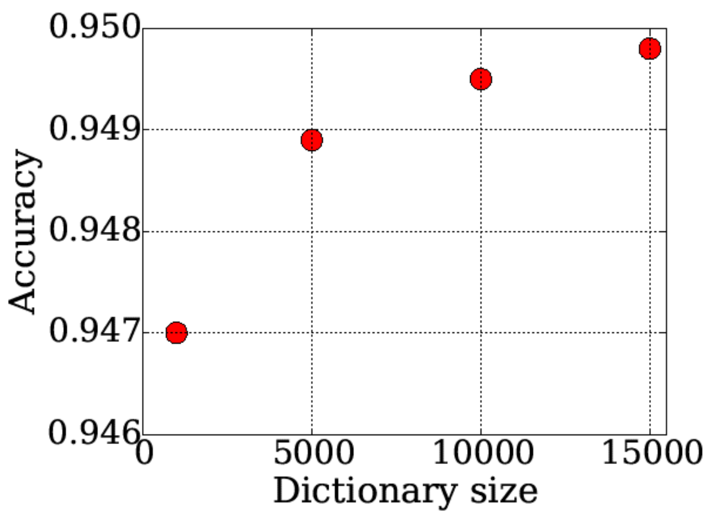
Original N^4 Settings



- neural code length = 16, dictionary size = 10000
- mean accuracy of 0.9491
- mean F1 score of 0.7844 (N^4 paper: 0.81)

Dictionary Size

dictionary entries	accuracy
1000	0.9470
5000	0.9489
10000	**0.9495**
15000	0.9498



- increase in performance with growing dictionary size
- problem with long runtime for very large dictionaries

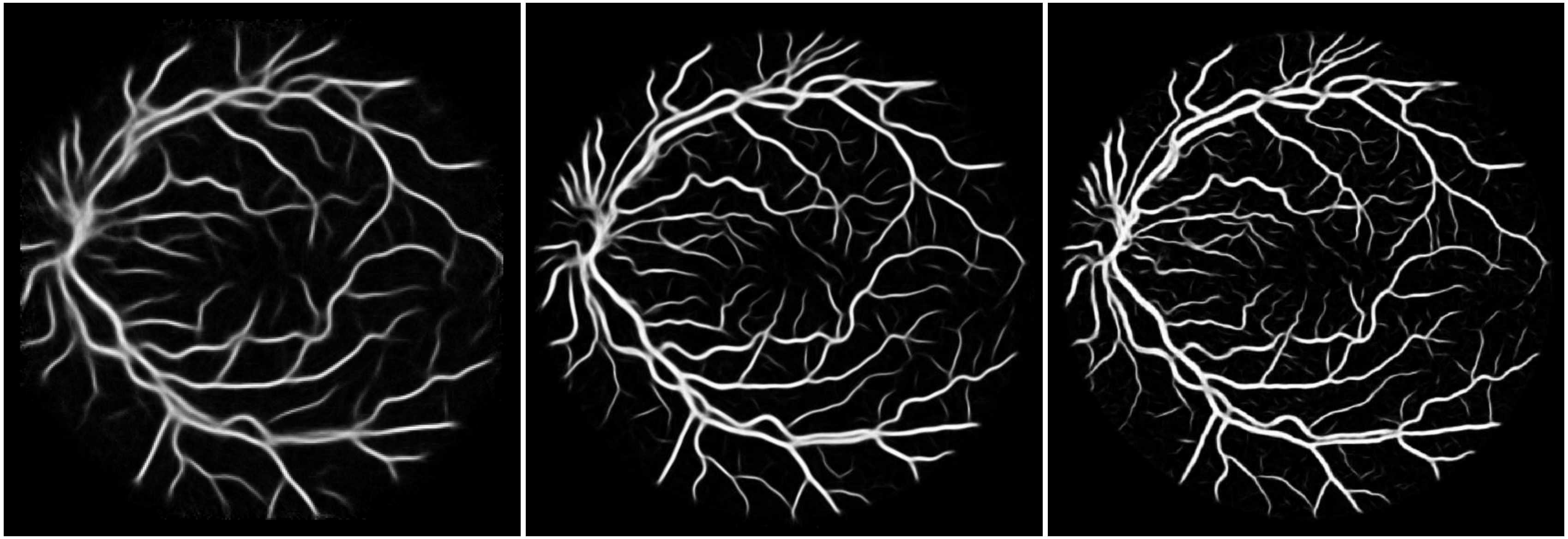
Neural Code Length

neural code length	final test error	accuracy
16	11241	0.9491
24	8600	0.9476
32	**8159**	**0.9495**
40	7502	0.9489

- test error for CNN decreases but no increase in performance
- 32 works well in other experiments

Multiple Scales

- train CNN on all 3 scales
- process input with 3 different resolutions each
- result: normal + 1.8 resolution (accuracy = 0.9506)

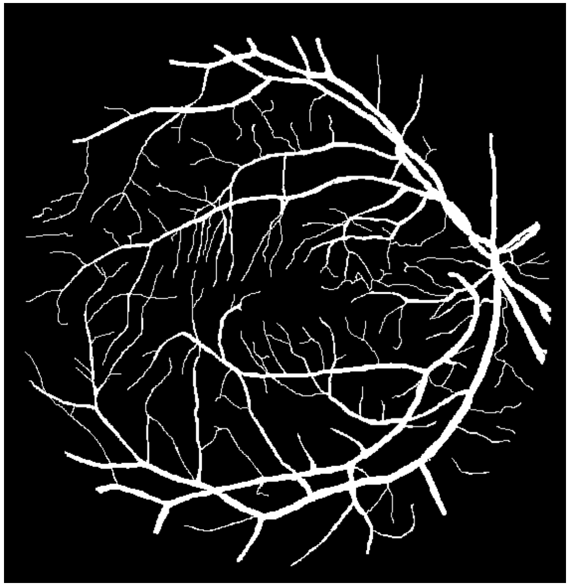


half resolution   normal resolution   1.8 resolution

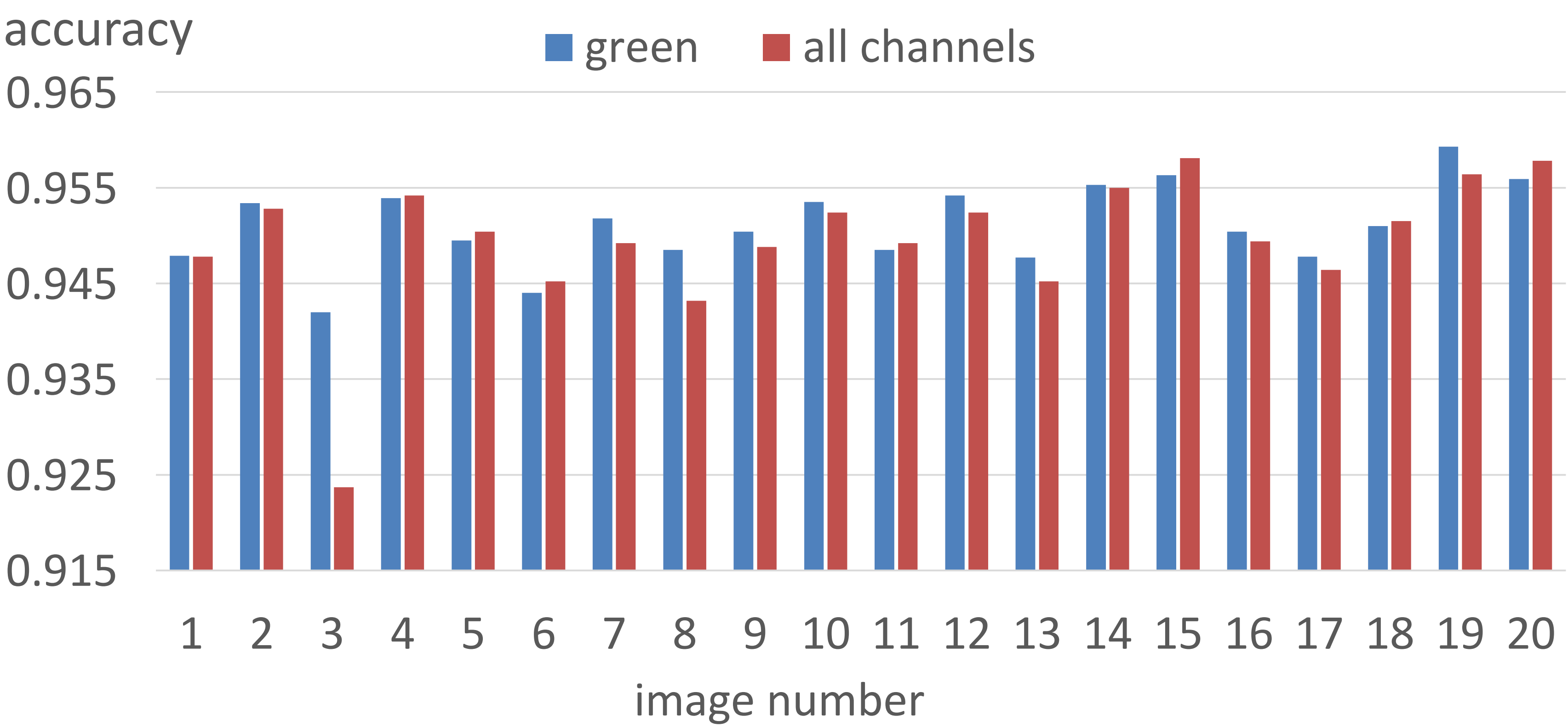
Normalized Neural Codes

- normalize neural codes in dictionary and from CNN
- accuracy = 0.9459 (unnormalized = 0.9495)
- many false positives, but thin vessels recognized


segmentation


ground truth

Green Colour Channel Only

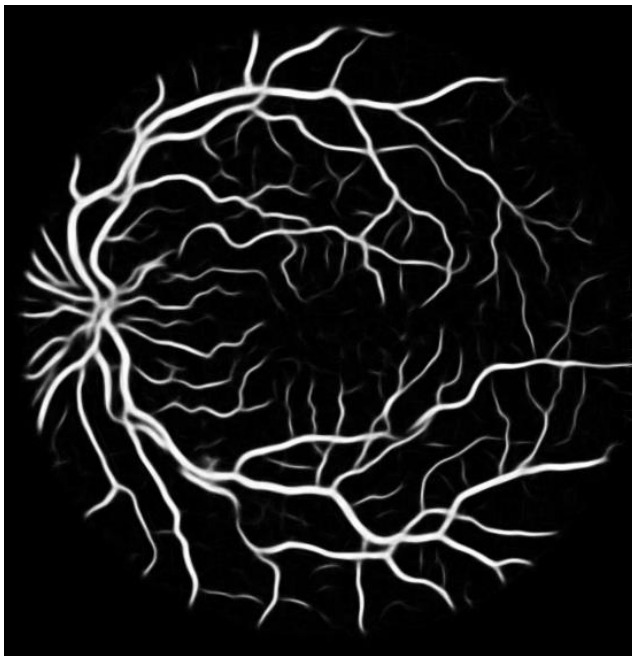


- mean accuracy green: 0.9511
- mean accuracy all channels: 0.9495

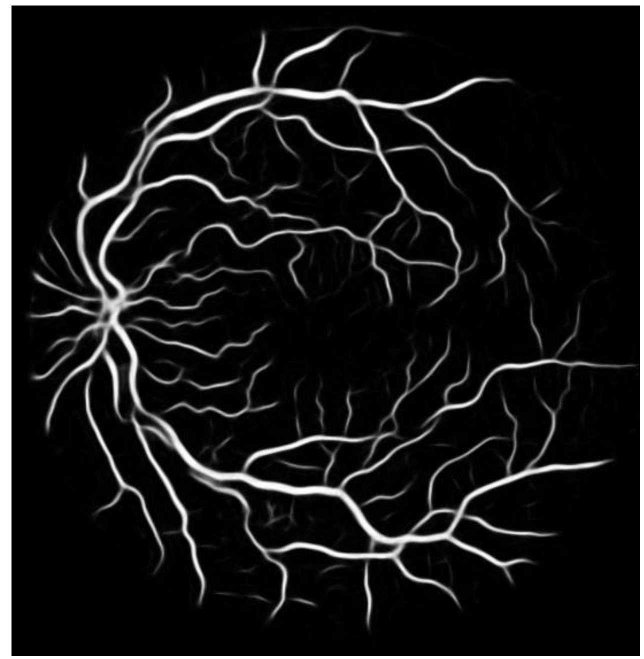
overall best model

Green Channel Only

$$\Delta_{accuracy} = 0.0183$$



green
accuracy = 0.942



all channels
accuracy = 0.9237

Ensemble

- combine output of multiple models

  → *green, 15000, 1.8* and *norm* (accuracy = 0.9521)

- adding more models does not necessarily increase performance

- disadvantage: long runtime, 4 CNNs have to be trained

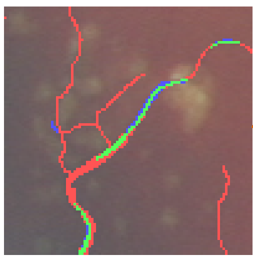
Experiments – Summary

model	neural code length	dictionary size	accuracy
original	16	10000	0.9491
dictionary size	32	15000	0.9498
neural code length	32	10000	0.9495
normal + 1.8 resolution	32	10000	0.9506
normalized neural codes	32	10000	0.9459
green channel only	**32**	**10000**	**0.9511**
green, 15000, 1.8* and *norm	**32**	**10000 15000**	**0.9521**

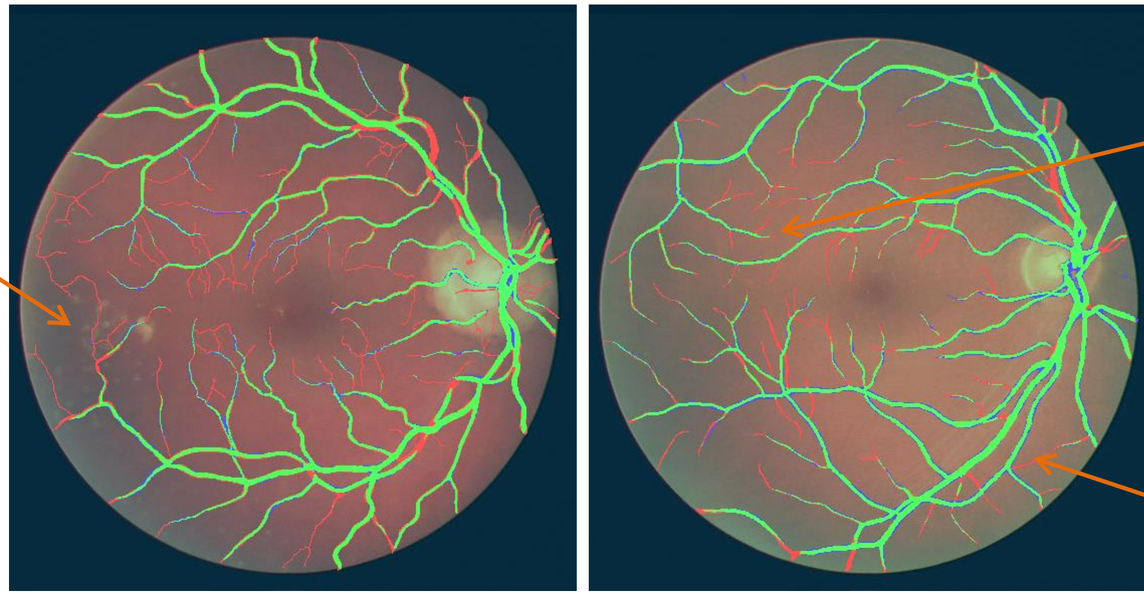
Outline

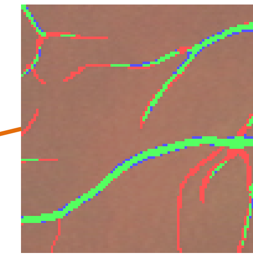
- Introduction
- Approach
    - Training
    - Testing
- Experiments
- **Discussion**
- Conclusion and Outlook

Discussion

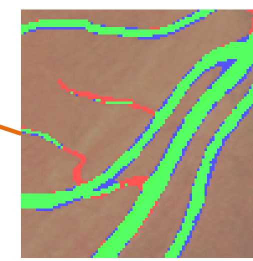


close to pathologies





bifurcations

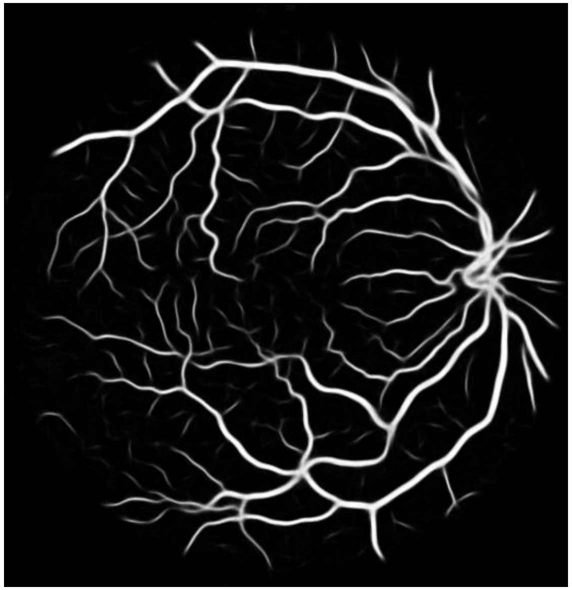


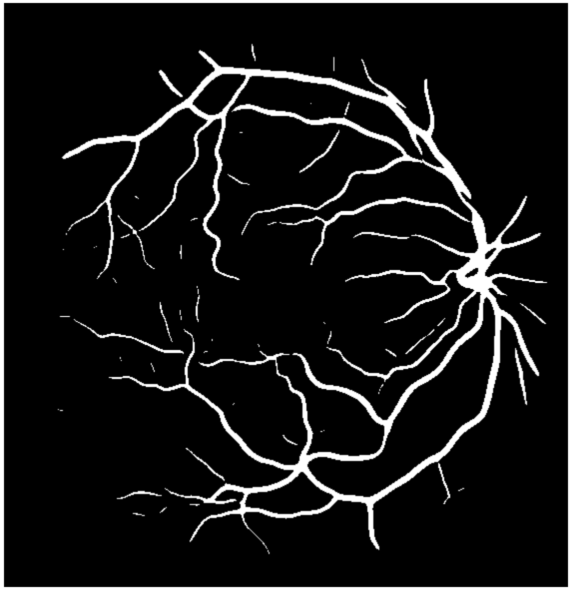
thicker vessels

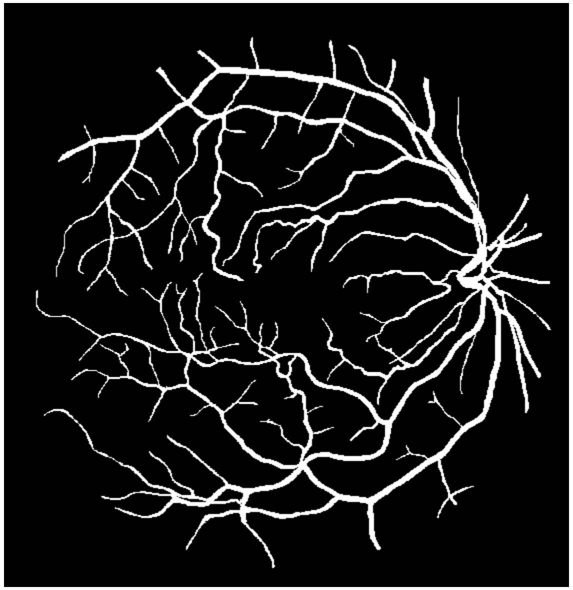
green = true positive
red = false negative
blue = false positive

Discussion - Binarization

- trade-off between thin vessels and correct vessel thickness


soft classification


binary segmentation


ground truth

Outline

- Introduction
- Approach
  - Training
  - Testing
- Experiments
- Discussion
- **Conclusion and Outlook**

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Conclusion and Outlook

- average accuracy 0.9511 on DRIVE database
- green colour channel only
- runtime 40 seconds, about 100 minutes training
- for optic disc, central vessel reflex and main vessels good performance
- very thin vessels and bifurcations still challenging

- more advanced thresholding technique (hysteresis, adaptive)
- higher resolution images
- reduce runtime
- different internal representation than PCA of patches

Thank you for your attention!

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT