Research Projects that can host a thesis

Giacomo Boracchi

DEIB, Politecnico di Milano

February, 2018

giacomo.boracchi@polimi.it

home.deib.polimi.it/boracchi/

Outline

Here is a list of the major reserach topics for a thesis related to "Image Classification".

In the next slides you’ll find a more detailed description

- Image Classification for Quality Assessment
- Classification on High-dimensional Images
- Robust Features For Object Detection
- Dictionary Learning
- Mutispectral X-Ray Analysis
- Convolutional Sparsity
- Foveated Image Features
- Anomaly Detection

For any enquiry, write me an email:

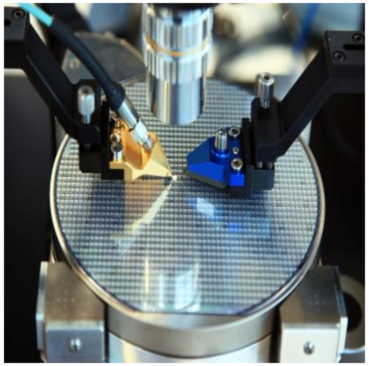
giacomo.boracchi@polimi.it

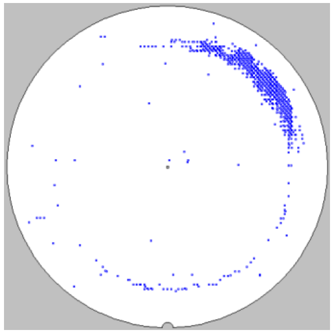
Image Classification for Quality Assessment

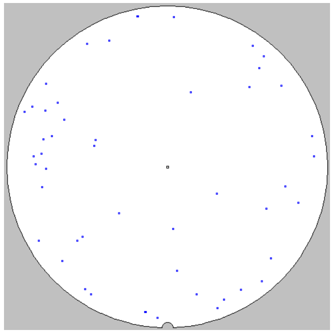
Monitoring Wafer Production

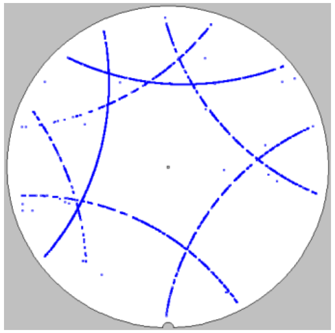
Automated Inspection Systems provide a **wafer defect map** containing the coordinates of each defective die.

The wafer defect map can be seen as an **high resolution** (20000 x 20000) **binary image**



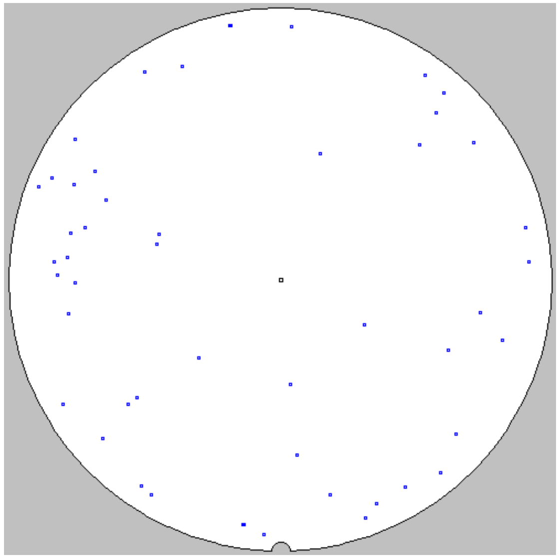


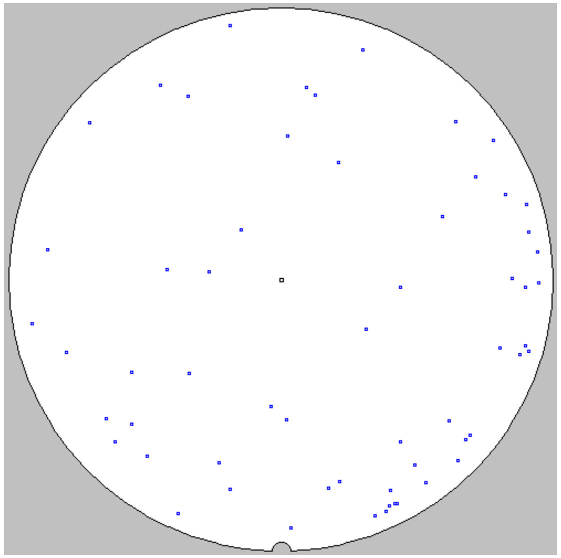




Normal Wafer Defect Maps

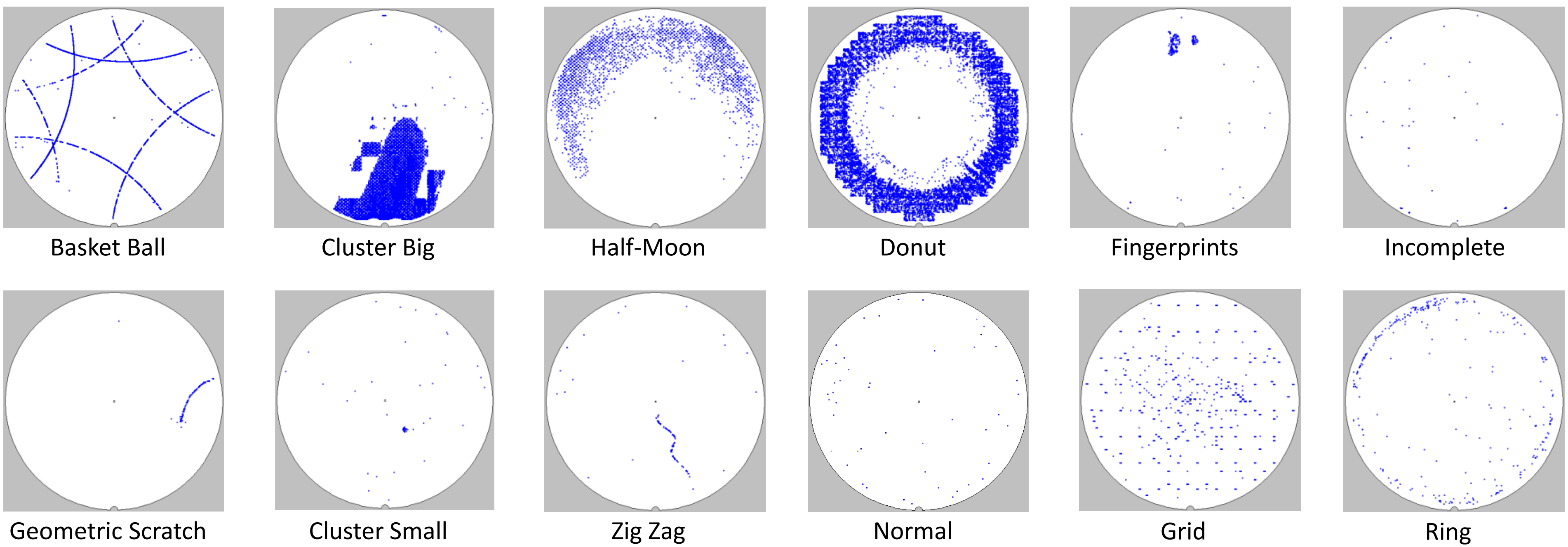
It is impossibile to produce a wafer that does not present any defect: a wafer defect map is considered **normal** when the number of defects is small and uniformly spread in the map.





Anomalous Wafer Defect Maps

Patterns in the wafer defect map indicate a **problem in the production line**



Goals and challenges

The goals of the project are

- **designing a deep convolutional neural network** architecture to address the classification problem
- Designing a specific **data augmentation** procedure to exploit the symmetry propoerties of the wafer defect maps

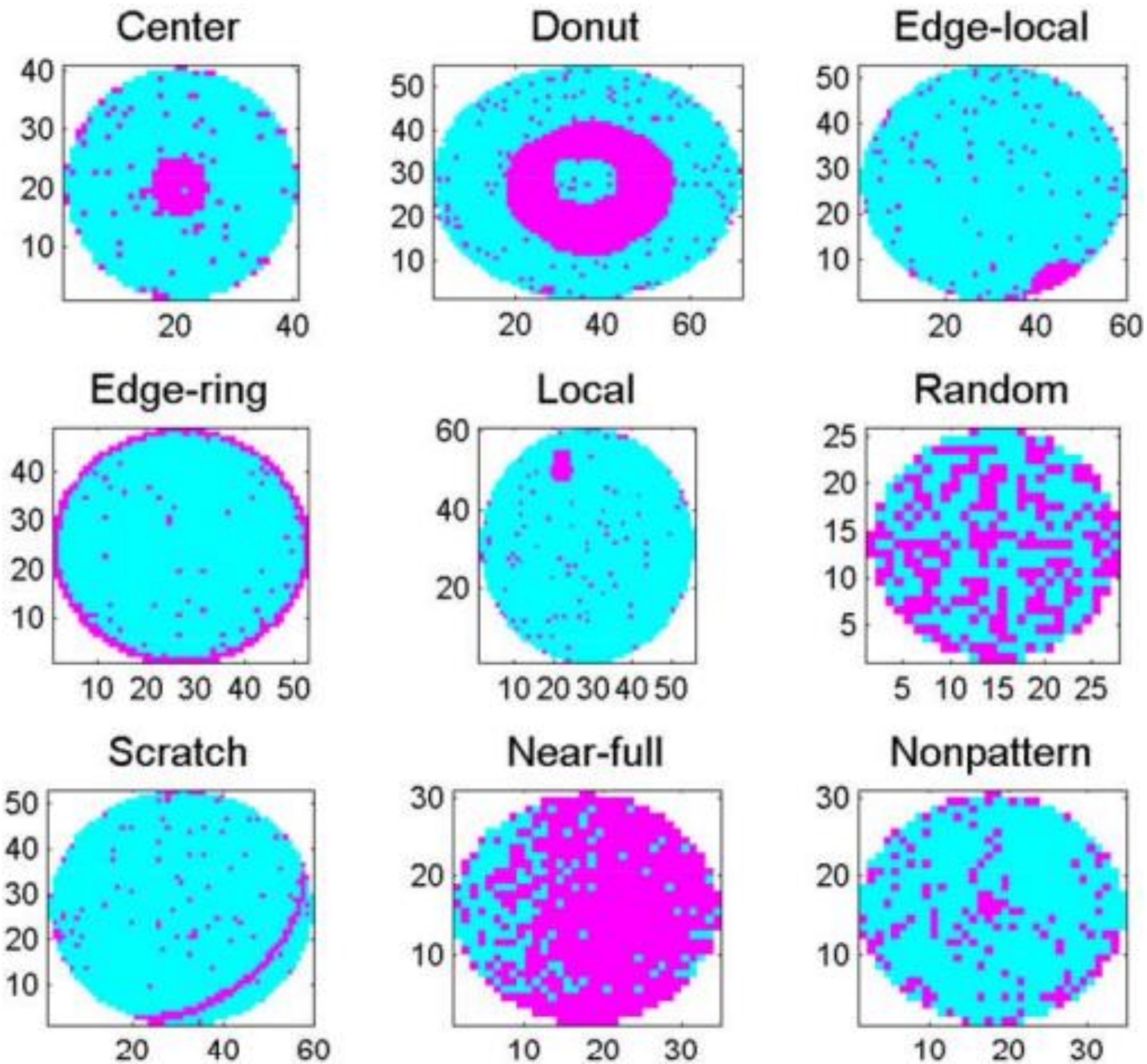
Challenges:

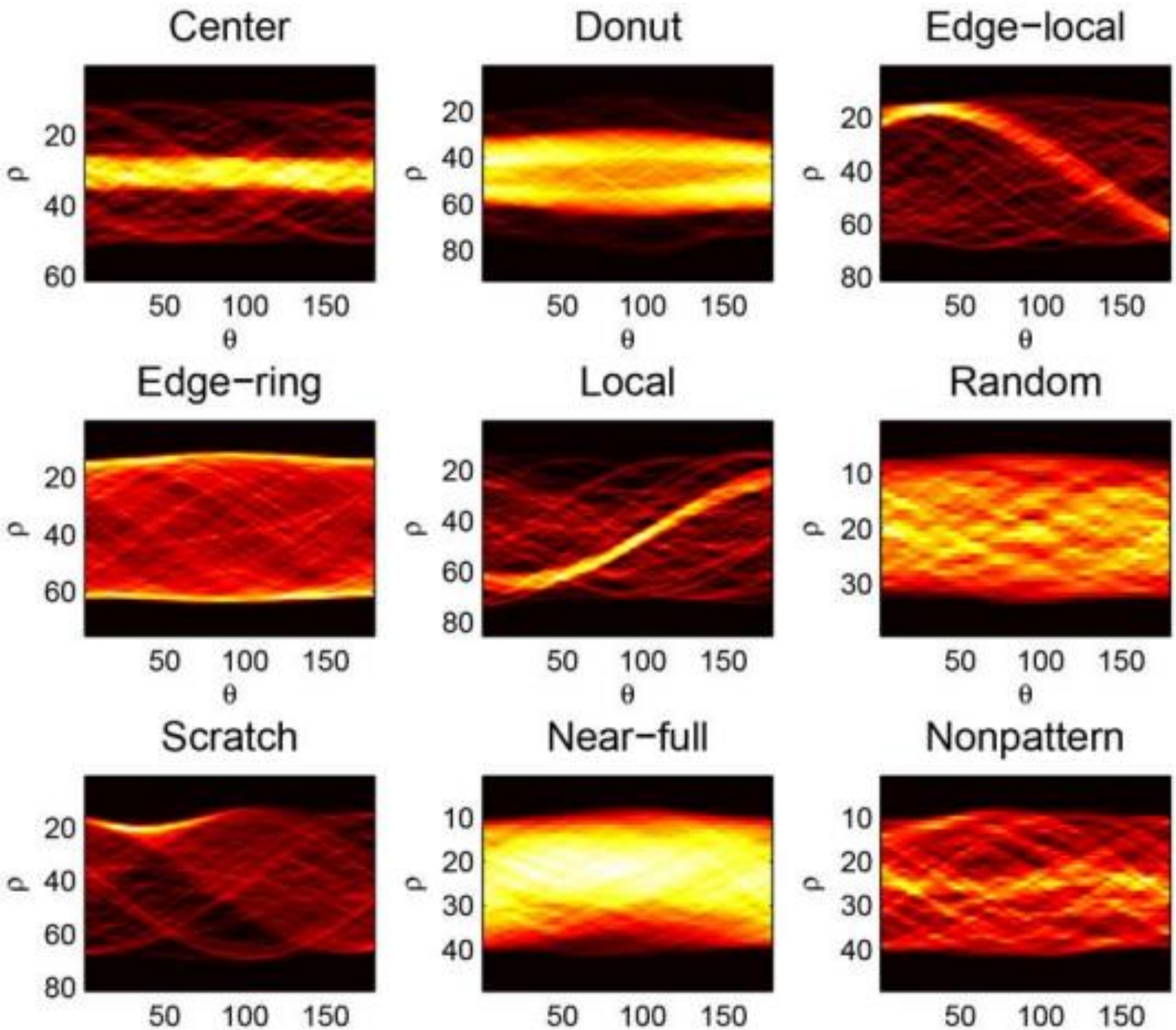
- **Imbalance dataset** (some patterns are very rare)
- **High dimensional data** (20000 x 20000 pixel)

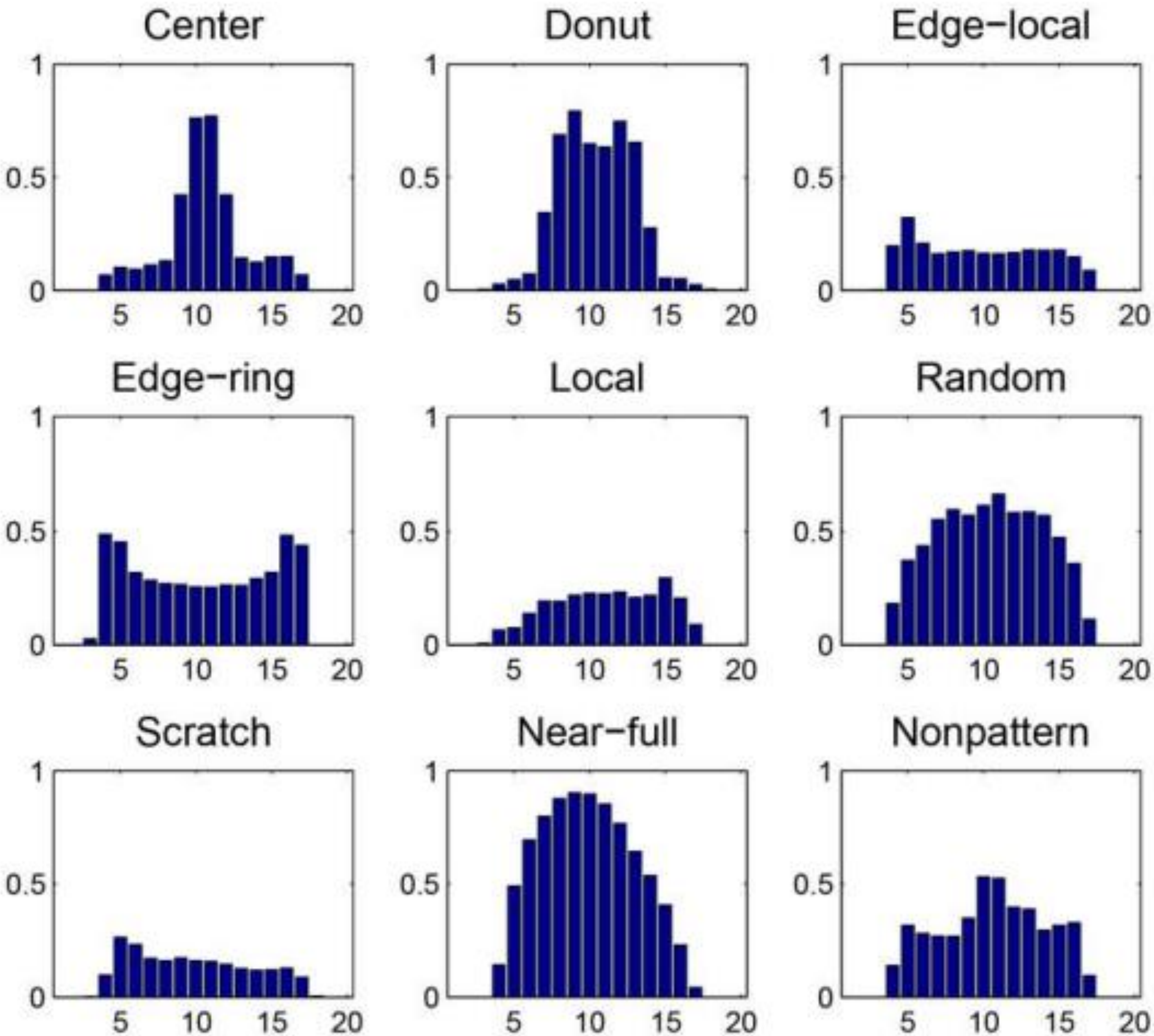
There is a public dataset (of much smaller images) to test the architecture http://mirlab.org/dataSet/public/

... and performance to compare with (hand-crafted features)

Example images of the public dataset




Center
Donut
Edge−local
Edge−ring
Local
Random
Scratch
Near−full
Nonpattern
ρ
θ
20
40
60
80
10
30
50
100
150



Center
Donut
Edge−local
Edge−ring
Local
Random
Scratch
Near−full
Nonpattern
1
0.5
0
5
10
15
20

Opportunities

The project offers the following opportunities

Thesis: a research thesis.

Thesis + Stage: the research thesis will be carried out under the joint supervision of the Applied Math team in the STMicroelectronics Labs (Agrate Brianza, via Camillo Olivetti, 2). Stage can last from 6 to 9 months, the net salary is 600 €/month, cantin + transportation from Gobba / Lambrate are included.

Stage for graduate students: as above, net salary of 750 €/month

PhD grant sponsored by STMicroelectronics on these topics: interested candidates, please apply!

Classification on High-dimensional Images

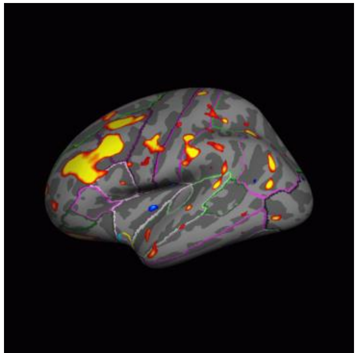
Biomarkers Huntington's Disease (HD)

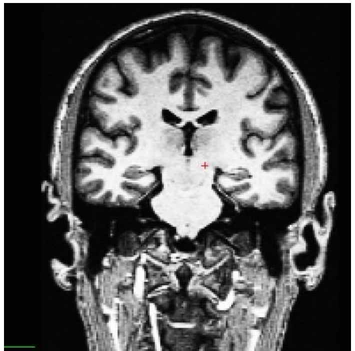
HD is a genetic, neurodegenerative condition causing loss of volumes and atrophy in the brain. Symptoms include motor, cognitive and emotional disorder.

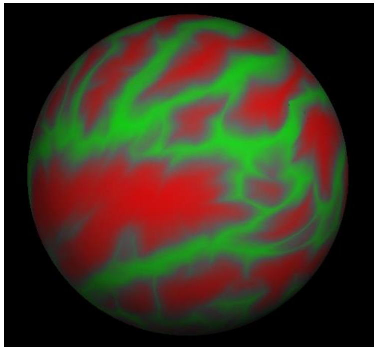
At the moment there exist no biomarkers able to assess the disease progression before the symptomatic stage.

MRI as well as fMRI and DTI have a great potential in the design of new data-driven and user-specific biomarkers.

This work was done in **collaboration with IBM TJ Watson** who provided us data from Huntington Fundation







MRI provides evidence of the disease when it is manifest

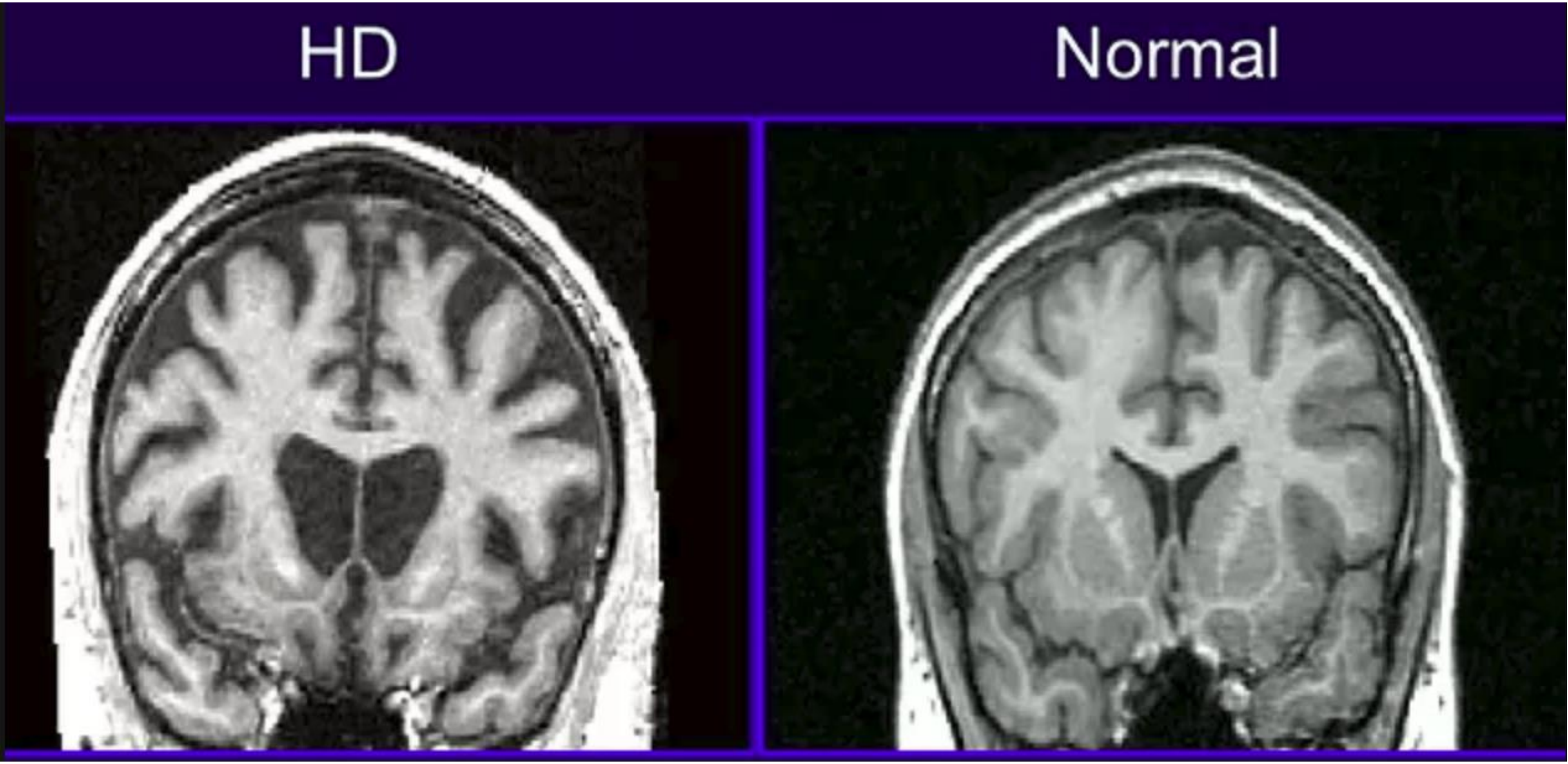
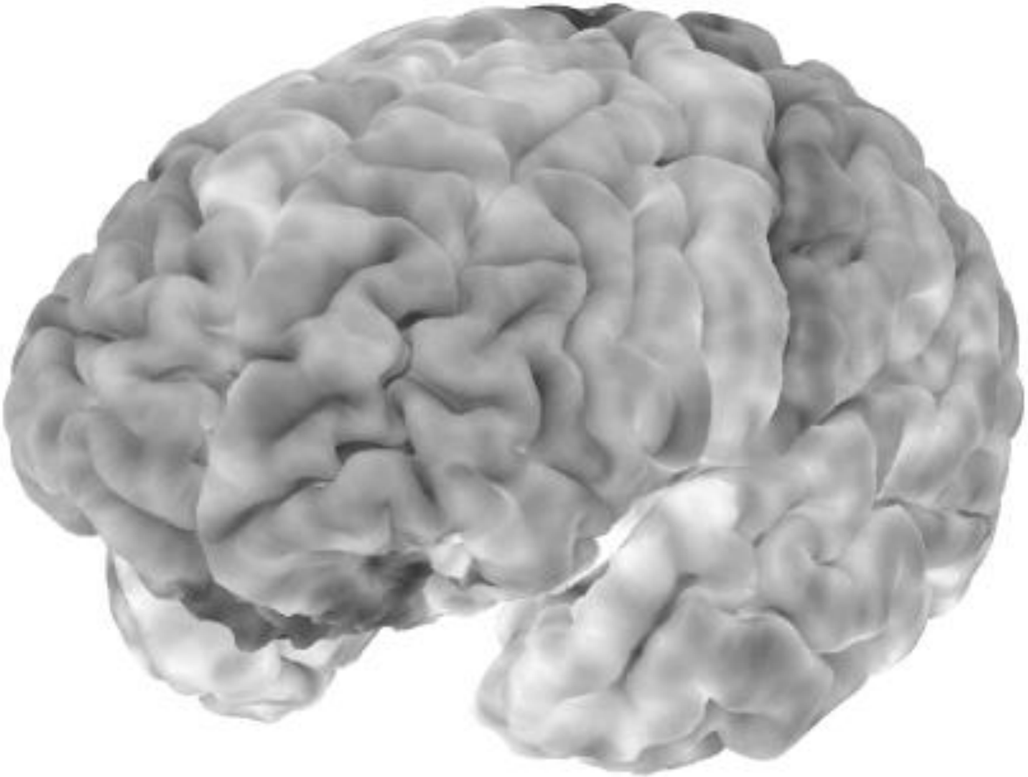


IMAGE CREDITS UC San Diego School of Medicine

Biomarkers Huntington's Disease (HD)

We designed a biomarker to analyze the structures characterizing the brain cortical surfaces:

- We pursued an anomaly-detection approach
- We exploit dictionaries yielding sparse representations for describing normal data



Biomarkers Huntington's Disease (HD)

We designed a biomarker to analyze the structures characterizing the brain cortical surfaces:

- We pursued an anomaly-detection approach
- We exploit dictionaries yielding sparse representations for describing normal data

Next steps are:

- work directly by comparing shapes, without need to pass through a data-driven model
- analyze different anatomical images (DTI, fMRI)
- design biomarkers to quantitively assess recovering from brain concussion
- Use GANs to visualize what is changing in these images over time (sequential monitoring)

See this interesting paper

Where GANs are used to learn a map $M(\boldsymbol{s})$ for an input image $\boldsymbol{s}$ such that the

$$\boldsymbol{y} = \boldsymbol{s} + M(\boldsymbol{s})$$

The output image $\boldsymbol{y}$ can not be correctly classified

So, if $\boldsymbol{s}$ is an image of a patient having HD, it provides a map that “occludes” regions that are reporting the biomarkers of the disease

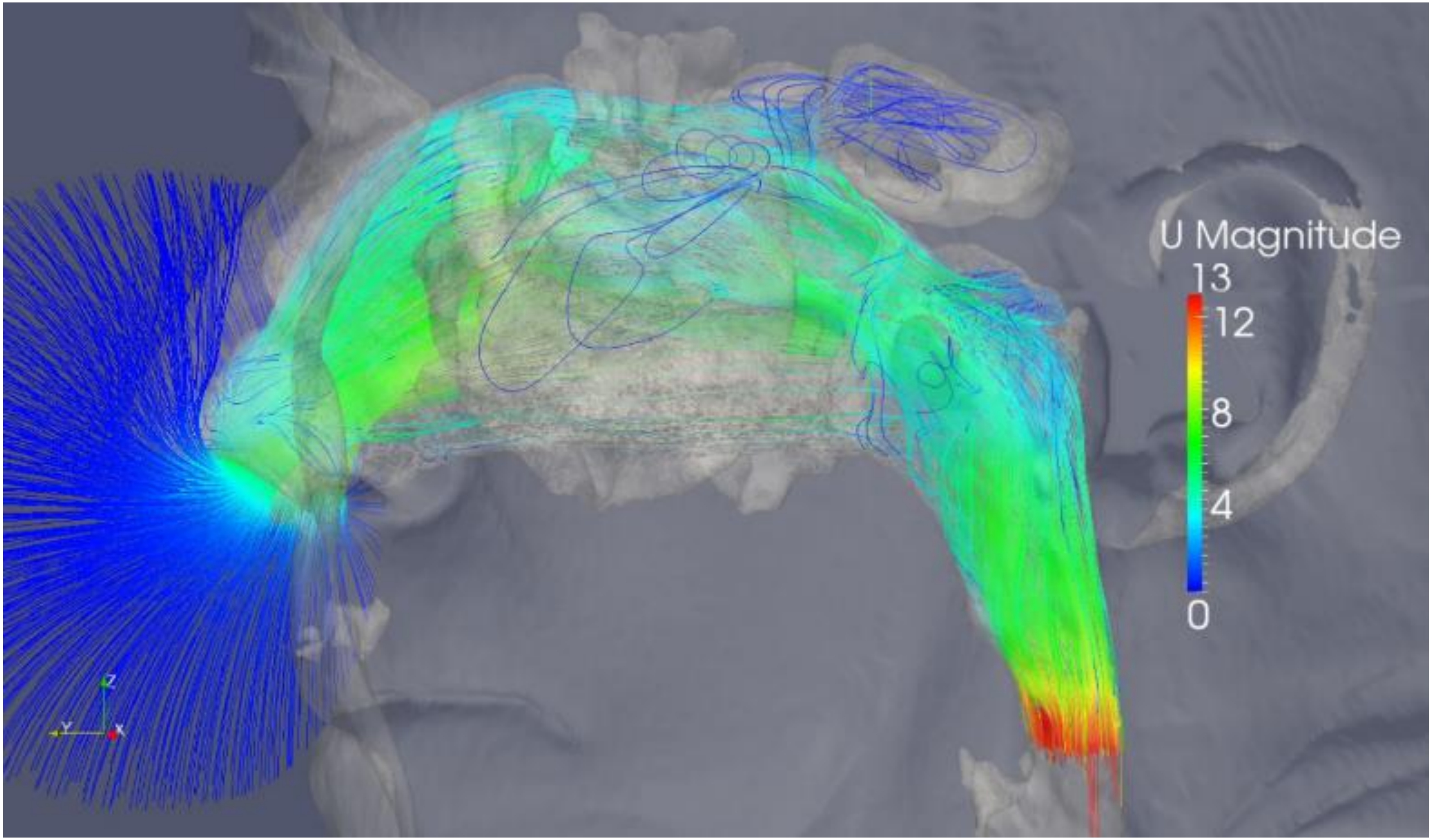
Visual Feature Attribution using Wasserstein GANs

Christian F. Baumgartner[1] Lisa M. Koch[2] Kerem Can Tezcan[1] Jia Xi Ang[1]
Ender Konukoglu[1] for the Alzheimer's Disease Neuroimaging Initiative*

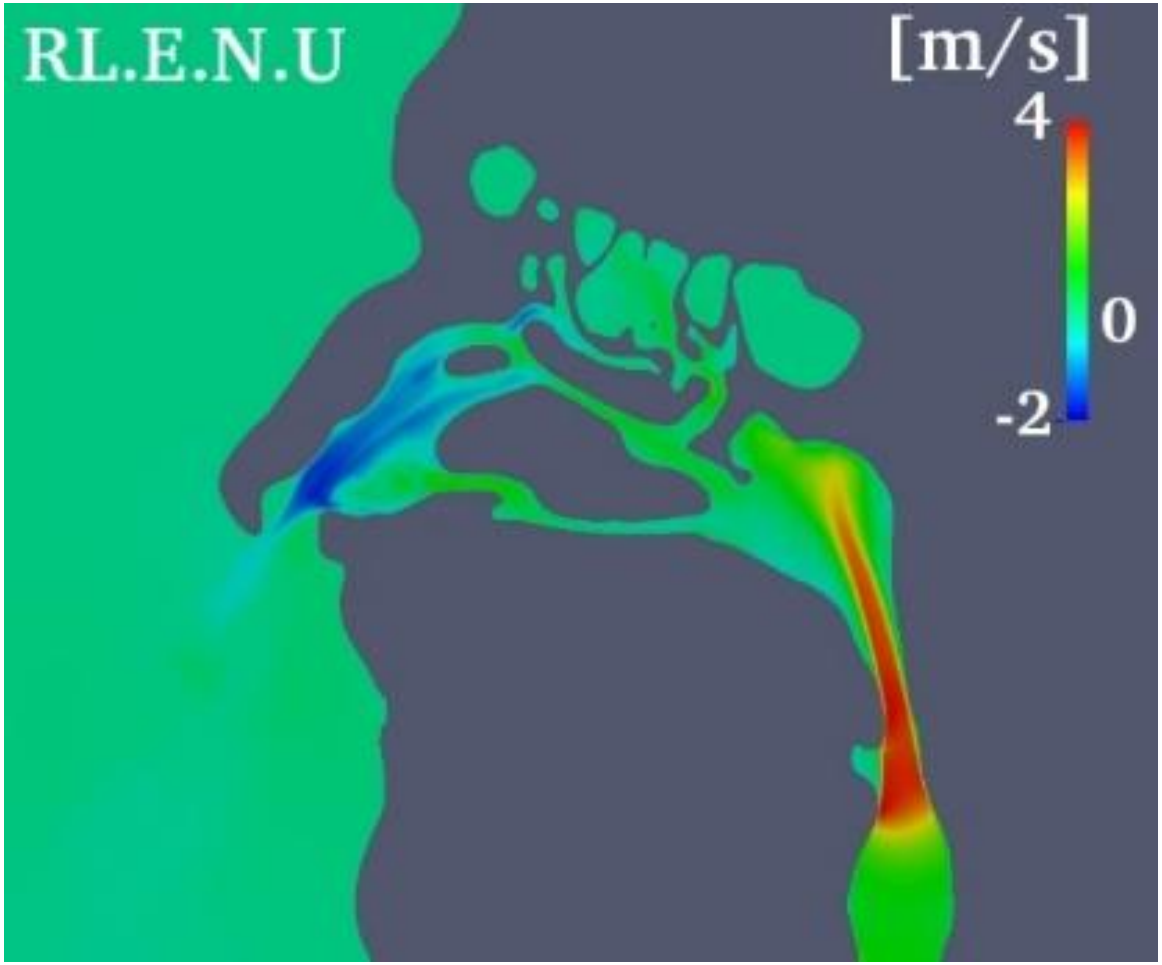
[1]Computer Vision Lab, ETH Zurich [2]Computer Vision and Geometry Group, ETH Zurich

Baumgartner, C. F., Koch, L. M., Tezcan, K. C., Ang, J. X., & Konukoglu, E. (2017). Visual Feature Attribution using Wasserstein GANs. *arXiv preprint arXiv:1711.08998*

Computational Fluid Dynamic and Machine Learning



Computational Fluid Dynamic and Machine Learning



Diagnosing nasal pathologies by analyzing airflows

Diagnosing nasal pathologies by automatically analyzing:

- Anatomical CT-scans of patients' face
- Airflows computed through Computational Fluid Dynamic (CFD)

Challenges:

- CFD yield a very high-dimensional 3d-vector field
- Extract features able that summarize properties of nasal cavities
- Selecting the most influential features for medical diagnosis

Methods and Materials:

- Features inspired to image analysis techniques.
- State-of-the-art feature selection methods from ML literature
- 50 CFD simulations of patients provided with the medical diagnosis

Expected outcomes:

- A classifier predicting the medical diagnosis from these data

Supervision:

- M. Restelli, G. Boracchi, Prof. M. Quadrio (Aero Polimi)

Robust Features for Object Recognition

Robust Detection by Feature Matching

Robust Detection by Feature Matching

Goal: detection by feature extraction and matching under geometrical constraints

Pro: it can handle occlusions and perspective deformation without need of training data but a template

Robust Detection by Feature Matching

Goal: detection by feature extraction and matching under geometrical constraints

Pro: it can handle occlusions and perspective deformation without need of training data but a template

Cons: ignores color information, requires some form of geometrical regularity, matching is based on a portion of the template, rather than the whole

Robust Detection by Feature Matching

Robust Detection by Feature Matching

Research Goals: impose local regularity to matches through Graph / Delanay representation of keypoint location.

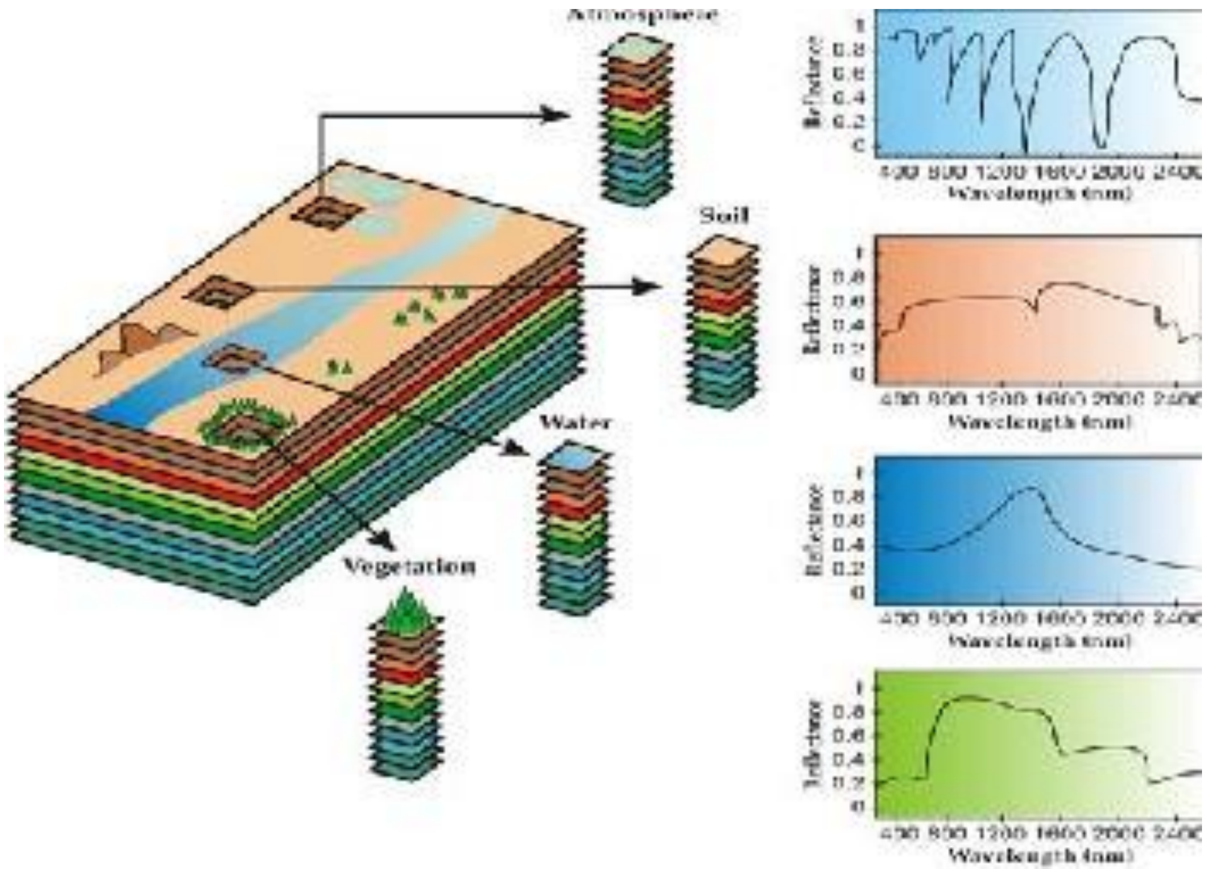
- These are weaker constraints than the homography (which is typically assumed when matching planar surfaces) but would allow us to detect non-planar surfaces.
- Pursue a multiscale approach
- Design eigen-template, as in face detection, to use a single template in case of many similar templates
- Analyze color and/or comparison against locally rectified image of the template

Multispectral X-ray Analysis

Multispectral X-ray images

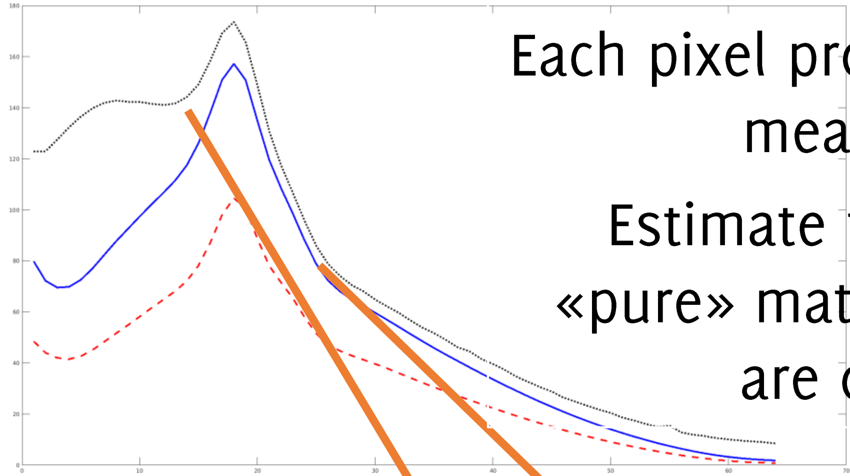
Analysis of multispectral X-ray images to identify hazardous materials

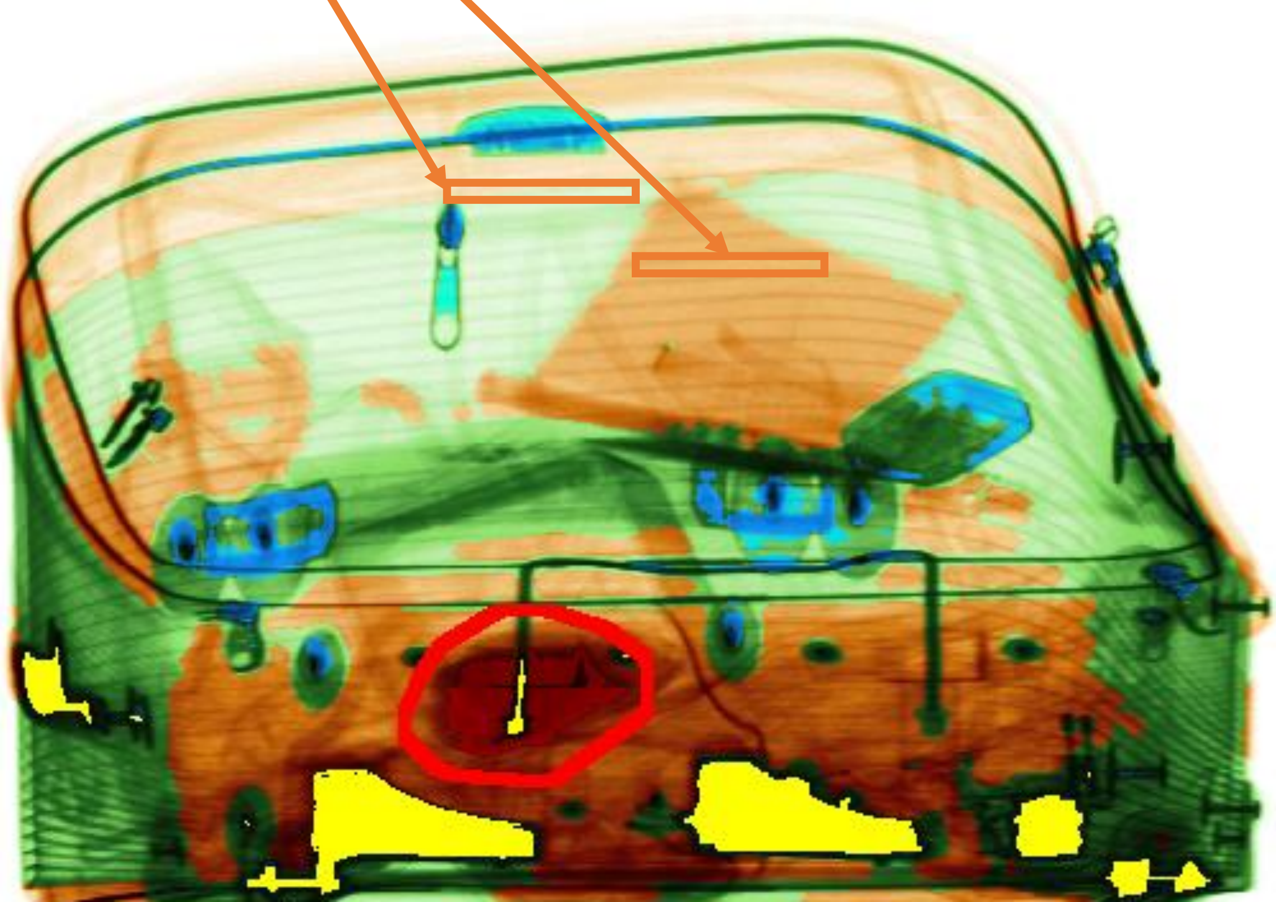
Each pixel provides a measurement over 64 spectral bands



Each pixel provide a spectrum of measurements

Estimate the spectrum of «pure» materials when these are overlapped





Spectral Unmixing for X-ray NDT

Analysis of multispectral X-ray images to identify hazardous materials

Each pixel provides a measurement **over 64 spectral bands**

Image Segmentation algorithms for multispectral images

Spectral unmixing algorithm to solve the overlap problem

- Provided the spectra from two overlapping materials, estimate the spectra of the two materials separately

Perform Classification over these sort of data

Dictionary Learning

Convolutional Sparsity

Convolutional sparse models are a recent development of sparse representations

$$\boldsymbol{s} \approx \sum_{i=1}^{n} \boldsymbol{d}_i \circledast \boldsymbol{\alpha}_i \,, \qquad \text{s.t. } \boldsymbol{\alpha}_i \text{ is sparse}$$

where a signal $\boldsymbol{s}$ is **entirely encoded** as the sum of n convolutions between a filter $\boldsymbol{d}_i$ and a coefficient map $\boldsymbol{\alpha}_i$

Pros:

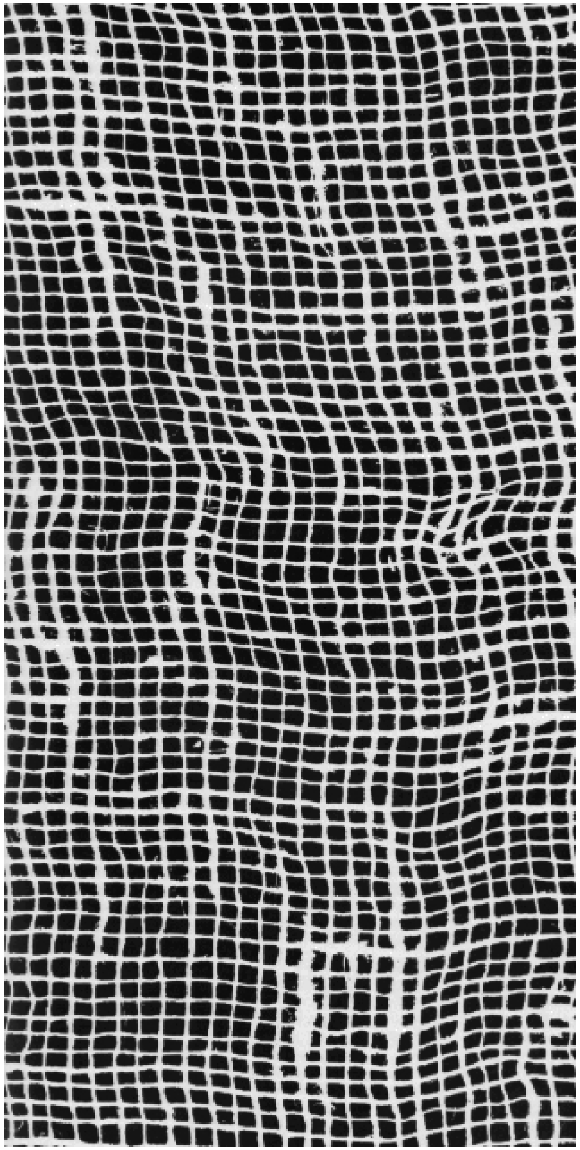
- Translation invariant representation
- Few small filters are typically required
- Filters exhibit very specific image structures
- Easy to use filters having different size

Collaboration with Los Alamos National Laboratory, NM, USA

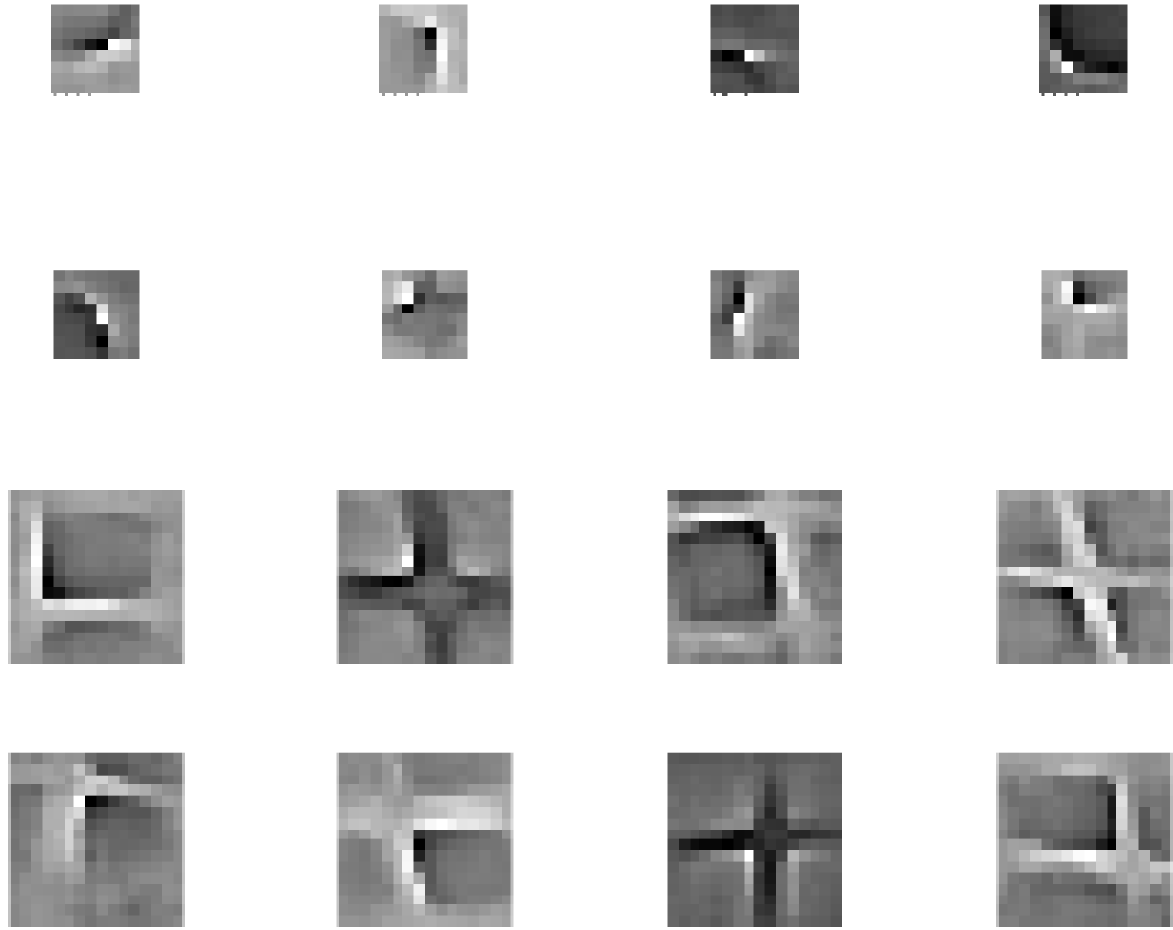
Los Alamos NATIONAL LABORATORY

Example of Learned Filters

Training Image



Learned Filters



Forgetting Mechanisms For Online Dictionary Learning

Dictionary learning is an unsupervised learning method that is very successful for images and signals

Goal: implement dictionary learning for online data. Introduce a forgetting mechanism for old data as it happens in adult-neurogenesis.

- Getting familiar with sparse representations and implement a simple forgetting mechanism (e.g., sliding window)
- Investigate more principled forgetting mechanisms where you remove samples that are less likely to be consistent with current data

Convolutional Sparsity

Convolutional Sparsity

Convolutional sparse models are a recent development of sparse representations

$$\boldsymbol{s} \approx \sum_{i=1}^{n} \boldsymbol{d}_i \circledast \boldsymbol{\alpha}_i \,, \qquad \text{s.t. } \boldsymbol{\alpha}_i \text{ is sparse}$$

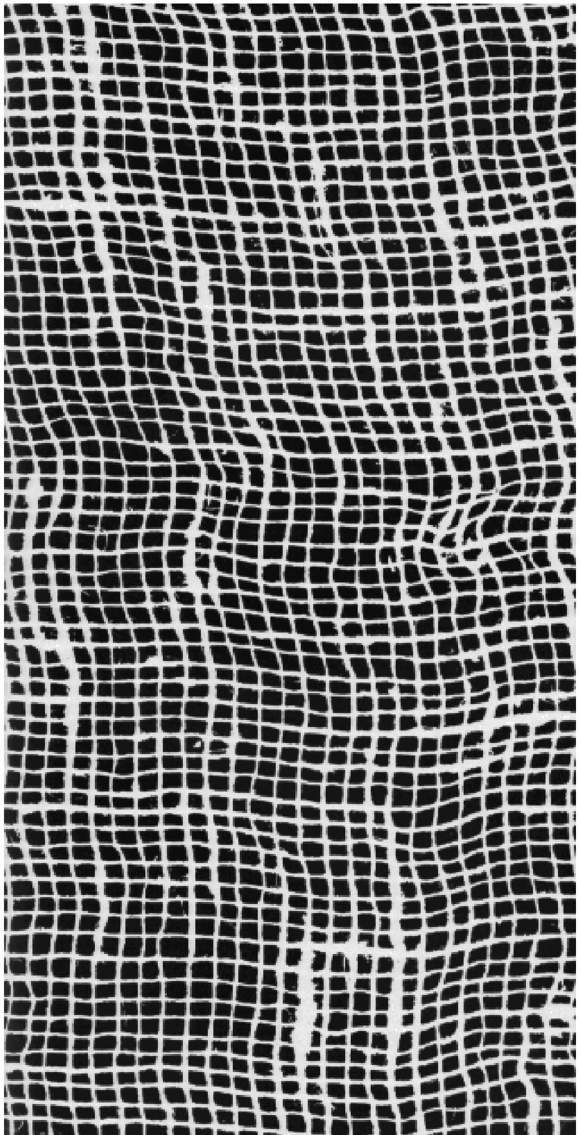
where a signal $\boldsymbol{s}$ is **entirely encoded** as the sum of n convolutions between a filter $\boldsymbol{d}_i$ and a coefficient map $\boldsymbol{\alpha}_i$

Pros:

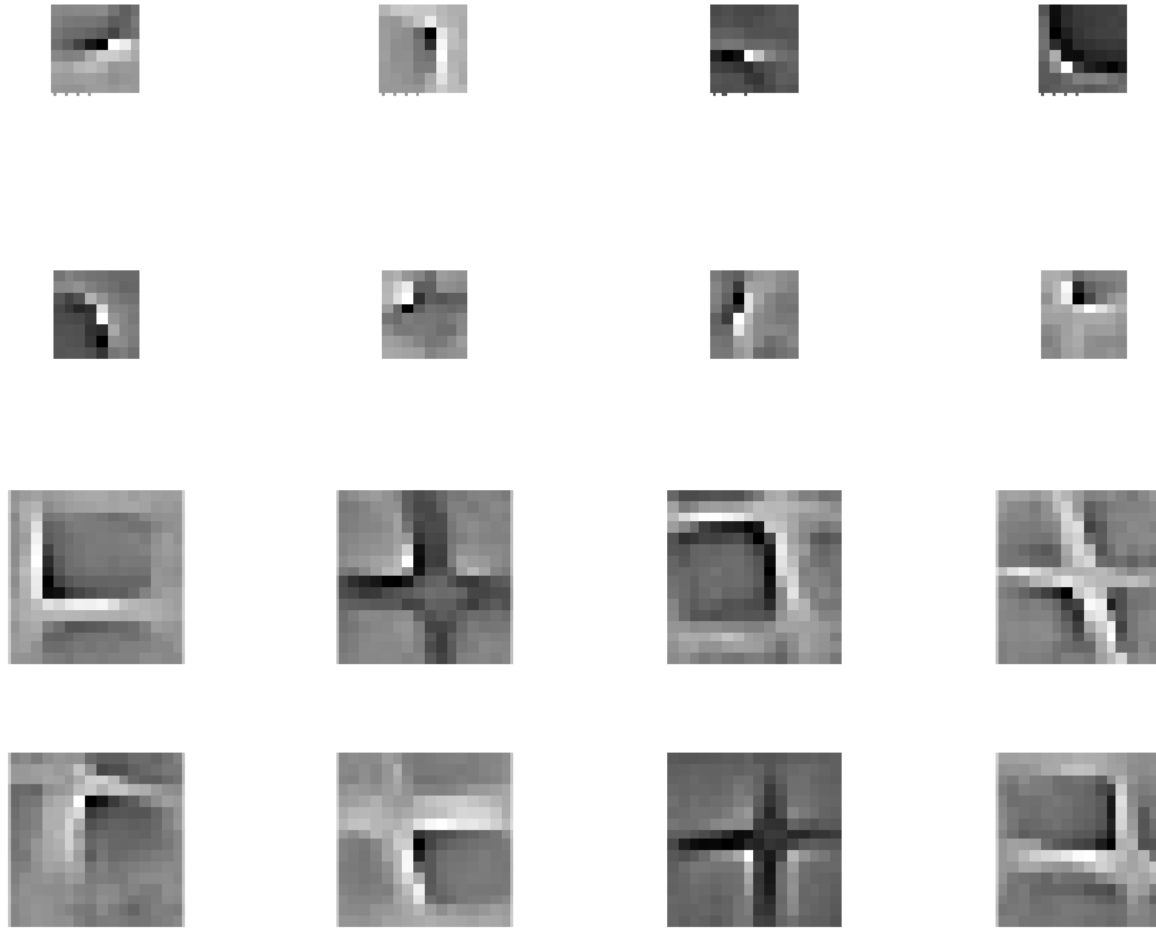
- Translation invariant representation
- Few small filters are typically required
- Filters exhibit very specific image structures
- Easy to use filters having different size

Example of Learned Filters

Training Image



Learned Filters



Convolutional Sparsity for Anomaly Detection

If we consider the convolutional sparse coding

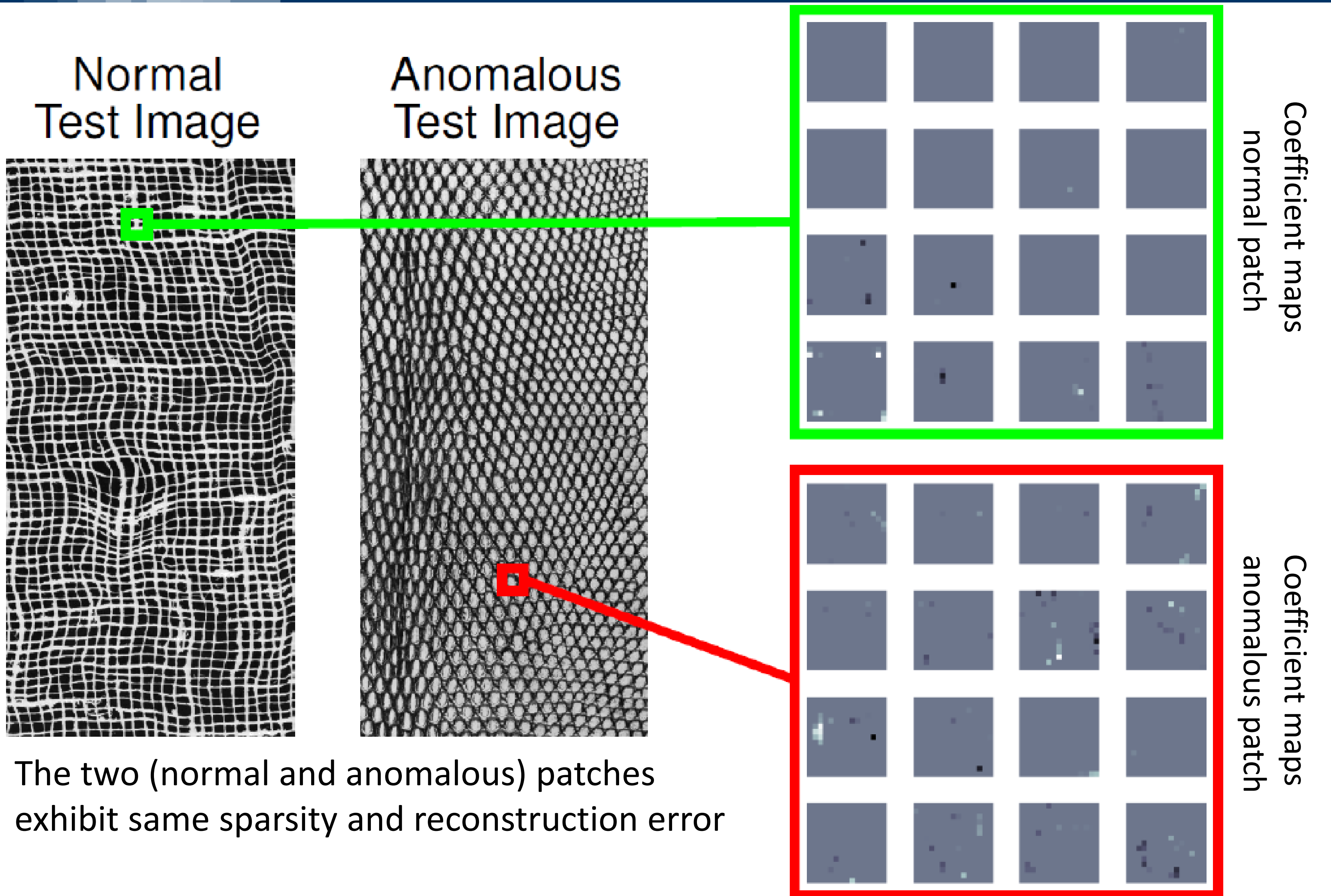
$$\{\widehat{\boldsymbol{\alpha}}\} = \underset{\{\alpha\}_n}{\operatorname{argmin}} \underline{\left\| \sum_{i=1}^{n} \boldsymbol{d}_i \circledast \boldsymbol{\alpha}_i - \mathbf{s} \right\|_2^2} + \lambda \underline{\sum_{i=1}^{n} \|\boldsymbol{\alpha}\|_1}$$

we can build the feature vector as:

$$\boldsymbol{x}_c = \begin{bmatrix} \left\| \prod_c \left(\sum_{i=1}^{n} \boldsymbol{d}_i \circledast \widehat{\boldsymbol{\alpha}}_i - \mathbf{s} \right) \right\|_2^2 \\ \sum_{i=1}^{n} \left\| \prod_c \widehat{\boldsymbol{\alpha}} \right\|_1 \end{bmatrix}$$

...but unfortunately, detection performance are rather poor

Sparsity is too loose a criterion for detection



The two (normal and anomalous) patches
exhibit same sparsity and reconstruction error

Contributions:

- Design a **feature vector** that accounts for the number of filters that are activated within each region

$$x_c = \begin{bmatrix} \left\| \prod_{\boldsymbol{c}} \left(\sum_{i=1}^{m} \boldsymbol{d}_i \circledast \widehat{\boldsymbol{\alpha}}_i - \mathbf{s} \right) \right\|_2^2 \\ \sum_{i=1}^{m} \left\| \prod_{\boldsymbol{c}} \widehat{\boldsymbol{\alpha}} \right\|_1 \\ \sum_{i=1}^{m} \left\| \prod_{\boldsymbol{c}} \widehat{\boldsymbol{\alpha}} \right\|_2 \end{bmatrix}$$

[IJCNN 2015] D. Carrera, G. Boracchi, A. Foi and B. Wohlberg , *"Detecting Anomalous Structures by Convolutional Sparse Models "* IEEE IJCNN 2015

Convolutional Sparsity for Anomaly Detection

Contributions:

- Design a **feature vector** that accounts for the number of filters that are activated within each region
- Design an **efficient sparse coding** algorithm that includes a term penalizing the local group sparsity

$$\{\widehat{\boldsymbol{\alpha}}\} = \underset{\{\boldsymbol{\alpha}\}_m}{\operatorname{argmin}} \left\| \sum_{i=1}^{m} \boldsymbol{d_i} \circledast \boldsymbol{\alpha_i} - \mathbf{s} \right\|_{\mathbf{2}}^{\mathbf{2}} + \lambda \sum_{i=1}^{m} \|\boldsymbol{\alpha}\|_1 + \xi \sum_{\boldsymbol{c}} \sum_{i=1}^{m} \left\| \prod_{\boldsymbol{c}} \boldsymbol{\alpha} \right\|_2$$

[IJCNN 2015] D. Carrera, G. Boracchi, A. Foi and B. Wohlberg , *"Detecting Anomalous Structures by Convolutional Sparse Models "* IEEE IJCNN 2015

Foveated Image Features

Foveated Image Features… a project/thesis

A new method to design image features, which are inspired by two remarkable properties of the human visual system:

- Foveation
- radial orientation preference

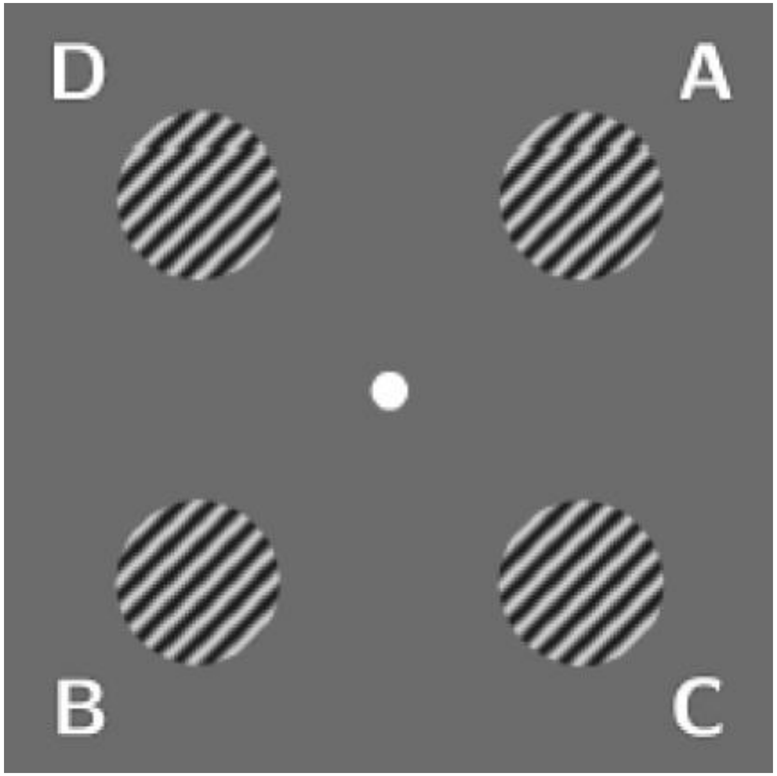
Foveated Image Features… a project/thesis

Foveation refers to the fact that the HVS is characterized by maximal acuity at the fixation point, and decreasing acuity towards the periphery of the visual field

(a) (b)

Lena foveated w.r.t. two different fixation points

Radial orientation preference refers to an increased sensitivity of the HVS to perceive patterns oriented towards the fixation point

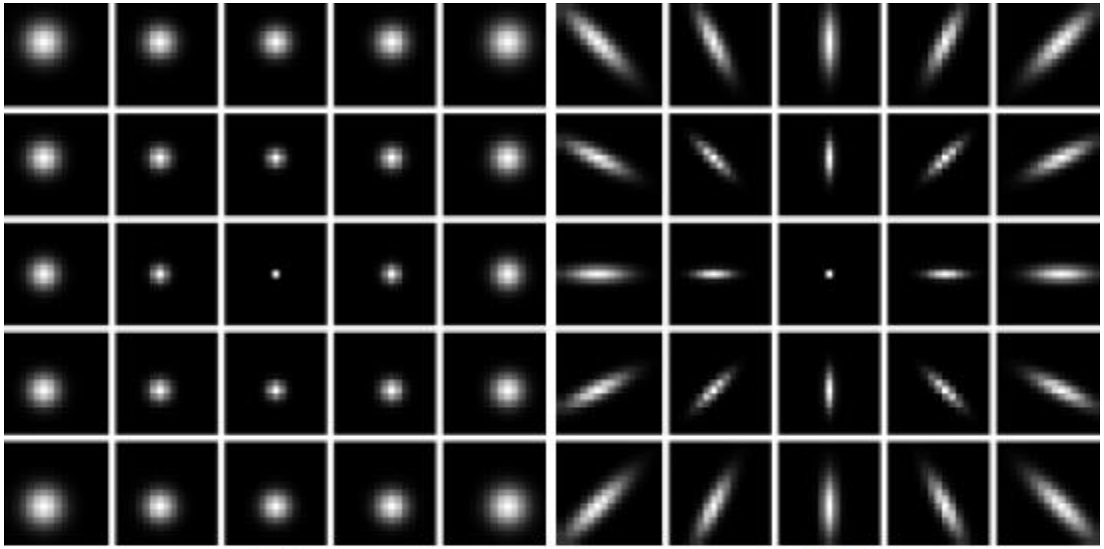


When gazing the white dot, the HVS is more sensitive to patterns like A and B that are oriented toward the fixation point, rather than C and D;

Foveated Image Features… a project/thesis

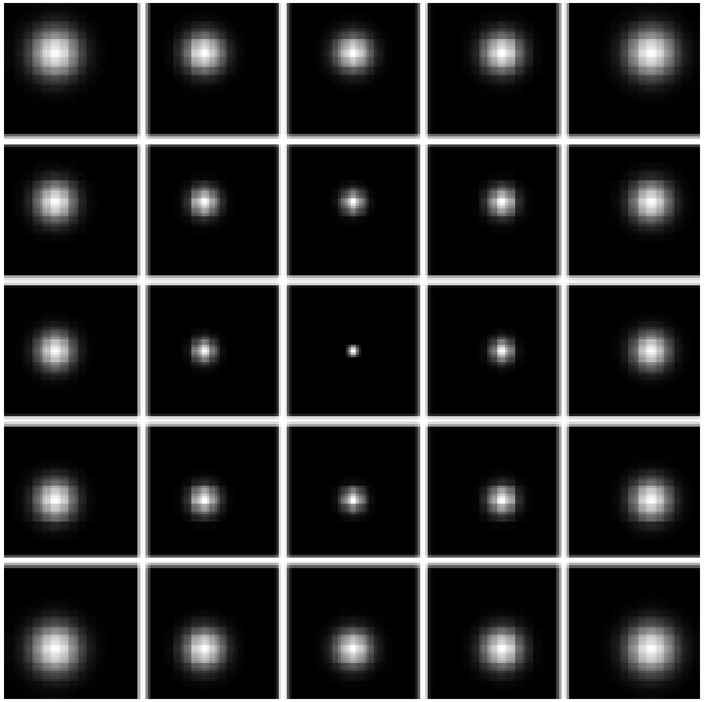
Foveation operators are linear operators that we have defined [Foi 2016] to yield foveation effects by means of blurring kernels yielding a spatially variant blur

Foveation operators enjoy directional and multiscale properties

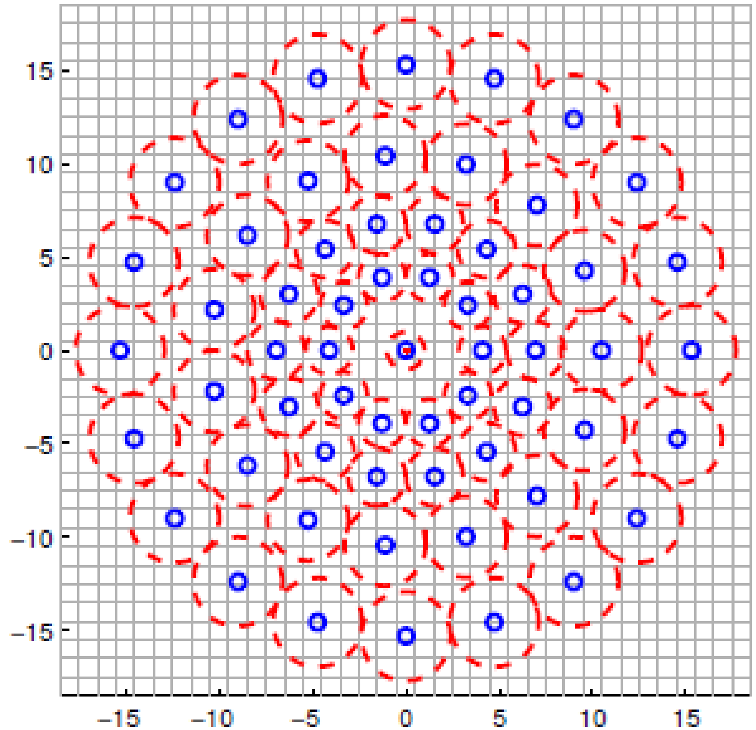


Blurring kernels of foveation operators: no blur at the center, the blur increases at the periphery. Elongated kernels of a radial foveation operators: edges directed towards the center are preserved, thus also structures and details centered at the keypoint.

[Foi 2016] Foi, Boracchi, "Foveated Nonlocal Self-Similarity," IJCV, 120 (1): 78–110, 2016.



Blurring kernel of Foveation Operator

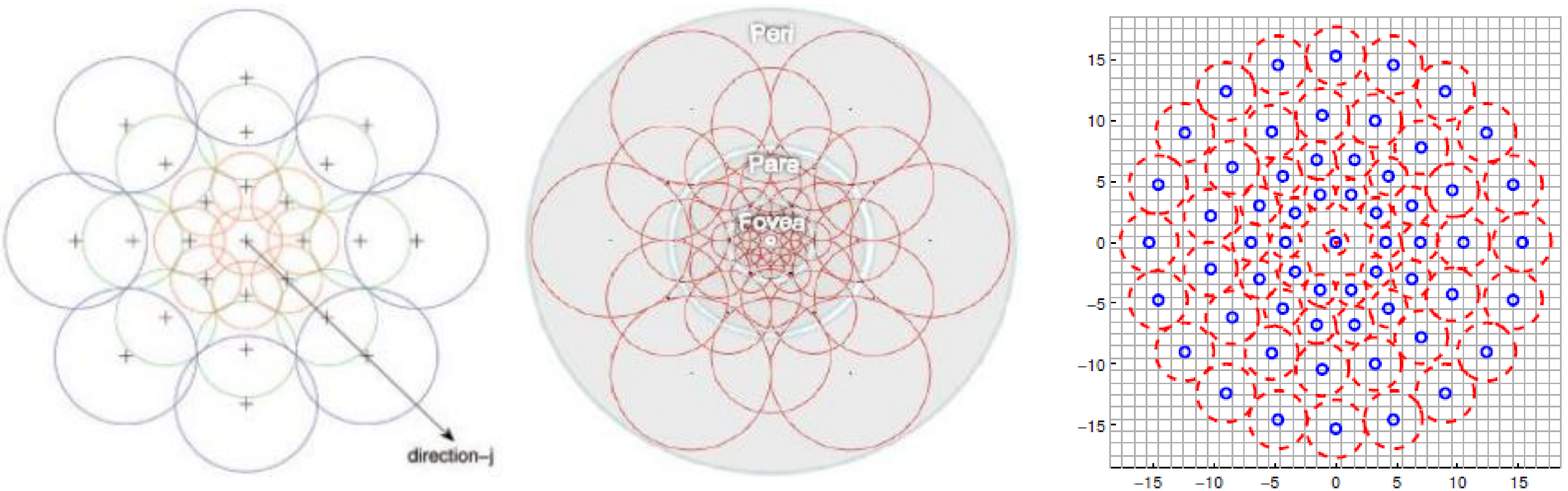


Averaging pattern of BRISK

Foveated Image Features… a project/thesis

The Idea: Superior performance achieved by radial operators and the radial orientation preference of the HVS suggest that **local image structures are better captured by filters preserving details directed towards the fixation point**

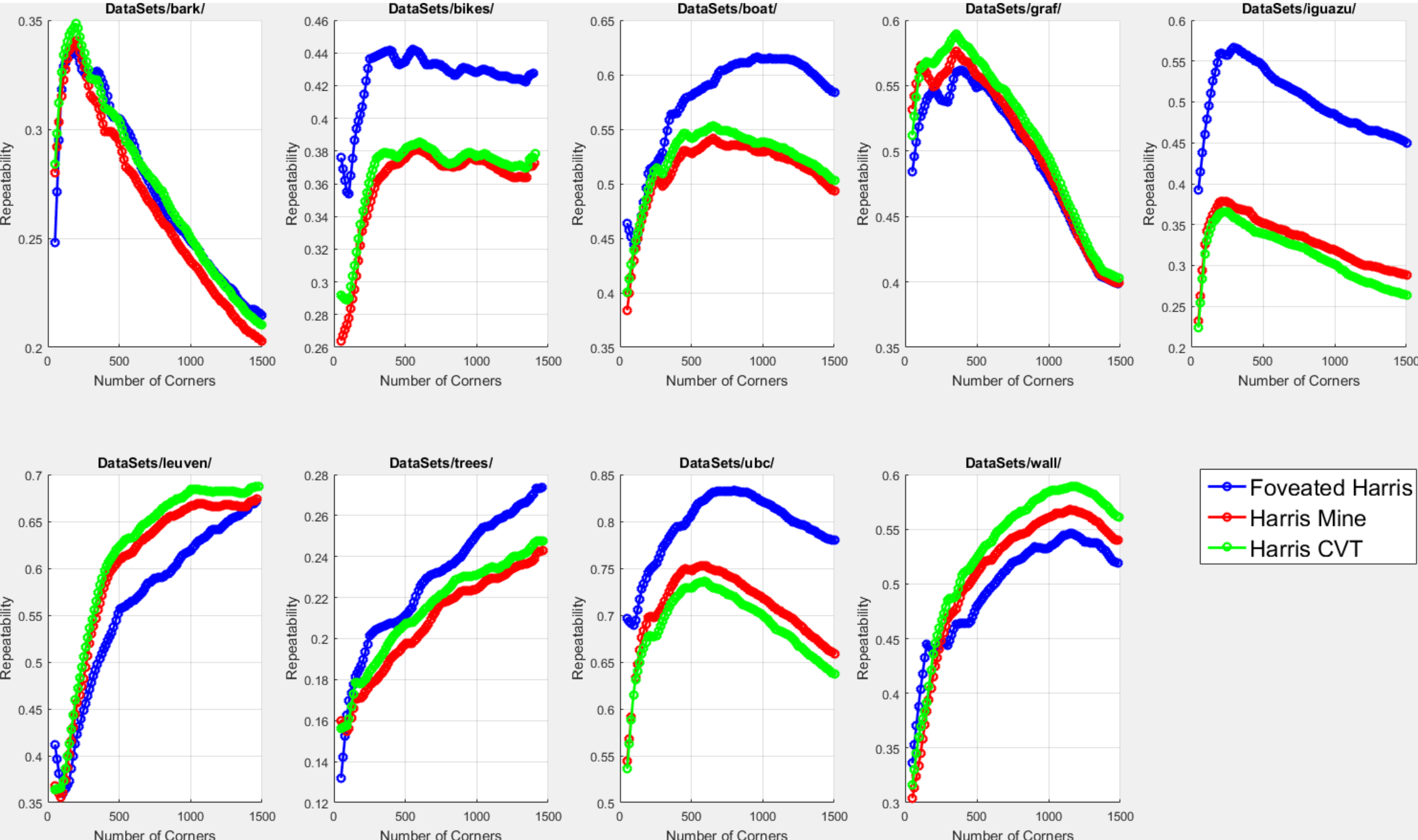
Existing features (neither foveated-like ones) have no orientation preference, as all the local structures are attenuated on circular averaging schemes



Sampling patterns of Daisy, Freak and Brisk features

Foveated Distance in Harris Corner

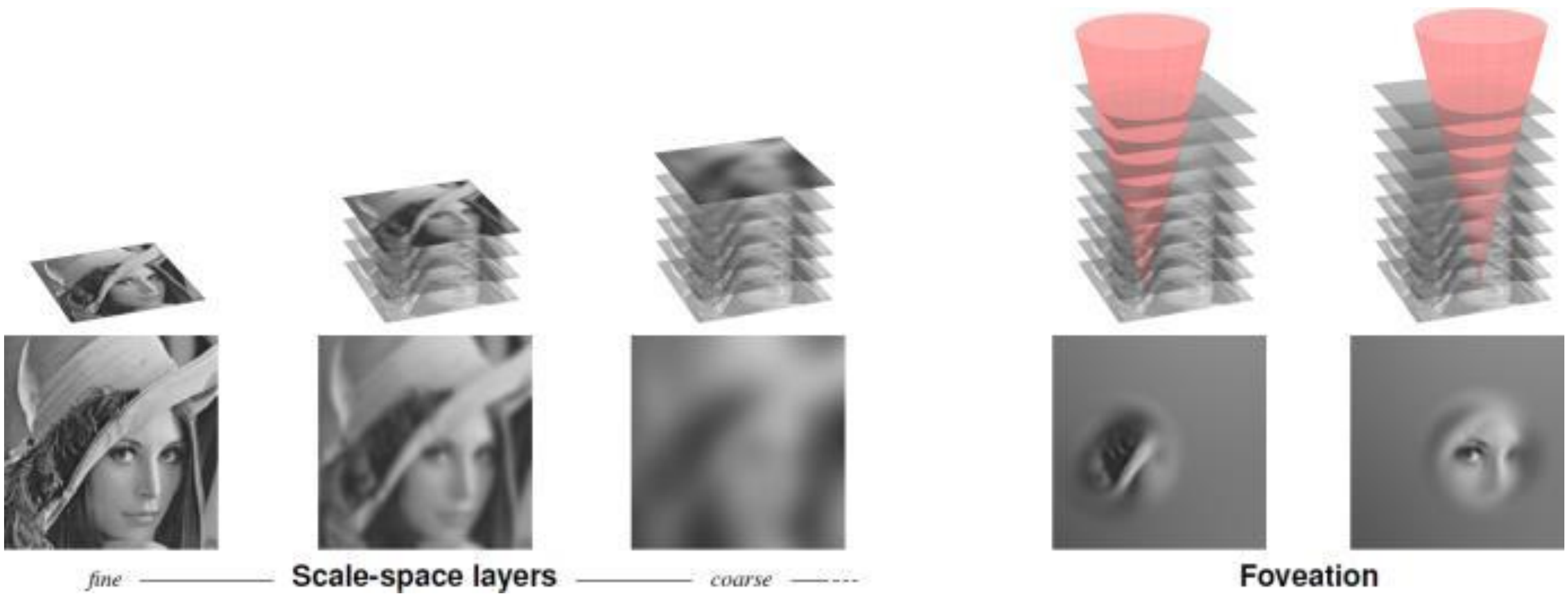
Harris corners using the Foveated distance to compare patches



Foveated Image Features

A Matlab package is available for implementing foveation operators in image filtering

http://home.deib.polimi.it/boracchi/Projects/Foveation.html



Foveated Image Features

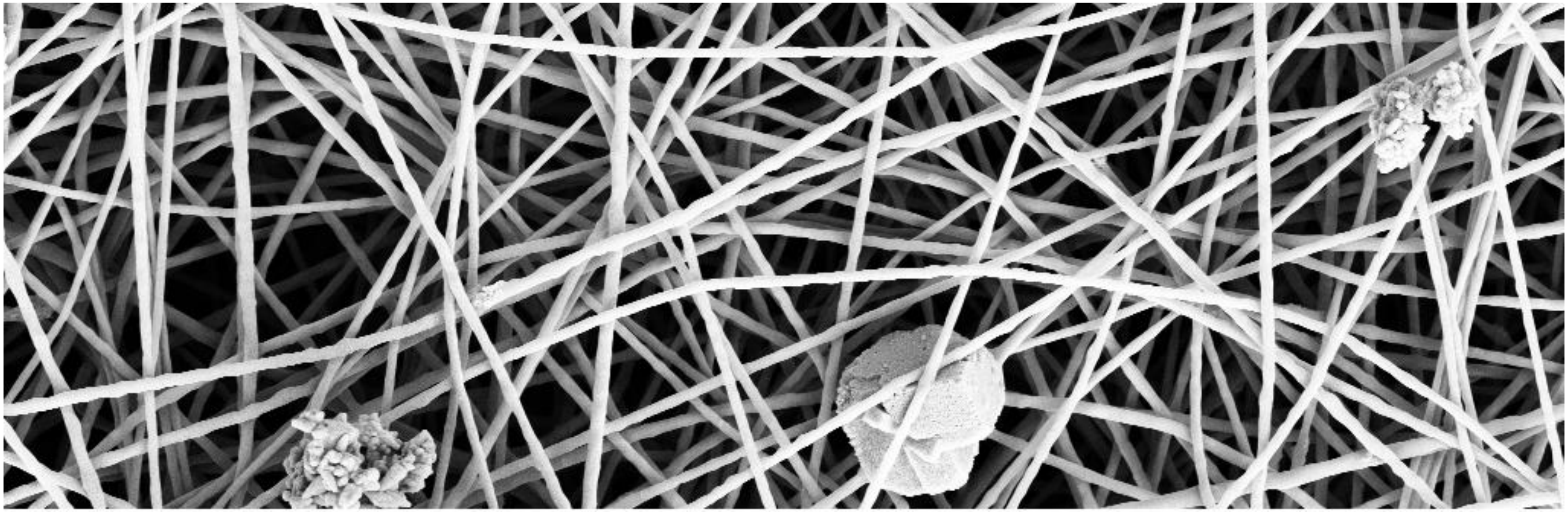
Project on Foveated Image Features: Test the effectiveness of foveated patches as feature vectors.

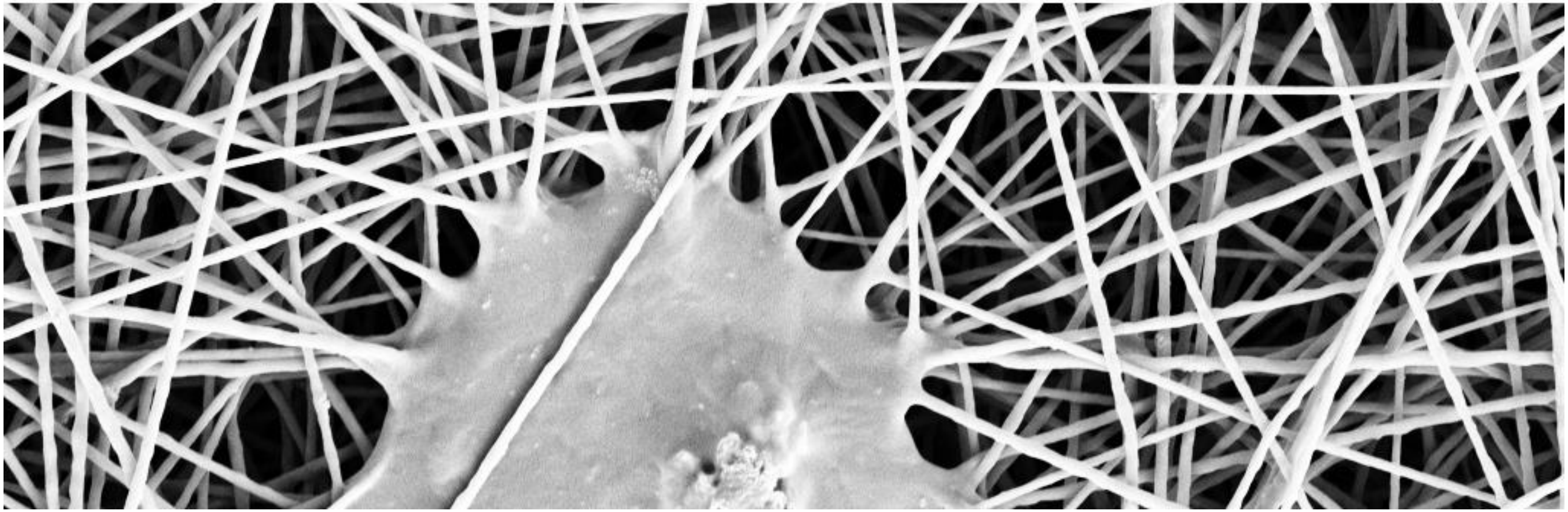
Thesis on Foveated Image Features: Design optimized BRISK (or FREAK) like features that leverage anisotropy of radial foveation operators. Optimized implementation.

Develop further foveated keypoints, adding scale / orientation information to the detected kepoints

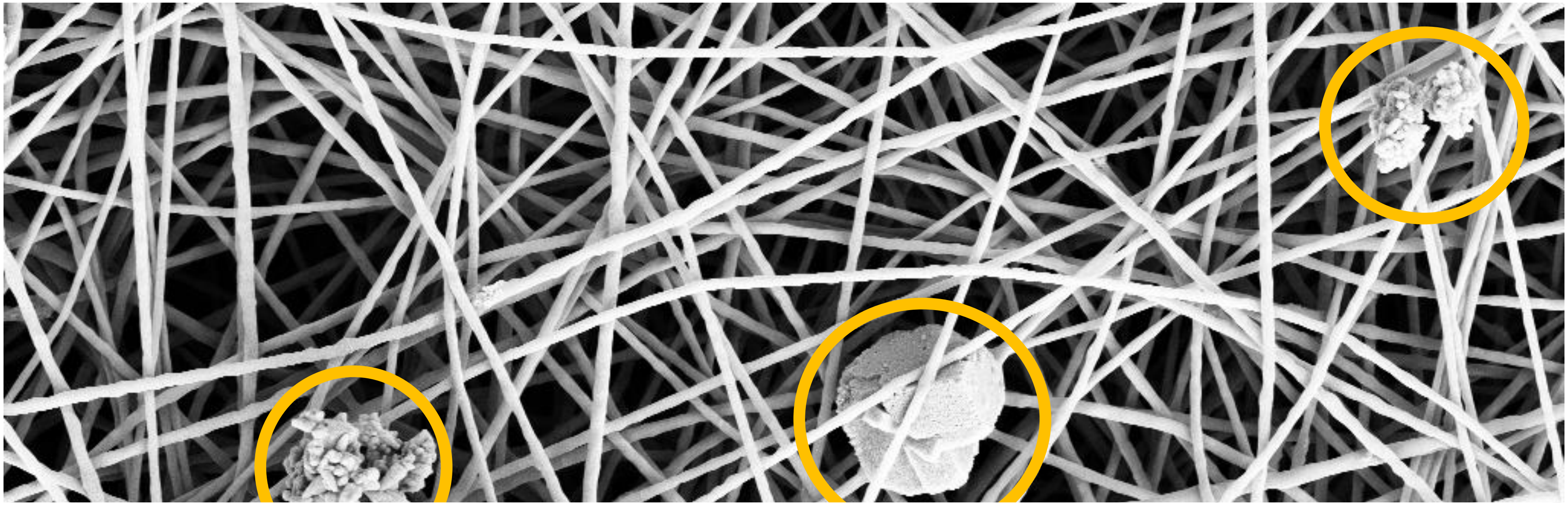
Anomaly Detection

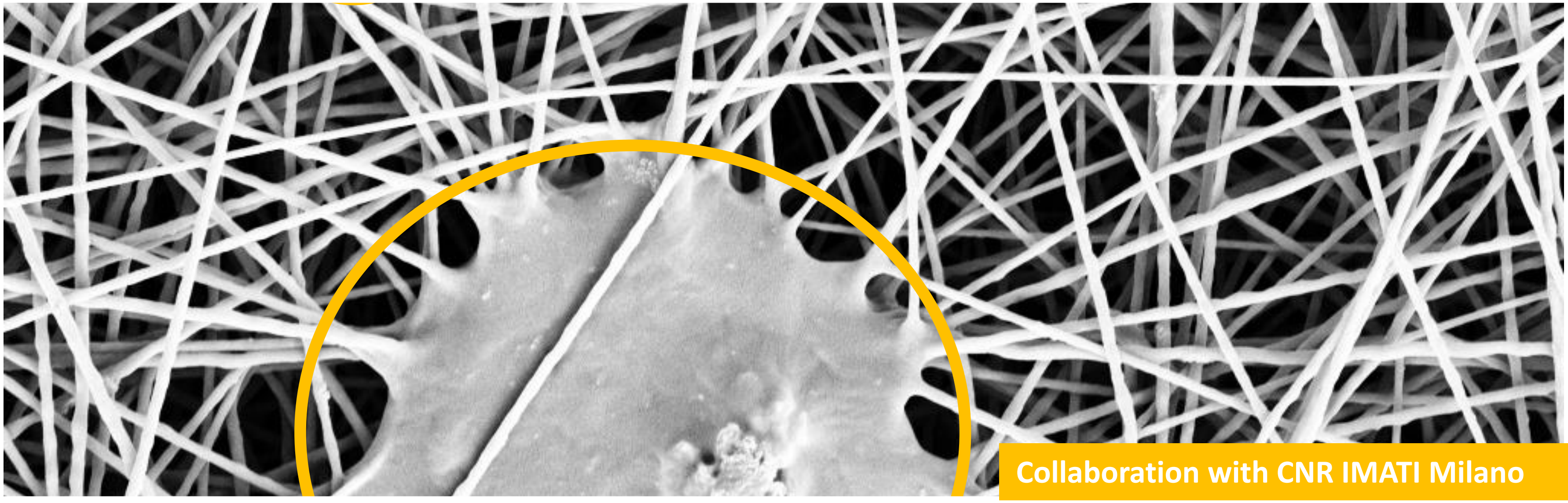
Examples of CD Problems: Anomaly Detection





Examples of CD Problems: Anomaly Detection





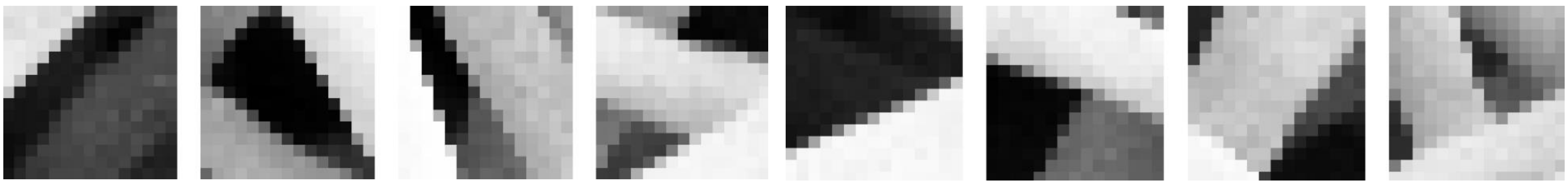
Collaboration with CNR IMATI Milano

Patch-based processing of nanofibers

Analyze each patch of an image s

$$\mathbf{s}_c = \{s(c+u), u \in \mathcal{U}\}$$

and determine whether it is normal or anomalous



Patches $\mathbf{s}_c \in \mathbb{R}^p$ are **too high-dimensional** $(p \gg 0)$ for modeling the distribution ϕ_0 generating normal paches

We need to **extract** suitable **features** to **reduce** the **dimensionality** of our anomaly-detection problem.

[TII 2017] D. Carrera, F. Manganini, G. Boracchi, E. Lanzarone "Defect Detection in SEM Images of Nanofibrous Materials", IEEE Transactions on Industrial Informatics 2017 11 pages, doi:10.1109/TII.2016.2641472

Feature Extraction

Expert-driven features: On each patch, compute

- the average,
- the variance,
- the total variation.

These are expected to **distinguish normal** and **anomalous** patches

Data-driven features: our approach consists in

1. Learning a model $\mathcal{D}$ that describes normal patches
2. Assessing the conformance of each patch $\mathbf{s}_c$ to $\mathcal{D}$

$\mathcal{D}$: Dictionary of patches

Sparse representations have shown to be a very useful method for **constructing signal models**

The underlying assumption is that

$$\mathbf{s} \approx D\boldsymbol{\alpha} \quad \text{i.e.,} \qquad \|\mathbf{s} - D\boldsymbol{\alpha}\|^2 \approx 0$$

and $\boldsymbol{\alpha} \in \mathbb{R}^n$ where:

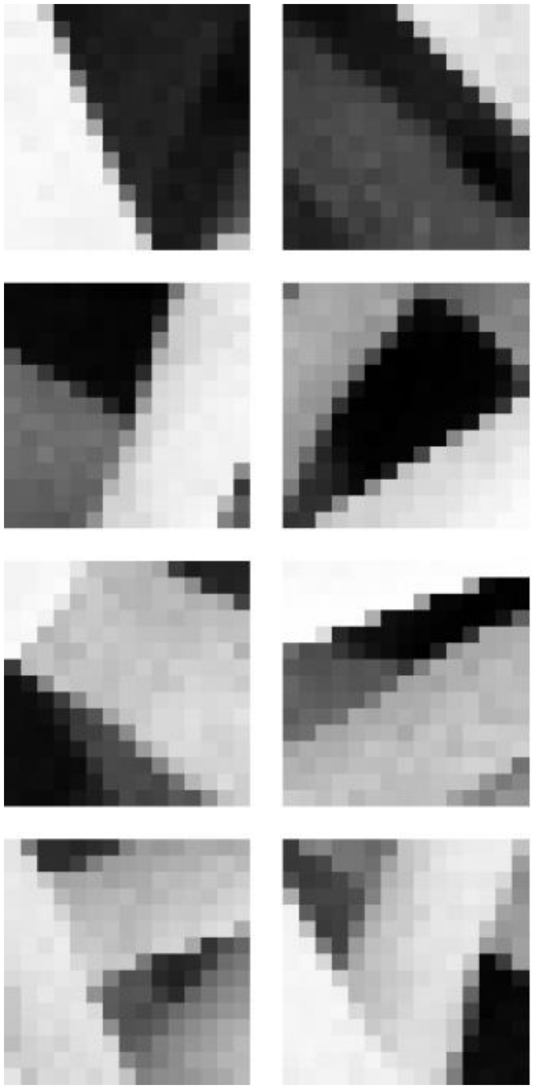
- $D \in \mathbb{R}^{p \times n}$ is the **dictionary**, columns are called **atoms**
- the coefficient vector $\boldsymbol{\alpha}$ is sparse
  - $\|\boldsymbol{\alpha}\|_0 = L \ll n$ or
  - $\|\boldsymbol{\alpha}\|_1$ is small

The **dictionary** is learned a training set of **normal patches**.

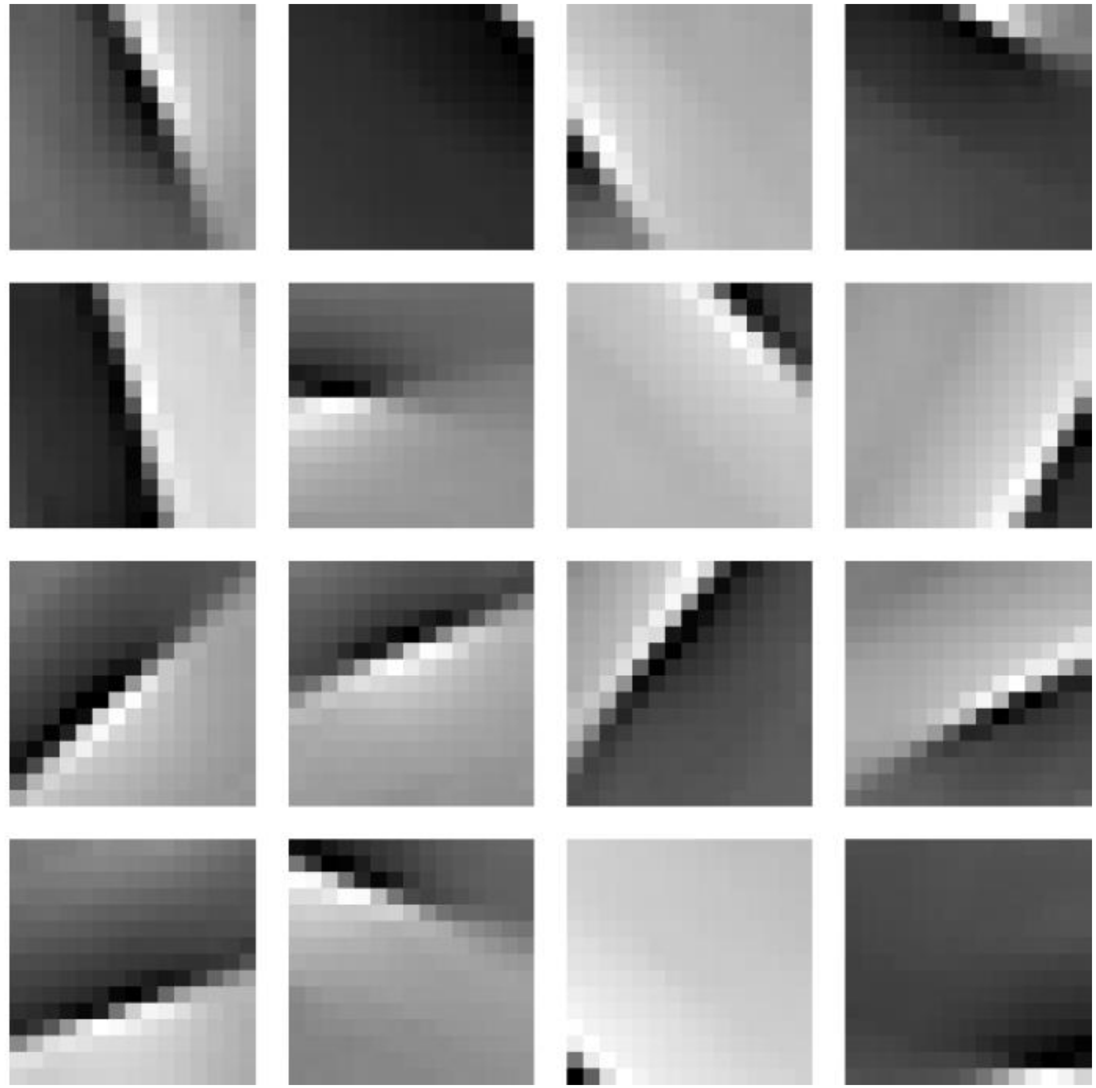
We learn a **union of low-dimensional sub-spaces** where **normal patches** live

The dictionary of normal patches

Example of training patches



Few learned atoms (ADMM Dictionary Learning)



Data-Driven Features

To assess the confrmance of $\boldsymbol{s}_c$ with $\mathcal{D}$ we perform the

Sparse coding:

$$\boldsymbol{\alpha} = \underset{\widetilde{\boldsymbol{\alpha}} \in \mathbb{R}^n}{\operatorname{argmin}} \|D\widetilde{\boldsymbol{\alpha}} - \mathbf{s}\|_2^2 + \lambda \|\widetilde{\boldsymbol{\alpha}}\|_1, \qquad \lambda > 0$$

which we solve using the BPDN problem (using ADMM).

We then measure

$$\|D\boldsymbol{\alpha} - \mathbf{s}\|_2^2$$

and

$$\|\boldsymbol{\alpha}\|_1$$

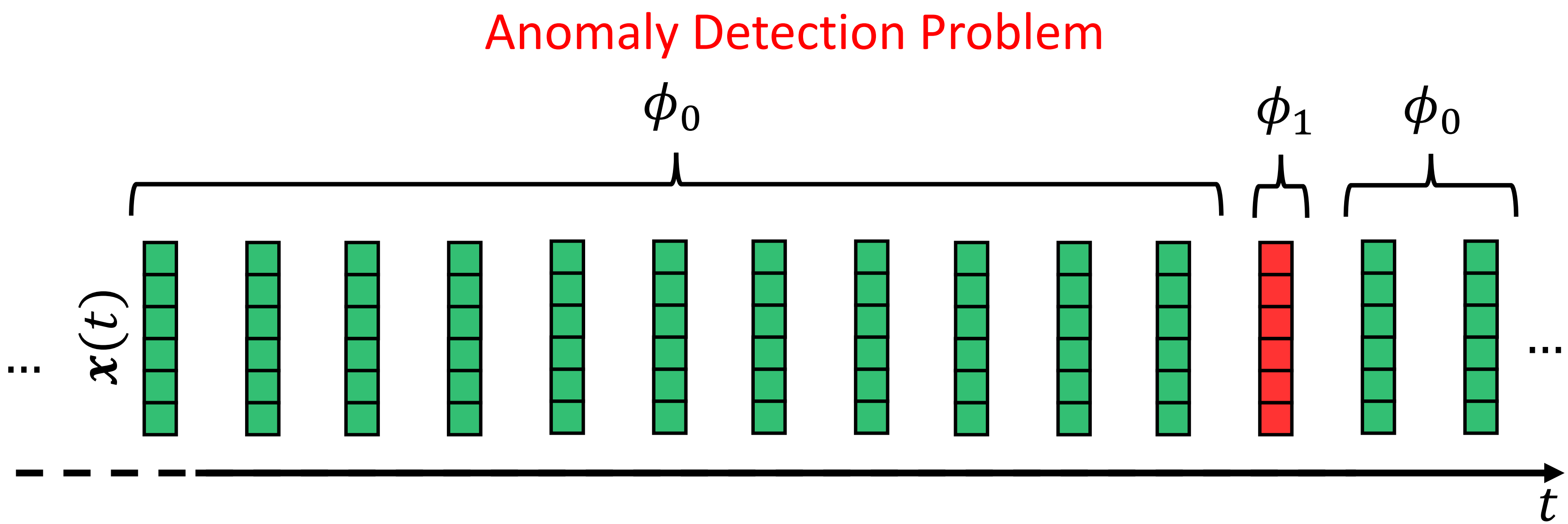
Data-driven features are $\mathbf{x} = \begin{bmatrix} \|D\boldsymbol{\alpha} - \mathbf{s}\|_2^2 \\ \|\boldsymbol{\alpha}\|_1 \end{bmatrix}$

Detecting Anomalies

Normal patches are expected to yield features **x** that are **i.i.d.** and that **follow a** (unknown) **distribution** ϕ_0, **anomalous patches do not,** as they follow $\phi_1 \neq \phi_0$

We are back to the original problem

"Determining whether a set of data $\{\boldsymbol{x}_c,\ c = 1, \dots\}$ ***is generated from*** $\boldsymbol{\phi_o}$ *and detect possible* ***outliers"***

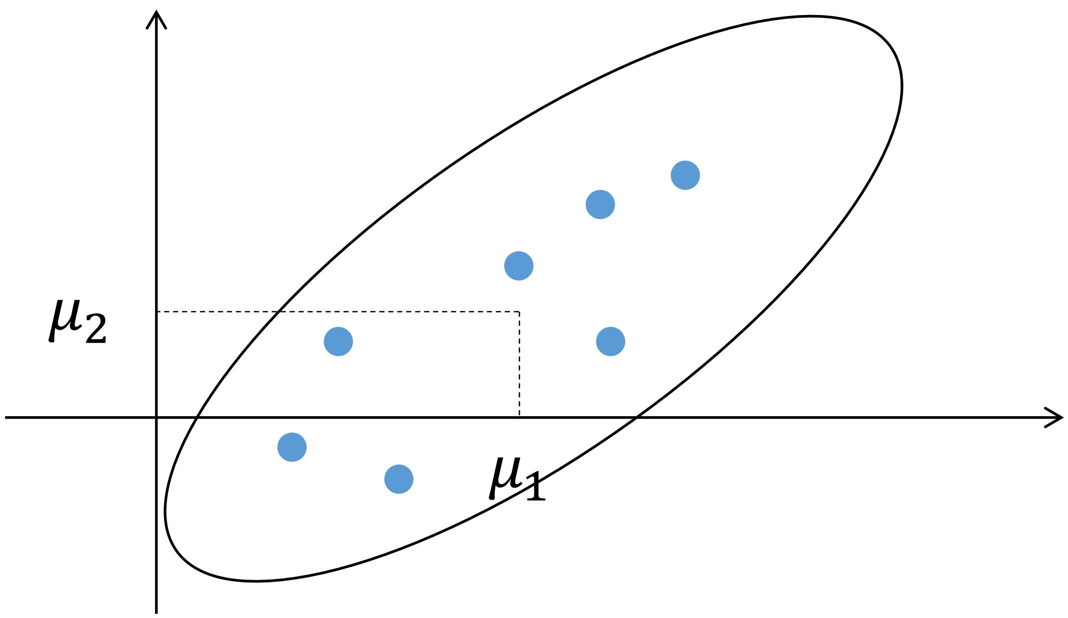


Anomaly Detection from 2D Change Indicators

For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{\xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi-\mu)'\Sigma^{-1}(\xi-\mu)} \leq \gamma\right\}$$

where μ and Σ are the sample mean and sample covariance of the change indicators from T.



Anomaly Detection from 2D Change Indicators

For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{\xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi-\mu)'\Sigma^{-1}(\xi-\mu)} \leq \gamma\right\}$$

where μ and Σ are the sample mean and sample covariance of the change indicators from T.

The Chebyshev's inequality says that a normal patch falls outside R_γ with probability $\leq 2/\gamma^2$

Anomalies are detected as

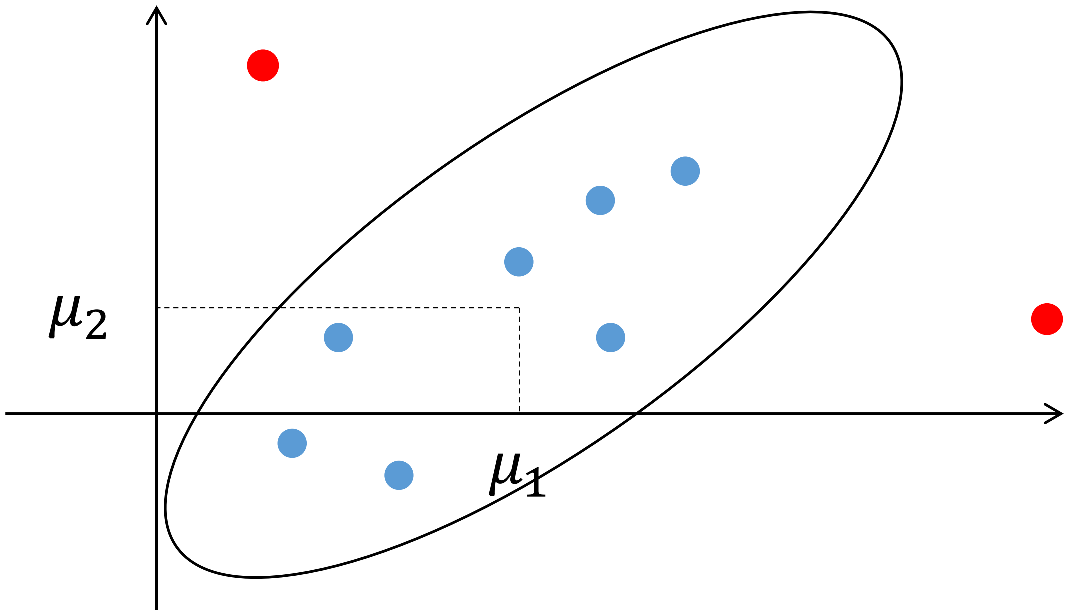
$$\mathbf{s} \text{ s.t. } \sqrt{(\boldsymbol{g}(\mathbf{s})-\mu)'\Sigma^{-1}(\boldsymbol{g}(\mathbf{s})-\mu)} > \gamma$$

Anomaly Detection from 2D Change Indicators

For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{\xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi-\mu)'\Sigma^{-1}(\xi-\mu)} \leq \gamma\right\}$$

where μ and Σ are the sample mean and sample covariance of the change indicators from T.

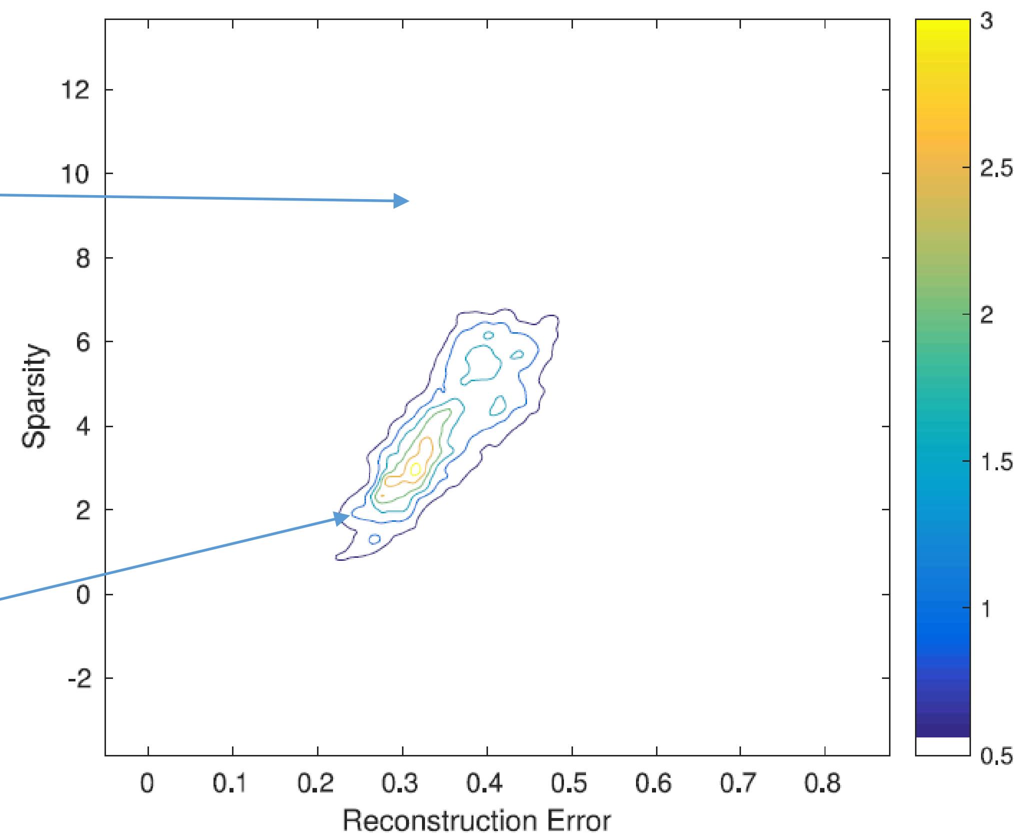


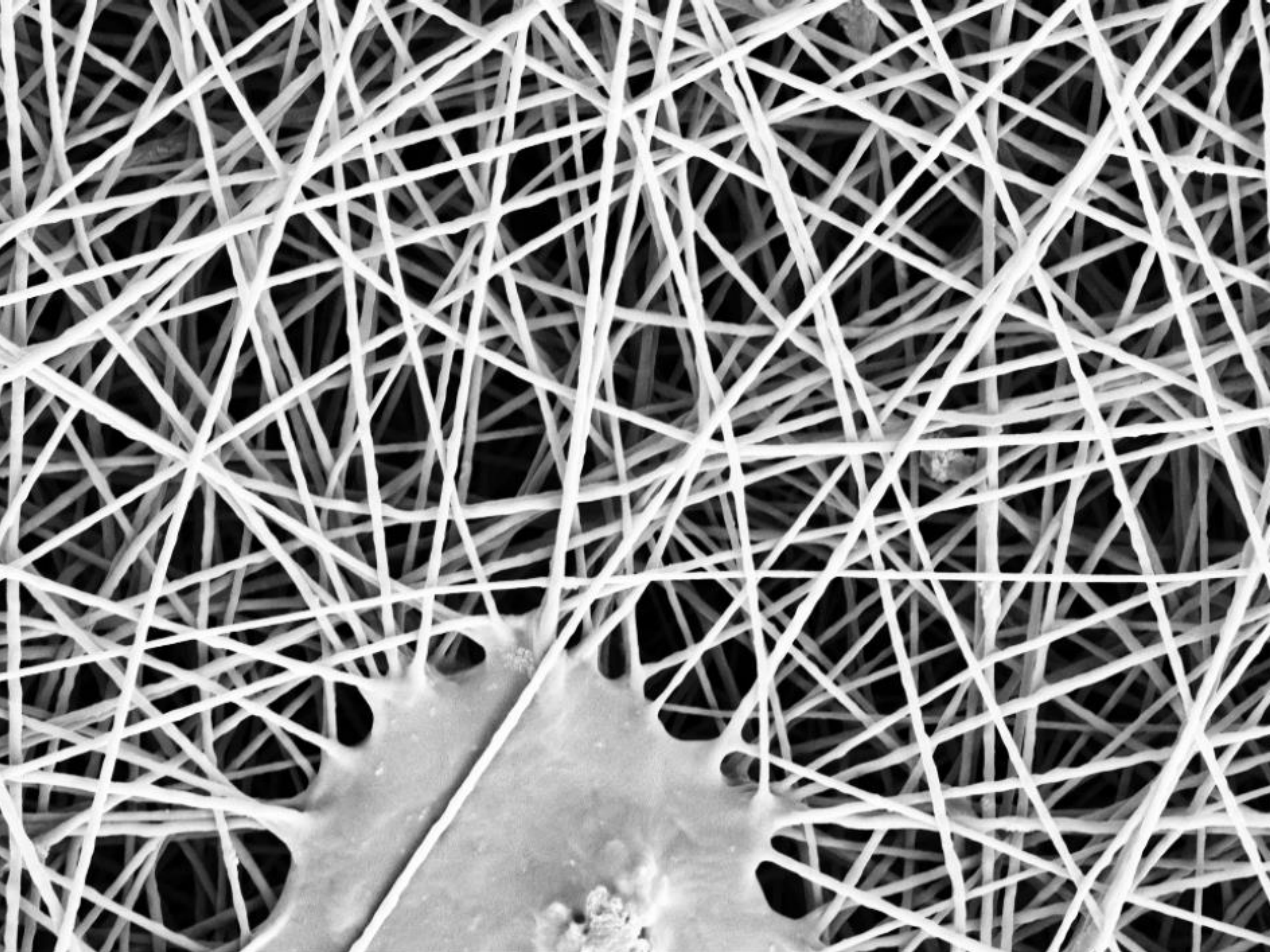
Fit a non-parametric density model to the bivariate indicator

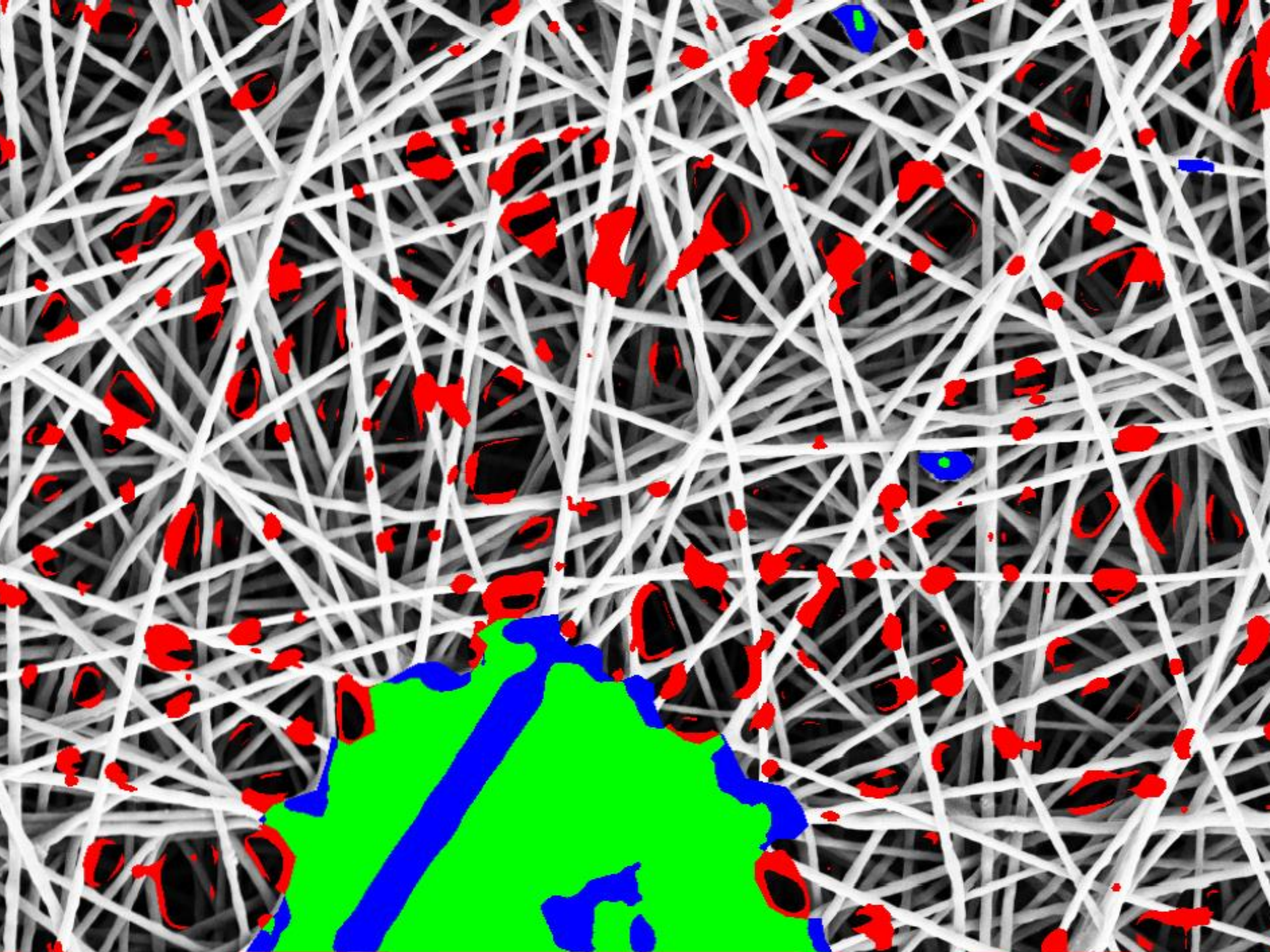
$$R_\gamma = \{\xi \in \mathbb{R}^2, \mathrm{s.t.}\ KDE(\xi) \geq \gamma\}$$

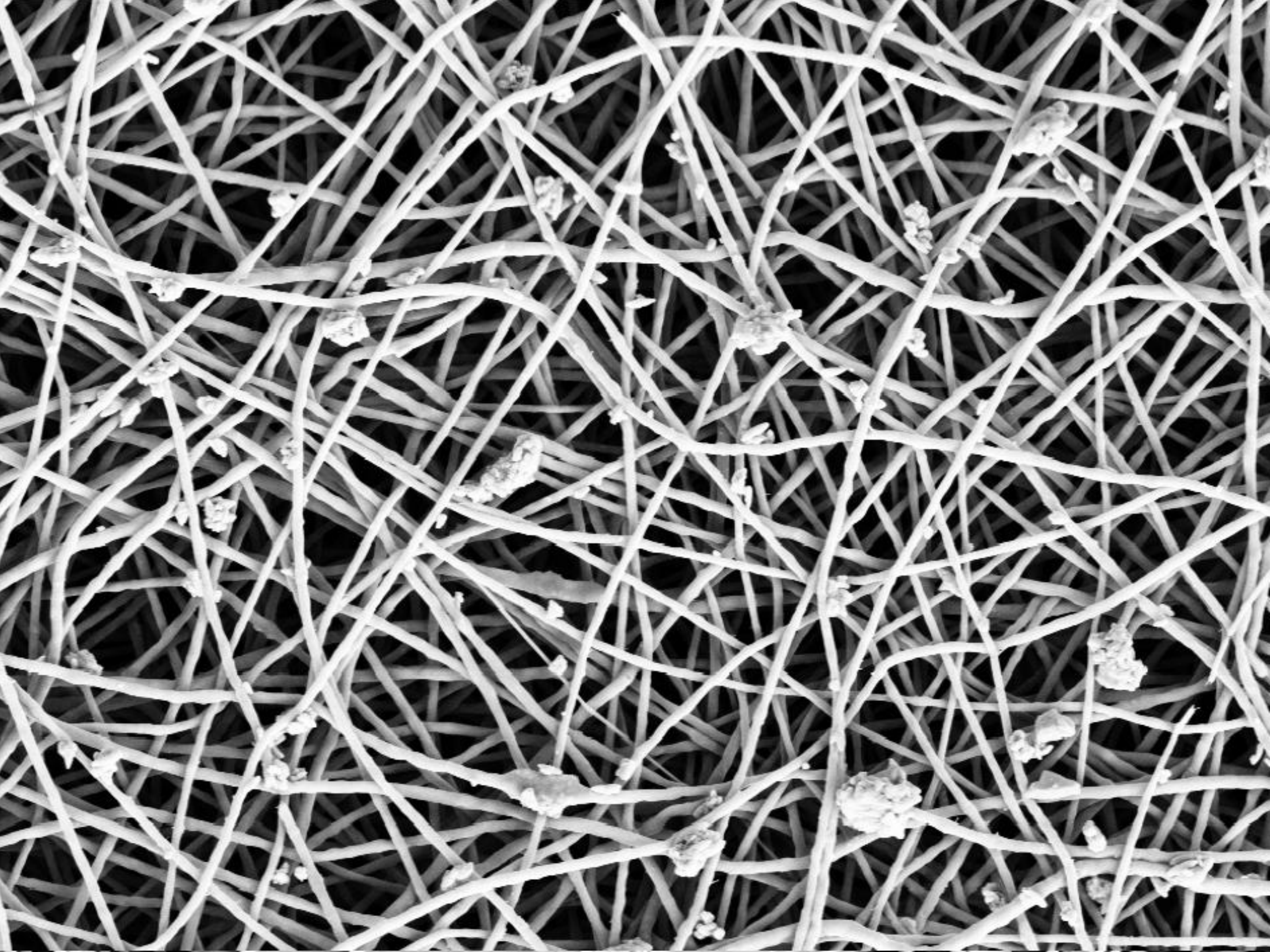
Anomalous

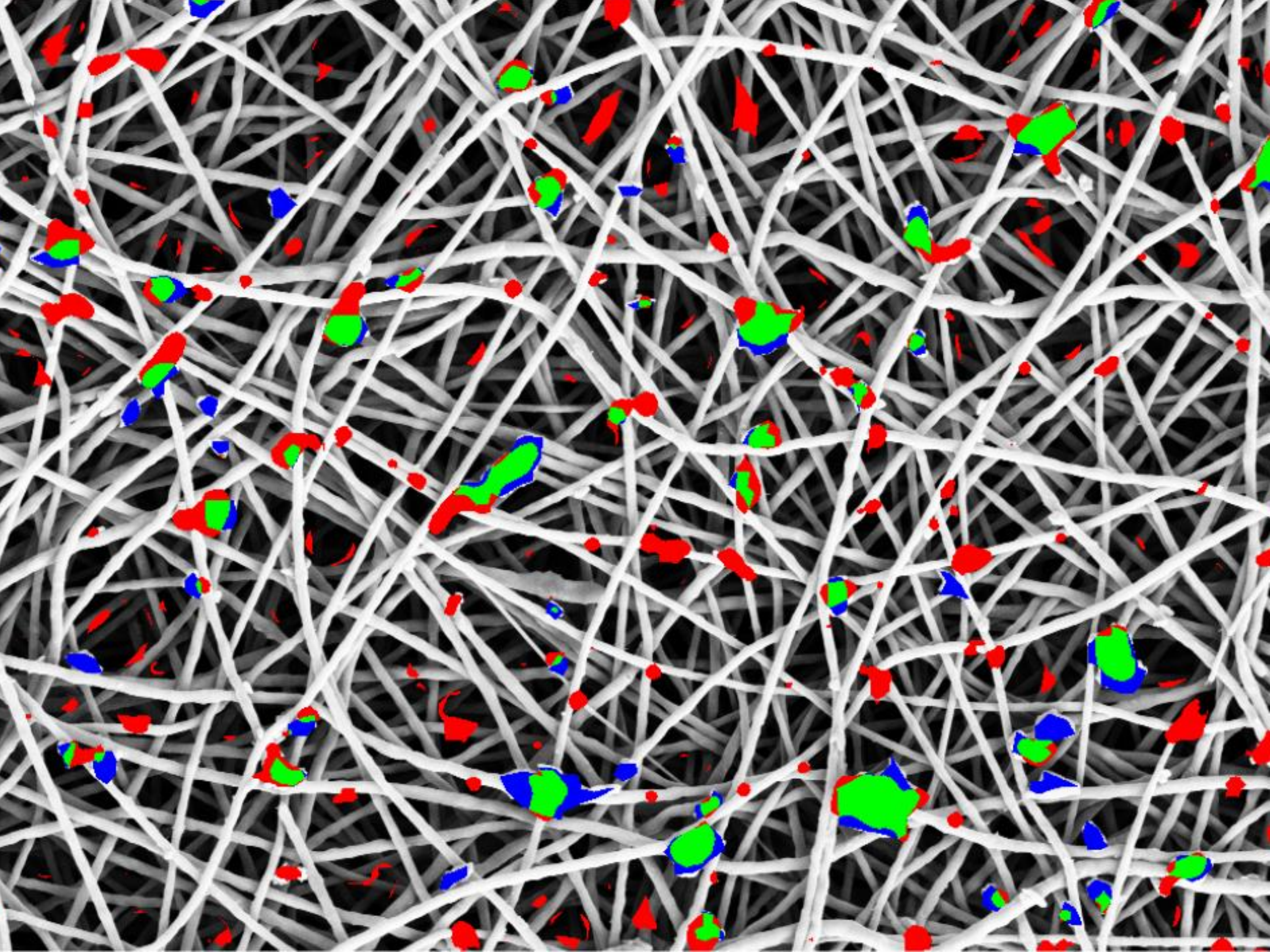
Normal











Anomaly-detection Performance

Anomaly detection performance:

- True positive rate: $TPR = \frac{\#\{\text{anomalies detected}\}}{\#\{\text{anomalies}\}}$
- False positive rate: $FPR = \frac{\#\{\text{normal samples detected}\}}{\#\{\text{normal samples}\}}$

You have probably also heard of

- False negative rate (or miss-rate): $FNR = 1 - TPR$
- True negative rate (or specificity): $TNR = 1 - FPR$
- Precision on anomalies: $\frac{\#\{\text{anomalies detected}\}}{\#\{\text{detections}\}}$
- Recall on anomalies (or sensitivity, hit-rate): TPR

Anomaly-detection Performance

There is always a **trade-off between $\boldsymbol{TPR}$ and $\boldsymbol{FPR}$** (and similarly for derived quantities), which is ruled by algorithm parameters

Thus, to correctly assess performance it is necessary to consider at least **two indicators** (e.g., TPR, FPR)

Unique indicators:

$$\text{Accuracy} = \frac{\#\{\text{anomalies detected}\} + \#\{\text{normal samples not detected}\}}{\#\{\text{samples}\}}$$

$$\text{F1 score} = \frac{2\#\{\text{anomalies detected}\}}{\#\{\text{detections}\} + \#\{\text{anomalies}\}}$$

These equal 1 in case of "ideal detector" which detects all the anomalies and has no false positives

Anomaly-detection Performance

Comparing different methods might be tricky since we have to make sure that both have been configured in their best conditions

Typically, one tests a large set of parameters and plot the **ROC** (receiver operating characteristic) **curve**

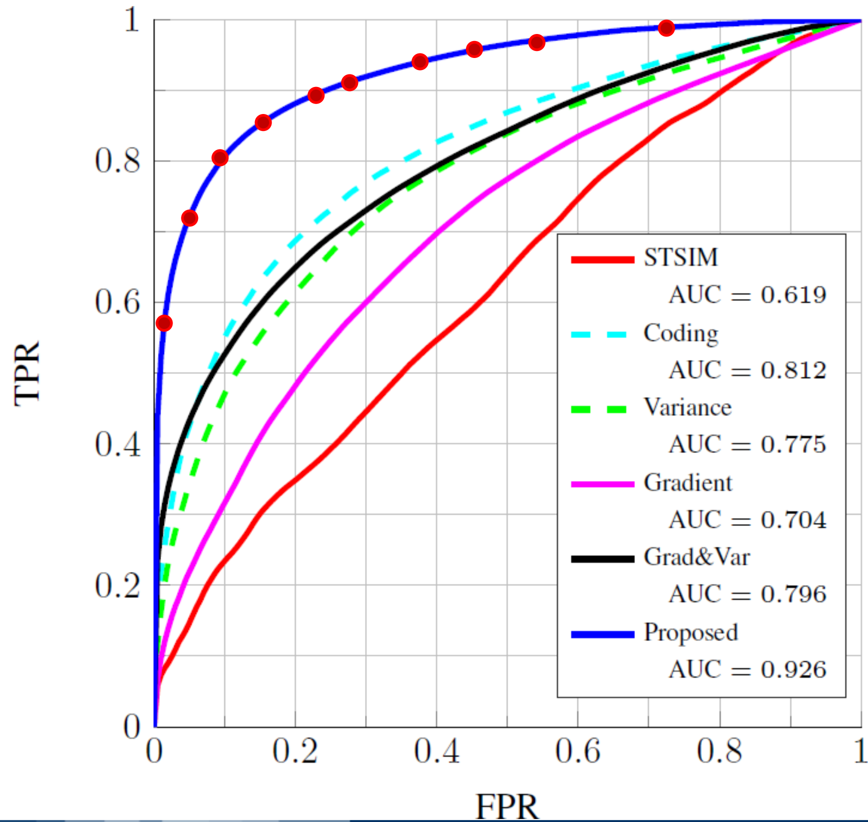
The ideal detector would achieve:

- $FPR = 0\%$,
- $TPR = 100\%$

Thus, the closer to (0,1) the better

The largest the **Area Under the Curve** (AUC), the better

The optimal parameter is the one yielding the point closest to (0,1)



Anomaly-detection Performance

Comparing different methods might be tricky since we have to make sure that both have been configured in their best conditions

Typically, one tests a large set of parameters and plot the **ROC** (receiver operating characteristic) **curve**

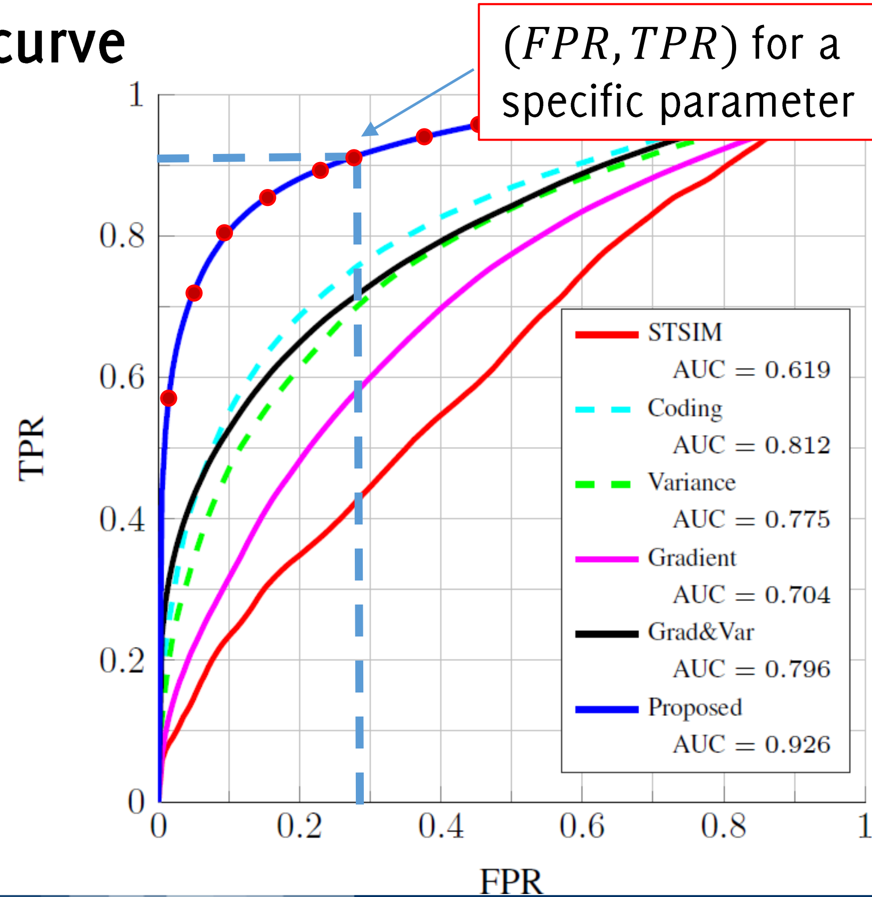
The ideal detector would achieve:

- $FPR = 0\%$,
- $TPR = 100\%$

Thus, the closer to (0,1) the better

The largest the **Area Under the Curve** (AUC), the better

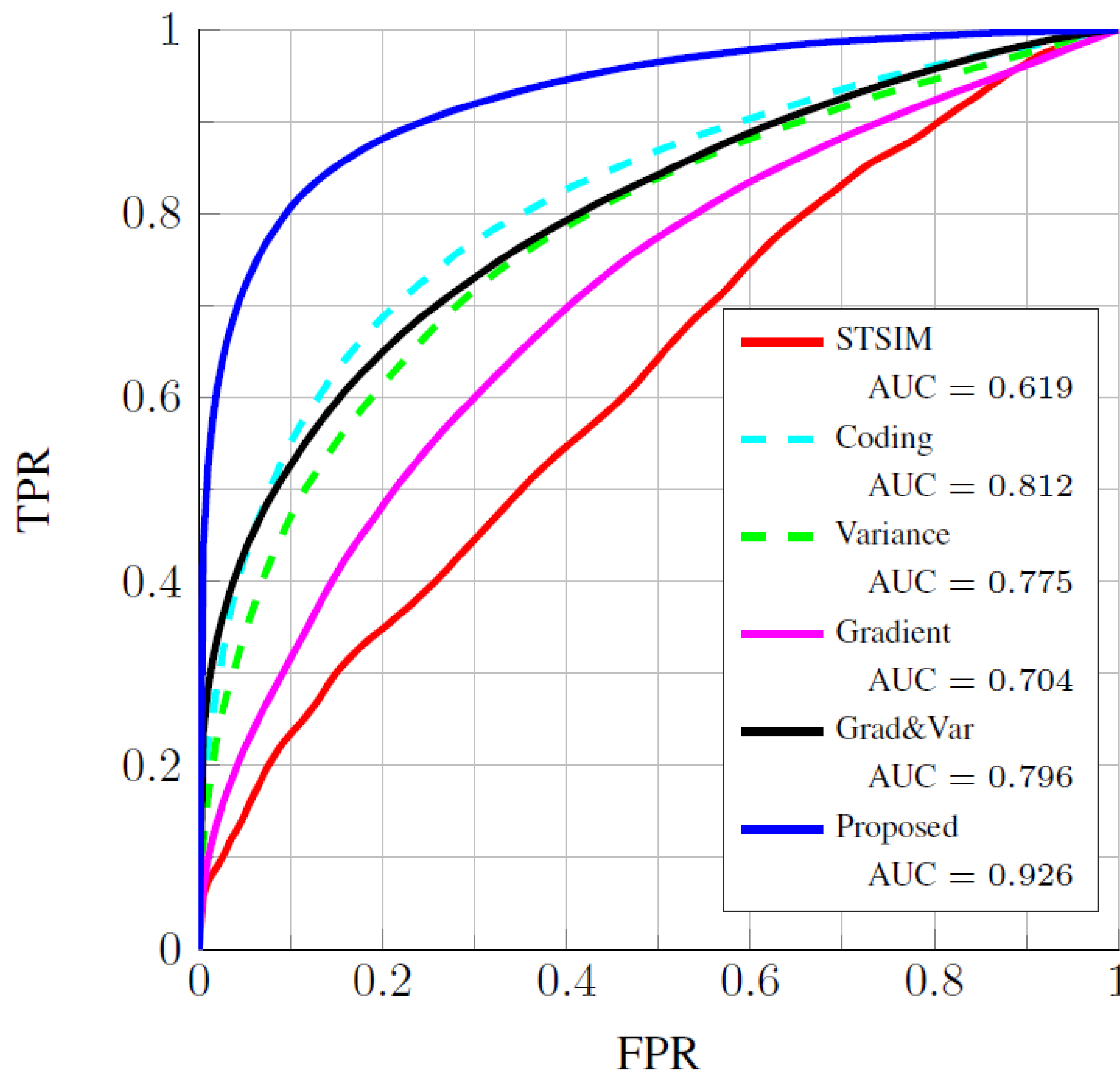
The optimal parameter is the one yielding the point closest to (0,1)



The ROC curves on the SEM image dataset

Tests on 40 images with anomalies manually annotated by an expert

The proposed anomaly detection algorithm outperforms expert-driven features and other methods based on sparse representations



Try it yourself

You can download an annotated dataset of SEM images from

https://home.deib.polimi.it/carrerad/codes/spad_v0.9.zip