DOI: https://doi.org/10.1109/ICIP.2018.8451059

REDUCING ANOMALY DETECTION IN IMAGES TO DETECTION IN NOISE

Axel Davy[†⋆], *Thibaud Ehret*[†⋆], *Jean-Michel Morel*[†], *Mauricio Delbracio*[§ *]

[†]CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France
[§]IIE, Universidad de la República, Uruguay

ABSTRACT

Anomaly detectors address the difficult problem of detecting automatically exceptions in an arbitrary background image. Detection methods have been proposed by the thousands because each problem requires a different background model. By analyzing the existing approaches, we show that the problem can be reduced to detecting anomalies in residual images (extracted from the target image) in which noise and anomalies prevail. Hence, the general and impossible background modeling problem is replaced by simpler noise modeling, and allows the calculation of rigorous thresholds based on the a contrario detection theory. Our approach is therefore unsupervised and works on arbitrary images.

Index Terms— Anomaly detection, Saliency, Self-similarity

1. INTRODUCTION

Anomalies are image regions not conforming with the rest of the image. Detecting them is a challenging image analysis problem, as there seems to be no straightforward definition of what is (ab)normal for a given image.

Anomalies in images can be high-level or low-level outliers. High-level anomalies are related to the semantic information presented in the scene. For example, human observers immediately detect a person inappropriately dressed for a given social event. In this work, we focus on the problem of detecting anomalies due to low or mid level rare local patterns present in images. This is an important problem in many industrial, medical or biological applications.

We introduce in this paper an unsupervised method for detecting anomalies in an arbitrary image. The method does not rely on a training dataset of normal or abnormal images, neither on any other prior knowledge about the image statistics. It directly detects anomalies with respect to residual images estimated solely from the image itself. We only use a generic, qualitative background image model: we assume that anything that repeats in an image is *not* an anomaly. In a nutshell, our method removes from the image its self-similar content (considered as being normal). The residual is modeled as colored Gaussian noise, but still contains the anomalies according to their definition: they do not repeat.

Detecting anomalies in noise is far easier and can be made rigorous and unsupervised by the *a-contrario* theory [1] which is a probabilistic formalization of the *non-accidentalness* principle [2]. The *a-contrario* framework has produced impressive results in many different detection or estimation computer vision tasks, such as, segment detection [3], spots detection [4], vanishing points detection [5], mirror-symmetry detection [6], among others. The fundamental property of the *a-contrario* theory is that it provides a way for automatically computing detection thresholds that yield a control on the number of false alarms (NFA). It favorably replaces the usual p-value when multiple testing is involved. It follows that not only one can detect anomalies in arbitrary images without complex modeling, but in addition the anomalies are associated an NFA which is often very small and therefore offers a strong guarantee of the validity of the detection. We shall show detections performed directly on the image residual, or alternatively on residuals extracted from dense low and mid-level features of the VGG neural net [7].

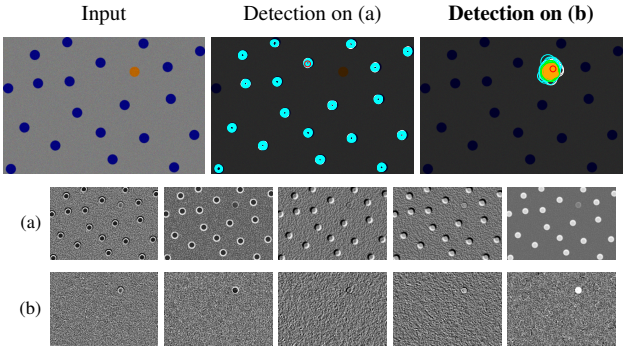

Fig. 1. Image anomalies are successfully detected by removing all self-similar content and then looking for structure in the residual noise. Top row: left, an image with a color anomaly (the red dot); middle, detections obtained from top five principal components of CNN features shown in (a); right, detections on features shown in (b), obtained after removing the self-similar content. Cyan corresponds to good detection and orange extremely salient detection.

The paper is organized as follows. Section 2 discusses previous work while Section 3 explains the proposed method and its implementation. Section 4 presents results of the proposed method on real/synthetic data, and a comparison to other state-of-the-art anomaly detectors. We finally close in Section 5.

2. RELATED WORK

The 2009 review [8] examining about 400 papers on anomaly detection considered allegedly all existing techniques and application fields. It is fairly well completed by the more recent [9] review. These reviews agree that classification techniques like SVM can be discarded, because anomalies are generally not observed in sufficient number and lack statistical coherence. There are exceptions

*Work supported by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, ONR grant N00014-17-1-2552, CNES MISS project, Agencia Nacional de Investigación e Innovación (ANII, Uruguay) grant FCE_1_2017_135458, DGA Astrid ANR-17-ASTR-0013-01, DGA ANR-16-DEFA-0004-01, Programme ECOS Sud – UdelaR - Paris Descartes U17E04, and MENRT.

⋆ Both authors contributed equally to this work

like the recent method [10] which defines anomalies as rare events that cannot be learned, but after estimating a background density model, the right detection thresholds are nevertheless learned from anomalies. A broad related literature exists on saliency measures, for which learning from average fixation maps by humans is possible [11]. Saliency detectors try to mimic the human visual perception and in general introduce semantic prior knowledge (e.g., face detectors). This approach works particularly well with neural networks trained on a base of detect/non-detect with ground truth obtained by for example, gaze trackers[12].

Anomaly detection has been generally handled as a "one class" classification problem. In [13] authors concluded that most research on anomaly detection was driven by modeling background data distributions, to estimate the probability that test data do not belong to such distributions [4, 14, 15, 16]. Autoencoders neural networks can be used to model background [17, 18]. The general idea is to compute the norm between the input and a reconstruction of the input. Another successful background based method is the detection of anomalies in periodic patterns of textile [19, 20]. In [21, 22], center surround detectors based on color, orientation and intensity filters are combined to produce a final saliency map. Detection in image and video is also done in [23] with center-surround saliency detectors which stem from [24] adopting similar image features. In [14], the main idea is to estimate the probability of a region conditioned on the surroundings. A more recent non parametric trend is to learn a sparse dictionary representing the background (i.e., *normality*) and to characterize outliers by their non-sparsity [25, 26, 27, 28, 29].

The self-similarity principle has been successfully used in many different applications [30, 31]. The basic assumption of this generic background model, is that in normal data, features are densely clustered. Anomalies instead occur far from their closest neighbors. This idea is implemented by clustering (anomalies being detected as far away from the centroid of their own cluster), or by simple rarity measurements based on nearest neighbor search (NN) [32, 33, 34].

Background probabilistic modeling is powerful when images belong to a restricted class of homogeneous objects, like textiles. But, regrettably, this method is nearly impossible to apply on generic images. Similarly, background reconstruction models based on CNNs are restrictive and do not rely on provable detection thresholds. Center-surround contrast methods are successful for saliency enhancement, but lack a formal detection mechanism. Being universal, the sparsity and the self-similarity models are tempting and thriving. But again, they lack a rigorous detection mechanism, because they work on a feature space that is not easily modeled.

We propose to benefit of the above methods while avoiding their mentioned limitations. To this aim, we do construct a probabilistic background model, but it is applied to a new feature image that we call the *residual*. This residual is obtained by computing the difference between a self-similar version of the target image and the target itself. Being not self-similar, this background is akin to a colored noise. Hence a hypothesis test can be applied, and more precisely multiple hypothesis testing (also called *a contrario* method), as proposed in [4]. In that way, we present a general and simple method that is universal and detects anomalies by a rigorous threshold. It does not require learning, and it is easily made multiscale.

3. METHOD

Our method is built on two main blocks: a removal of the self-similar image component, and a simple statistical detection test on the residual based on the *a contrario* framework.

Algorithm 1 Computation of the unstructured residual

Require: Multichannel Image u, n the number of nearest neighbors
Ensure: Model $\hat{u}$ of u based on $\mathcal{D}$, residual $r(u) = \hat{u} - u$.

1: **for all** Multichannel patch P of u **do**
2: Compute n near.neigh. $\{P_i\}$ of P **(outside square region)**.
3: Reconstruct the patch (using (1))
4: **for all** pixels j in u **do**
5: $\hat{u}(j) = \frac{\sum_{i \in \{s | j \in W_s, s \in [\![1,N]\!]\}} \hat{P}_i(j)}{\#\{s | j \in W_s, s \in [\![1,N]\!]\}}$

Notation convention. W_s : set of pixels in the patch centered at s. $\hat{P}_i(j)$: value at pixel j of the reconstructed patch centered at i.

3.1. Construction of the residual image

The proposed self-similarity based background subtraction is inspired from patch-based non-local denoising algorithms, where the estimate is done from a set of similar patches [31]. This search is generally performed locally around each patch [35, 31] to keep computational cost low and to avoid noise overfitting. The main difference with non-local denoisers is that we *forbid* local comparisons. The nearest neighbor search is performed *outside a square region surrounding each query patch*. This square region is defined as the union of all the patches intersecting the query patch. Otherwise any anomaly with some internal structure might be considered a valid structure. What matters is that the event represented by the anomaly is unique, and this is checked away from it.

For each patch P in the image the n most similar patches denoted by P_i are searched and averaged to give a self-similar estimate,

$$\hat{P} = \frac{1}{Z} \sum_{i=1}^{n} \exp\left(-\frac{\|P - P_i\|_2^2}{h^2}\right) P_i \tag{1}$$

where $Z = \sum_{i=1}^{n} \exp\left(-\frac{\|P-P_i\|_2^2}{h^2}\right)$ is a normalizing constant, and h is a parameter.

Since each pixel belongs to several different patches, they will therefore receive several distinct estimates that can be averaged. Algorithm 1 gives a generic pseudocode for this process, which ends with the generation of a residual image $r(u)$ allegedly containing only noise and the anomalies (see Figure 1). The intuition is that it is much easier to detect anomalies in $r(u)$ than in u.

3.2. Statistical detection by the *a contrario* approach

Our goal is to detect structure in the residual image $r(u) = \hat{u} - u$. We are in a much better situation modeling $r(u)$ than u. Indeed, contrarily to u, $r(u)$ is by construction *unstructured* and akin to a colored noise (as illustrated in Fig. 1). In what follows we assume that $r(u)$ is a spatial stationary random process and follow [4], who proposed automatic detection thresholds in any colored Gaussian noise.

Given a set of random variables $(X_i)_{i \in [|1,N|]}$ a function f is called an NFA if it guarantees a bound on the expectation of its number of false alarms under the null-hypothesis, namely, $\forall \epsilon > 0, \mathbb{E}[\#\{i, f(i, X_i) \leq \epsilon\}] \leq \epsilon$. In other words, thresholding all the $f(i, X_i)$ by ϵ should give up to ϵ false alarms when $(X_i)_{i \in [|1,N|]}$ verifies the null-hypothesis. In our case, we consider

$$f(i, \mathbf{x}) = N\mathbb{P}(|X_i| \geq |x_i|), \tag{2}$$

Where i index among the N executed tests (detailed below), X_i is a random variable distributed as the residual at position i, and x_i the actual measured value (pixel or feature value) at position i. The null-

hypothesis is that the residual, represented by $(X_i)_{i\in[|1,N|]}$, verifies that each X_i follows a standard normal distribution. Independence is not required.

Residual distribution. In practice the distribution of the residual $r(u)$ is not necessarily Gaussian. A careful study of the residual distribution lead us to consider that it follows a generalized Gaussian distribution (GCD). We approximately estimate the GCD parameters, and then apply a non-linear mapping to make it normally distributed.

Choice of NFA. The choice of the NFA given in (2) enables to detect anomalies in both tails of the Gaussian distribution (i.e., very bright or very dark spots). To detect anomalies of all sizes, the detection is carried out independently at N_{scales} scales computed from the residual at the original resolution (by Gaussian subsampling of factor two). Let us denote by Ω_s the set of pixels in the residual image at scale s having N_{feat} number of features. When working with colored noise, Grosjean and Moisan [4] propose to convolve the noise with a measure kernel to detect spots of a certain size. This corresponds to the generation of new image features $\bar{r}(u) = r(u) * K$, where K is a disk of a given radius. This idea is used in our framework, where the residual is convolved with kernels of small sizes. Since we apply the detection at all dyadic scales, the tested radii are limited to a small set of N_{kernel} values (1,2 to 3) at each scale. Because the residual is assumed to be a stationary Gaussian field, the result after filtering is also Gaussian. The variance is estimated and the filtered residuals are normalized to have unit variance. This is the input to the NFA (2) computation (i.e., $\mathbf{x}_i$). Thus, the inputs to the detection phase are multi-channel images of different scales, where each pixel channel, representing a given feature, follows a standard normal distribution.

Then, the number of tests is $N = N_{\text{kernel}} \cdot N_{\text{feat}} \cdot \sum_{i=0}^{N_{\text{scales}}-1} |\Omega_s|$.

3.3. Choice of the image features

Anomaly detectors work either directly on image pixels or on some feature space but the detection in the residual, which is akin to unstructured noise, is fairly independent of the choice of the features. We used with equal success the raw image color pixels, or some intermediate feature representation extracted from the VGG convolutional neural network [7]. To compress the dynamical range of the feature space we apply a square root function to the network features.

In order to reduce the feature space dimension, we compute the principal components (PCA) and keep only the first five. This is done per input image independently.

Parameters. The main method parameter is the number of allowed false alarms in the statistical test. In all presented experiments, we set NFA=10^{-2}. Hence, an anomaly is detected at pixel $\mathbf{x}$ in channel i iff the NFA function $f(i,\mathbf{x})$ is below $\epsilon = 10^{-2}$. This implies a (theoretical) expectation of less than 10^{-2} "casual" detection per image under the null hypothesis that the residual image is noise. Obviously the lower the NFA the better. Most anomalies have a much lower NFA. For the basic method working on image pixels we used two disks of radius one and two, while for the neural network features, we add a third disk of radius three. The number of scales is set to $N_{scale} = 4$ in all tests. The patch size in Alg. 1 is $8\times8\times3$ for the pixels variant, while when using neural nets features, we use a patch size of $5\times5\times5$. The number of nearest patches is always set to $n = 16$, and $h = 10$. Results presented herein use the outputs from VGG-19 layers `conv1_1`, `conv2_1` and `conv3_1`.

4. EXPERIMENTS

In absence of a valid test image database for anomalies, we used the most common images proposed in the literature (see Fig. 2) and we adopted the following comparison methodology, that was applied to our method and to other four state-of-the-art ones for comparison:

a) *Sanity check:* verifying that for toy examples proposed in the literature the sole detection is the anomaly;

b) *Theoretical sanity check:* verify the *a contrario* principle: "no detection in white noise"

c) *Classic challenging images:* we verify the detector power on classic challenging images of the literature: side scan sonar, textile, mammography and natural images. In the case of the mammography where one paper computed an NFA, we verify crucially that by computing the NFA on the residual instead of the image, we gain a huge factor, the NFA being divided by eleven orders of magnitude.

We tested our proposed anomaly detector on two different input image representations: the basic one, `pixels`, directly applies the anomaly detection procedure to the residuals obtained from the color channels, and three different variants using as input features extracted at different levels from the VGG network [7], namely, very low level (`conv1_1`), low level (`conv2_1`), and medium level (`conv3_1`) features. As we shall check the four detections are similar and can be fused by a mere pixel union of all detections.

Existent anomaly detectors are often tuned for specific applications, which probably explains the poor code availability. We compared to Mishne and Cohen [36], a state-of-the-art anomaly detector with available code, to the salient object detector DRFI [37] (which is state-of-the-art according to [40]), and to the state-of-the-art human gaze predictor SALICON [12]. We also compared to the Itti *et al.* salient object detector [21], which works reasonably well for anomaly detection. All methods produce saliency maps where anomalies have the highest score. Anomalies for Mishne and Cohen are red-colored, while the other methods don't have a threshold for anomalies. More results are available in the supplementary materials.

Synthetic images. The proposed method performs well on synthetic examples as shown in Figure 2). Some weak false detections are found when using as input features extracted at different layers of the VGG net. All the other compared methods miss some detections. SALICON successfully detects the anomalous density on the fourth example but misses several anomalies in others or introduces numerous wrong detections. Itti *et al*. method successfully detects the anomalous color structure in the first example, but fails to detect the other ones. Mishne and Cohen and DRFI methods do not perform well on any of the five synthetic examples.

Real images. The comparison on real images is more intricate and requires looking in detail to find out whether detections make sense (Figure 2). In the garage door (fourth row), there are two detections that stand out (lens flare and red sign), some others – less visible – can be found (door scratches or holes in the brick wall). For our method, the main detections are present in all the variants. There are also specific anomalies that can be detected only at a given layer of the neural network. For example, `conv1_1` detects the holes in the brick wall and the gap between the garage door and the wall, in addition to the ones detected with `pixels` input. The variants `conv2_1` and `conv3_1` detect a missing part of a brick in the wall. Saliency methods detect the red sign but not the lens flare. Mishne and Cohen one only detects the garage door gap. The second real example is a man walking in front of some trees. Our method detects the man with `pixels` and `conv1_1`. DRFI and SALICON detect

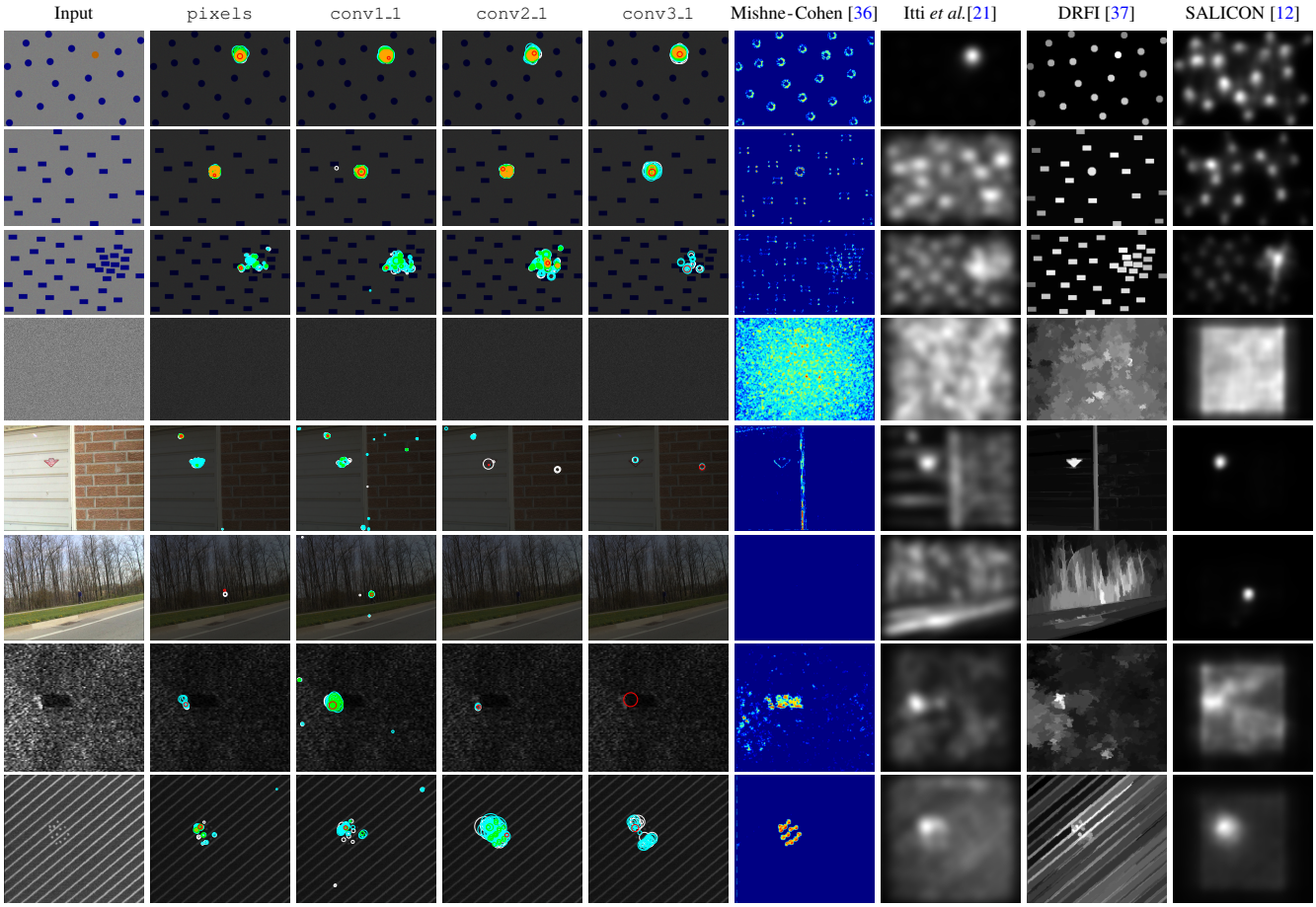

Fig. 2. Detection results on synthetic (top four rows) and real (bottom four rows) images. Detections represented by circles, with radius proportional to detected scale and color to detection strength (NFA). White: weak detection - NFA $\in [10^{-3}, 10^{-2}]$, cyan: mild detection - NFA $\in [10^{-8}, 10^{-3}]$, green: strong detection - NFA $\in [10^{-21}, 10^{-8}]$, and orange: very strong detection - NFA $\leq 10^{-21}$. Red: detection with lowest NFA. Examples in rows 5th and 6th are from the Toronto dataset [38] while 7th and 8th from [36] and [39] respectively.

the man while Mishne and Cohen and Itti *et al*. do not. The third real example is a radar image showing a mine, while the last example is a defect in a periodic textile. All methods detect the anomalies, with more or less precision. Note that the detection in the top right corner for both `pixels` and `conv1_1` (and only these) correspond to a defect inside the periodic pattern.

Comparison to the *a contrario* method of Grosjean and Moisan [4]. This *a contrario* method is designed to detect spots in colored noise textures, and was applied to the detection of tumors in mammographies. This detection algorithm is the only other one computing NFAs, and we can directly compare them to ours. The detection results on a real mammography (having a tumor) are shown in Figure 3. With our method the tumor is detected with a much significant NFA (NFA of 10^{-12} whereas in [4] NFA of 0.15). Our self-similar anomaly detection method shows fewer false detections, actually corresponding to rare events like the crossings of arterials.

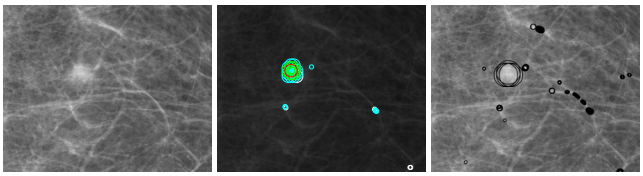

Fig. 3. The region represented by the large white spot in the left image is a tumor. The proposed self-similarity anomaly detector successfully detects the tumor with a much significant NFA than the one from Grosjean and Moisan [4] (an NFA of 10^{-12} versus their reported NFA of 0.15), while making fewer false detections.

5. CONCLUSION

We have shown that anomalies are easier detected on the residual image, computed by removing the self-similar component, and then performing hypothesis testing. It is reassuring to see that our method finds all anomalies proposed in the literature with very low NFA. In addition, we have experimentally shown that the method verifies the non-accidentalness principle: no anomalies are detected in white noise. We plan to build a database of test images with anomalies to run extensive validation and comparison. We also plan to extend the method to videos, by analyzing anomalies in the motion field.

6. REFERENCES

[1] A. Desolneux, L. Moisan, and J.-M. Morel, *From gestalt theory to image analysis: a probabilistic approach*, vol. 34, Springer Science & Business Media, 2007.

[2] D. Lowe, *Perceptual organization and visual recognition*, Kluwer Academic Publishers, 1985.

[3] R. Grompone Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, 2010.

[4] B. Grosjean and L. Moisan, "A-contrario detectability of spots in textured backgrounds," *J. Math. Imaging Vis.*, vol. 33, no. 3, pp. 313–337, 2009.

[5] J. Lezama, R. Grompone von Gioi, G. Randall, and J.-M. Morel, "Finding vanishing points via point alignments in image primal and dual domains," in *CVPR*, 2014.

[6] V. Patraucean, R. Grompone von Gioi, and M. Ovsjanikov, "Detection of mirror-symmetric image patches," in *CVPR*, 2013.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15, 2009.

[9] M. Pimentel, D. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[10] X. Ding, Y. Li, A. Belatreche, and L. Maguire, "An experimental evaluation of novelty detection methods," *Neurocomputing*, vol. 135, pp. 313–327, 2014.

[11] H. Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in *Scandinavian Conf. on Ima. Anal.*, 2011.

[12] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *ICCV*, 2015.

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

[14] T. Honda and S. Nayar, "Finding "anomalies" in an arbitrary image," in *ICCV*, 2001.

[15] A. Goldman and I. Cohen, "Anomaly detection based on an iterative local statistics approach," *Signal Processing*, vol. 84, no. 7, pp. 1225–1229, 2004.

[16] D. Aiger and H. Talbot, "The phase only transform for unsupervised surface defect detection," in *CVPR*, 2010.

[17] J. An, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability," *CoRR*, 2016.

[18] T. Schlegl, P. Seeböck, S. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," in *IPMI*, 2017.

[19] D. Tsai and T. Huang, "Automated surface inspection for statistical textures," *Image Vis. Comput.*, vol. 21, no. 4, pp. 307–323, 2003.

[20] D. Perng, S. Chen, and Y. Chang, "A novel internal thread defect auto-inspection system," *Int. J. Adv. Manuf. Tech.*, vol. 47, no. 5-8, pp. 731–743, 2010.

[21] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[22] N. Murray, M. Vanrell, X. Otazu, and C. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *CVPR*, 2011.

[23] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *NIPS*, 2008.

[24] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10, pp. 1489–1506, 2000.

[25] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?," in *CVPR*, 2013.

[26] G; Boracchi, D. Carrera, and B. Wohlberg, "Novelty detection in images by sparse representations," in *IES*, 2014.

[27] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *CVPR*, 2012.

[28] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," *J. Signal Process. Syst.*, vol. 79, no. 2, pp. 179–188, 2015.

[29] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, "Detecting anomalous structures by convolutional sparse models," in *IJCNN*, 2015.

[30] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *ICCV*, 1999.

[31] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *CVPR*, 2005.

[32] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *IJCV*, vol. 74, no. 1, pp. 17–31, 2007.

[33] H. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of vision*, vol. 9, no. 12, pp. 15–15, 2009.

[34] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[36] G. Mishne and I. Cohen, "Multiscale anomaly detection using diffusion maps," *IEEE J. Sel. Topics Signal Process*, vol. 7, no. 1, pp. 111–123, 2013.

[37] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *CVPR*, 2013.

[38] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2006.

[39] D.-M. Tsai and C.-Y. Hsieh, "Automated surface inspection for directional textures," *Image Vis. Comput.*, vol. 18, no. 1, pp. 49–62, 1999.

[40] A. Borji, M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, 2015.

SUPPLEMENTARY MATERIALS FOR: REDUCING ANOMALY DETECTION IN IMAGES TO ANOMALY DETECTION IN NOISE

Axel Davy[†]*, Thibaud Ehret*[†]*, Mauricio Delbracio*[§]*, Jean-Michel Morel*[†]

[†]CMLA, ENS-Cachan, France
[§]IIE, Universidad de la República, Uruguay

1. ADDITIONAL EXPERIMENTS

In this supplementary section, we show the result of our algorithm on new inputs, we illustrate some parts of the algorithm with intermediate results, and show more results with different Neural Networks layers as the source of dense input features. It is advised to zoom-in on the input figures to see the small anomalies.

1.1. Additional results

Figures 1, 2 and 5 show additional results to highlight the performance of our algorithm on a wide variety of examples, and the comparison with other algorithms. A placeholder (white image) is shown when the a compared algorithm failed to produce a result. Figure 9 shows a larger version of one of these examples with a detection of a tank hidden in a cluttered landscape.

1.2. The residual highlights what is not self-similar

Figure 6 shows an example of the residual using the `pixels` version of the algorithm. In the self-similar regions of the image the residual has very little structure thus can be easily assimilated to noise. Non-similar details such as a couple of small white strings appear perfectly in the residual. They are the only structures in this "noise" and are what needs to be detected (*cf* Figure 5 for the detections). Figures 3 and 4 show the residuals for `pixels` at different scales for images of Figures 1, 2. Figure 8 highlights the residuals for `conv_21` on the input of Figure 7.

1.3. What does the algorithm see?

On Figure 7, we highlight one particular detection shown on Figure 5. While at first sight, the detection could look like a false detection, the input image after a zoom-out and an affine transform reveals a stain at the detection location. The stain is even more strongly detected when using a bigger image (Figure 7 shows only the result on a crop).

Many detections are made at small scale and can only be verified by looking carefully at the zoomed-in images. The NFA threshold might be a valuable parameter to tune (application dependent). For example, one may want to ignore weak detections (by using a larger NFA threshold).

1.4. Impact of the choice of the layer of the pre-trained Neural Network

Figure 10 displays the detections for several different layers of VGG [9] as input features of the algorithm. The two most salient anomalies, the door sign and the lens flare, are detected with most Neural Network layers tested. However the low-level features of `conv1_1` are required to detect small sized anomalies like small holes in the cement, while the brick with a side stained by cement on the right of the image is detected by higher levels features of `conv2_1` and `conv3_1`, and the scratch on the door is detected with layers `conv3_3` and `conv3_4`.

2. REFERENCES

[1] Gal Mishne and Israel Cohen, "Multiscale anomaly detection using diffusion maps," *IEEE J. Sel. Topics Signal Process*, vol. 7, no. 1, pp. 111–123, 2013.

[2] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, and Shipeng Li, "Salient object detection: A discriminative regional feature integration approach," in *CVPR*, 2013.

[3] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[4] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *ICCV*, 2015.

[5] Neil Bruce and John Tsotsos, "Saliency based on information maximization," in *NIPS*, 2006.

[6] Laurent Itti and Christof Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10, pp. 1489–1506, 2000.

[7] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba, "Mit saliency benchmark," .

[8] Bénédicte Grosjean and Lionel Moisan, "A-contrario detectability of spots in textured backgrounds," *J. Math. Imaging Vis.*, vol. 33, no. 3, pp. 313–337, 2009.

[9] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

Work supported by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, ONR grant N00014-17-1-2552, CNES MISS project, DGA Astrid ANR-17-ASTR-0013-01, DGA ANR-16-DEFA-0004-01, and MENRT.

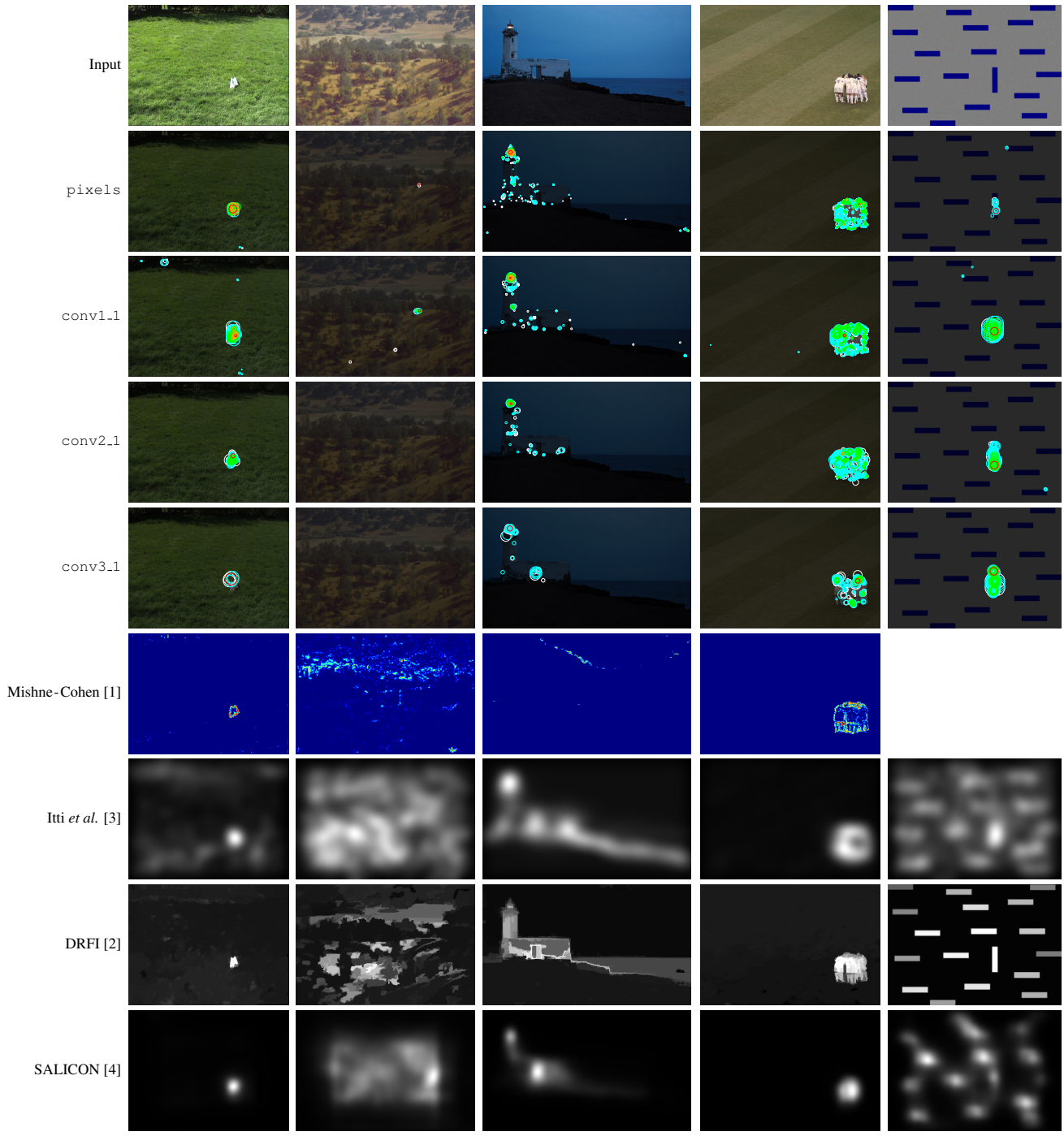

Fig. 1. Detection results of our method when using as image representation (input) directly the image pixels (`pixels`) or the activation maps of the VGG neural Network at different layers (`conv1_1`, `conv2_1`, `conv3_1`) and a comparison to [1], [3], [2] and [4] on multiple examples. Each detection is represented by a circle, where the circle radius represents the detection scale and the color the strength of the detection (NFA). White corresponds to a weak detection (NFA test value between 10^{-3} and 10^{-2}), cyan to a good detection (NFA between 10^{-8} and 10^{-3}), green to a very strong detection (NFA between 10^{-21} and 10^{-8}) and orange to an extremely salient detection (NFA smaller than 10^{-21}). Red corresponds to the detection with lowest NFA. The first image (left column) is part of the Toronto dataset [5]. The second image is extracted from [6]. The third and fourth images come from the dataset MIT300 [7].

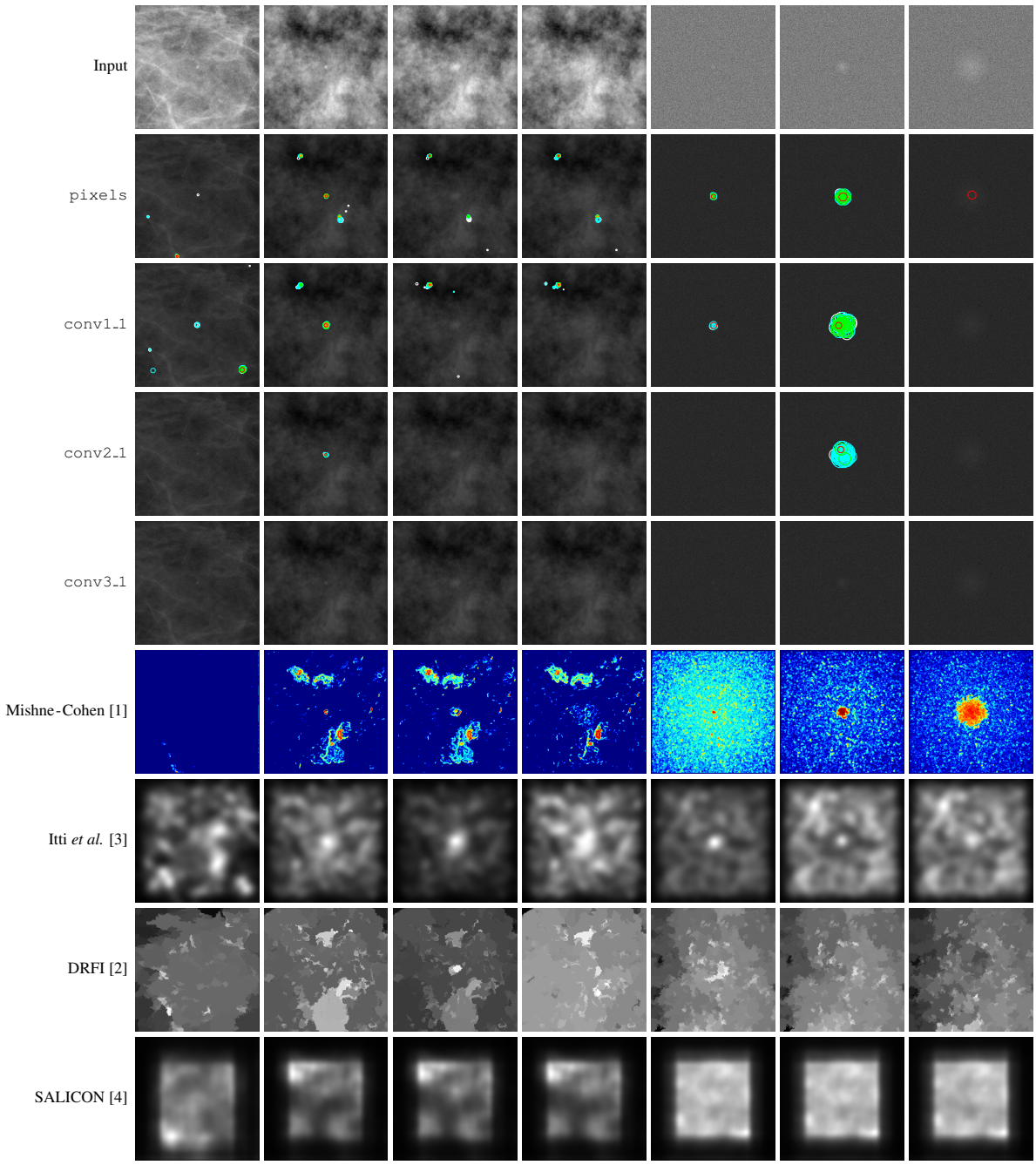

Fig. 2. Detection results of our method when using as image representation directly the image pixels (`pixels`) or the activation maps of the VGG neural Network at different layers (`conv1_1`, `conv2_1`, `conv3_1`) and a comparison to [1], [3], [2] and [4] on multiple examples. Each detection is represented by a circle, where the circle radius represents the detection scale and the color the strength of the detection (NFA). White corresponds to a weak detection (NFA test value between 10^{-3} and 10^{-2}), cyan to a good detection (NFA between 10^{-8} and 10^{-3}), green to a very strong detection (NFA between 10^{-21} and 10^{-8}) and orange to an extremely salient detection (NFA smaller than 10^{-21}). Red corresponds to the detection with lowest NFA. The images are extracted from [8].

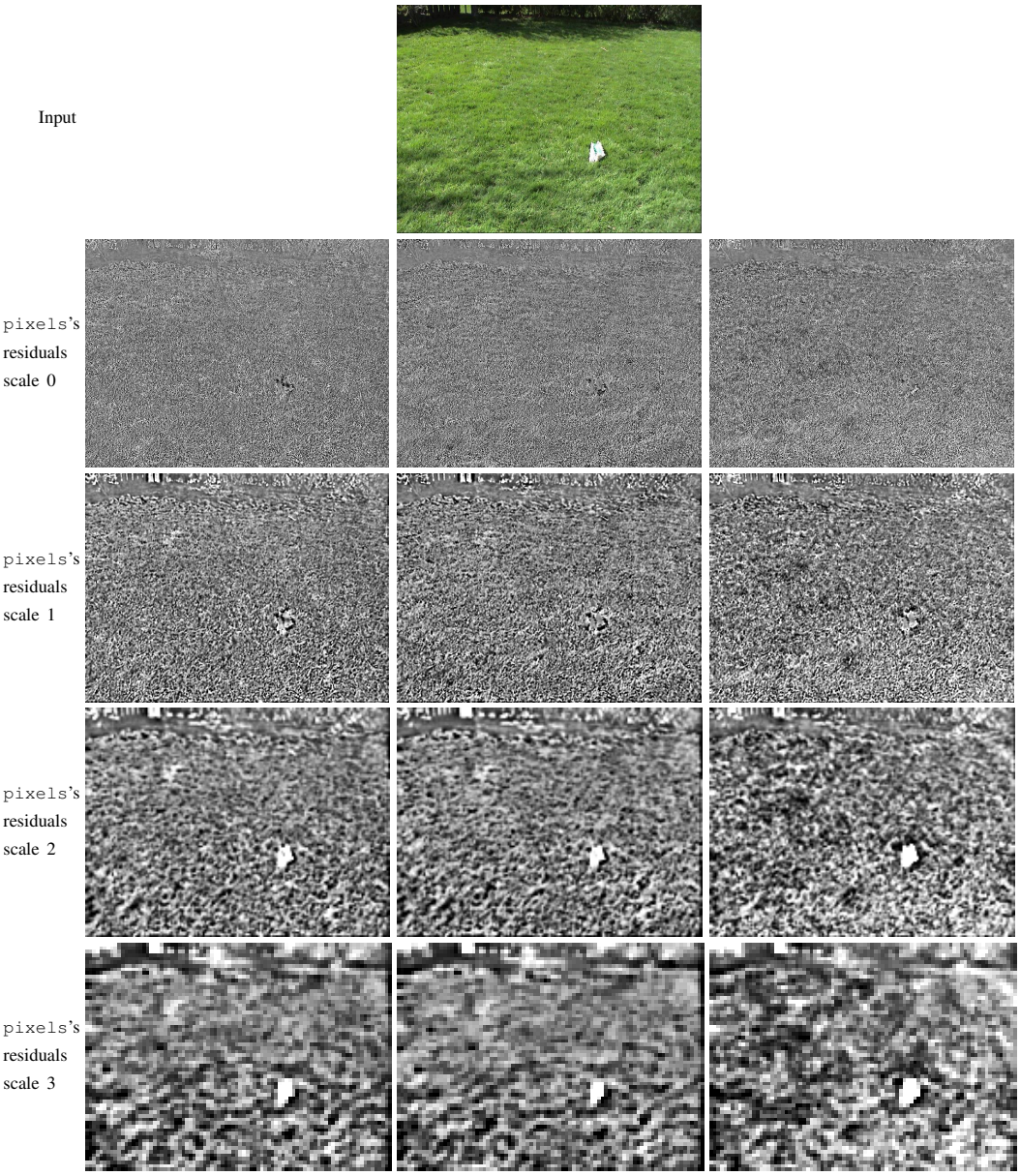

Fig. 3. Residuals for the method `pixels` for one image of Figure 1. The three columns correspond to R, G and B channels, which the rows correspond to the scales.

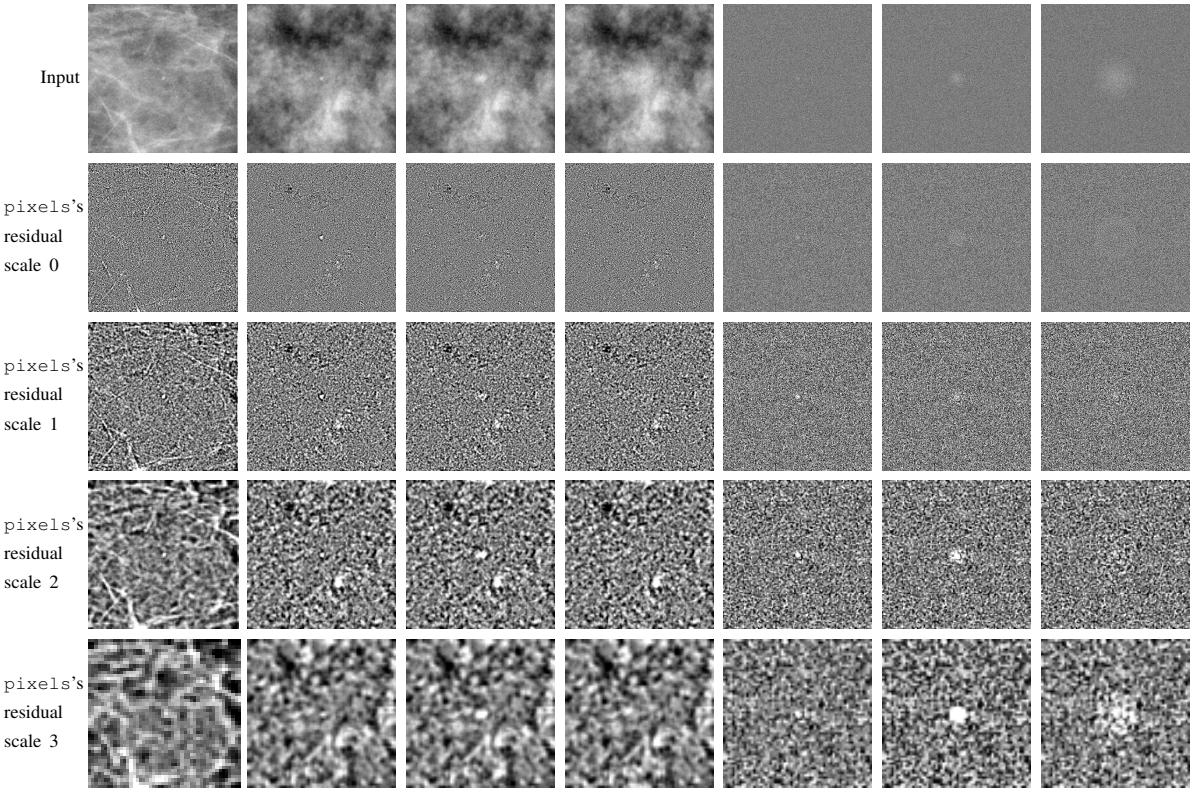

Fig. 4. Residuals for the method `pixels` for the images of Figure 2. The inputs being grayscale, there is only one residual per scale.

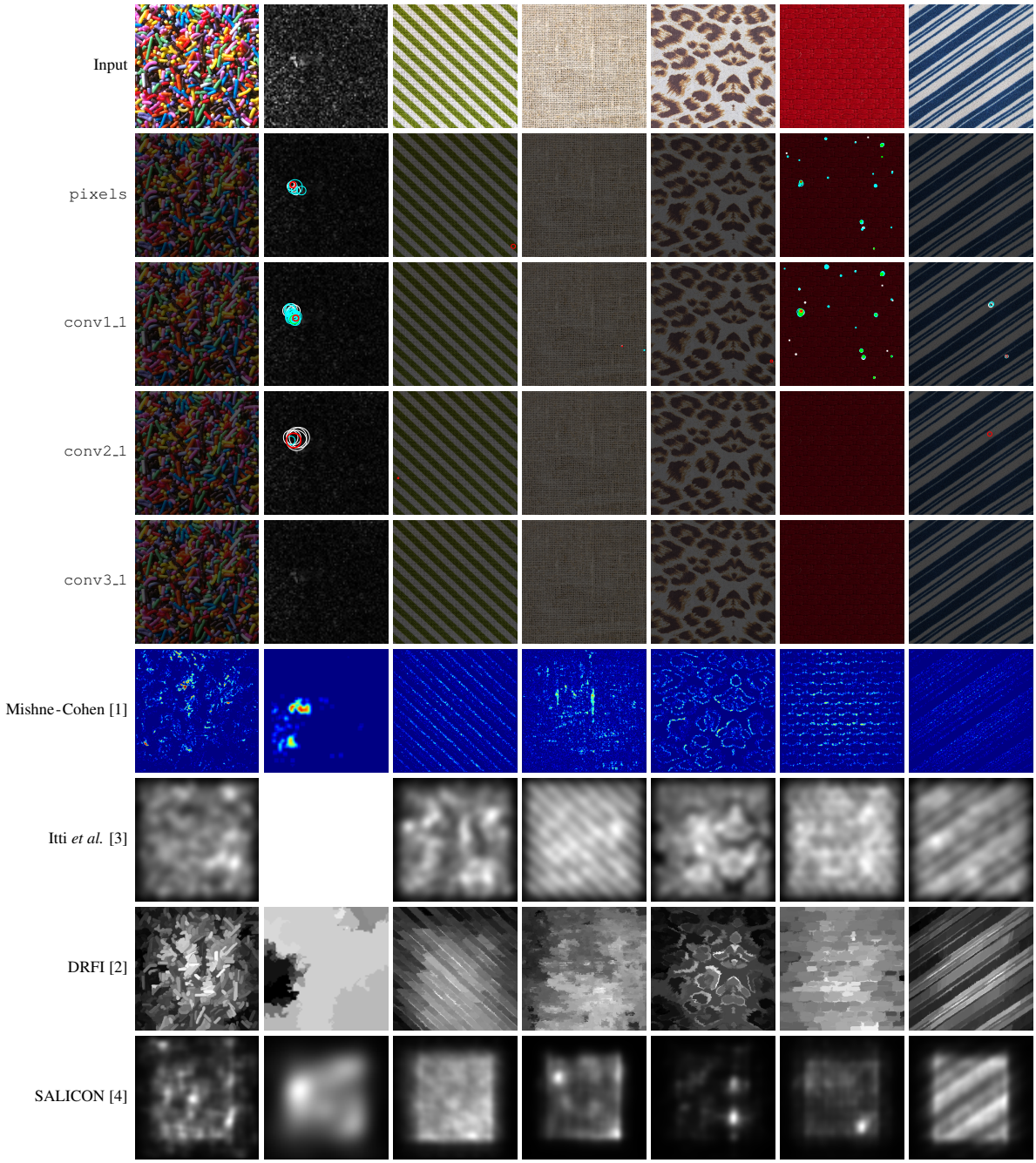

Fig. 5. Detection results of our method when using as image representation directly the image pixels (`pixels`) or the activation maps of the VGG neural Network at different layers (`conv1_1`, `conv2_1`, `conv3_1`) and a comparison to [1], [3], [2] and [4] on multiple examples. Each detection is represented by a circle, where the circle radius represents the detection scale and the color the strength of the detection (NFA). White corresponds to a weak detection (NFA test value between 10^{-3} and 10^{-2}), cyan to a good detection (NFA between 10^{-8} and 10^{-3}), green to a very strong detection (NFA between 10^{-21} and 10^{-8}) and orange to an extremely salient detection (NFA smaller than 10^{-21}). Red corresponds to the detection with lowest NFA. The second image comes from [1]. Notice on the first input no detection whatsoever is performed by our method, and on the fourth and fifth which are also complex textures without obvious anomalies only `conv1_1` detects one or two tiny defects.

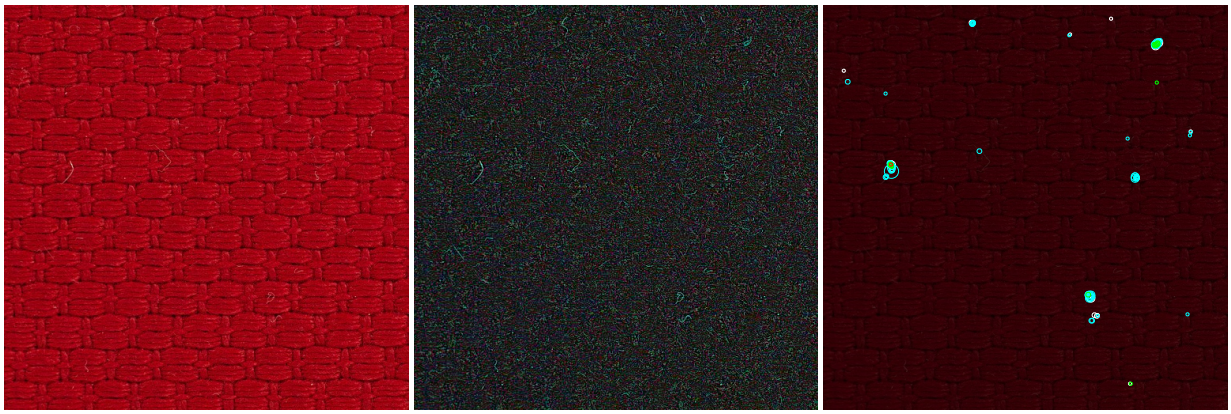

Fig. 6. Left: Picture of textile, right: The residual at scale 0 for `pixels` and the detections. All the textile impurities are highlighted on the residual.

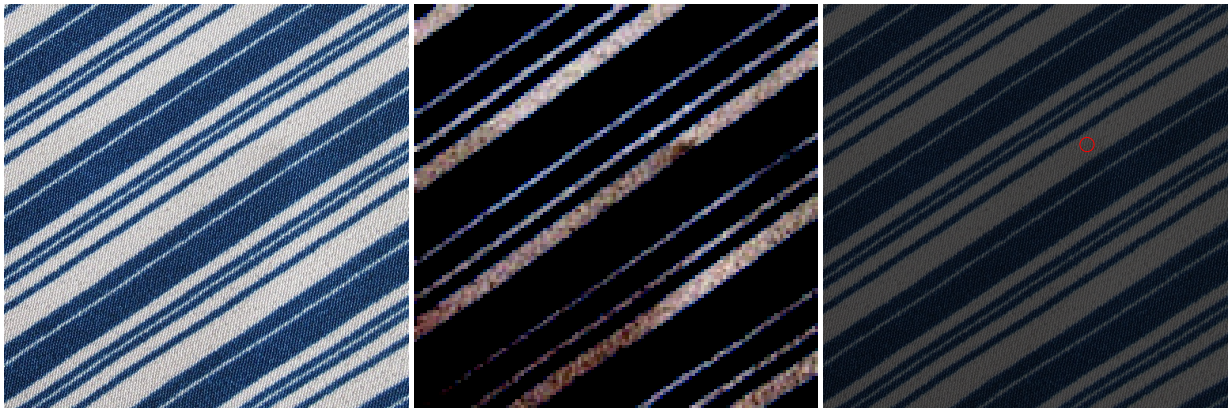

Fig. 7. Left: Input image, middle: transformed image to highlight the stain (zoomed out + affine transform), right: detections with `conv_21`. While a human does not spot the stain at first glance (indeed, one needs to look very carefully), the system is able to raise a weak detection (NFA of 6.10^{-3} here).

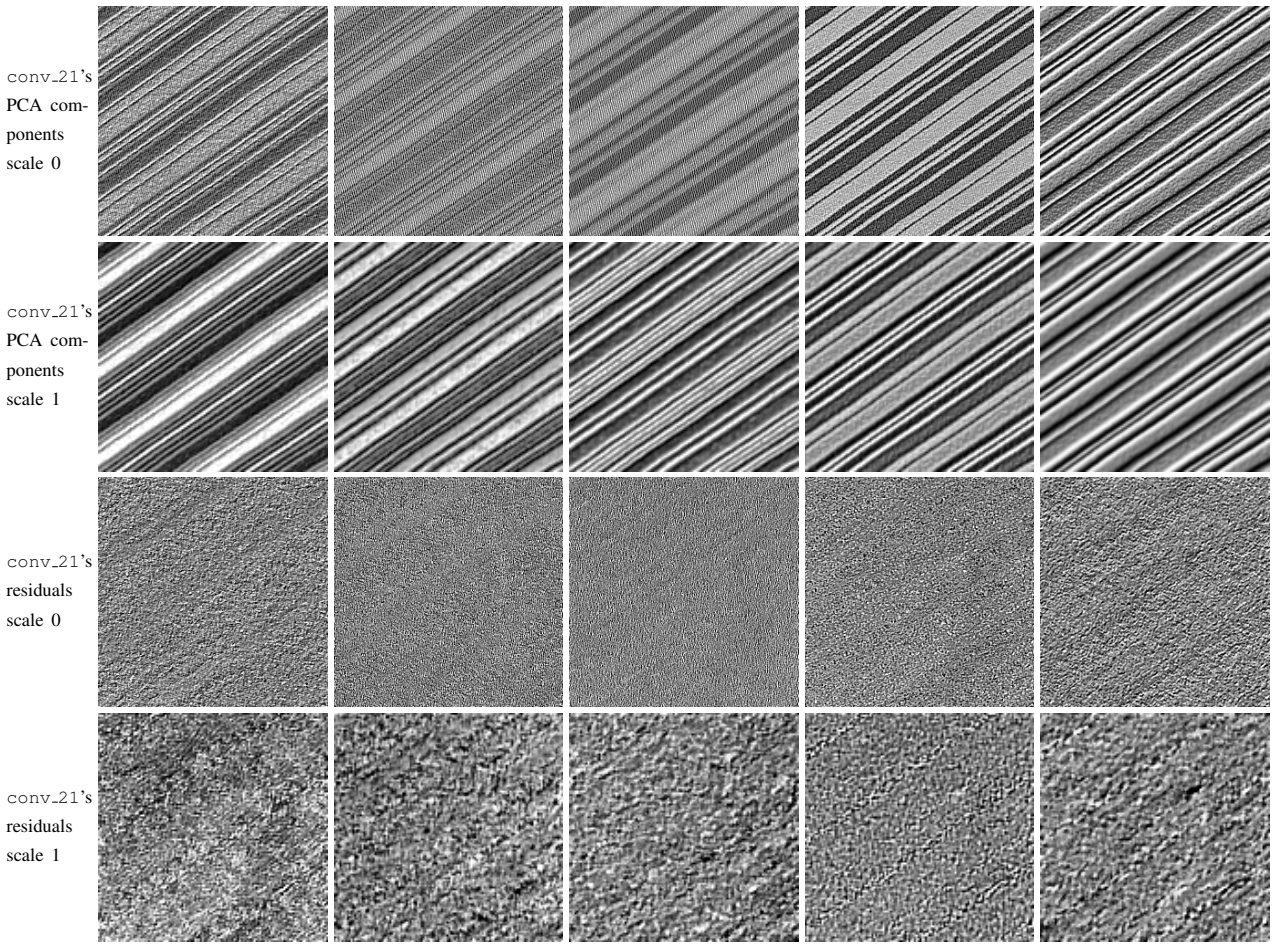

Fig. 8. Residuals for `conv_21` before (two first rows) and after (two last rows) self-similarity filtering for scales 0 (first and third rows) and 1 (second and fourth rows) for figure 7.

Fig. 9. Left: Input image, Right: detections with `pixels`. The method successfully detects a tank hidden in the landscape. This example is one of the examples provided by Itti *et al.* [6]

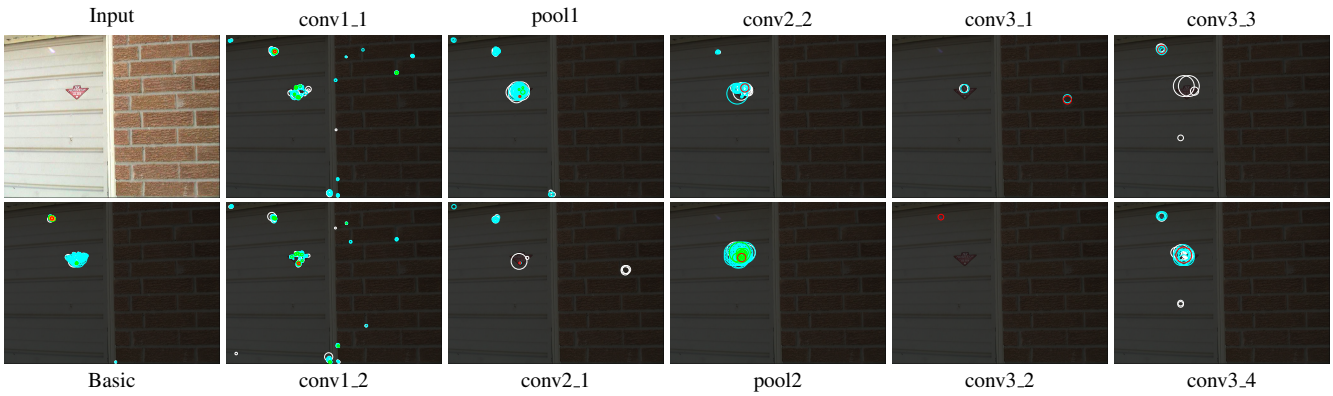

Fig. 10. Detections using different layers of Neural Network as input, or directly the image pixels (`pixels`). Almost all layers detect the two main anomalies: the door sign and the lens flare. The first layers, with low-level features, have more success detecting small scale anomalies (e.g., small holes in the cement), while deeper layers can detect a brick side stained with cement or a scratch on the door.