scientific reports

OPEN

Multiscale relevance of natural images

Samy Lakhal[1,2,3], Alexandre Darmon[4], Iacopo Mastromatteo[1,5], Matteo Marsili[6] & Michael Benzaquen[1,2,5]✉

We use an agnostic information-theoretic approach to investigate the statistical properties of natural images. We introduce the Multiscale Relevance (MSR) measure to assess the robustness of images to compression at all scales. Starting in a controlled environment, we characterize the MSR of synthetic random textures as function of image roughness H and other relevant parameters. We then extend the analysis to natural images and find striking similarities with critical (H ≈ 0) random textures. We show that the MSR is more robust and informative of image content than classical methods such as power spectrum analysis. Finally, we confront the MSR to classical measures for the calibration of common procedures such as color mapping and denoising. Overall, the MSR approach appears to be a good candidate for advanced image analysis and image processing, while providing a good level of physical interpretability.

Recent advances in image processing have benefited from the emergence of powerful learning frameworks combining efficient architectures[1–3] with large high-quality databases[4,5]. In particular, neural networks, layering simple linear and non-linear operators such as convolution matrices or activation functions, have proven to be very efficient to classify or generate high dimensional data. They are now able to capture similarities between images with unprecedented success. However, while their performance increases with the depth of the architecture, it is generally at the cost of physical interpretation in terms of informational content or maximisation of meaningful measures. Understanding the learning dynamics and the statistical features of the resulting images remains a challenge for the community[6,7].

Before the advent of machine learning algorithms, tasks such as compression[8,9], denoising[10] or edge detection were (and in some cases still are) performed using signal processing methods. Among the classical approaches, the first kind is based on specific measures, such as the widely used Peak Signal-to-Noise Ratio (PSNR)[11], that are built upon common signal processing metrics (Euclidian distance, power spectrum, etc.). The second family uses vision based experiments to construct semi-empirical measures of similarities, such as the Structural Similarity Index (SSI)[12]. As such, in both cases the approach relies on strong underlying assumptions.

In the context of statistical physics, the problem of high dimensional data inference has recently been addressed using a novel, fully agnostic, approach. Developed to measure specific properties of finite size samples[13], the approach consists in assessing the influence of a prescribed compression procedure over simple entropy measures. Applications in biological inference[14], finance[15], language models[13] or optimal machine learning[16,17] have already shown exciting results. In this paper, we adapt the latter formalism to image analysis and image processing, focusing specifically on the case of natural images. Natural scenes or landscapes have long been studied as they display distinguishable statistical features such as scale invariance[18–20], non-Gaussianity[21], or patch criticality[22].

The outline of the paper is as follows. In Section I, we introduce the Resolution/Relevance formalism using an illustrative example, and adapt it to the purpose of image analysis. In Section II, we analyse a class of parameterizable images, that is random $1/f^\alpha$ Gaussian fields, and introduce the Multiscale Relevance (MSR). In Section III, we extend the analysis to natural images and their gradient magnitudes. We discuss meaningful statistical similarities with the synthetic Gaussian fields. In Section IV, we show how the MSR approach can be used in the context of common image processing tasks.

[1]Chair of Econophysics and Complex Systems, Ecole Polytechnique, 91128 Palaiseau Cedex, France. [2]LadHyX, UMR CNRS 7646, Ecole Polytechnique, 91128 Palaiseau Cedex, France. [3]Institut Jean Le Rond d'Alembert, UMR CNRS 7190, Sorbonne Université, 75005 Paris, France. [4]Art in Research, 33 Rue Censier, 75005 Paris, France. [5]Capital Fund Management, 23 Rue de l'Université, 75007 Paris, France. [6]Quantitative Life Sciences Section, The Abdus Salam International Centre for Theoretical Physics, 34151 Trieste, Italy. ✉email: michael.benzaquen@polytechnique.edu

The resolution/relevance framework

Here we present the information-theoretic framework that was recently built by one us[13] for agnostic analysis of high-dimensional data samples and their behaviour under compression procedures. Relevant metrics are derived from simple statistics of the compressed samples.

Tradeoff between precision and interpretability. Let us consider the problem of binning, namely clustering samples of a random variable X into groups characterized by a similar value of X. If the sampled data points $\mathcal{S} = \{x_1, \ldots, x_N\}$ all take different states (e.g. when the distribution of X is continuous) the empirical distribution is a Dirac comb. In order to gain insight into the sampled variable, one can visualize the data by using histograms with well chosen bins/boxes. Indeed, this procedure enforces the emergence of structure by reducing data resolution through compression, allowing for more interpretability. One can then make assumptions on the underlying process and find the optimal parameters to best describe the data.

We illustrate this intuition by sampling $N = 100$ realizations of a Gaussian variable $X \sim \mathcal{N}(0, 1)$ in Fig. 1. The data are binned into n identical boxes between -4 and 4, for three different values of $n = 5, 23$ and 400. We also define the bin width ℓ as a compression parameter transforming the original sample $\mathcal{S}$ into a compressed sample $\mathcal{S}^\ell$. The compression step consists in replacing each data point by its corresponding histogram bar index. Figure 1a1 (large ℓ) displays a situation of *oversampling*. With only five bins a considerable amount of data resolution is lost. On the contrary, Fig. 1a3 (small ℓ) corresponds to an *undersampling* regime, with very narrow bins (mostly containing only one data point) and a resulting distribution close to a Dirac comb. Figure 1a2 (intermediate ℓ) appears as a reasonable compromise in which the histogram is visually close to the generator density, indicating we might be close to the optimal level of data compression. From the latter observation, one is tempted to go for a Gaussian model, with suitable estimators for the mean and variance. However such decision solely relies on a specific compression level, and thus does not make full use of the sample at play.

The formalism that we introduce in the next section provides a principled framework to connect the choice of the compression level with an optimality criterion that is agnostic to the nature of the generative model from which the data is sampled.

Resolution and relevance. Previous work from Marsili *et al.*[15] addressed the issue of the *overampling/undersampling* transition by introducing observables that allow one to monitor changes in a reduced sample $\mathcal{S}^\ell = \{s_1^\ell, \ldots, s_N^\ell\}$ obtained by compressing $\mathcal{S}$ with a parameter ℓ. First, let us consider k_s^ℓ the number of data points of identical state s and m_k^ℓ the number of states appearing k times in $\mathcal{S}^\ell$. It follows that $\sum_s k_s^\ell = \sum_k k m_k^\ell = N$. For example, in the compressed sample displayed in Fig. 1a1, values taken by k_s^ℓ are $\{2, 26, 49, 23, 0\}$, and since each bar in the histogram has a different height, one has $m_2 = m_{26} = m_{49} = m_{23} = m_0 = 1$ and $m_k = 0$ otherwise.

One can then define the *Resolution $\hat{H}^\ell[s]$* and *Relevance $\hat{H}^\ell[k]$* as:

$$\begin{cases} \hat{H}^\ell[s] &= -\sum_s \frac{k_s^\ell}{N} \log_2 \frac{k_s^\ell}{N}, \\ \hat{H}^\ell[k] &= -\sum_k \frac{k m_k^\ell}{N} \log_2 \frac{k m_k^\ell}{N}. \end{cases} \tag{1}$$
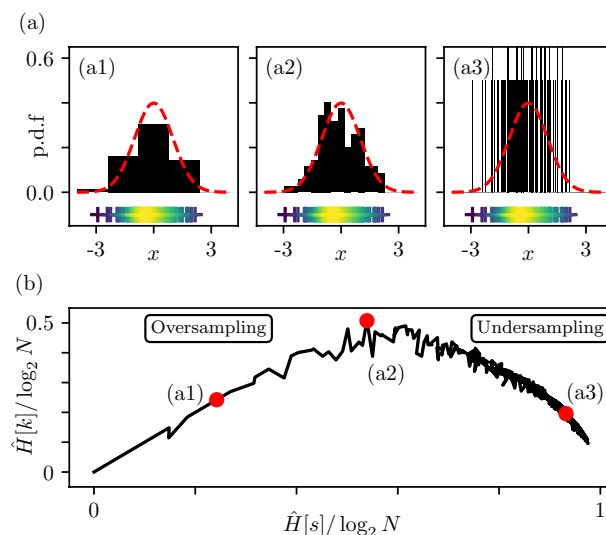

Figure 1. Relevance analysis of a Gaussian distribution sample ($N = 100$). (**a**) Influence of the number of bins n on the normalized histogram (black bars), for (**a1**) $n = 5$, (**a2**) $n = 23$ and (**a3**) $n = 400$. The red curve corresponds to the underlying distribution. The bottom markers ($+$) represent the initial sample data points with color indicating local data density. (**b**) Resolution/Relevance curve.

The *Resolution* is the entropy of the empirical distribution $\{p_s^\ell = k_s^\ell/N\}_s$ and describes the average amount of bits needed to code a state probability in $\mathcal{S}^\ell$. The compression clusters data points together hence reducing the average coding cost. The *Resolution* is maximal for raw data and monotonically decreases with ℓ, until it reaches the minimally entropic fully compressed sample. The *Relevance* is the entropy of the distribution $\{q_k^\ell = km_k^\ell/N\}_k$, that is the probability that a data point sampled from $\mathcal{S}^\ell$ appears k times in the sample. This is a compressed version of p_s^ℓ, where identical frequency states are clustered, dropping their label s in the process. Knowing q_k^ℓ is then sufficient to build a histogram without labels, and is equivalent to assuming indistinguishability of states sampled the same number of times. Sorting them in decreasing frequency values would yield the famous Zipf plot. In the end, the *Relevance* encodes the height of each bar and is maximal when $\{km_k^\ell/N\}_k$ is uniformly distributed, leading to $m_k \propto k^{-1}$. We reported in Table 1 the typical sampling situations and their corresponding value in Resolution/Relevance.

Coming back to the Gaussian sampling example, Fig. 1b displays $\hat{H}^\ell[k]$ as function of $\hat{H}^\ell[s]$, obtained by varying ℓ. Corresponding values for $n = 5$ (a1), $n = 23$ (a2) and $n = 400$ (a3) are highlighted. Note that (a2) maximizes Relevance while (a1) and (a3) respectively correspond to oversampling and undersampling. Let us emphasize at this point, that, despite the visual impression in this specific example, the sample (a2) does not necessarily minimize the distance between the underlying and empirical distributions. Interestingly, the Resolution/Relevance properties are only dependent on the raw sample $\mathcal{S}$ and the compression parameter ℓ, making the overall approach agnostic to the generating process. What is most interesting is thus the way in which the sample evolves with compression, while transitioning from undersampling to oversampling. As a result, one must choose a compression procedure that allows to crossover between these two regimes.

Application to images. Images are usually described as fields $h(\boldsymbol{r})$ where $\boldsymbol{r} \in \{1, \ldots, N_X\} \times \{1, \ldots, N_Y\}$. This is equivalent to a sample made of $\mathcal{S} = \{(\boldsymbol{r}, h(\boldsymbol{r})\}$ of size $N = N_X N_Y$, describing the position and color of each pixel. Naturally, $\mathcal{S}$ lies in the full undersampling regime as each data point is unique.

To compress grayscale images, we therefore propose a simple procedure consisting in two steps: (i) segmentation, and (ii) spatial compression, as illustrated in Fig. 2. Segmentation means grayscale levels are transformed

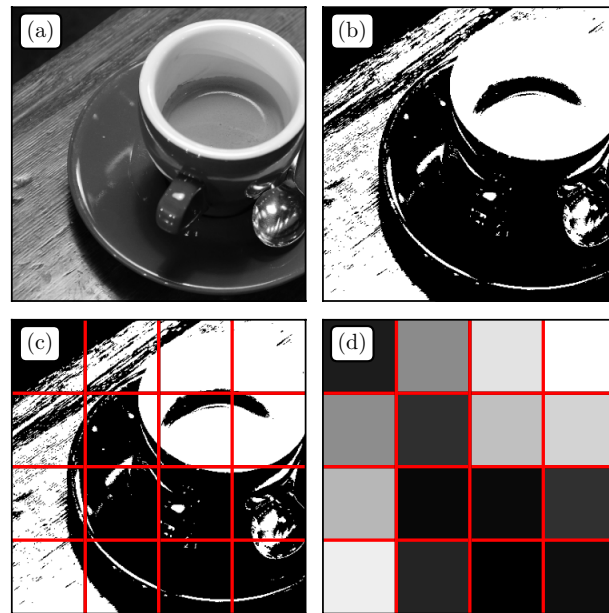| Situation	Sampled states	$\{m_k\}_k$	$\hat{H}[s]$	$\hat{H}[k]$
Full oversampling	Identical	$m_N = 1$ $m_k = 0$	0	0
Full undersampling	Distinct	$m_1 = N$ $m_k = 0$	$\log N$	0
Intermediate sampling	Intermediate	$m_k \propto k^{-1}$	$\hat{H}_0$	$\max \hat{H}[k]$

Table 1. Typical sampling situations.

Figure 2. Illustration of the segmentation/compression procedure on a classic benchmark image. (**a**) Original Image. (**b**) Thresholded image at a given quantile value a. (**c**) Thresholded image with reduced grid. (**d**) Reconstructed image from the reduced sample $\mathcal{S}_a^\ell$ where each grid cell is replaced by the average pixel color.

into black (b) and white (w) pixels using a threshold level a (fraction of black pixels), leading to the binary image $h_a(\boldsymbol{r})$ (Fig. 2b). This lowers the amount of possible color states in the sample, a necessary condition to reach the full oversampling regime. Note that one can reconstruct the original image by averaging over all segmentations. This step is generalisable for colors, for example by using a triplet (a_R, a_G, a_B) in the RGB space. The second step consists in the compression of pixel positions (Fig. 2c). One replaces each coordinate $\boldsymbol{r}$ by the index $\boldsymbol{r}^\ell$ of its position on a grid of stepsize ℓ. One ends up with a compressed sample:

$$\mathcal{S}_a^\ell = \{(\boldsymbol{r}^\ell, h_a(\boldsymbol{r}))\}. \tag{2}$$

A visualization of $\mathcal{S}_a^\ell$ is displayed in Fig. 2d where each cell is given the grayscale level corresponding to its proportion of black and white pixels. Finally, $k_{(\boldsymbol{r}_\ell,\mathrm{b})}^\ell$ and $k_{(\boldsymbol{r}_\ell,\mathrm{w})}^\ell$ would be defined as the number of black and white pixels in cell $\boldsymbol{r}^\ell$, and m_k^ℓ as the number of cells with k black or white pixels at scale ℓ. Using Eqs. (1), one can compute the values of $\hat{H}^\ell[s]$ and $\hat{H}^\ell[k]$ that will be used in the sequel.

One can make a direct analogy between this compression procedure and image processing architectures such as Convolution Neural Networks (CNN)[1]. First, their constitutive layers usually combine a spatial compression procedure, that is a first linear convolution, with a trainable or prescribed layer. Then, a segmentation step is performed using a nonlinear transformation on pixel values called *activation function*. In a similar fashion, our procedure consists in a one layer network, taking $\mathcal{S}$ as input and giving $\mathcal{S}_a^\ell$. Interestingly, we do not need to specify a particular convolution matrix as an input to the algorithm, but only a size parameter, by that making our approach more agnostic. Ultimately, note that any compression procedure allowing the undersampling/oversampling transition could have been selected. For example, one could use Discrete Fourier or Wavelet coefficients, classically used in JPEG or JPEG 2000 compression algorithms[8,9]. Another approach would consist in using intermediate representations of trained or untrained networks with binary activation functions (perceptron-like) and tunable layer size, as in the Resolution/Relevance trade-offs of deep neural architectures[16].

Relevance of random textures

In this section we illustrate the use of the metrics $(\hat{H}^\ell[s], \hat{H}^\ell[k])$ on a simple yet widely encountered class of processes: two-dimensional $1/f^\alpha$ random Gaussian fields. These are notably found in the spectral analysis of turbulence[23], water waves[24,25], and fracture mechanics[26,27]. We first recall the properties of such fields and then study the influence of α on Resolution and Relevance.

On $1/f^\alpha$ Gaussian fields. $1/f^\alpha$ Gaussian fields consist in the linear filtering of an initially uncorrelated 2D white noise (Supplementary Material). The latter presents a flat Fourier spectrum that is then multiplied by $1/f^\alpha$, therefore leading to a power spectrum scaling as $1/f^{2\alpha}$. This leads to the forcing of spatial correlations in the direct space. Such power law filter introduces scaling properties that are usually described by the roughness *Hurst* exponent $\mathrm{H} := \alpha - d/2$ where d is the field dimension (here $d = 2$). Depending on the
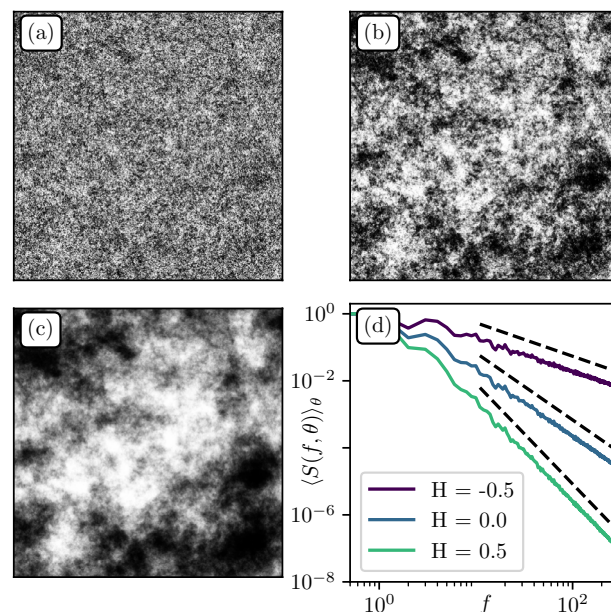

Figure 3. $1/f^\alpha$ textures generated from the same white Gaussian noise seed. (**a–c**) Quantile representations over 255 levels of $1/f^\alpha$ random fields with respective roughness $\mathrm{H} = -0.5, 0, 0.5$ and spatial resolution 512×512. (**d**) Azimuthally averaged power spectrum $\langle S(f, \theta) \rangle_\theta$. Black dashed lines indicate the theoretical power spectrum decay $1/f^{2\alpha}$ with $\alpha = 1 + \mathrm{H}$.

sign of H, one can recover two types of processes. When H < 0 the random field is stationary, that is with fixed mean and correlations $C(\delta r) \propto \delta r^{2H}$ at lag distance δr. The specific case H = $-d/2$ corresponds to an unmodified spectrum (white noise). When H > 0, the process is no longer stationary but possesses stationary increments with scaling $\langle [h(\mathbf{r} + \delta \mathbf{r}) - h(\mathbf{r})]^2 \rangle \propto \delta r^{2H}$. We generate three samples of distinct roughness values H $\in \{-0.5, 0, 0.5\}$, shown in Fig. 3a–c respectively, see e.g.[28]. The Hurst exponent influences the visual aspect of roughness, with images getting smoother as H increases. Figure 3d shows the azimuthally averaged power spectrum $\langle S(f, \theta) \rangle_\theta = \langle |\tilde{h}(f, \theta)|^2 \rangle_\theta$ allowing to check that the generating method is robust as the expected scaling behavior and exponents are recovered.

Multiscale relevance of random textures. We now perform the segmentation described above on the fields presented in Fig. 3. The resulting textures for threshold value $a = 0.5$ are displayed in Fig. 4a–c and the corresponding Resolution/Relevance curves $(\hat{H}^\ell[s], \hat{H}^\ell[k])_{\ell \in \{1, ..., N\}}$ are plotted in Fig. 4d.

One can see that while the patterns remain quasi-identical for H = -0.5 (Fig. 4a) and H = 0 (Fig. 4b), this is not the case for H = 0.5 (Fig. 4c) where large areas of uniform tint are created by the segmentation procedure. This is due to the presence of stronger spatial correlations, inducing more persistence of patterns and less fluctuations around the average. Further, one can see that the H = 0 texture displays interesting visual features at all scales, as reported in visual quality assessment experiments[29], while they appear limited to small scales for H = -0.5. It is not straightforward to connect these observations with the Relevance curves in Fig. 4d, as the relative Relevance varies with Resolution. It thus seems more natural to consider the Relevance across all levels of compression.

To do so, we introduce a measure that quantifies the overall robustness of a sample to compression called *Multiscale Relevance* (MSR) and defined as:

$$\text{MSR} := \int \hat{H}^\ell[k] d\hat{H}^\ell[s], \tag{3}$$

which is nothing other than the area under the Resolution/Relevance curve. This measure was introduced in Reference[14] as an order parameter characterizing neuronal activity time series, and was successful at distinguishing useful information from ambient noise, as expected from a complexity measure[30]. Note that while several measures of complexity based on multi-scale entropy contributions have already been introduced in the literature[31,32], the MSR differs in that the contribution of each scale is naturally weighted by the Resolution. Other measures generally give identical weights to each compression level.

For the images in Fig. 4, one obtains MSR(H = 0.5) < MSR(H = -0.5) < MSR(H = 0). This is consistent with our previous visual impression that the texture in Fig. 3b seems to contain more information at different scales.

Most relevant segmentation(s). One naturally expects the segmentation threshold a to influence the Relevance. Indeed, at given H < 0, most relevant representations do not seem to correspond to $a = 0.5$. This is confirmed in Fig. 5a where the Relevance curve for H = -0.8 is higher for $a = 0.66$ than $a = 0.5$. Figure 5b
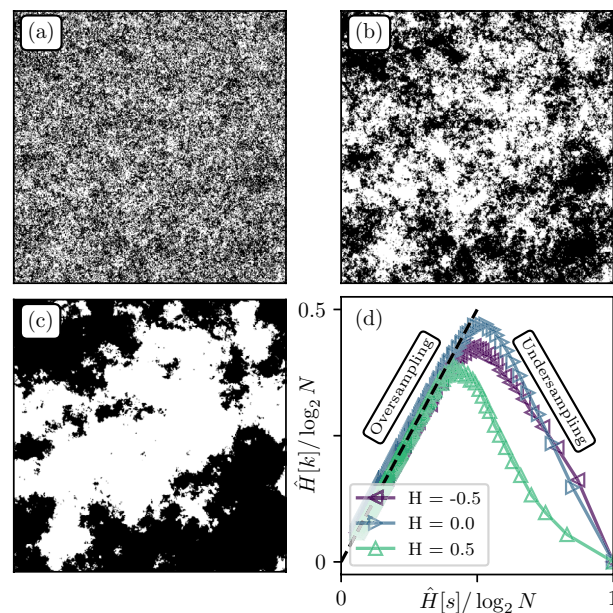

Figure 4. (**a–c**) Segmented versions of the textures of Fig. 3, with H = $-0.5, 0, 0.5$ respectively, and threshold value $a = 0.5$. (**d**) Resolution/Relevance curves normalized by the maximum entropy $\log_2 N$.
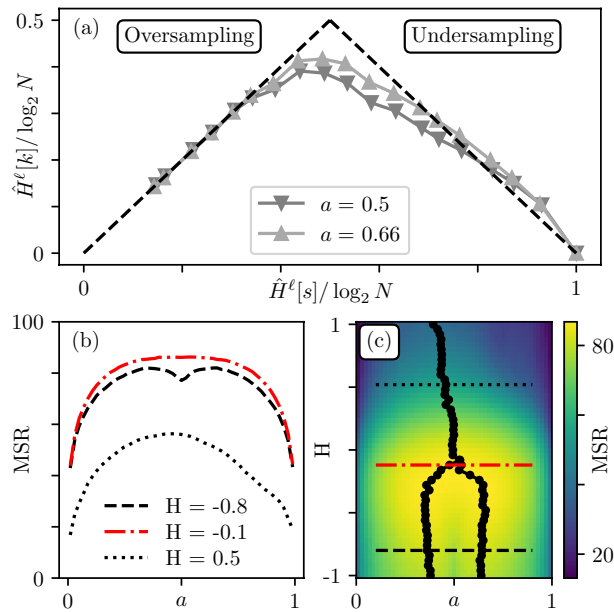
Figure 5. Influence of the segmentation value a. (**a**) Relevance curves for H $= -0.8$ for two values of a. (**b**) MSR as function of a for H $= -0.8$ (black dashed line), H $= -0.1$ (red dashed dotted line) and H $= 0.5$ (black dotted line). (**c**) Density plot MSR(H, a). The maxima are signified with black markers.

displays the MSR as function of a for three values of H. For H $= -0.8$ (dashed curve) one observes two symmetric maxima at $a_c = 0.5 \pm .13$, consistent with Fig. 5a. Interestingly, breaking the symmetry in the distribution of pixels by choosing a "background canvas" leads to more interesting samples in terms of Resolution/Relevance. As one can see in Fig. 5c, there is a bifurcation at H ≈ 0 below which two maxima of MSR coexist. The obtained values of a_c for H $< -1/2$ fall close to the classic percolation threshold $a^* = 0.59$ on the 2D square lattice[33]. Indeed, our segmented images are equivalent to samples of the correlated percolation site problem. In particular, Prakash et al.[34] observed, as we do here, that when H $\rightarrow 0$ from below both maxima continuously meet at $a_c = 0.5$ while flattening the MSR(a) curve around such value (see Fig. 5b). At this critical point, the information content of images becomes less sensitive to the segmentation process.

When H $\gtrsim 0$, MSR(a) displays one unique maximum at $a_c = 0.5$. However, as H increases further, so does the range of correlations, leading to finite-size effects. The resulting a_c becomes very noise dependent as different samples lead to different critical thresholds. Interestingly, such behavior was also reported in the percolation of 2D Fractional Brownian Motion[35].

Relevance of natural images

We now focus on *natural images*, namely pictures of natural scenes and landscapes. These have long been studied in the literature[18–21,36], as they display robust statistical features, such as scale invariance and criticality.

On the grayscale field $h(r)$.   Figure 6a shows the photograph from Tkacik et al.[37] in the Okavango Delta in Botswana, described as a "[...] *tropical savanna habitat similar to where the human eye is thought to have evolved*". The image is subdivided into fifteen patches of size 512×512 pixels. One can observe a wide variety of patterns, ranging from uniform shades of light gray in the sky to strong discontinuities with tree branches and noisy vegetation textures.

A power spectrum analysis for all patches is shown in Fig. 6b. The shape in the high frequency limit is due to camera calibration, optical blurring, or post-processing procedures, which are independent of the patch content. At low frequency we observe a decaying power law with exponent -2.0 ± 0.1. Note that, although there are small fluctuations that may be related to patch features[36], the power spectrum analysis seems rather unable to capture the visual heterogeneity from one patch to another mentioned above.

This being said, $S(f) \sim 1/f^2$ translates to H $= 0.0 \pm 0.1$ in terms of roughness exponents, which is precisely the range in which the MSR displayed critical and nontrivial behaviour for random textures in Sec. II.

We thus expect that the MSR approach may allow for a finer characterization of each patch. Another issue with classical spectral analysis is that the power spectrum of the image is expected to be extremely sensitive to non-linear transformations of its color histogram, even monotonous, that keep the visuals identical. With the MSR method, there is no such issue as the segmentation parameter a defines the proportion of black and white pixels, regardless of the shape of the color histogram.

Figure 6c shows the MSR curves for all patches. First observation is that the range of MSR values is similar in magnitude to that of H ≈ 0 textures in Sec. II. Then, one clearly sees significant differences between the MSRs of each patch. Patches containing mainly bushy textures with no abrupt changes in patterns display a unique
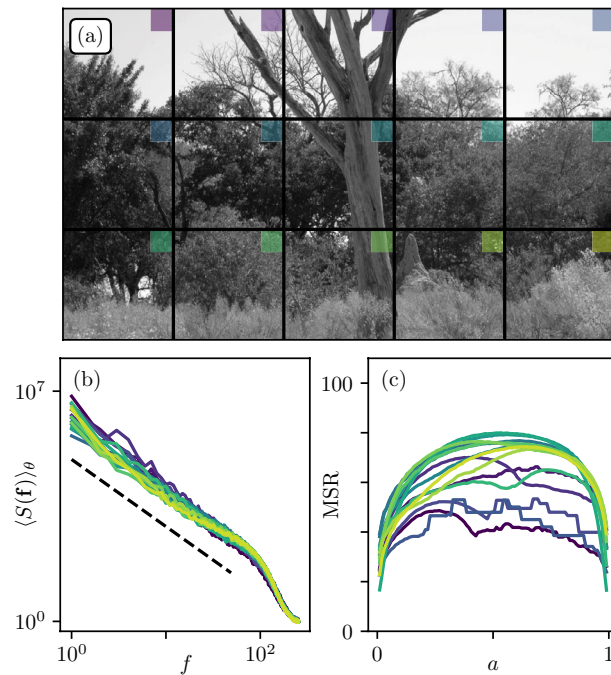
Figure 6. (**a**) Natural grayscale image from Reference[37], segmented in patches of size 512× 512. (**b**) Power spectrum for each patch. Dotted line is a decaying power law with exponent −2. (**c**) MSR as function of a for each patch.

maximum in the MSR(a) curve. Note that the irregularities that appear in some cases are due to specific colors being disproportionate in the histogram (uniform sky). Patches containing heterogeneous shades, or physical objects of different sizes combining tree trunks, branches and bush (e.g. bottom left in 6a) tend to display two maxima, similarly to H < 0 (see Sec. II).

Figure 7 focuses on the bottom-left patch of Fig. 6a. This sub-image seems to display two distinct dominant color levels. Such levels actually correspond to the maxima of the MSR curve in Fig. 7b. This is visually confirmed from the segmentations Fig. 7c,d which capture best the fluctuations at the top and bottom of the image respectively. We emphasize that the latter representations together constitute the most informative segmentations of (a). Superimposing them (Fig. 7e) indeed leads to a good approximation of the original image with only three color levels {0,127,255}. The MSR method thus seems to account well for the diversity of content of natural images, inaccessible through classical power spectrum analysis.

On the gradient magnitude |∇*h*|. To understand further the architecture of natural images, we now focus on the gradient magnitude field intended to capture strong spatial irregularities such as contours or borders. In addition, taking the gradient has the advantage of stationarizing the initial field. The gradient analysis is a fundamental block of various image processing procedures, from classic edge detection[38], to supervised[39] or unsupervised[1] classification architectures in machine learning. From a more perception-based psychophysical perspective, it has been shown that essential information such as orientations, geometries and positions could be directly inferred from the visual assessment of the gradient field[40–42]. We compute the gradients |∇*h*| from wavelet convolutions. This method is now extensively used as it shows excellent robustness for signal processing tasks[43–47]. One has:

$$|\nabla h| = |h * \boldsymbol{\psi}_j(\boldsymbol{r})|^2, \tag{4}$$

where $\boldsymbol{\psi}_j := (\psi_{j,x}, \psi_{j,y})$ is a wavelet gradient filter of characteristic dyadic size 2^j. This wavelet consists in mixing gradient and Gaussian windows, the latter being of standard deviation $\sigma_j = 2^j$ pixels. The procedure with $j = 0$ yields the image in Fig. 8a. As expected, one obtains a strong signal (bright shades) for fluctuating textures of vegetation or sharp contours like branches, and low values (dark shades) for smooth and uniform regions like the sky.

We then conduct the MSR analysis on these new patches (Fig. 8b), and observe that most patches give flat MSR curves. This is tantamount to the critical H ≈ 0 case with logarithmic correlations described in Section II (see Fig. 5). One may indeed think of natural images as a patchwork of objects of various sizes; such superposition of patterns is reminiscent of additive cascades processes[48] that also display logarithmic correlations.

We now explore the effect of changing the wavelet size (see Fig. 9). We chose the top middle patch in Fig. 8a as it contains large objects and small scale details. As one can see in Fig. 9c,d, increasing j has the effect of coarse-graining small fluctuations to only leave larger ones. This translates into smaller Relevance at low compression, which in turn reduces the overall MSR (Fig. 9b). Finally, the segmented gradient fields at critical threshold values
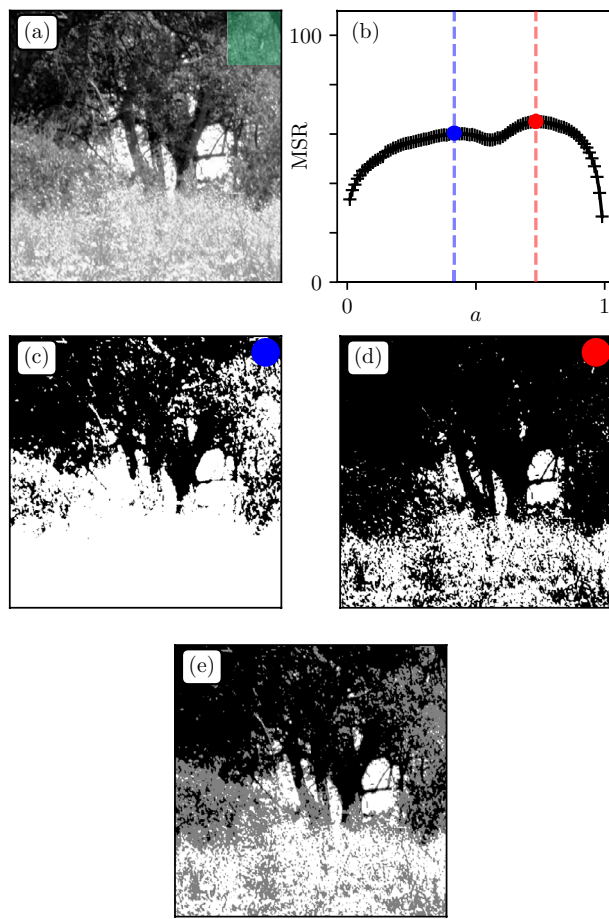
Figure 7. (**a**) Bottom-left patch of Fig. 6a. (**b**) MSR as function of *a* with highlighted critical thresholds $(a_1, a_2) = (0.42, 0.73)$. (**c,d**) Corresponding segmented patches. (**e**) Image obtained by adding (**c,d**), with three color levels {0,127,255}.
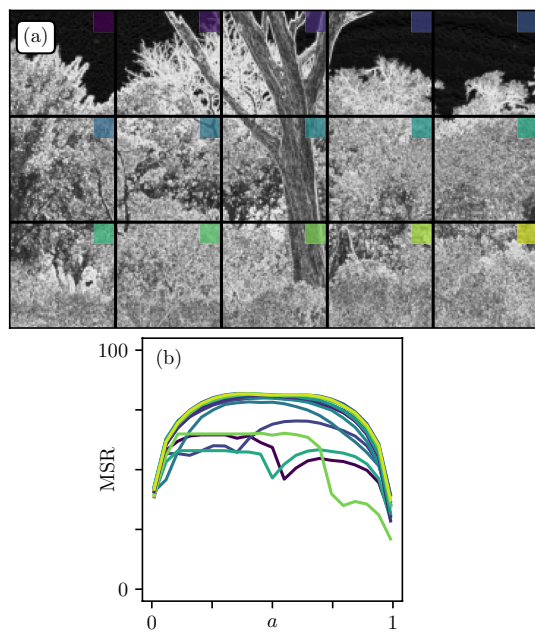

Figure 8. (**a**) Quantile representation over 255 levels of the Gradient Magnitude field of Fig. 6a, with $j = 0$, divided in 512×512 patches. (**b**) MSR as function of *a* for the different patches.
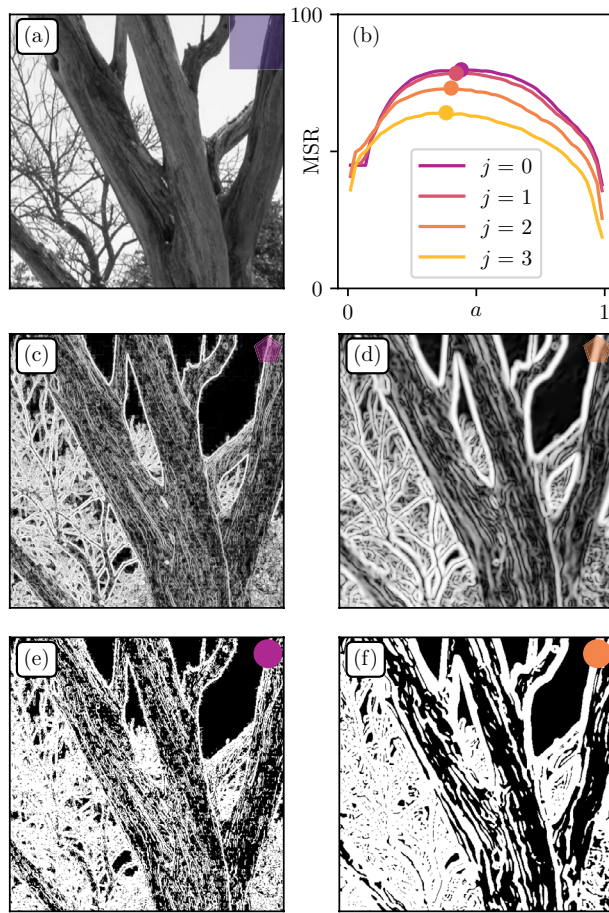
Figure 9. Influence of the Gradient Wavelet size. (**a**) Original patch from Fig. 6a. (**b**) MSR as function of *a* for gradient wavelets of dyadic size $(2^j), j \in \{0, 1, 2, 3\}$. (**c,d**) Gradient magnitudes for $j = 0$ and $j = 2$ respectively. (**e,f**) Segmented gradient magnitudes at critical threshold values a_c for $j = 0$ and $j = 2$ respectively.

(Fig. 9e,f) remain visually close to initial fields (Fig. 9c,d). This is indeed expected as gradient magnitudes already show a large proportion of black and white pixels at the contours of physical objects.

Application to image processing

Here we illustrate the potential of MSR in the context of common digital image processing tasks, namely color mapping and denoising.

Color mapping. Consider the color mapping problem consisting in projecting pixel values onto a reduced palette. For the sake of simplicity, let us consider the case of an initial grayscale palette projected on binary values $\{0, 255\}$ (b&w). We implement a stochastic mapping procedure using the Boltzmann distribution $\mathbb{P}(c|h_{ij}) \propto e^{-(h_{ij}-c)^2/T}$, where h_{ij} is the original color of pixel with coordinates ij, $c \in \{0, 255\}$ the color in the reduced palette, and T a temperature parameter, see[49]. Note that this probability density is obtained from the maximal entropy distribution related to the minimization of the Mean-Squared Error (MSE) between the original and mapped images. $T = 0$ corresponds to the choice of the closest color in the reduced palette, while $T \to \infty$ leads to uniform noise.

Optimizing the procedure consists in calibrating T to maximize some advanced similarity measure between the original and reduced images, in the hope that it will capture more interesting properties than a simple pixel-pixel Euclidean distance minimization. Here we propose an alternative approach consisting in maximizing an information measure, the MSR, and compare it to classical metrics, namely the Peak Signal-to-Noise Ratio (PSNR)[11] and the Structural Similarity Index (SSI)[12]. PSNR is directly related to the Mean Squarred Error (MSE) between original and mapped images through $\text{PSNR} = 10 \log_{10} \left(\Delta^2/\text{MSE}\right)$ where Δ is the range of the signal, that is 255 for typical grayscale encoding. SSI is based on the comparison of patches between two images and takes into account properties such as luminance and contrast. Both are widely used in the digital image processing community.

Figure 10a displays the original patch extracted from Fig. 6a. Figure 10b shows the evolution of each metrics with temperature T. One sees that the PSNR between the original and mapped images is maximized at $T = 0$. This
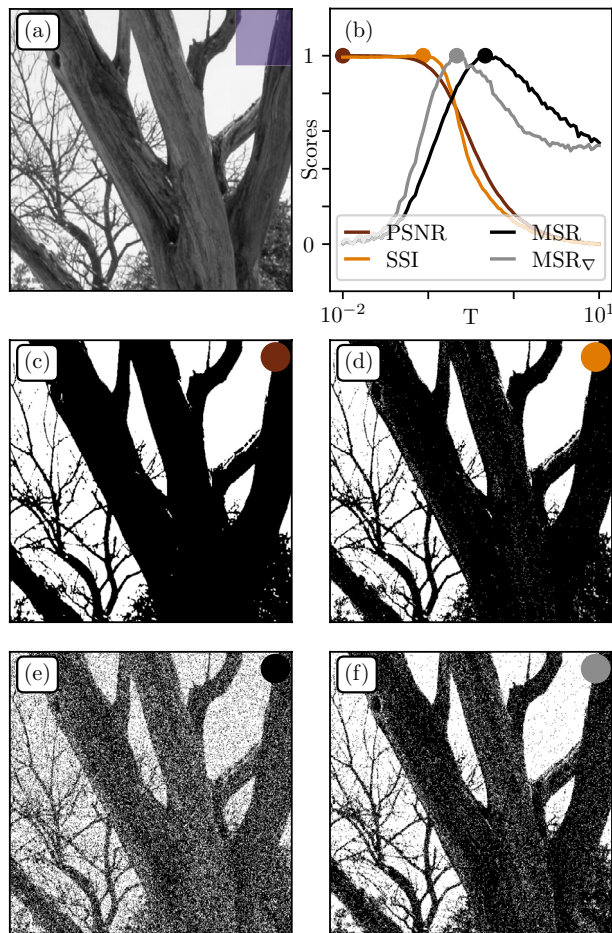
Figure 10. Color mapping. (**a**) Original patch from Fig. 6a. (**b**) Rescaled scores as function of temperature for different performance measures: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSI), direct Multiscale Relevance (MSR), and MSR over the gradient field MSR$_\nabla$. (**c–f**) Color mapping at optimal temperatures T^* for PSNR, SSI, MSR and MSR$_\nabla$ respectively.

is not surprising as the PSNR is monotonously related to the MSE by definition. The corresponding mapping in Fig. 10c appears too sharp and contrasted, clearly separating vegetation from sky while introducing thresholding artifacts. Optimization of the SSI yields a non-zero yet small temperature $T = 0.1$, barely improving the resulting image (see Fig. 10d). We then compute the MSR for both direct and gradient fields. The maximization of MSR(T) leads to the image shown in Fig. 10e, which contains more faithful visual features and a decent similarity to the original image at large scales, at the cost of artificial small scale features.

Now recall that conducting the MSR analysis on the gradient magnitude of natural images provided better results (consistency between different patches in Fig. 8b). Here, we introduce the *Gradient Magnitude MSR*, henceforth denoted MSR$_\nabla$, computed by segmenting the grayscale images obtained from the gradient procedure, and averaging over a. The maximization of MSR$_\nabla$ is shown in Fig. 10f. This seems like a good compromise between (c),(d) and (e) as it also displays medium scale features (tree trunk details) without blurring finer ones (small branches). Note however that MSR brings more noise than PSNR or SSI which might make it less desirable if noiselessness is a strong constraint.

Hence, for strong color reduction, a Multiscale Relevance approach can bring better visuals than classical metrics such as the Structural Similarity Index which, in addition, requires an *a priori* semantic knowledge of the original image. Note that the analysis could be extended to more elaborate color mapping procedure such as error diffusion[50,51] or Monte-Carlo based algorithms[52].

Denoising with Rudin–Osher–Fatemi algorithm.

We now focus on a denoising procedure which consists in correcting unwanted noise caused by signal processing or camera artefacts. To tackle this problem, a classic algorithm is the Rudin-Osher-Fatemi (ROF)[10] which minimizes the following functional:

$$\mathcal{L}[f] = \lambda \|\nabla f\|_{2,1} + \|h - f\|_2^2, \tag{5}$$

where h is the original noisy image, f the target denoised image and λ a *regularization/penalty* term preventing gradient explosion and allowing for smooth solutions. Note that the first term in the RHS of Eq. (5) comprises
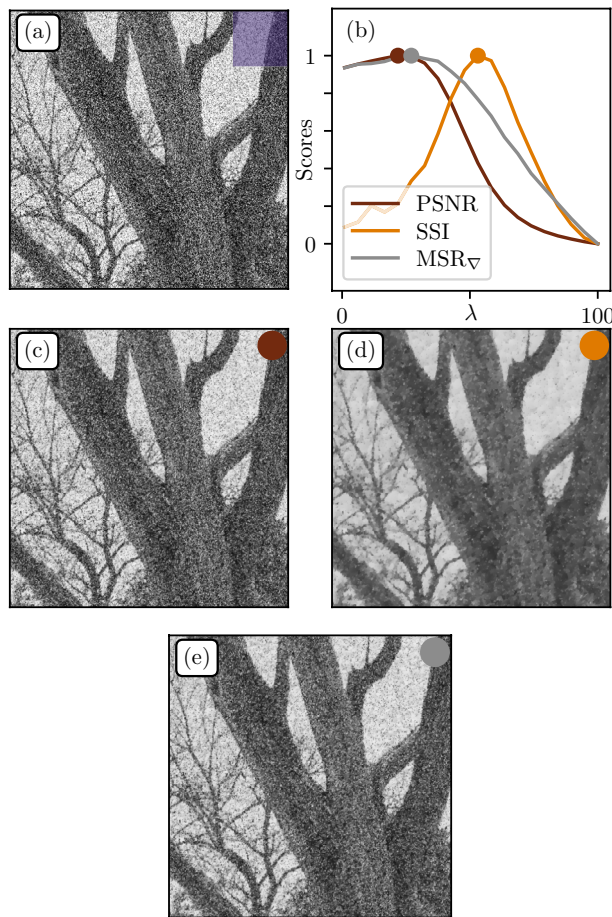
Figure 11. Denoising. (**a**) Noisy patch obtained from adding a Gaussian noise ($\sigma = 100$) to the same patch from Fig. 10a. (**b**) Rescaled scores as function of λ for different performance measures: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSI) and MSR over the gradient field MSR$_\nabla$. (**c–e**) Denoising at optimal regularization parameter λ^* for PSNR, SSI and MSR$_\nabla$ respectively.

a mixed norm, consisting in the L2 norm of each gradient vector integrated over the image domain, classically called *Total Variation*[10]. The free parameter λ is generally chosen by the operator through visual assessment.

Here we propose to calibrate such a model using again the PSNR, SSI and MSR$_\nabla$ metrics. We consider the image in Fig. 11a obtained by adding a Gaussian white noise to the patch in Fig. 6a. We intentionally choose a high noise value to make the denoising procedure difficult, such that some details from the original image may never be recovered. Our goal is to seek the optimal λ^* leading to the best visual. The scores obtained for each method as function of λ are displayed in Fig. 11b. Optimally denoised images using PSNR, SSI and MSR$_\nabla$ are shown in Fig. 11c–e respectively. With PSNR, one is left with a rather high level of noise, while details on the trunk surface or in the branches are conserved. In contrast, SSI removes a significant part of the noise, but at the cost of blurring small scale details. One might argue that optimal denoising with MSR$_\nabla$ seems like a good compromise between a too noisy PSNR image and an overly smoothed SSI image. But again, as argued in the colour mapping section above, assessing which result is "best" clearly depends on the context and target constraints.

Conclusion

Let us summarize what we have achieved. We first introduced the Resolution/Relevance framework through a simple illustrative example. We showed how such formalism can be applied to image analysis. With the aim of investigating the framework in a controlled environment, we started by studying random textures. We then defined the Multiscale Relevance (MSR) which measures the entropy contribution at all compression scales, and obtained statistical features reminiscent of the correlated percolation problem. In particular, we highlighted the existence of a critical roughness parameter H$_c \approx 0$, corresponding to logarithmic correlations, and discussed optimal segmentation. We then extended the analysis to natural images and drew a successful comparison with random textures; we observed strong similarities with critical random Gaussian fields. Looking at gradient magnitude fields revealed an even stronger similarity to roughness criticality. Finally, we confronted the MSR procedure to classical signal processing measures in the context of simple image processing tasks: color mapping and denoising. We obtained interesting results thereby demonstrating the potential of the agnostic MSR approach for image processing.

This last section would benefit from an extension to more elaborate image processing techniques, beyond the scope of the present paper. Future research should also focus on analytically tractable developments of Relevance and Resolution in simple cases, e.g. Gaussian white noise with well chosen cascading processes. Also note that we considered a straightforward compression procedure on the direct space but equivalent representations, for example Discrete Cosine[8] or Wavelet harmonics[9], could be used to define the reduced sample $\mathcal{S}$. Finally, we have seen that the MSR is able to capture the most relevant segmentation values, which may be used as a pre-processing method for learning frameworks.

Data availability

The datasets generated and/or analysed during the current study are available in the Lakhal2023 repository: https://github.com/SamyL2/Lakhal2023. The image used in Fig. 2 has been taken from the scikit-image Python library. It has no copyright restrictions and is under the CC0 licence by the photographer (Rachel Michetti). The image used in Fig. 6, and subsequently used in Figs. 7, 9, 10 and 11, has been taken from the database article[37] and is distributed under a Creative Commons Attribution-Noncommercial Unported licence to facilitate research in computer vision, psychophysics of perception, and visual neuroscience.

References

1. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84 (2017).
2. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278 (1998).
3. Vaswani, A. *et al.* Attention is all you need *Adv. Neural Inf. Process. Syst.* **30** (2017).
4. Krizhevsky, A. Learning multiple layers of features from tiny images Department of Computer Science, University of Toronto (2009).
5. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database *IEEE Conf. Comput. Vis. Pattern Recognit.* **248** (2009).
6. Saxe, A. M., McClelland, J. L. & Ganguli, S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv preprint* arXiv:1312.6120 (2013).
7. Goldt, S., Advani, M., Saxe, A. M., Krzakala, F. & Zdeborová, L. Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup *Adv. Neural Inf. Process. Syst.* **32** (2019).
8. Wallace, G. K. The JPEG still picture compression standard. *IEEE Trans. Consum. Electron.* **38**, xviii (1992).
9. Skodras, A., Christopoulos, C. & Ebrahimi, T. The JPEG 2000 still image compression standard. *IEEE Signal Process. Mag.* **18**, 36 (2001).
10. Rudin, L. I., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259 (1992).
11. Korhonen, J. & You, J. Peak signal-to-noise ratio revisited: Is simple beautiful? *Fourth International Workshop on Quality of Multimedia Experience* **37** (2012).
12. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600 (2004).
13. Marsili, M. & Roudi, Y. Quantifying relevance in learning and inference. *Phys. Rep.* **963**, 1 (2022).
14. Cubero, R. J., Marsili, M. & Roudi, Y. Multiscale relevance and informative encoding in neuronal spike trains. *J. Comput. Neurosci.* **48**, 85 (2020).
15. Haimovici, A. & Marsili, M. Criticality of mostly informative samples: A Bayesian model selection approach. *J. Stat. Mech Theory Exp.* **2015**, P10013 (2015).
16. Song, J., Marsili, M. & Jo, J. Resolution and relevance trade-offs in deep learning. *J. Stat. Mech. Theory Exp.* **2018**, 123406 (2018).
17. Duranthon, O., Marsili, M. & Xie, R. Maximal relevance and optimal learning machines. *J. Stat. Mech. Theory Exp.* **2021**, 033409 (2021).
18. Balboa, R. M. & Grzywacz, N. M. Power spectra and distribution of contrasts of natural images from different habitats. *Vis. Res.* **43**, 2527 (2003).
19. Ruderman, D. L. The statistics of natural images. *Netw. Comput. Neural Syst.* **5**, 517 (1994).
20. Ruderman, D. L. & Bialek, W. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* **73**, 814 (1994).
21. Zoran, D. & Weiss, Y. Scale invariance and noise in natural images *IEEE 12th International Conference on Computer Vision*, 2209 (2009).
22. Stephens, G. J., Mora, T., Tkačik, G. & Bialek, W. Statistical thermodynamics of natural images. *Phys. Rev. Lett.* **110**, 018701 (2013).
23. Kolmogorov, A. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Akademiia Nauk. SSSR Doklady.* **30**, 301 (1941).
24. Moskowitz, L. Estimates of the power spectrums for fully developed seas for wind speeds of 20 to 40 knots. *J. Geophys. Res.* **69**, 5161 (1964).
25. Toba, Y. Local balance in the air-sea boundary processes: III. On the spectrum of wind waves. *J. Oceanogr. Soc. Jpn.* **29**, 209 (1973).
26. Bouchaud, E. Scaling properties of cracks. *J. Phys.: Condens. Matter* **9**, 4319 (1997).
27. Schmittbuhl, J., Schmitt, F. & Scholz, C. Scaling invariance of crack surfaces. *J. Geophys. Res. Solid Earth* **100**, 5953 (1995).
28. Pesquet-Popescu, B. & Véhel, J. L. Stochastic fractal models for image processing. *IEEE Signal Process. Mag.* **19**, 48 (2002).
29. Lakhal, S., Darmon, A., Bouchaud, J.-P. & Benzaquen, M. Beauty and structural complexity. *Phys. Rev. Res.* **2**, 022058 (2020).
30. Aaronson, S., Carroll, S. M. & Ouellette, L. Quantifying the rise and fall of complexity in closed systems: the coffee automaton. *arXiv preprint* arXiv:1405.6903 (2014).
31. Zhang, Y.-C. Complexity and 1/f noise. A phase space approach. *J. Phys. I* **1**, 971 (1991).
32. Humeau-Heurtier, A. Multiscale entropy approaches and their applications. *Entropy* **22** (2020).
33. Isichenko, M. B. Percolation, statistical topography, and transport in random media. *Rev. Mod. Phys.* **64**, 961 (1992).
34. Prakash, S., Havlin, S., Schwartz, M. & Stanley, H. E. Structural and dynamical properties of long-range correlated percolation. *Phys. Rev. A* **46**, R1724 (1992).
35. Du, C. & Satik, C. Percolation in a fractional Brownian motion lattice *AIChE J.* **42** (1996).
36. Van der Schaaf, A. & van Hateren, J. Modelling the power spectra of natural images: Statistics and information. *Vis. Res.* **36**, 2759 (1996).
37. Tkačik, G. *et al.* Natural images from the birthplace of the human eye. *PLoS ONE* **6**, e20409 (2011).
38. Peli, T. & Malah, D. A study of edge detection algorithms. *Comput. Graph. Image Process.* **20**, 1 (1982).
39. Andreux, M. *et al.* Kymatio: Scattering transforms in python. *J. Mach. Learn. Res.* **21**, 1 (2020).

40. Keil, M. S., Cristobal, G. & Neumann, H. Gradient representation and perception in the early visual system-a novel account of Mach band formation. *Vis. Res.* **46**, 2659 (2006).
41. Keil, M. S. Gradient representations and the perception of luminosity. *Vis. Res.* **47**, 3360 (2007).
42. Kilpeläinen, M. & Georgeson, M. A. Luminance gradient at object borders communicates object location to the human oculomotor system. *Sci. Rep.* **8**, 1593 (2018).
43. Morel, R., Rochette, G., Leonarduzzi, R. Bouchaud, J.-P. & Mallat, S. Scale Dependencies and Self-Similarity Through Wavelet Scattering Covariance *arXiv preprint* arXiv:2204.10177 (2022).
44. Mallat, S. G. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoust. Speech Signal Process.* **37**, 2091 (1989).
45. Antoine, J.-P., Carrette, P., Murenzi, R. & Piette, B. Image analysis with two-dimensional continuous wavelet transform. *Signal Process.* **31**, 241 (1993).
46. Abry, P., Jaffard, S., Roux, S., Vedel, B. & Wendt, H. Wavelet decomposition of measures: Application to multifractal analysis of images *Unexploded Ordnance Detect. Mitig.* **1** (2009).
47. Wendt, H., Abry, P., Roux, S. G., Jaffard, S. & Vedel, B. Multifractal analysis for images: The wavelet leaders contribution. *Traitement du Signal* **26**, 47 (2009).
48. Duplantier, B., Rhodes, R., Sheffield, S. & Vargas, V. Log-correlated Gaussian fields: An overview. *Geom. Anal. Probab. Honor Jean-Michel Bismut* **191** (2017).
49. Lakhal, S., Darmon, A. & Benzaquen, M. A new spin on color quantization. *J. Stat. Mech. Theory Exp.* **2023**, 033401 (2023).
50. Jarvis, J. F., Judice, C. N. & Ninke, W. A survey of techniques for the display of continuous tone pictures on bilevel displays. *Comput. Graphics Image Process.* **5**, 13 (1976).
51. Floyd, R. W. & Steinberg, L. An adaptive algorithm for spatial greyscale. *Proc. Soc. Inf. Display* **17** (1976).
52. Puzicha, J., Held, M., Ketterer, J., Buhmann, J. M. & Fellner, D. W. On spatial quantization of color images. *IEEE Trans. Image Process.* **9**, 666 (2000).

Acknowledgements

Author contributions

M.M., I.M., A.D. and M.B. designed research. S.L. performed research. S.L., A.D. and M.B. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-41714-0.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.