Novelty Detection in Images by Sparse Representations

Giacomo Boracchi, Diego Carrera

Dipartimento di Elettronica
Informazione e Bioingegneria,
Politecnico di Milano, Italy

Brendt Wohlberg

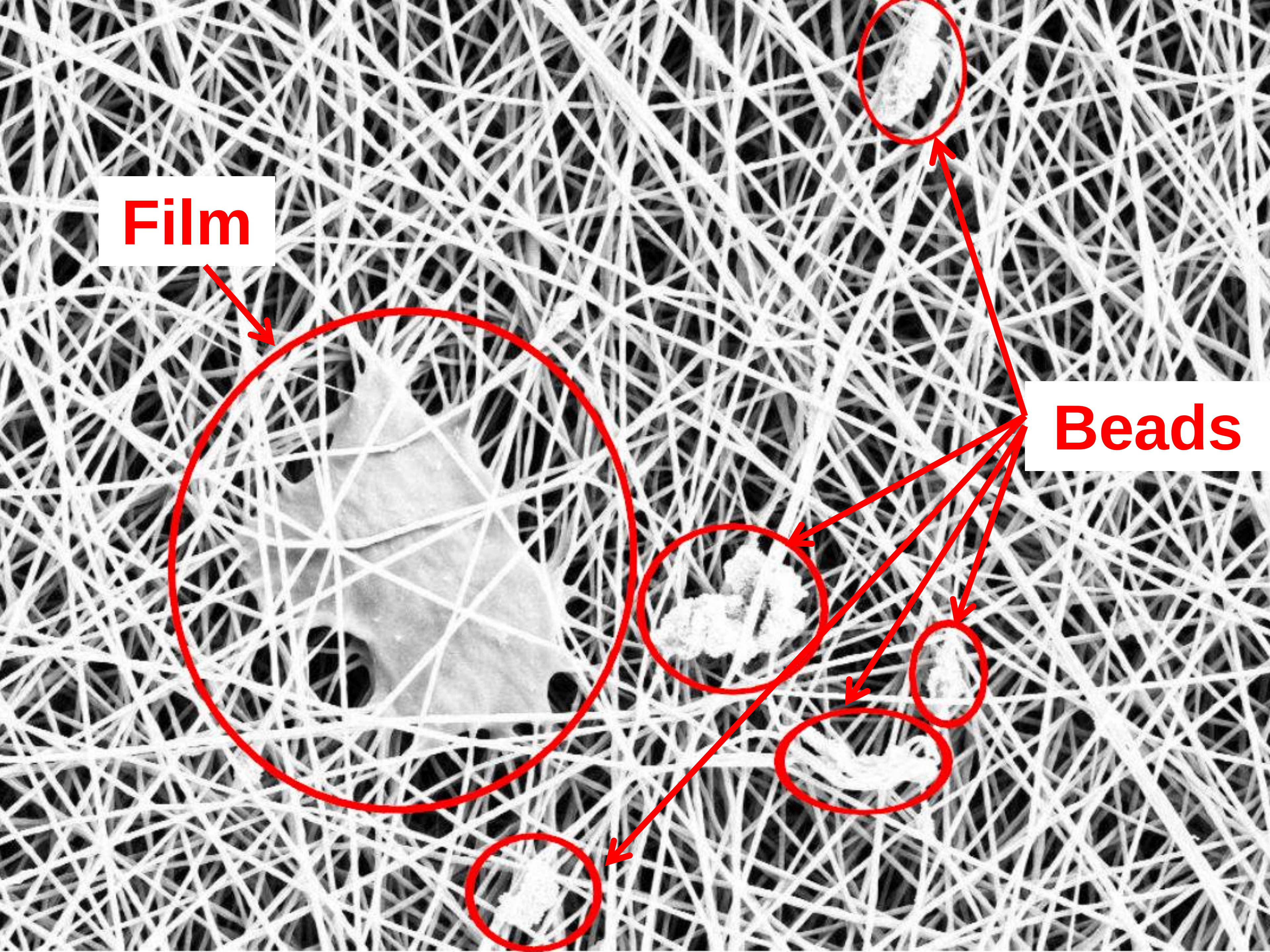
Theoretical Division,
Los Alamos National Laboratory,
NM, USA

Dec. 10, 2014

Intelligent System for Novelty Detection

- We consider **monitoring systems** acquiring and processing **images**, such as those employed in biomedical or industrial control applications.
- We assume that **images** acquired under **normal conditions** are characterized by **specific structures**
- **Regions** that **do not conform** to these structures are considered **anomalies**
- **An intelligent system has to automatically detect anomalous regions**

- As «running example» we consider scanning electron microscope (SEM) images for monitoring the production of nanofibers


Film
Beads

Outline

- Problem Formulation
- Sparse Representations for Novelty Detection
- Anomaly indicators
- Experiments
  - Texture Images
  - SEM images for nanofiber production

PROBLEM FORMULATION

Patch-Generating Process

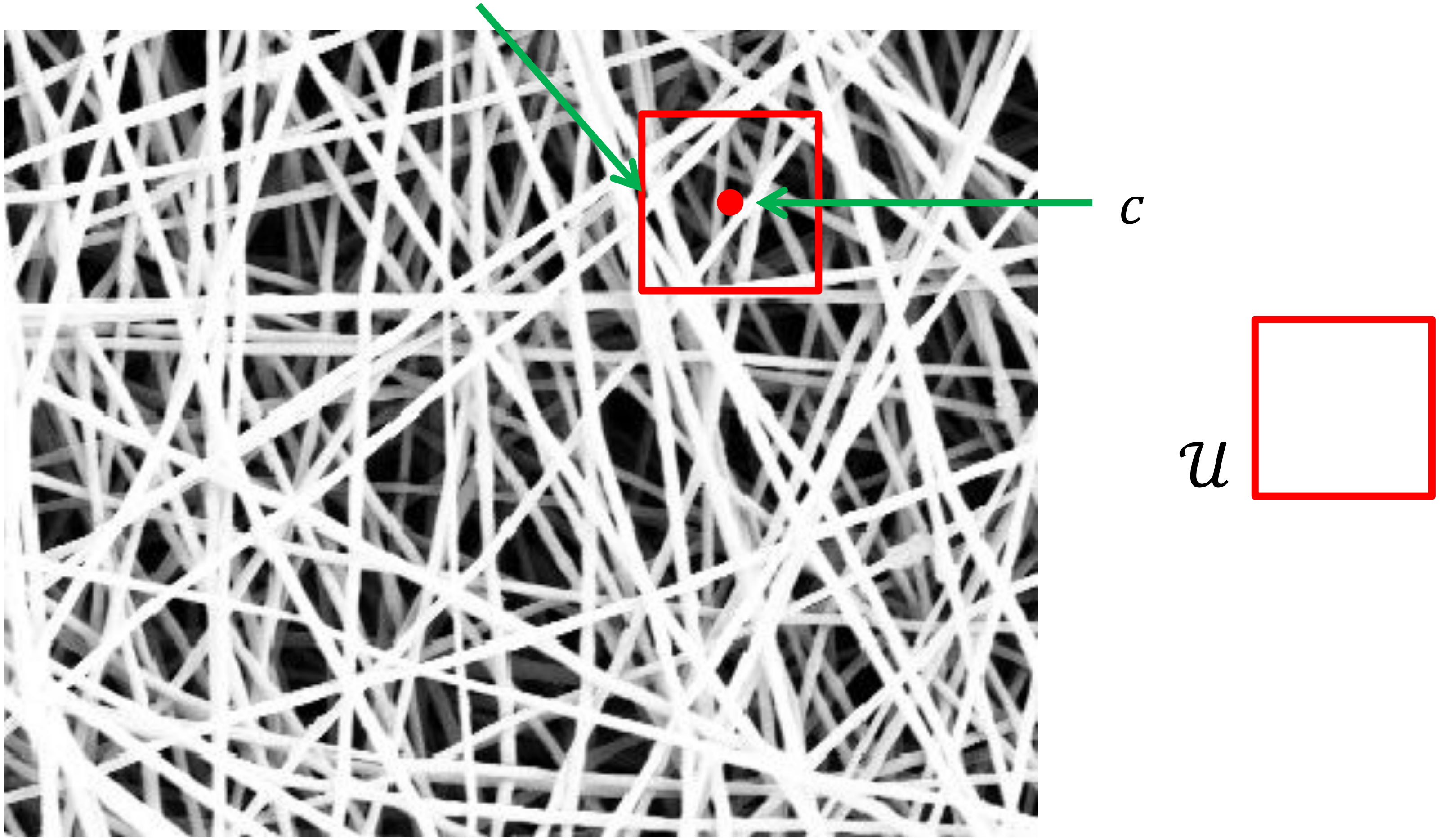
- Patches are small image regions of a predefined shape $\mathcal{U}$,

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

Patch-Generating Process

- Patches are small image regions of a predefined shape $\mathcal{U}$,

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



Patch-Generating Process

- Patches are small image regions of a predefined shape $\mathcal{U}$,

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

- We assume that in **nominal** conditions, patches $\mathbf{s}_c \in \mathbb{R}^m$ are i.i.d. realizations from a stochastic process $\mathcal{P}_N$

$$\mathbf{s}_c \sim \mathcal{P}_N$$

- A training set of l normal patches $T \in \mathbb{R}^{m \times l}$ is given to learn a model $\widehat{D}$ approximating normal patches

The Novelty-Detection Problem

- We assume that anomalous patches are generated by $\mathcal{P}_A$

$$\mathbf{s}_\mathrm{c} \sim \mathcal{P}_A$$

- The process generating anomalies $\mathcal{P}_A \neq \mathcal{P}_N$ is unknown
- Anomalies have to be detected as patches that **do not conform** the model learned to describe normal patches
  - We define **anomaly indicators** $f(\mathbf{s}_i)$ that measure the degree to which the learned model fits each patch $\mathbf{s}_i$
  - We detect anomalies as outliers in the anomaly indicators
- Peculiarity of the proposed approach is **to leverage models** $\widehat{D}$ **yielding sparse representation** of image patches

SPARSE REPRESENTATIONS

for novelty detection

Sparse Representations

- **Sparse representations** have shown to be a very useful method for **constructing signal models**
- The underlying assumption is that

$$\mathbf{s} \approx D\mathbf{x}$$

and $\|\mathbf{x}\|_0 = L \ll n$, where:

  - $D \in \mathbb{R}^{m \times n}$ is the **dictionary**, columns are called **atoms**
  - the coefficient vector $\mathbf{x}$ is assumed to be sparse

- Sparse signals live in a union of **low-dimensional subspaces** of $\mathbb{R}^m$, each having maximum dimension L, defined by dictionary atoms.

Learning a Dictionary for Modeling Stationarity

- Learning $\widehat{D}$ corresponds to learning the union of subpaces where patches in T – the normal ones- live.
- Solution is a joint optimization over the dictionary and coefficients of a sparse representation of T

$$\widehat{D} = \underset{D \in \mathbb{R}^{m \times n}, X \in \mathbb{R}^{n \times l}}{\operatorname{argmin}} \|DX - T\|_F$$

such that $\|\mathbf{x}_k\|_0 \leq L, \forall k$

- We consider here the KSVD algorithm [Aharon 06]

[Aharon 06] M. Aharon, M. Elad, and A. M. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” Transactions on Signal Processing vol. 54, no. 11, November 2006, pp. 4311–4322.

Sparse Coding

- Given the dictionary $\widehat{D}$ we use it for computing the sparse representation of a patch to be tested
- There are efficient tools for computing $\mathbf{x}$, the sparse approximation of a patch $\boldsymbol{s}$ w.r.t. a given dictionary $\widehat{D}$

$$\widehat{D}\mathbf{x} \approx \mathbf{s}$$

in a sense that $\left\|\widehat{D}\mathbf{x} - \mathbf{s}\right\|_{\mathbf{2}}$ is small

- This operation is referred to as the **sparse coding**

Sparse Coding - ℓ^0 norm problem

- Sparse coding solving the constrained problem

$$\text{P0:} \quad \hat{\mathbf{x}}_{\mathbf{0}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\| \widehat{D}\mathbf{x} - \mathbf{s} \right\|_{\mathbf{2}} \text{ s.t. } \|\mathbf{x}\|_0 \leq L$$

- The sparsity of the solution is constrained to be at most L
- Exact solutions are computationally intractable.
- Typically solved by means of Greedy Algoritms, such as the Orthogonal Matching Pursuit (OMP).

Sparse Coding - ℓ^1 norm problem

- Sparse coding solving the unconstrained problem

$$\text{P1:} \quad \hat{\mathbf{x}}_{\mathbf{1}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \; J_\lambda\left(\mathbf{x}, \widehat{D}, \mathbf{s}\right)$$

where the functional is

$$J_\lambda\left(\mathbf{x}, \widehat{D}, \mathbf{s}\right) = \left\|\widehat{D}\mathbf{x} - \mathbf{s}\right\|_{\mathbf{2}}^{\mathbf{2}} + \lambda\|\mathbf{x}\|_1$$

- The sparsity requirement is relaxed by a penalization term on the ℓ_1- norm of the coefficients
- This is a Basis Pursuit Denoising (BPDN) problem: there are several optimization methods in the literature.
- We adopt Alternating Direction Method of Multipliers (ADMM)

ANOMALY INDICATORS

Anomaly Indicators

- In order to measure the extent to which a given patch $\boldsymbol{s}$ is **consistent with the nominal conditions** we compute the **sparse coding** of $\boldsymbol{s}$ w.r.t. $\widehat{\boldsymbol{D}}$

$$\mathbf{s} \rightarrow \hat{\mathbf{s}}, \quad \text{where} \;\; \hat{\mathbf{s}} = \widehat{D}\hat{\mathbf{x}} \text{ and } \hat{\mathbf{s}} \approx \mathbf{s}$$

- We need suitable **anomaly-indicators** that **quantitatively assess** how close $\boldsymbol{s}$ is to nominal patches.
  - In the specific case of sparse representations, the **anomaly indicators** have to take into account both **accuracy** and **sparsity** of the representation

Anomaly Indicators

- The **following anomaly indicators** have been considered:

  - When solving $\mathrm{P0}$ the reconstruction error

    $e(\mathbf{s}) = \left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{0}}\right\|_{\mathbf{2}}$, being $\hat{\mathbf{x}}_{\mathbf{0}}$ the solution of $\mathrm{P0}$

  - When solving $\mathrm{P1}$, the value of the functional

    $f(\mathbf{s}) = \left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{1}}\right\|_{\mathbf{2}} + \lambda\|\hat{\mathbf{x}}_{\mathbf{1}}\|_{\mathbf{1}}$, being $\hat{\mathbf{x}}_{\mathbf{1}}$ the solution of $\mathrm{P1}$

  - When solving $\mathrm{P1}$, jointly the sparsity and the error

    $g(\mathbf{s}) = [\left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{1}}\right\|_{\mathbf{2}};\ \lambda\|\hat{\mathbf{x}}_{\mathbf{1}}\|_{\mathbf{1}}]$, being $\hat{\mathbf{x}}_{\mathbf{1}}$ the solution of $\mathrm{P1}$

Anomaly Detection from 1D Anomaly Indicators

- We treat **anomaly indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of anomaly indicators from T
- In case of 1D-anomaly indicators, such a region is

$$\mathcal{J}_{\alpha}^{e} = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

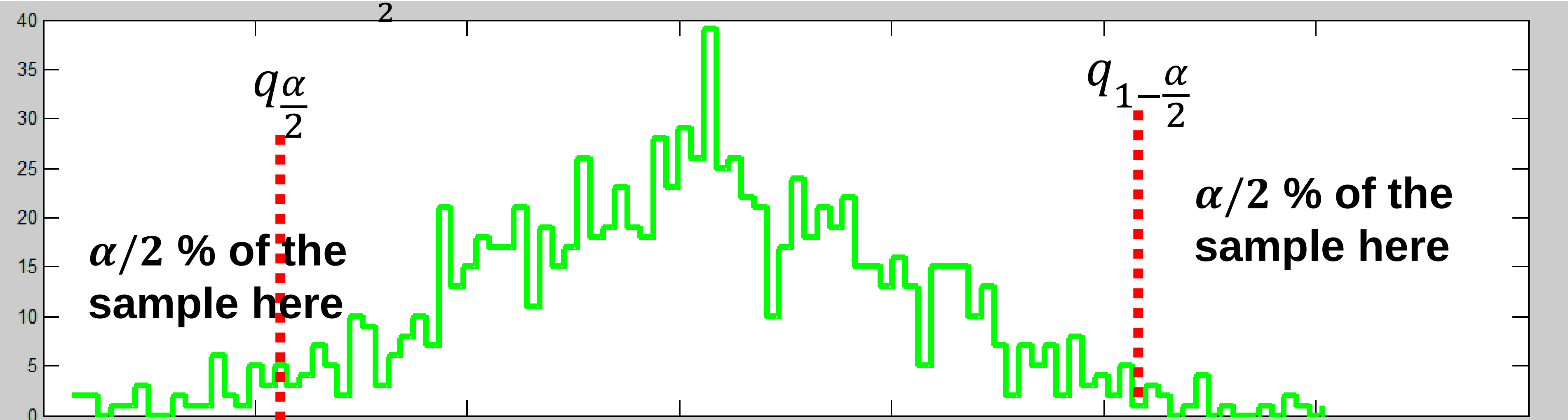
where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution

Anomaly Detection from 1D Anomaly Indicators

- We treat **anomaly indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of anomaly indicators from T
- In case of 1D-anomaly indicators, such a region is

$$\mathcal{J}^e_\alpha = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution



Anomaly Detection from 1D Anomaly Indicators

- We treat **anomaly indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of anomaly indicators from T
- In case of 1D-anomaly indicators, such a region is

$$\mathcal{J}_\alpha^e = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution

- We detect anomalies as data yielding anomaly indicators, out of high-density regions (outliers)

$$e(\mathbf{s}) \notin \mathcal{J}_\alpha^e$$

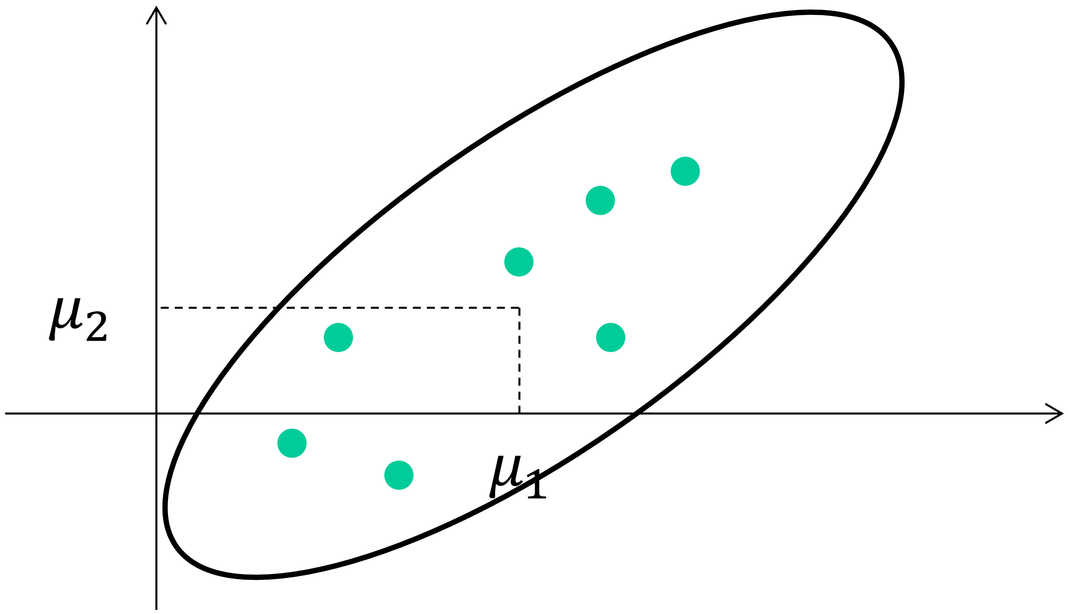
- The same for anomaly indicator $f(\cdot)$

Anomaly Detection from 2D Anomaly 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 anomaly indicators from T.



Anomaly Detection from 2D Anomaly 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 anomaly indicators from T.

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

- Anomalies are detected as

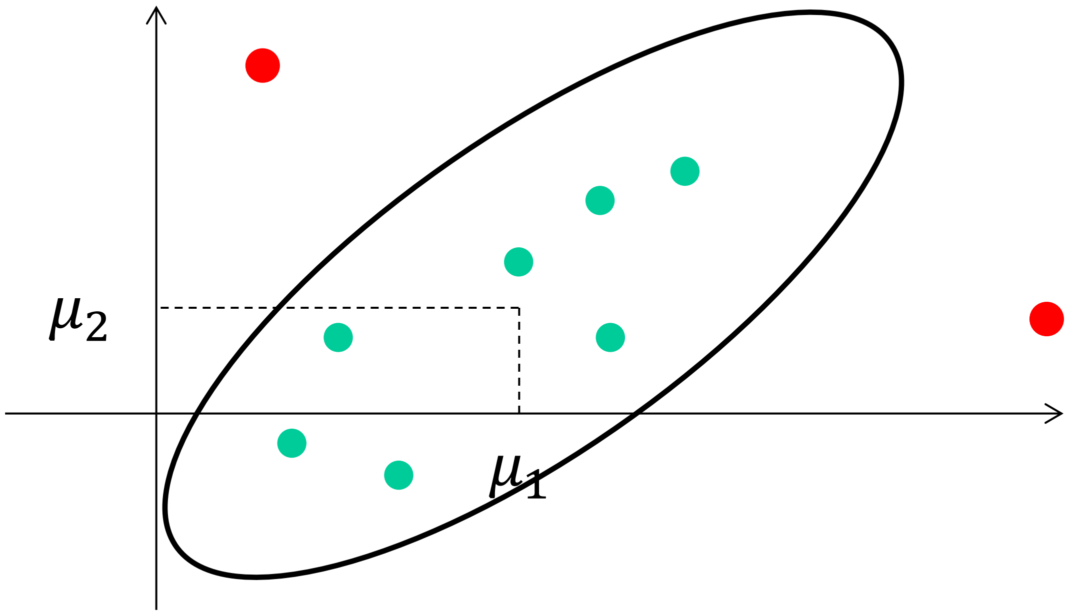
$$\mathbf{s} \quad \text{s.t.} \sqrt{(\boldsymbol{g}(\mathbf{s}) - \mu)'\Sigma^{-1}(\boldsymbol{g}(\mathbf{s}) - \mu)} > \gamma$$

Anomaly Detection from 2D Anomaly 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 anomaly indicators from T.



EXPERIMENTS

Performing change/anomaly detection using sparse representations

Anomaly detection in images

- We extract 15×15 patches from textured images, each characterized by a specific structure

Test on Synthetic Images

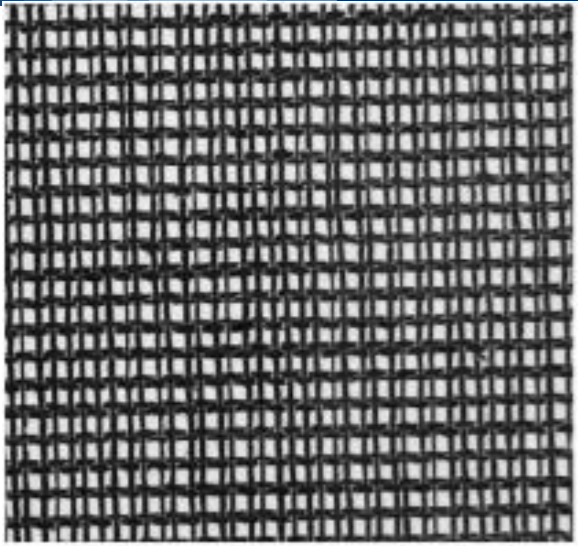


Image 1

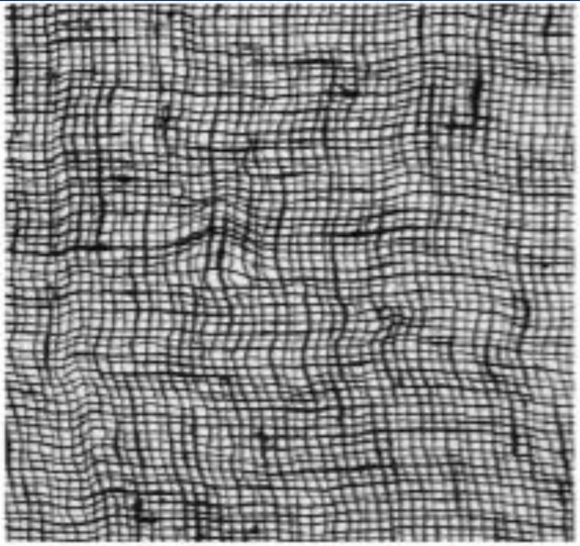


Image 2

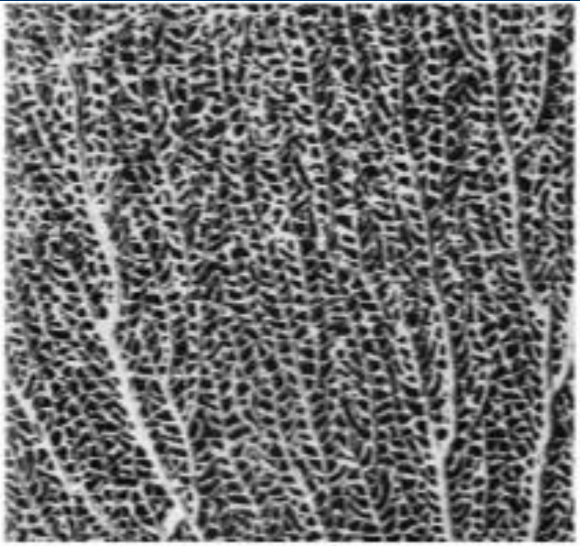


Image 3

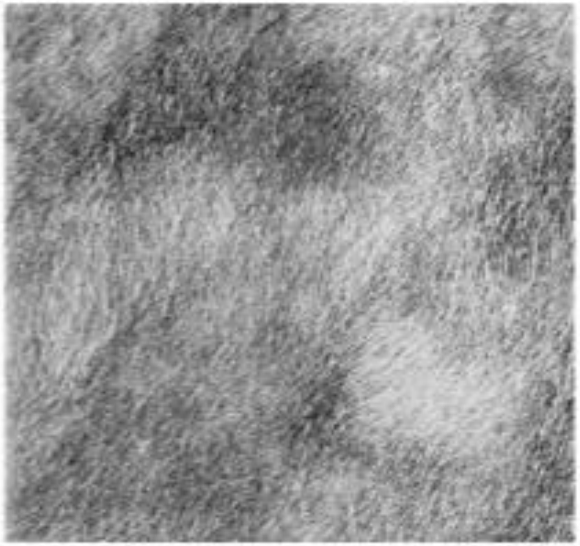


Image 4

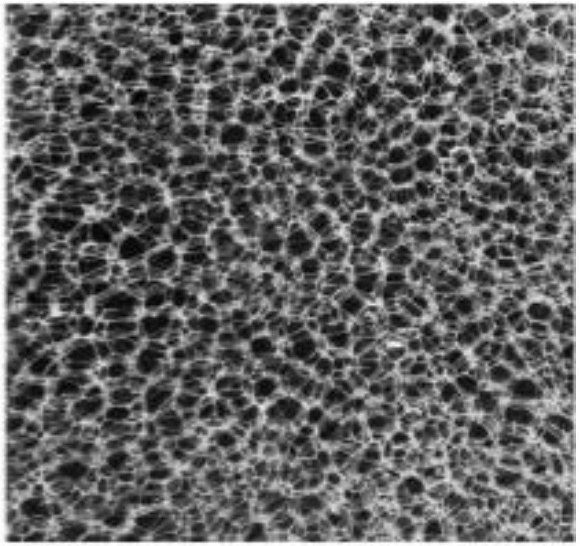


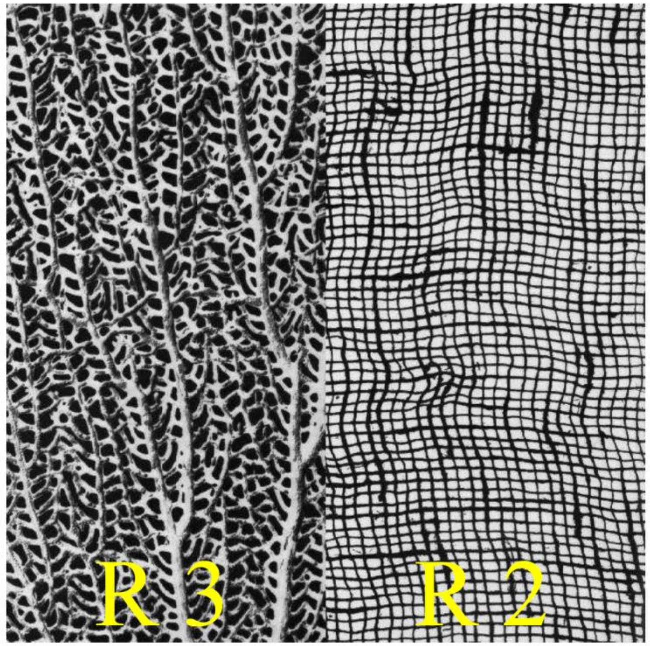
Image 5

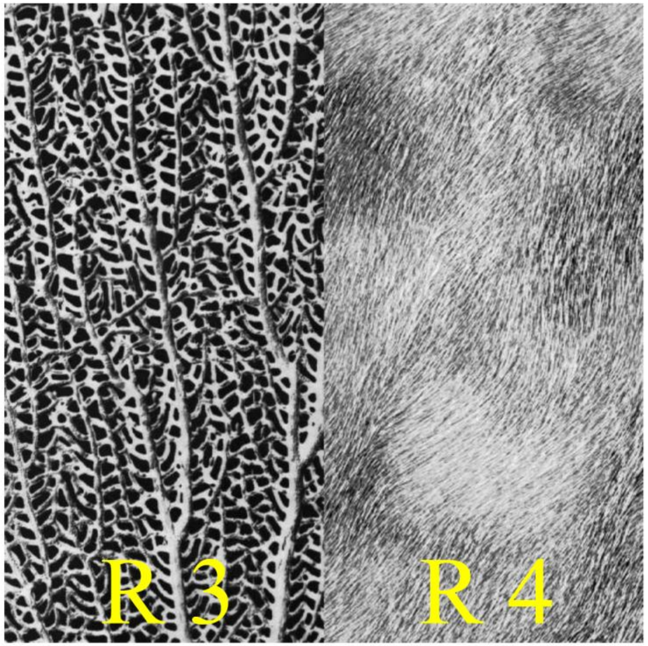
Anomaly detection in images

- Data are 15×15 patches extracted from textured images characterized by a specific structure
- Anomaly detection problems are simulated by assembling test images that contains patches from different texture
  - The left half of each image is used to learn $\widehat{D}$
  - The right half is used for testing and juxtaposed with other half images

Test Images

Test images





We learn a dictionary from L3

Anomaly detection in images

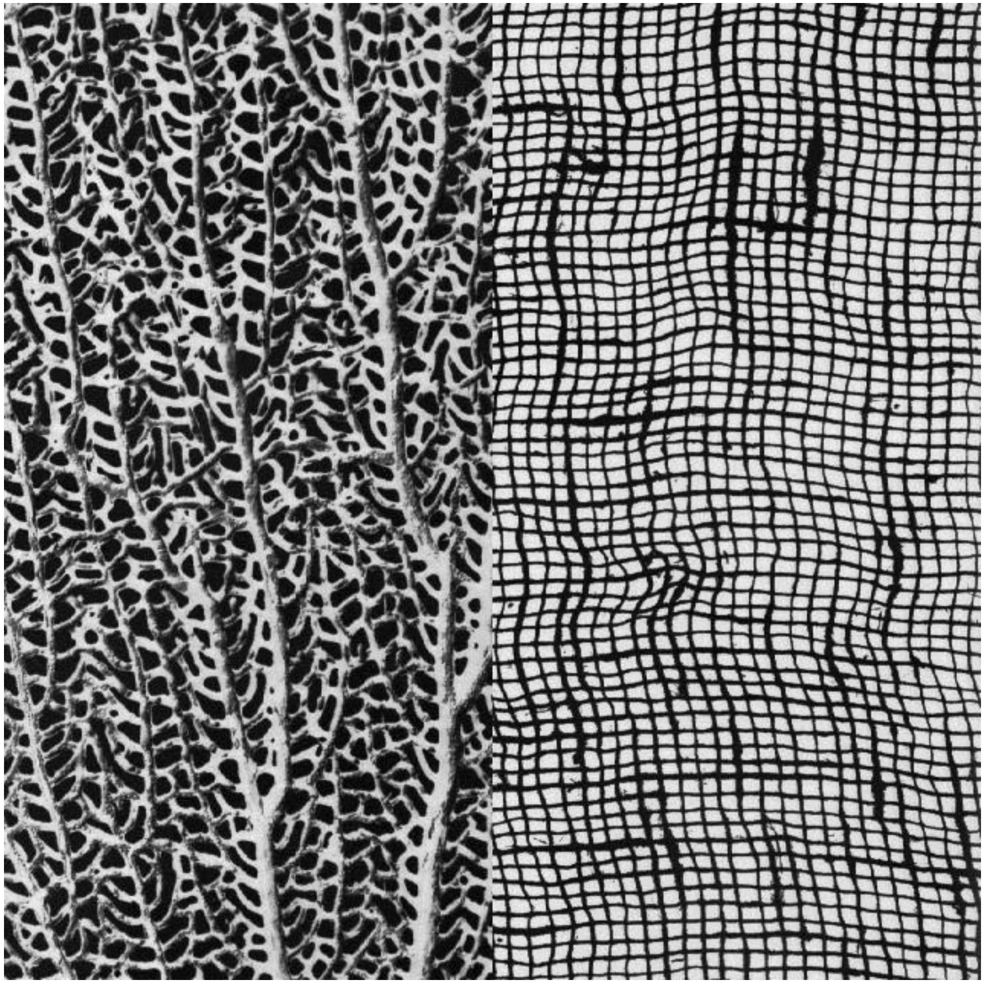
- Data are 15×15 patches extracted from textured images characterized by a specific structure
- Anomaly detection problems are simulated by syntetically creating test images gathering patches from different texture
- Each patch is **pre-processed** by subtracting its mean
- **No post-processing** to aggregate decision spatially is performed
- For further details, please refer to [Boracchi 2014]

[Boracchi 2014] Giacomo Boracchi, Diego Carrera, Brendt Wohlberg «Anomaly Detection in Images By Sparse Representations» SSCI 2014

Figures of Merit

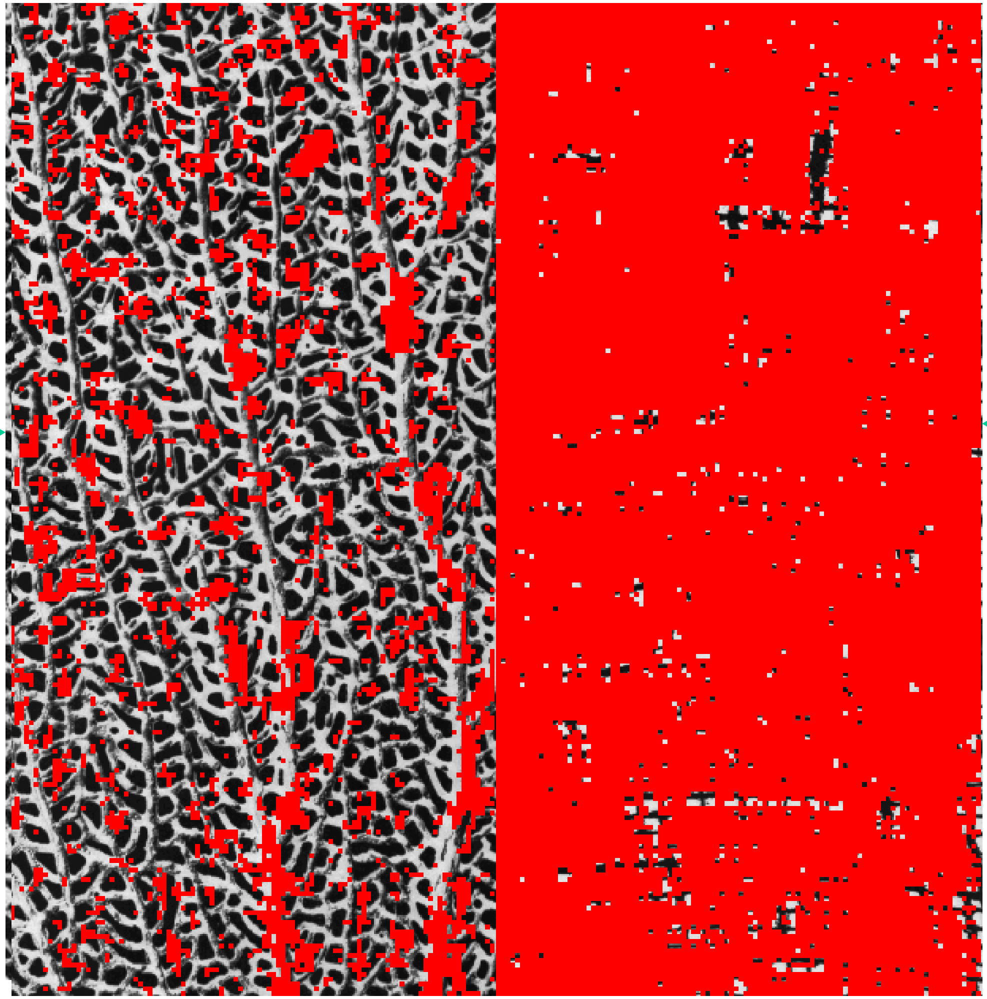
- FPR: the false positive rate, i.e. the percentage of normal patches labelled as anomalous
- TPR: the true positive rate, i.e., the percentage of anomalies correctly detected

Figures of Merit



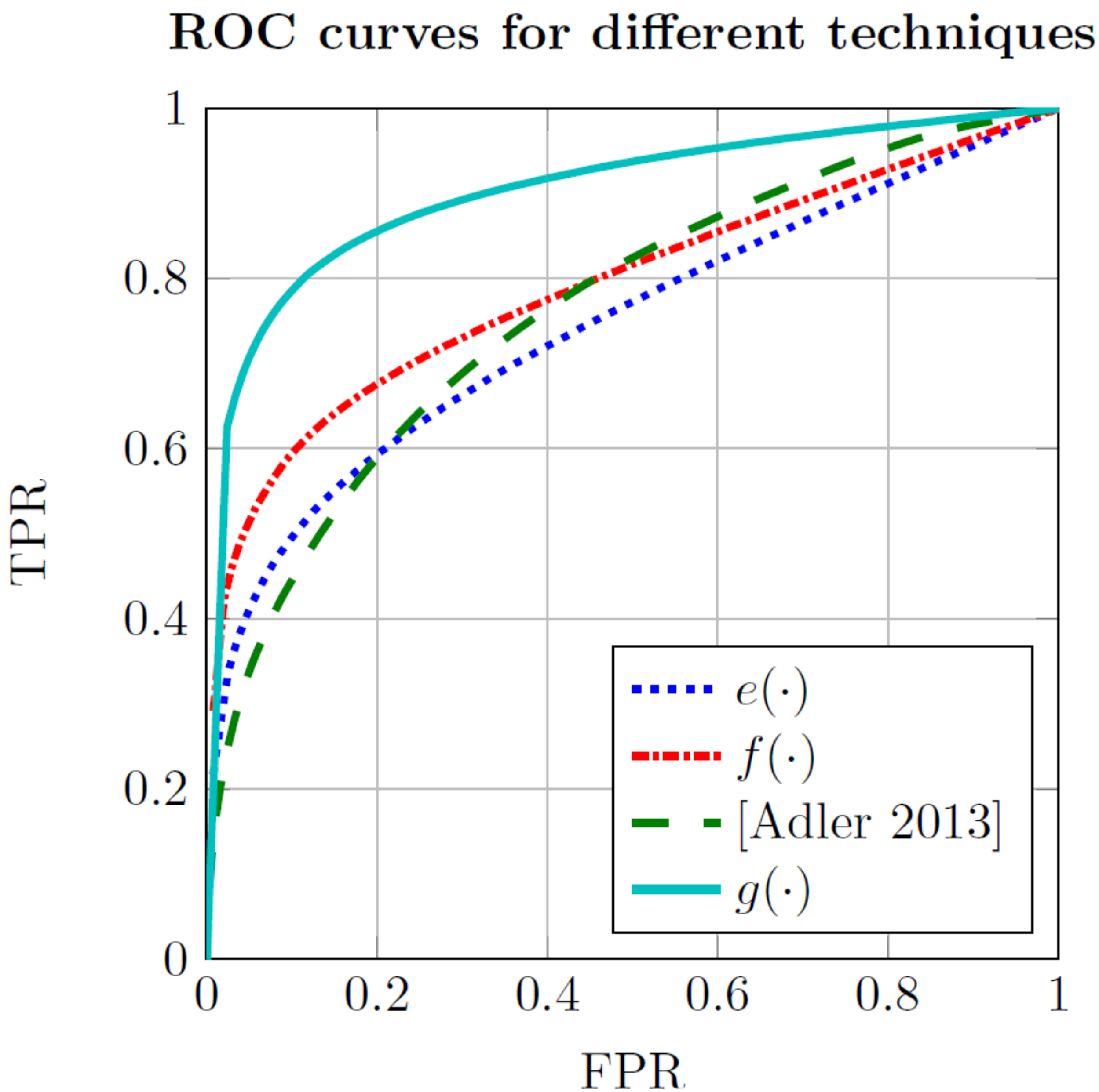
Figures of Merit

False Positives



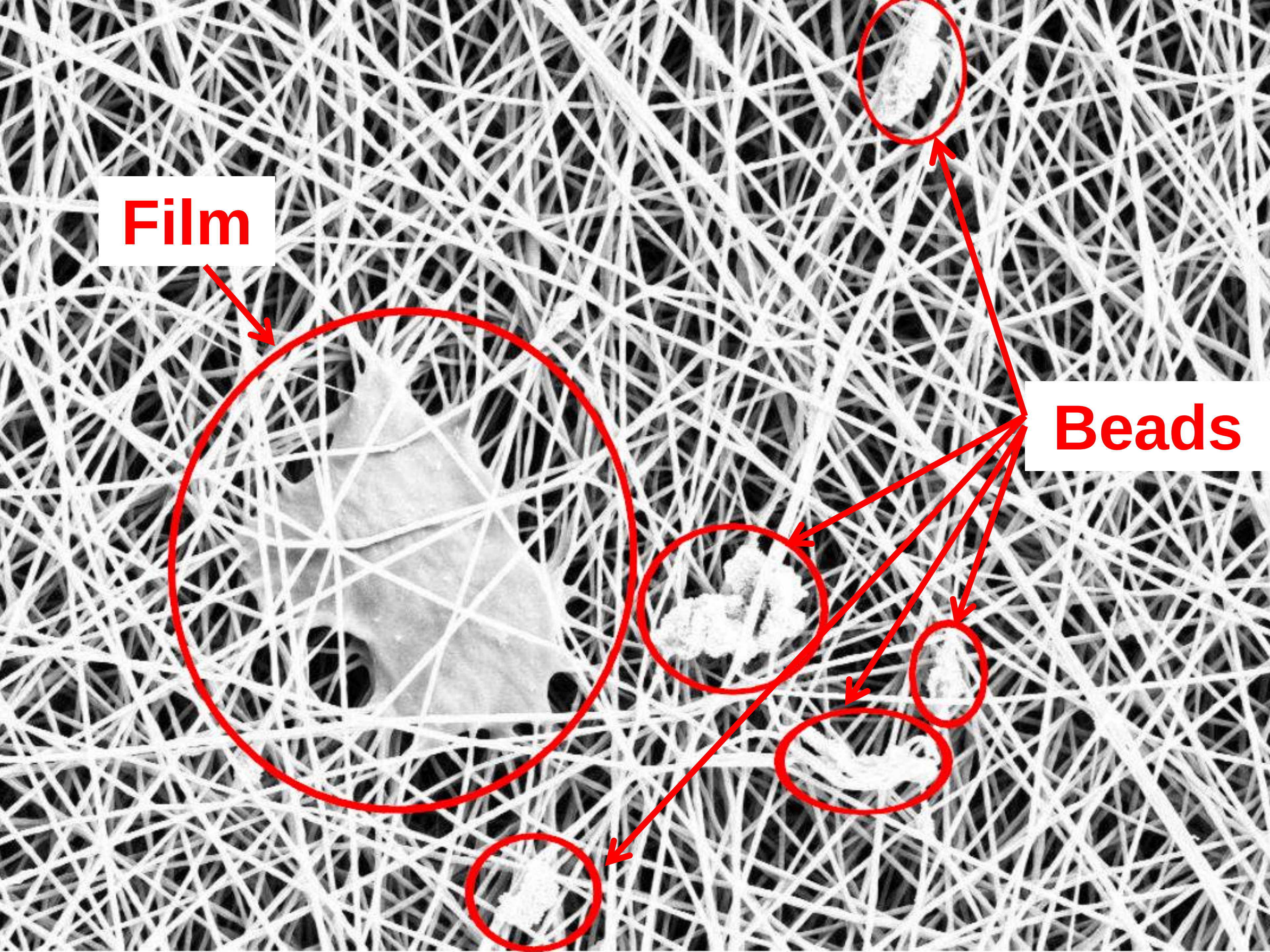
True Positives

Performance evaluation of the considered indicators



Anomaly detection in SEM images

- Problem Description: we consider the production of nanofibrous materials by an electrospinning process
- An scanning electron microscope (SEM) is used to monitor the production process and detect the presence of
  - Beads
  - Films
- Detecting anomalies and assessing how large they are is very important for supervising the monitoring process

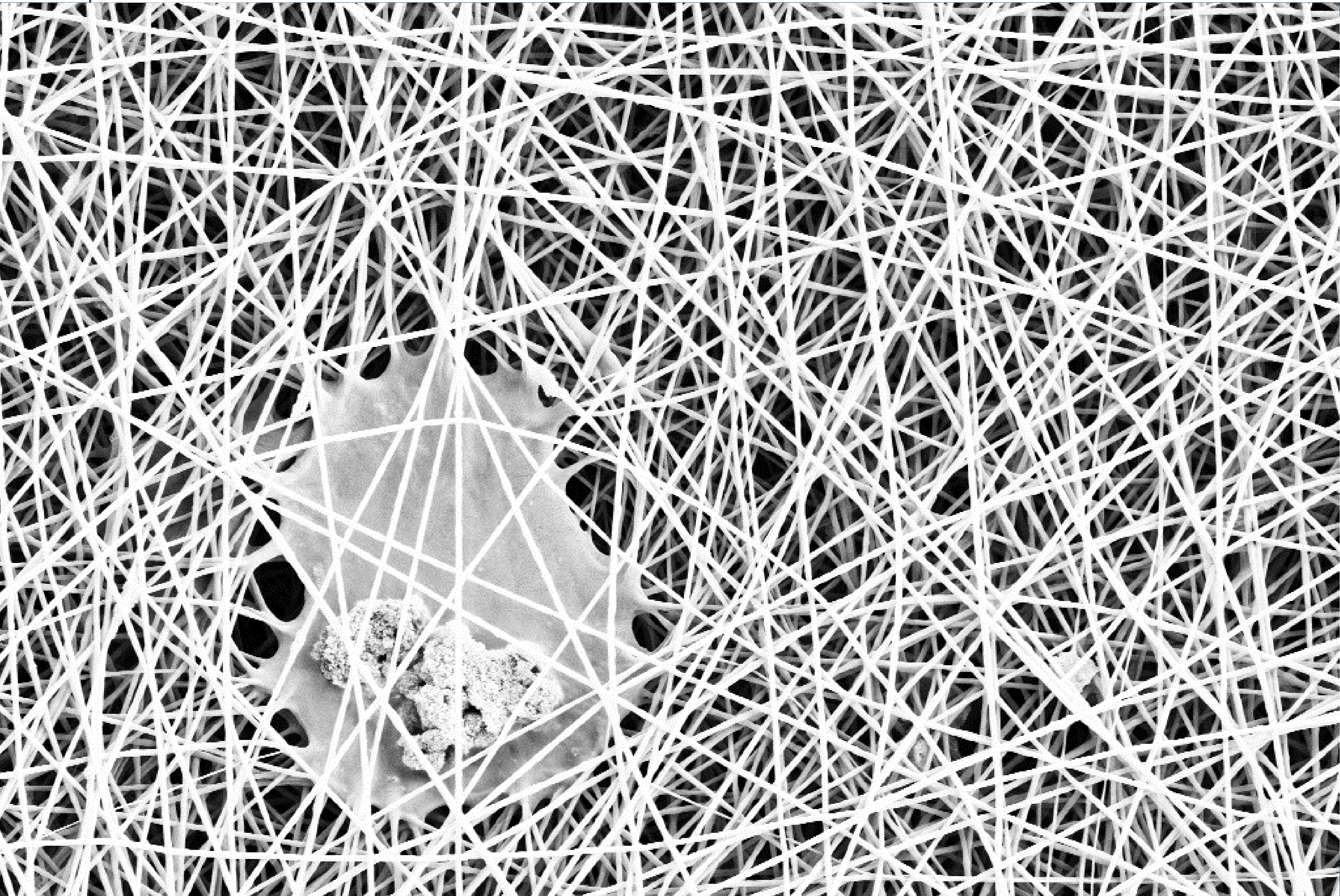

Film
Beads

Anomaly detection in SEM images

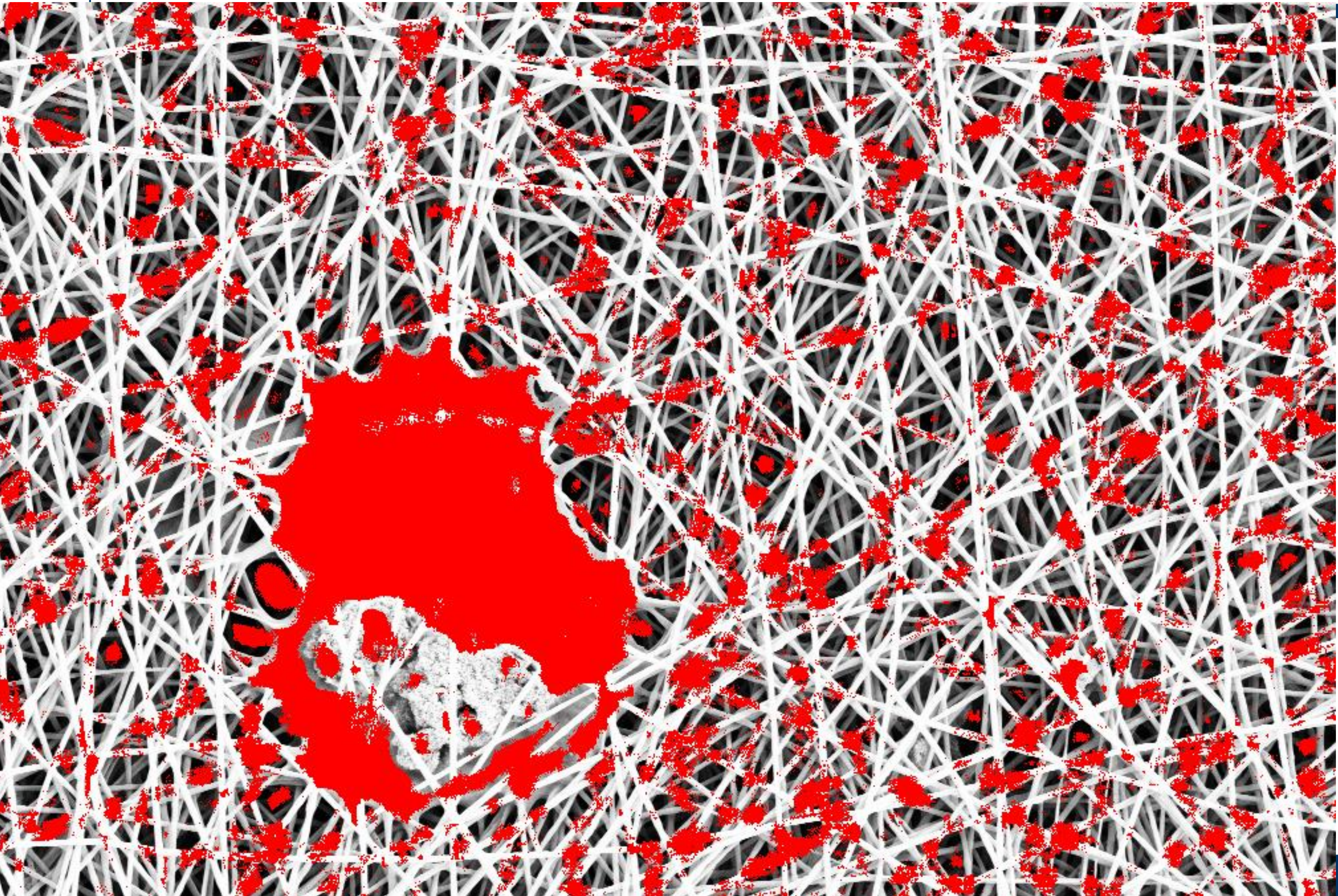
- Problem Description: we consider the production of nanofibrous materials by an electrospinning process
- An scanning electron microscope (SEM) is used to monitor the production process and detect the presence of
  - Beads
  - Films
- Detecting anomalies and assessing how large they are is very important for supervising the monitoring process
- Each anomaly detection method has been manually tuned to operate at its best performance
- Further details can be found in [Boracchi 2014]

[Boracchi 2014] Giacomo Boracchi, Diego Carrera, Brendt Wohlberg «Anomaly Detection in Images By Sparse Representations» SSCI 2014

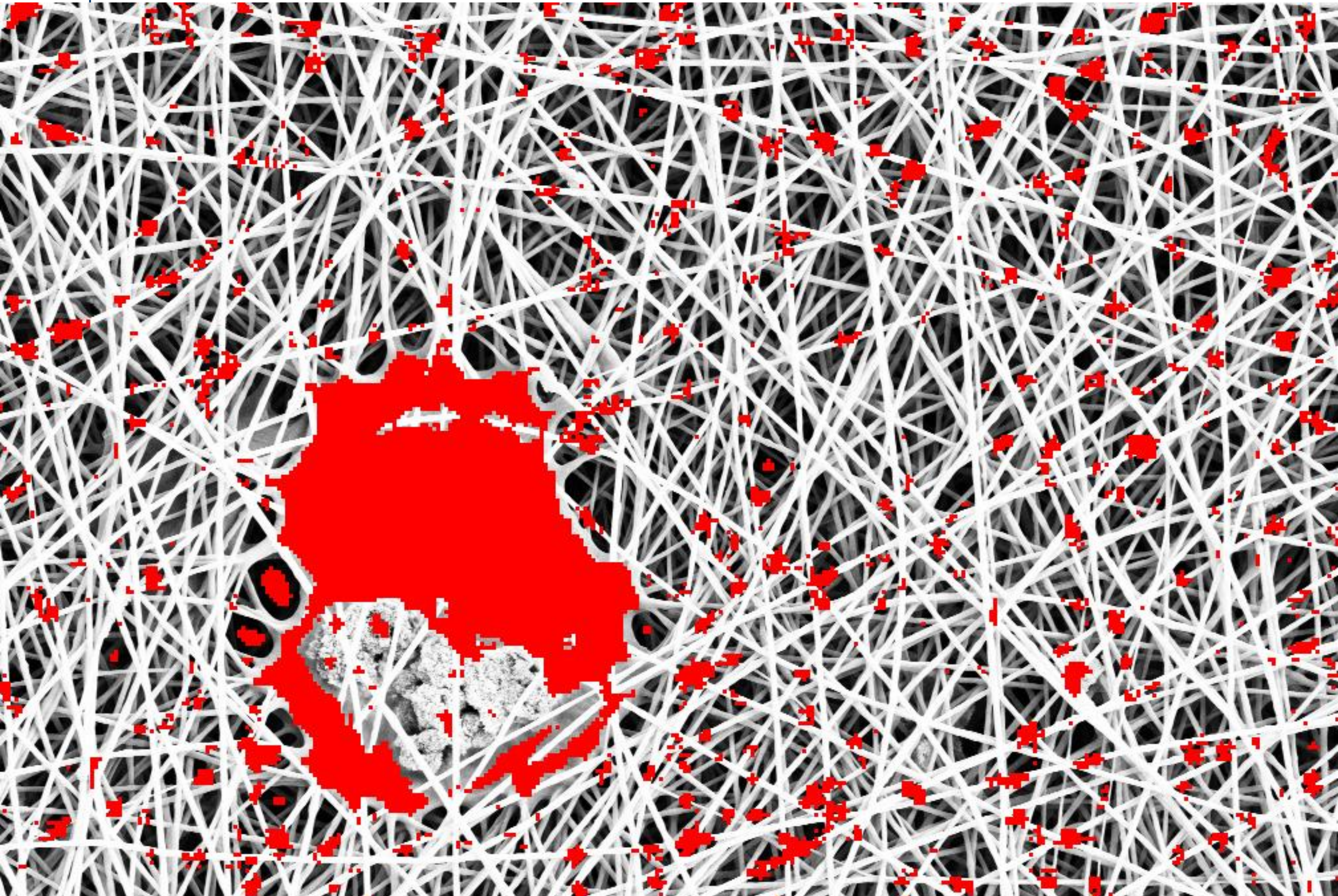
Original Image



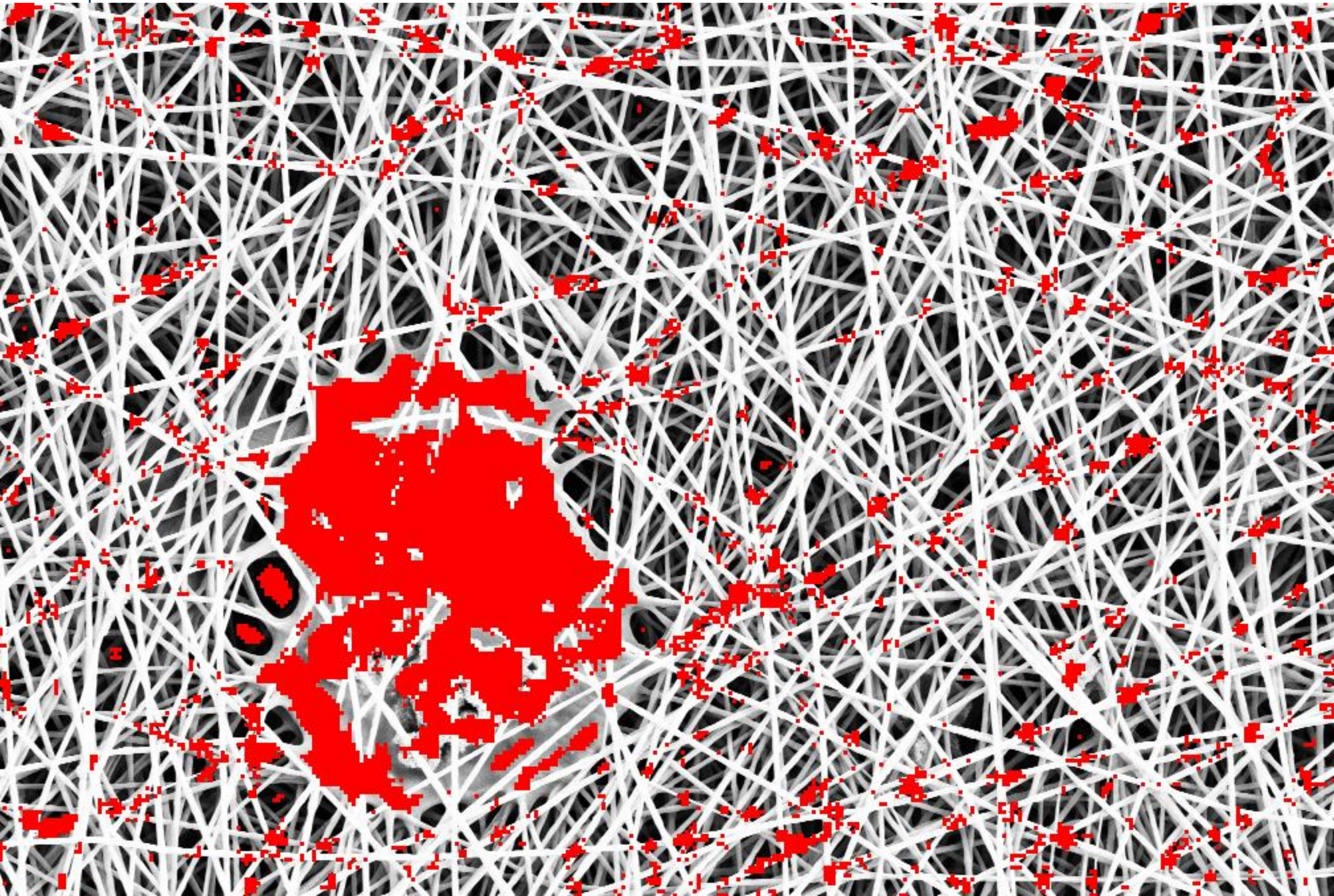
Anomaly detection by means of $e(\cdot)$



Anomaly detection by means of $f(\cdot)$



Anomaly detection by means of $g(\cdot)$



CONCLUDING REMARKS

Conclusions

- Our experiments show that **sparse representation** allows to build effective models for detecting data characterized by anomalous structures
  - Jointly monitoring the reconstruction error and the sparsity of the solution to the unconstrained BPDN problem provides best performance
- Sparse representations provide models able to describe data that in stationary conditions yield heterogenous signals (e.g. belonging to different classes): atoms of $\widehat{D}$ might be from different *classes.*

- Ongoing works include:
  - the application of these results to the sequential monitoring scenario
  - the study of customized dictionary learning metods for performing change/anomaly detection
  - the application of the proposed system to other application domains such as EGC analysis to detect arrhythmia.