Detecting Anomalous Structures By Convolutional Sparse Models

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

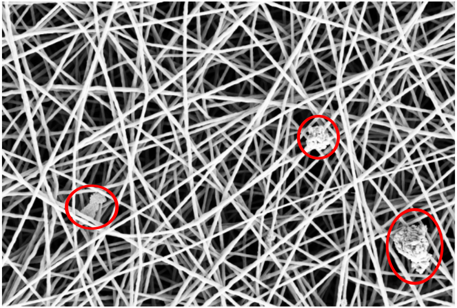
July 15, 2015

Joint work with:

- D. Carrera from Politecnico di Milano,
- A. Foi from Tampere Univerisity of Technology
- B. Wohlberg from Los Alamos National Laboratory

Anomaly Detection in Images

- We address the problem of detecting anomalies in images
- We assume that images acquired under normal conditions are characterized by specific structures
- Regions that do not conform to these structures are considered anomalies



Outline

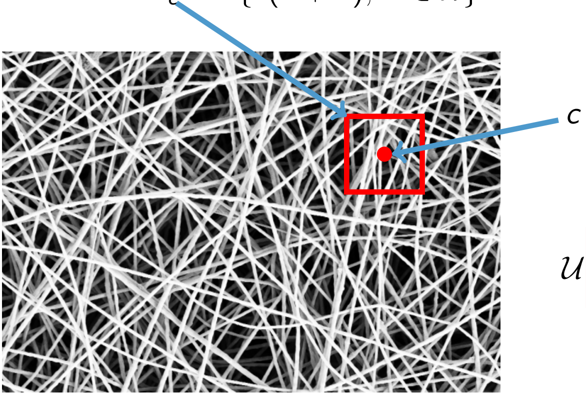
- Problem Formulation
- Convolutional Sparse Models
- Anomaly Detection
- Experiments

PROBLEM FORMULATION

Patch-Generating Process

- Patches are *small* regions having shape $\mathcal{U}$ extracted from an image s

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



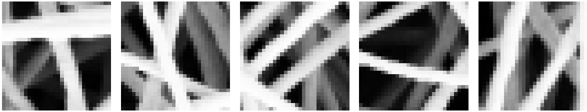
Patch-Generating Process

- Patches are *small* regions having shape $\mathcal{U}$ extracted from an image s

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

- We assume that in normal conditions, patches $\Pi_c s \in \mathbb{R}^P$ are i.i.d. realizations from an unknown stochastic process $\mathcal{P}_N$

$$\Pi_c s \sim \mathcal{P}_N$$



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

$$\Pi_c s \sim \mathcal{P}_A$$

- The process generating anomalies $\mathcal{P}_A \neq \mathcal{P}_N$ is unknown

Our Approach to Anomaly Detection

Our Assumptions

- A training set $\mathcal{T}$ of normal images is available
- No training samples for anomalous data are given

Our Approach

- Learn a model $\mathcal{M}$ approximating patches generated by $\mathcal{P}_N$
- Detect as anomalous any patch that does not conform to $\mathcal{M}$
  - We define an indicator vector $\mathbf{g}(c)$ that measure the degree to which $\mathcal{M}$ fits each patch $\Pi_c s$
  - We detect anomalies as outliers w.r.t. the distribution of indicator-vectors

CONVOLUTIONAL SPARSE MODELS

Convolutional Sparse Models

Main Assumptions

For each anomaly-free image s:

$$s \approx \sum_{m=1}^{M} d_m * x_m \quad (1)$$

- $\{d_m\}$ are small filters that approximate normal images
- $\{d_m\}$ have to be learned to describe any normal image s
- the coefficient maps $\{x_m\}$ are sparse
- the number of filters M is small

Convolutional Sparse Models

Main Assumptions

For each anomaly-free image s:

$$s \approx \sum_{m=1}^{M} d_m * x_m \quad (1)$$

- $\{d_m\}$ are small filters that approximate normal images
- $\{d_m\}$ have to be learned to describe any normal image s
- the coefficient maps $\{x_m\}$ are sparse
- the number of filters M is small

Convolutional Sparse Models

Main Assumptions

For each anomaly-free image s:

$$s \approx \sum_{m=1}^{M} d_m * x_m \quad (1)$$

- $\{d_m\}$ are small filters that approximate normal images
- $\{d_m\}$ have to be learned to describe any normal image s
- the coefficient maps $\{x_m\}$ are sparse
- the number of filters M is small

Convolutional Sparse Models

Main Assumptions

For each anomaly-free image s:

$$s \approx \sum_{m=1}^{M} d_m * x_m \quad (1)$$

- $\{d_m\}$ are small filters that approximate normal images
- $\{d_m\}$ have to be learned to describe any normal image s
- the coefficient maps $\{x_m\}$ are sparse
- the number of filters M is small

Convolutional Sparse Models

To make the approximation (1) accurate for small values of M, each image s is preprocessed as

$$s = s_l + s_h$$

- s_l: low-frequency components
- s_h: high-frequency components

Main Assumptions

For each anomaly-free image s:

$$s_h \approx \sum_{m=1}^{M} d_m * x_m \quad (1)$$

- $\{d_m\}$ are small filters that approximate normal images
- $\{d_m\}$ have to be learned to describe any normal image s
- the coefficient maps $\{x_m\}$ are sparse
- the number of filters M is small

Learning the Filters for Modeling Normal Patches

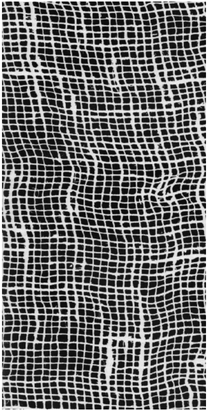
- Filter $\{d_m\}$ are learned by solving the optimization problem over a training set $\{s^{(t)}\}$ of normal images

$$\underset{\{d_m\},\{x_m^{(t)}\}}{\arg\min} \sum_{s^{(t)}\in\mathcal{T}} \left[\frac{1}{2} \left\| \sum_m d_m * x_m^{(t)} - s_h^{(t)} \right\|_2^2 + \lambda \sum_m \left\| x_m^{(t)} \right\|_1 \right]$$

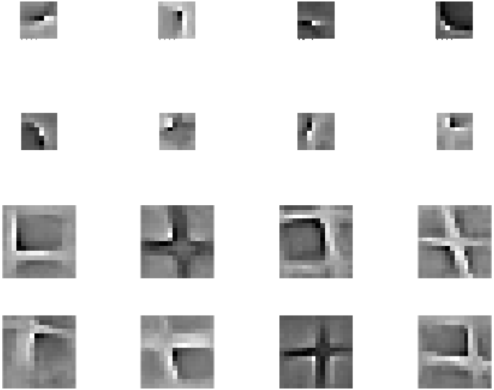
- Parameter $\lambda > 0$ balances the reconstruction error and the sparsity of the coefficient maps
- This problem is solved via an efficient formulation in the Fourier Domain by the Alternating Direction Method of Multipliers

Learning the Filters for Modeling Normal Patches

Training images



Learned filters



Sparse Coding

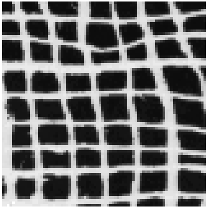
- The learned filters $\{d_m\}$ are used to compute the coefficient maps $\{x_m\}$ corresponding to an image s
- This operation is referred to as the sparse coding and it is performed by solving the optimization problem

$$\underset{\{x_m\}}{\arg\min} \frac{1}{2} \left\| \sum_m d_m * x_m - s_h \right\|_2^2 + \lambda \sum_m \|x_m\|_1$$

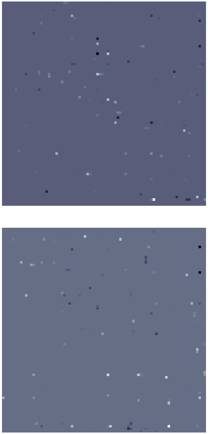
- This problem is solved via an efficient formulation in the Fourier Domain of Alternating Direction Method of Multipliers

Sparse Coding

Test Image



Feature maps



Filters

ANOMALY DETECTION

Indicator Vector

For each patch of an image s we compute an indicator vector $\mathbf{g}(c)$ to quantitatively assess the extent to which $\Pi_c s$ is consistent with normal patches:

$$s = s_l + s_h$$

$$\mathbf{g}(c) = \begin{bmatrix} \mathbf{g}_l(c) \\ \mathbf{g}_h(c) \end{bmatrix}$$

Sample Moments

Convolutional Sparse Models

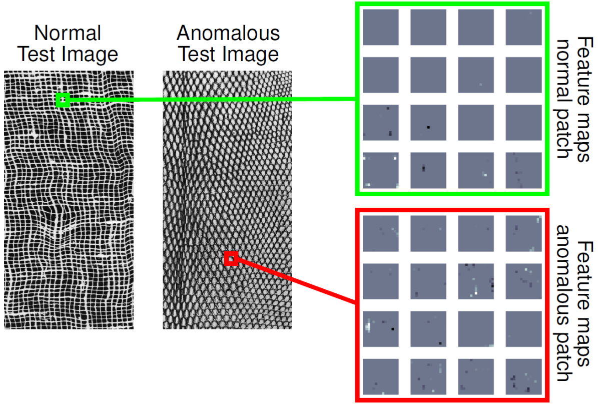
Indicator Vector in High Frequencies

Given the coefficient maps $\{x_m\}$ of an image s,
the most straightforward indicators are

- Reconstruction Error $\left\|\Pi_c \left(\sum_m d_m * x_m - s_h\right)\right\|_2^2$

- Sparsity $\sum_m \left\|\Pi_c x_m\right\|_1$

Indicator Vector in High Frequencies



- The coefficient maps corresponding to the highlighted regions have the same ℓ^1 norm
- The spread of the coefficients across the feature maps has to be taken into account

Indicator Vector in High Frequencies

Given the coefficient maps $\{x_m\}$ of an image s,
the most straightforward indicators are

- Reconstruction Error $\left\|\Pi_c\left(\sum_m d_m * x_m - s_h\right)\right\|_2^2$
- Sparsity $\sum_m \|\Pi_c x_m\|_1$

We experience that these indicators are not sufficient for discriminating anomalous regions

- Group Sparsity $\sum_m \|\Pi_c x_m\|_2$

Indicator Vector in High Frequencies

Given the coefficient maps $\{x_m\}$ of an image s, the most straightforward indicators are

- Reconstruction Error $\left\|\Pi_c \left(\sum_m d_m * x_m - s_h\right)\right\|_2^2$

- Sparsity $\sum_m \|\Pi_c x_m\|_1$

We experience that these indicators are not sufficient for discriminating anomalous regions

- Group Sparsity $\sum_m \|\Pi_c x_m\|_2$

$$\mathbf{g}_h(c) = \begin{bmatrix} \left\|\Pi_c \left(\sum_m d_m * x_m - s_h\right)\right\|_2^2 \\ \sum_m \|\Pi_c x_m\|_1 \\ \sum_m \|\Pi_c x_m\|_2 \end{bmatrix}$$

Indicator Vector in Low Frequencies

- To detect anomalies affecting the low frequencies components s_l we do not learn a model
- We simply consider the sample moments of the patches from s_l

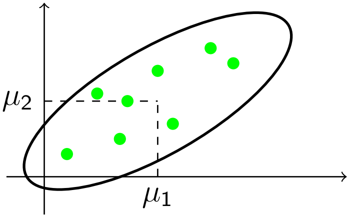
$$\mathbf{g}_l(c) = \begin{bmatrix} \hat{\mu}(\Pi_c s_l) \\ \hat{\sigma}(\Pi_c s_l) \end{bmatrix}$$

Anomaly Detection

- For the indicator $\mathbf{g}(\cdot)$ we build a confidence region

$$R_\gamma = \left\{ \xi \in \mathbb{R}^5, \text{ s.t. } \sqrt{(\xi - \mu)^T \Sigma^{-1} (\xi - \mu)} \leq \gamma \right\}$$

where μ and Σ are the sample mean and the sample covariance of the anomaly indicators computed on the training images



Anomaly Detection

- For the indicator $\mathbf{g}(\cdot)$ we build a confidence region

$$R_\gamma = \left\{\xi \in \mathbb{R}^5, \text{ s.t. } \sqrt{(\xi-\mu)^T\Sigma^{-1}(\xi-\mu)} \leq \gamma\right\}$$

  where μ and Σ are the sample mean and the sample covariance of the anomaly indicators computed on the training images
- The Chebyshev's inequality ensures that a normal patch falls outside R_γ with probability $\leq 2/\gamma^2$
- Anomalies are detected as

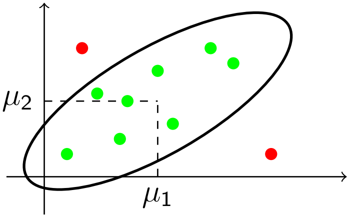
$$\Pi_c s \text{ s.t. } \sqrt{(\xi-\mu)^T\Sigma^{-1}(\xi-\mu)} > \gamma$$

Anomaly Detection

- For the indicator $\mathbf{g}(\cdot)$ we build a confidence region

$$R_\gamma = \left\{ \xi \in \mathbb{R}^5, \text{ s.t. } \sqrt{(\xi - \mu)^T \Sigma^{-1} (\xi - \mu)} \leq \gamma \right\}$$

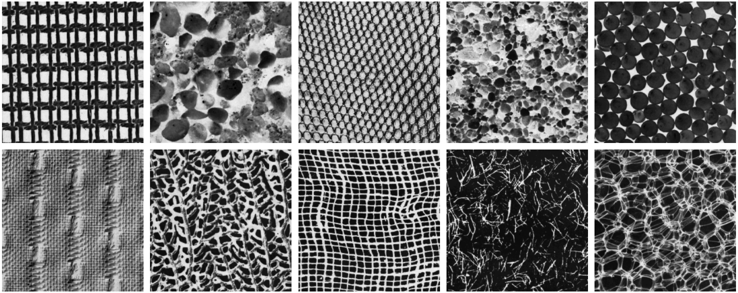
where μ and Σ are the sample mean and the sample covariance of the anomaly indicators computed on the training images



EXPERIMENTS

Dataset

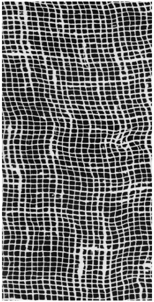
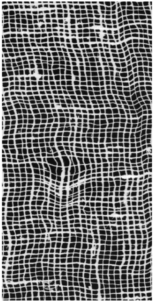
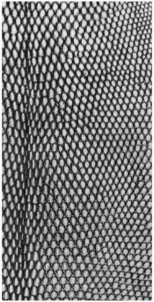
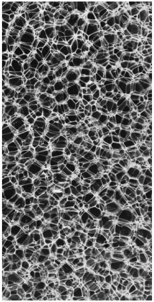
- Data are 15×15 patches extracted from 25 texture images characterized by a specific structure



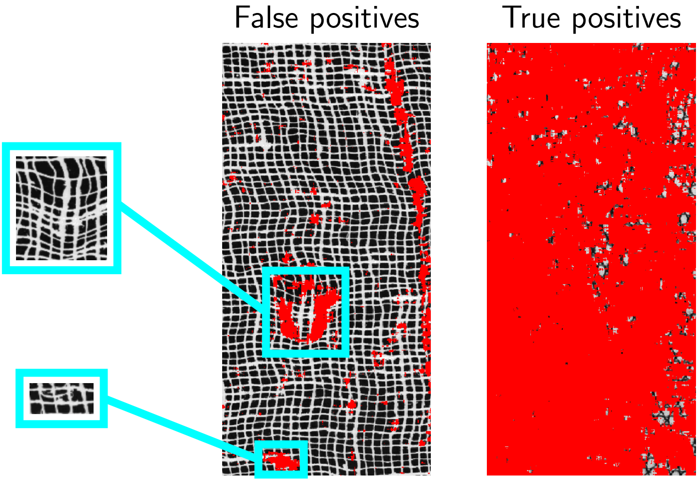
Dataset

- Data are 15×15 patches extracted from 25 texture images characterized by a specific structure
- Each texture image is split in two halves
  - Each left half is used to learned filters $\{d_m\}$
  - The right halves are used as test images

Test Images

Training Image	Normal Test Image	Anomalous Test Images	
			

Anomaly Detection Example



Figures of Merit

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

Both FPR and TPR depend on parameter γ used in the definition of the confidence region R_γ

$$R_\gamma = \left\{ \xi \in \mathbb{R}^5, \text{ s.t. } \sqrt{(\xi - \mu)^T \Sigma^{-1} (\xi - \mu)} \leq \gamma \right\}$$

Considered Methods

The following approaches are considered to analyze the high-frequency components s_h and build the vector $\mathbf{g}_h$:

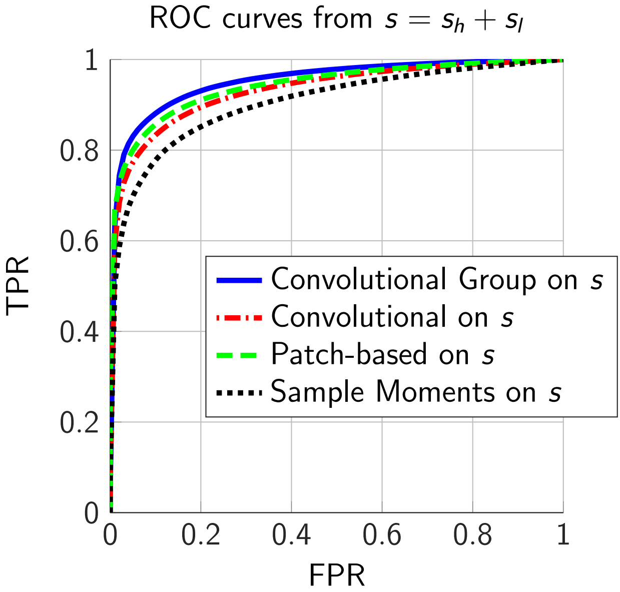
- Convolutional group: the proposed approach
- Convolutional: only the reconstruction error and sparsity are monitored
- Patch-Based: a standard sparse model instead of a convolutional one is exploited. The indicator vector includes the reconstruction error and the sparsity
- Sample Moments: no model is used to approximate normal patches. We monitor only the sample moments of the patch from s_h

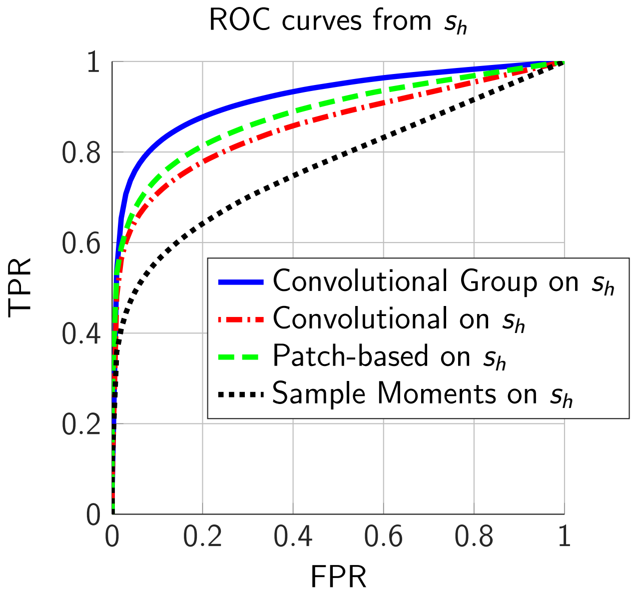
The low-frequency components are analyzed by monitoring the sample moments of the patch from s_l

Experiments performed

- In Experiment 1 we consider the whole spectrum and we detect anomalies by monitoring both s_l and s_h
- In Experiment 2 we analyze only the high frequency components s_h, to assess the performance of the detector when it is based exclusively on convolutional sparse models

Experiment 1




ROC curves from s_h
TPR
FPR
0
0.2
0.4
0.6
0.8
1
Convolutional Group on s_h
Convolutional on s_h
Patch-based on s_h
Sample Moments on s_h

CONCLUSIONS

Conclusions

- Our experiments show that
  - convolutional sparse models are very effective at capturing local structures in high frequency components of images
  - the local-group sparsity term is extremely effective in discriminating normal patches from anomalous ones
- Our approach has two main advantages with respect to the one based on patch-based sparsity:
  - the filters may have different size, while it is not trivial exploits multiscale dictionary in a sparsity model
  - the patch size can be arbitrarily increase at a negligible computational overhead

QUESTIONS?