Scale-invariant anomaly detection with multiscale group-sparse models

SCALE-INVARIANT ANOMALY DETECTION WITH MULTISCALE GROUP-SPARSE MODELS

Diego Carrera[a,b] *Giacomo Boracchi*[b] *Alessandro Foi*[a] *Brendt Wohlberg*[c]

[a] Department of Signal Processing, Tampere University of Technology, Finland
[b] Dipartimento di Elettronica, Informazione e Bioingegeria, Politecnico di Milano, Italy
[c] Theoretical Division, Los Alamos National Laboratory, NM, USA

ABSTRACT

The automatic detection of anomalies, defined as patterns that are not encountered in representative set of *normal* images, is an important problem in industrial control and biomedical applications. We have shown that this problem can be successfully addressed by the sparse representation of individual image patches using a dictionary learned from a large set of patches extracted from normal images. Anomalous patches are detected as those for which the sparse representation on this dictionary exceeds sparsity or error tolerances. Unfortunately, this solution is not suitable for many real-world visual inspection-systems since it is not scale invariant: since the dictionary is learned at a single scale, patches in normal images acquired at a different magnification level might be detected as anomalous.
We present an anomaly-detection algorithm that learns a dictionary that is invariant to a range of scale changes, and overcomes this limitation by use of an appropriate sparse coding stage. The algorithm was successfully tested in an industrial application by analyzing a dataset of Scanning Electron Microscope (SEM) images, which typically exhibit different magnification levels.

Index Terms— Anomaly detection, image analysis, sparse representations, dictionary learning, group sparsity

1. INTRODUCTION

We address the problem of automatically detecting anomalies in images, defined as regions that do not conform to the structures in a reference training set of *normal* images. This is a very important issue in high-throughput industrial control or biomedical applications, where normal images are characterized by their own peculiar structures, and any region departing from these might require further inspection by technicians or doctors. Moreover, in these applications it is often necessary to provide a quantitative assessment of the acquired images and measure the area covered by anomalies.

As a meaningful example of applications that we are targeting, consider Scanning Electron Microscope (SEM) images like those in Figure 1(a), which are acquired to monitor the production of nanofibers [1, 2]. Inside normal regions, fibers appear as very thin filaments (having diameter below 0.5 micron), while anomalous regions might exhibit very different structures, like those in Figure 1(b).

There are two main approaches to this sort of anomaly-detection problems [3]: (*i*) designing features that discriminate between normal and anomalous regions, and (*ii*) designing features that provide a known response to normal data, while anomalous data are identified as those yielding an unusual response. The former is a viable approach only when potential anomalies are known beforehand, and yields solutions that are highly application-specific. The latter is a more challenging approach because it does not exploit any information about the anomalies to be detected, however, it is also more practical as it requires only a training set of normal data which is often easy to collect. This problem is also referred as *novelty detection* [4, 5, 6], and is formulated as a one-class classification problem [7].

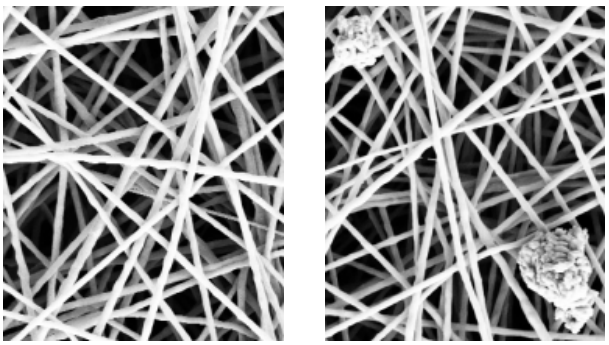
We pursue the latter approach and tackle anomaly detection by learning a model that approximates patches belonging to normal images. Then, during operation, we analyze test images in a patch-wise manner, and we determine whether each patch is normal or anomalous by assessing the goodness of fit of the learned model. A patch-wise analysis is necessary since the anomalous regions may be small, and thus not detectable by analyzing global statistics.

We observe that the content of patches extracted from normal images may be very different (even when they do not overlap with anomalous regions), also because in real-world applications images might be acquired at different magnification levels. Consider, for example, the images in Figure 2, acquired from different specimens of a nanofibers. While patches in these images are perceptually similar, their content is very different because the images are acquired at different magnification levels (i.e. scales). This means that an anomaly detector trained on an image like Figure 2(a) would possibly detect as anomalous patches from images like Figure 2(d), which should instead be considered as normal since only the scale has changed. To successfully address this issue it is necessary to design anomaly detectors which are invariant with respect to change in the scale.

Here, we propose a scale-invariant anomaly-detection algorithm, which uses a multiscale dictionary to describe patches belonging to normal images acquired at different scales. Our contribution does not concern dictionary learning, which is here performed by juxtaposing atoms of dictionaries learned from images at different scales, but rather the anomaly-detection algorithm that effectively uses this powerful model. In particular, we show that, to successfully detect anomalies, it is necessary to introduce a group-sparsity penalization term when encoding each patch with respect to the learned dictionary. By doing so, we promote representations using atoms belonging to one or a few scales. Our experiments show that our scale-invariant anomaly detector can correctly identify anomalous regions in SEM images from a real-world industrial-monitoring application, even when these are acquired at a scale that does not appear among the training images. Most importantly, the proposed anomaly detector achieves a performance close to that of an ideal single-scale detector that is trained on the exact scale of the test images.

1.1. Related works

Anomaly-detection problems are quite common in imaging applications, like the identification of masses in mammograms [8], the



(a) An example of SEM image that is considered normal.

(b) An example of SEM image containing anomalous clots.

Fig. 1. SEM images for monitoring nanofibers production.

detection of sea mines in side-scan sonar images [9], or of defects in industrial monitoring applications [10], to name a few examples.

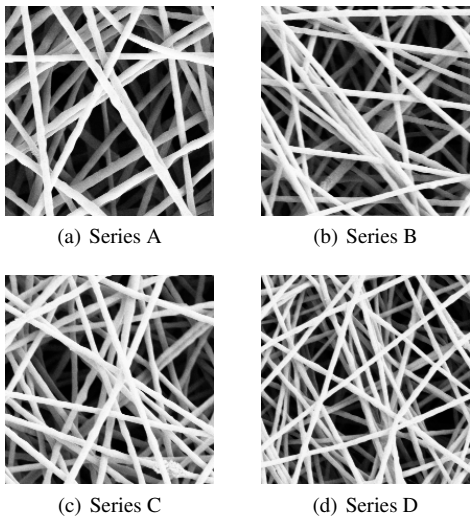
All the above methods either assume that a reference defect-free image is provided [10], or specific features are available to characterize normal data [8], and this may not be the case in most applications. Here we consider solutions where a model describing normal structures has to be learned from training images, and in particular solutions based on sparse representations [11, 12, 15]. Recently, sparse models have been successfully used for detecting anomalies in images [11, 12]. In practice, patches from test images are represented (sparse coding) with respect to a dictionary learned from normal images. Then, [11] identifies anomalies as outliers in the distribution of indicators computed from such representations, while [12] embeds the detection phase inside the sparse coding and does not implement separate outlier detection. Convolutional-sparse models [13, 14] were also successfully used to detect anomalies in images [15]. Unfortunately, because of a known limitation of patch-based sparse representations, none of these methods is able to handle scale change. The solution proposed here solves this issue, as it is able to analyze normal images acquired at different scales without any reference image or know feature that describes normal data.

2. PROBLEM FORMULATION

Let us denote by $s : \chi \to \mathbb{R}^+$ a grayscale image, where $\chi \subset \mathbb{Z}^2$ is the regular pixel grid representing the image domain. We formulate the anomaly-detection problem in terms of image patches, where we define a patch $\mathbf{s}_c$ as a square region of $\sqrt{P} \times \sqrt{P}$ pixels centered at the pixel c of the image s. To simplify our notation, in what follows we omit the center c of each patch and refer to image patches as $\mathbf{s}$, which are organized in column vectors.

We assume that normal patches, i.e. those in normal images, are generated by an unknown stochastic process $\mathcal{P}_N$. Anomalous patches are instead generated from a different process $\mathcal{P}_A$ that is also unknown and feature structures that do not conform to those characterizing normal images. Since images may be acquired at different magnification levels, patches generated by either $\mathcal{P}_N$ and $\mathcal{P}_A$ may exhibit very different content, although they are perceptually similar, as in Figure 2. In the following we assume that training images contain only patches generated by $\mathcal{P}_N$ and that these have been acquired at the maximum scale available. Thus, that test images are acquired at the same or at a lower magnification. This is not a restrictive assumption in applications where the maximum magnification level is known beforehand.

Our goal is to locate those regions in a test image where the structures do not conform $\mathcal{P}_N$. In particular, we want to detect anomalies even when the test images are acquired at a different scale than the training ones.



(a) Series A

(b) Series B

(c) Series C

(d) Series D

Fig. 2. Example of normal images acquired at different magnification levels. Patches extracted from images are perceptually similar, but their content is very different.

3. PROPOSED SOLUTION

For simplicity, in the following we illustrate the proposed solution assuming a single training image s is provided, even though multiple training images can be easily handled. Our solution is based on a dictionary D which is able to approximate any patch $\mathbf{s} \in \mathbb{R}^P$ taken from an anomaly-free image as

$$\mathbf{s} \approx D\mathbf{x}, \tag{1}$$

where the coefficients vector $\mathbf{x} \in \mathbb{R}^M$ is sparse, i.e. has few nonzero or non-negligible components. In the following we explain how to compute the dictionary $D \in \mathbb{R}^{P \times M}$ and the coefficient vector $\mathbf{x}$ for a given patch $\mathbf{s}$, then we illustrate the anomaly-detection method.

3.1. Multiscale Dictionary

The dictionary D is expected to provide suitable representations of normal images acquired at different scales. Various algorithms for learning multiscale dictionaries have been proposed in the literature [16, 17, 18, 19, 20]. However, to better investigate the role of the sparse coding and the choice of suitable indicators to achieve scale-invariance, we adopt a simple design of the multiscale dictionary.

Let us denote by s_σ the image s with support rescaled by a factor σ. Consider now a set of L scaling factors σ_i, $i \in \{1, \dots, L\}$ and construct, for each image s, a set of rescaled images s_{σ_i} to simulate normal data at different scales. Since we assume the scale of the training image is higher than in test images, we consider scaling factors $\sigma \leq 1$. For each rescaled images s_{σ_i} we extract a suitable set of patches and subtract their mean, then assemble them as the columns of matrix $T_i \in \mathbb{R}^{P \times N}$. The dictionary $D_i \in \mathbb{R}^{P \times M_i}$ corresponding to the scale σ_i is thus learned solving the Basis Pursuit DeNoising (BPDN) [21, 22] problem

$$D_i = \arg\min_{D,X} \frac{1}{2} \|T_i - DX\|_2^2 + \lambda \|X\|_1 \ , \tag{2}$$

where $\lambda > 0$ balances the reconstruction error $\|T_i - DX\|_2$ and the sparsity, assessed by the ℓ^1 norm of the coefficient $X \in \mathbb{R}^{M_i \times N}$.

The multiscale dictionary $D \in \mathbb{R}^{P \times M}$ representing the training image at multiple scales is constructed by collecting all the learned dictionaries D_i into a single matrix

$$D = [D_1 \mid D_2 \mid \cdots \mid D_L]. \tag{3}$$

In principle, the dictionaries D_i may have different number of columns M_i, however here we consider D_i having the same size $P \times M/L$.

3.2. Multiscale Sparse Coding

The sparse coding of each patch $\mathbf{s}$ with respect to the dictionary D corresponds to computing a sparse vector $\mathbf{x} \in \mathbb{R}^M$ of coefficients that properly approximate $\mathbf{s}$. Given the specific form of D, each coefficient vector has the form:

$$\mathbf{x} = \left[\mathbf{x}_1^T \ \mathbf{x}_2^T \ \cdots \ \mathbf{x}_L^T\right]^T, \tag{4}$$

where $\mathbf{x}_i$ is a column vector that collects the coefficients corresponding to dictionary D_i learned from the image at scale σ_i. For anomaly-detection purposes, it is not desirable to approximate a patch $\mathbf{s}$ by mixing atoms from different dictionaries D_i, as this mixture could possibly match anomalous structures. Therefore, we expect that in each sparse representation $\mathbf{x}$, only one, or possibly a few, groups $\mathbf{x}_i$ are active, i.e. $\mathbf{x}$ should be group sparse. This goal is achieved by formulating the sparse coding as a BPDN problem that includes an $\ell^{2,1}$-norm regularization term [23]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{s} - D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \xi \sum_{j=1}^{L} \|\mathbf{x}_j\|_2 \ . \tag{5}$$

When solving (5), the group-sparsity term penalizes representations involving atoms belonging to different dictionaries. Both (2) and (5) can be solved via Alternating Direction Method of Multipliers [24]. In our experiments, we use the MATLAB implementation provided in the SPORCO library [25].

3.3. Anomaly Detection

We use the learned dictionary D and the sparse coding procedure in (5) to define, for each patch $\mathbf{s}$, an indicator vector $\mathbf{g}(\mathbf{s})$ that assesses the extent to which $\mathbf{s}$ is consistent with $\mathcal{P}_N$.

As in [11], $\mathbf{g}(\mathbf{s})$ is the vector stacking all the summands of the cost function minimized during the sparse coding (5) of $\mathbf{s}$: the reconstruction error $\|\mathbf{s} - D\mathbf{x}\|_2$, the sparsity $\|\mathbf{x}\|_1$, and the group sparsity $\sum_i \|\mathbf{x}_i\|_2$. The group-sparsity term is used to assess the spread of significant coefficients among different scales of the dictionary atoms. Since normal patches are expected to involve atoms from one or few scales σ_i, this term is expected contribute to discriminate normal ad anomalous patches. Therefore, for each patch $\mathbf{s}$ we obtain an indicator vector having three components:

$$\mathbf{g}(\mathbf{s}) = \begin{bmatrix} \|\mathbf{s} - D\mathbf{x}\|_2 \\ \|\mathbf{x}\|_1 \\ \sum_i \|\mathbf{x}_i\|_2 \end{bmatrix}. \tag{6}$$

To detect whether a patch $\mathbf{s}$ is normal or anomalous, we build a confidence region $\mathcal{R}_\gamma$ from the values of $\mathbf{g}$ computed from the normal patches in the training set:

$$\mathcal{R}_\gamma = \left\{ \phi \in \mathbb{R}^3 \ : \ \sqrt{(\phi - \overline{\mathbf{g}})' \Sigma^{-1} (\phi - \overline{\mathbf{g}})} \leq \gamma \right\}, \tag{7}$$

where $\overline{\mathbf{g}}$ and Σ are the average and the sample covariance matrix of $\mathbf{g}$ computed on few normal patches, respectively. Then, $\mathbf{s}$ is labeled anomalous if $\mathbf{g}(\mathbf{s})$ falls outside the region $\mathcal{R}_\gamma$.

Since we analyze test images in a patch-wise manner and we consider overlapping patches, in practice we assign to each pixel one label (normal/anomalous) for each patch including it. To aggregate all these labels, we consider a majority-voting scheme: a pixel is considered anomalous when the majority of the patches containing that pixel are labeled anomalous.

4. EXPERIMENTS

We assess the performance of the proposed anomaly-detection algorithm on a dataset of SEM images acquired in a quality control application to monitor the industrial production of nanofibers. High variability affects this industrial process, introducing defects like the one shown in Figure 1(b). The detection of these anomalies is very important to determine whether the produced nanofibers conform to the desired standards and eventually adjust the production.

We consider 20 SEM images acquired at different magnification levels, and we group them into 4 series, each sharing the same magnification. An example of a normal image from each series is shown in Figure 2. It can be seen that images from Series A have been acquired at the highest scale (maximum magnification), thus they are used to train the proposed anomaly detector. In the test phase, we consider images containing anomalies from all these series, thus from different scales. The performance of the anomaly-detector can be assessed thanks to a binary mask that labels each pixel as normal or anomalous and that is provided for each image.

In this experiment we use patches having size 32×32 and, to speed up the dictionary learning (2) and the sparse coding (5) stages, we project patches in the Discrete Cosine Transform (DCT) domain and consider only the first 225 coefficients, ordered in a zig-zag fashion, so that $\mathbf{s}$ in (1) is a vector of 225 DCT coefficients.

We consider the following anomaly detection techniques:

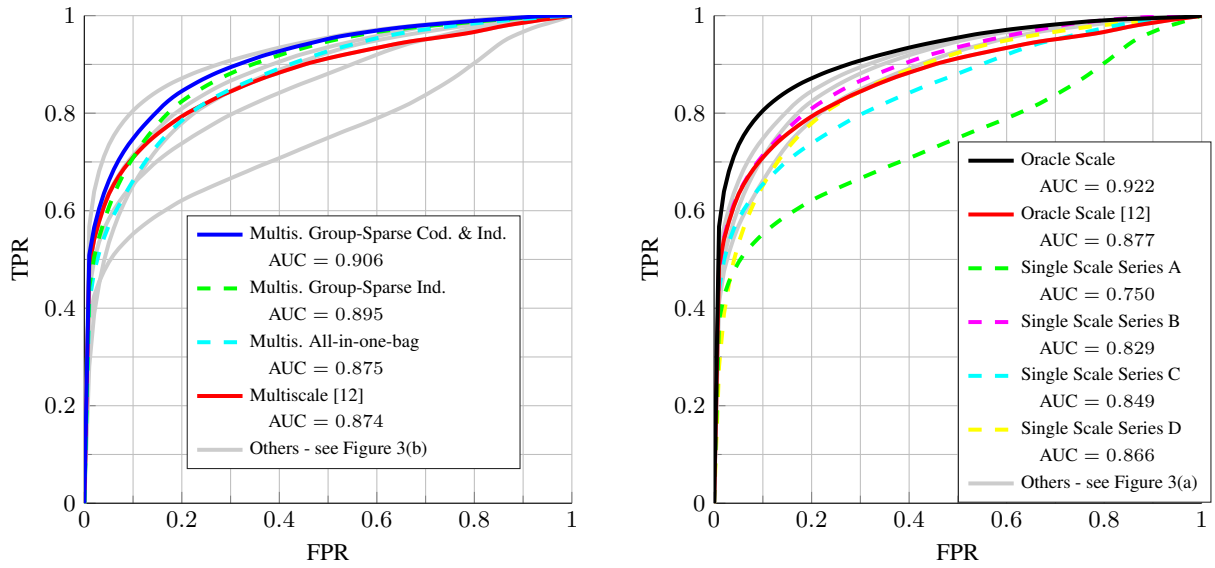
Multiscale Group-Sparse Coding and Indicator: this is our proposed solution, described in Section 3. The multiscale dictionary D is learned by scaling the training images of a factor $\sigma \in \{1, 0.75, 0.5\}$. This method differs from the following ones because it is multiscale in all its parts: dictionary learning, sparse coding and computed indicators.

Multiscale Group-Sparse Indicator: here we perform the sparse coding via the standard BPDN without the group sparsity term, i.e. we set $\xi = 0$ in (5). We use the same multiscale dictionary D (3), learned in the above solution. Then, we monitor the whole indicator vector (6) that includes also the group sparsity term, which is however ignored in the sparse coding.

Multiscale All-in-one-bag: this method is presented in [11], and here it uses the multiscale dictionary D. In practice, the group sparsity is neither taken into account in the sparse coding nor in the indicator vector.

Multiscale [12]: we use the same multiscale dictionary D (3) as in the above solutions, and we perform the sparse coding following the procedure described in [12], which embeds an anomaly detection procedure. This anomaly detector is also controlled by a threshold $\gamma > 0$.

Oracle Scale: we learn 4 different dictionaries, one for each series, and process test images using the dictionary learned from images acquired at the same (correct) scale. This is considered ideal solution since the correct scale is rarely known, exactly. This method is the same as [11], which does not consider a group sparsity term.



(a) Solutions based on multiscale dictionary D.

(b) Solutions based on single-scale dictionaries D_i.

Fig. 3. ROC curves for all methods presented in Section 4. (a) the ROC curves of solutions based on the multiscale dictionary D (3); (b) the ROC curves of solutions based on single scale dictionary D_i. There are 4 ROC curves for single scale dictionaries: one for each series of images used in the training phase. The area under the curve (AUC) of each ROC curve is reported in the legend.

Oracle Scale [12]: as in the Oracle Scale, we use the dictionary learned from training images at the same scale of the test images, but the sparse coding and the anomaly detection are performed as in [12].

Single Scale: we learn 4 dictionaries, one from each series, and use each of them to detect anomalies in images from all the series. Anomalies are detected as in [11], which is not multiscale. Obviously, the performance of this solution might vary according to the series used for training.

To assess the performance of the considered solutions we consider the following figures of merit:

- **FPR**, the False Positive Rate, i.e. the percentage of normal pixels detected as anomalous.
- **TPR**, the True Positive Rate, i.e. the percentage of anomalous pixels correctly detected.

These figures of merit depend on the threshold γ which defines the confidence region $\mathcal{R}_\gamma$ in (7) or the promptness of the anomaly detector [12]. Therefore, different methods have to be compared by means of the Receiver Operating Characteristic (ROC) curves, which are computed by varying the value of γ in a suitable range. Figure 3 shows the ROC curves averaged over all the test images: in Figure 3(a) we report the results of the solutions based on the multiscale dictionary D (3), while Figure 3(b) shows the ROC curves of the solutions that exploit single-scale dictionaries.

The ROC curves in Figure 3 and the corresponding AUC values indicate that using multiscale dictionaries is beneficial, as these provide better performance than single scale dictionaries in all the considered solutions. As expected, the *Oracle Scale* solution outperforms all the others, since test images are analyzed by a dictionary that was learned on normal images acquired at the same scale. However, these settings might not be realistic in all the practical applications. The *Multiscale Group-Sparse Coding and Indicator* achieves the best performance. In particular, it outperforms *Multiscale All-in-one-bag* and *Multiscale [12]*, demonstrating that simple sparsity with respect to a multiscale dictionary is not enough to handle test images at a different scales, and that the group sparsity term is instead necessary in the design of the anomaly detector. Moreover, the comparison between *Multiscale Group-Sparse Coding and Indicator* and *Multiscale Group-Sparse Indicator* confirms that is not enough to measure the group sparsity in the indicator vector, but this has to be taken into account also during the sparse coding.

5. CONCLUSIONS

We present an anomaly-detection algorithm that is able to correctly analyze test images that have been acquired at scales that are different from those of training images. Our anomaly-detector uses a multiscale dictionary that aggregates atoms learned from synthetically resized normal images, and performs sparse coding by enforcing the group sparsity of the representations. This regularization term turns to be essential to achieve superior anomaly-detection performance. We test our solution on a dataset of SEM images acquired to monitor the industrial production of nanofibers, and demonstrate it can effectively handle changes in magnification level that typically occurs in industrial/medical imaging applications.

Acknowledgments

This work was supported by the Academy of Finland (project no. 252547, Academy Research Fellow 2011-2016), by the U.S. Department of Energy through the LANL/LDRD Program, and by UC Lab Fees Research grant 12-LR-236660. The authors would like to thank the Institute of Applied Mathematics and Information Technology (IMATI), Milan, Italy, and the Institute of Science and Technology for Ceramics (ISTEC), Faenza, Italy, for providing them the SEM images.

6. REFERENCES

[1] Wee E. Teo and Seeram Ramakrishna, "A review on electrospinning design and nanofibre assemblies," *Nanotechnology*, vol. 17, no. 14, pp. R89, 2006.

[2] Leon M. Bellan and Harold G. Craighead, "Applications of controlled electrospinning systems," *Polymers for Advanced Technologies*, vol. 22, no. 3, pp. 304–309, 2011.

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 15, 2009.

[4] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[5] Markos Markou and Sameer Singh, "Novelty detection: a review - part 1: statistical approaches," *Signal processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[6] Markos Markou and Sameer Singh, "Novelty detection: a review - part 2: neural network based approaches," *Signal processing*, vol. 83, no. 12, pp. 2499–2521, 2003.

[7] Bernhard Schölkopf, Robert C. Williamson, Alex J. Smola, John Shawe-Taylor, and John C. Platt, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems*, 1999, pp. 582–588.

[8] Lionel Tarassenko, Paul Hayton, Nicholas Cerneaz, and Michael Brady, "Novelty detection for the identification of masses in mammograms," in *Proceedings of International Conference on Artificial Neural Networks (ICANN)*, 1995, pp. 442–447.

[9] Gal Mishne and Israel Cohen, "Multiscale anomaly detection using diffusion maps," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 111–123, 2013.

[10] Maria Zontak and Israel Cohen, "Defect detection in patterned wafers using anisotropic kernels," *Machine Vision and Applications*, vol. 21, no. 2, pp. 129–141, 2010.

[11] Giacomo Boracchi, Diego Carrera, and Brendt Wohlberg, "Novelty detection in images by sparse representations," in *Proceedings of IEEE Symposium on Intelligent Embedded Systems (IES)*, 2014, pp. 47–54.

[12] Amir Adler, Michael Elad, Yacov Hel-Or, and Ehud Rivlin, "Sparse coding with anomaly detection," in *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013, pp. 1–6.

[13] Matthew D. Zeiler, Dilip Krishnan, Graham W. Taylor, and Robert Fergus, "Deconvolutional networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2528–2535.

[14] Brendt Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, Jan. 2016.

[15] Diego Carrera, Giacomo Boracchi, Alessandro Foi, and Brendt Wohlberg, "Detecting anomalous structures by convolutional sparse models," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland, July 2015, pp. 1–8.

[16] Julien Mairal, Guillermo Sapiro, and Michael Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[17] Rémi Gribonval, "Fast matching pursuit with a multiscale dictionary of gaussian chirps," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 994–1001, 2001.

[18] Bruno A. Olshausen, Phil Sallee, and Michael S. Lewicki, "Learning sparse image codes using a wavelet pyramid architecture," in *Advances in Neural Information Processing Systems*, 2001, pp. 887–893.

[19] Phil Sallee and Bruno A. Olshausen, "Learning sparse multiscale image representations," in *Advances in Neural Information Processing Systems*, 2002, pp. 1327–1334.

[20] James Michael Hughes, Daniel N. Rockmore, and Yang Wang, "Bayesian learning of sparse multiscale image representations," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4972–4983, 2013.

[21] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[22] Ivana Tošić and Pascal Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[23] Ming Yuan and Yi Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[24] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[25] Brendt Wohlberg, "SParse Optimization Research COde (SPORCO)," Matlab library available from `http://math.lanl.gov/~brendt/Software/SPORCO/`, 2015, Version 0.0.3.