Image and texture synthesis

From natural image statistics to deep generative models

First part

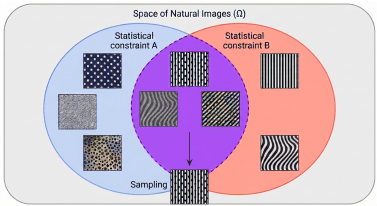
Yann GOUSSEAU

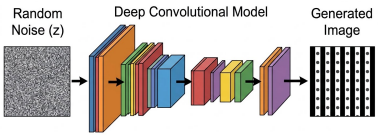
Télécom Paris - Institut Polytechnique de Paris

MVA
2026-2027

From natural image statistics to generative models

1. Natural image statistics
2. Texture synthesis methods
   - Patch-based methods
   - Statistical constraints
3. Deep generative models





Part I

Natural image statistics

Natural image statistics

- What regularity for visual signals ?
  Which statistics to capture this regularity ?





- Motivations :
  - Neurosciences (efficient coding hypothesis)
  - Image compression (leveraging image redundancy and sparsity of transform codings)
  - Bayesian (and other statistical) approaches
    To estimates a quantity a (labels, pixels, etc.) from an image I, Bayes rule is

$$Pr(\{a\}|I) = \frac{Pr(I|\{a\})Pr(\{a\})}{Pr(I)}$$

    so that estimation of a (e.g. MAP, maximum a posteriori) requires the prior $Pr(a)$

- What is the shape of natural image distribution ?
  Huge dimensionality : restrict to marginals, filter responses, patches, etc.

Neurosciences / efficient coding hypothesis

- **Efficient coding hypothesis** in neurosciences (Barlow, 1961[1])
  Efficient representation : few spikes
  Neural coding leverage signal redundancy (link with information theory)
- Similarity between efficient image algorithms and human visual processing
  e.g. :
  - Sparse coding yields filters that are *localized, oriented, selective in frequency* sharing similarity with receptive fields in V1[2]
  - PCA on natural image color values[3] or patches[4] yields *opponent color spaces* similar to preprocessing on the retina

1 Horace B. Barlow. "Possible Principles Underlying the Transformations of Sensory Messages". In: *Sensory Communication*. Cambridge, MA: MIT Press, 1961.

2 Bruno A. Olshausen and David J. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* (1996).

3 Yuichi Ohta, Takeo Kanade, and Toshiyuki Sakai. "Color information for region segmentation". In: *Computer Graphics and Image Processing* (1980).

4 Daniel L. Ruderman, Thomas W. Cronin, and Chuan-Chin Chiao. "Statistics of cone responses to natural images: implications for visual coding". In: *Journal of the Optical Society of America A* (1998).

Neurosciences / efficient coding hypothesis

- **Efficient coding hypothesis** in neurosciences (Barlow, 1961[1])
  Efficient representation : few spikes
  Neural coding leverage signal redundancy (link with information theory)
- Similarity between efficient image algorithms and human visual processing
  e.g. :
  - Sparse coding yields filters that are *localized, oriented, selective in frequency* sharing similarity with receptive fields in V1[2]
  - PCA on natural image color values[3] or patches[4] yields *opponent color spaces* similar to preprocessing on the retina

[1] Horace B. Barlow. "Possible Principles Underlying the Transformations of Sensory Messages". In: *Sensory Communication*. Cambridge, MA: MIT Press, 1961.

[2] Bruno A. Olshausen and David J. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* (1996).

[3] Yuichi Ohta, Takeo Kanade, and Toshiyuki Sakai. "Color information for region segmentation". In: *Computer Graphics and Image Processing* (1980).

[4] Daniel L. Ruderman, Thomas W. Cronin, and Chuan-Chin Chiao. "Statistics of cone responses to natural images: implications for visual coding". In: *Journal of the Optical Society of America A* (1998).

Overcomplete sparse coding

Sparse coding of 7×7 patches, minimizing

$$\min_{a_i, \phi_i} ||I - \sum a_i \phi_i|| + \lambda S(a),$$

with $S(a)$ a sparsity term

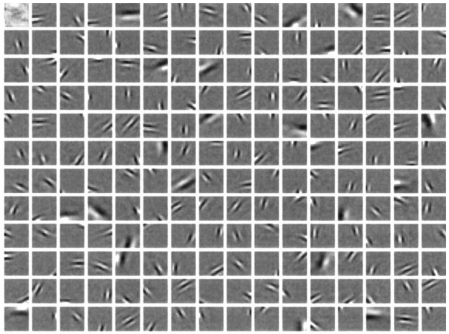


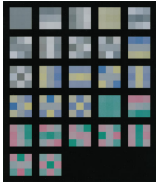
Illustration from Olshausen-Field 1996
196 basis functions obtained from 16×16 patches from natural scenes
localized, oriented and band-pass

Opponent color spaces

- PCA on pixel values
  (after conversion to a (L, M, S) space similar to CIE-RGB, see lecture from Arthur Leclaire)
  yields (roughly) the new color space :

$$(L + M + S, L + M - 2S, L - M)$$

- Same color opposition with PCA on 3×3 patches



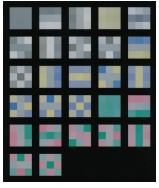
- Similar to neural pre-processing made by the retina
- Similar to the so-called *opponent color spaces* :
  Luminance, red-green, blue-yellow

Opponent color spaces

- PCA on pixel values
  (after conversion to a (L, M, S) space similar to CIE-RGB, see lecture from Arthur Leclaire)
  yields (roughly) the new color space :

$$(L + M + S, L + M - 2S, L - M)$$

- Same color opposition with PCA on 3×3 patches



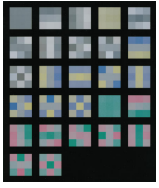
- Similar to neural pre-processing made by the retina
- Similar to the so-called *opponent color spaces* :
  Luminance, red-green, blue-yellow

Opponent color spaces

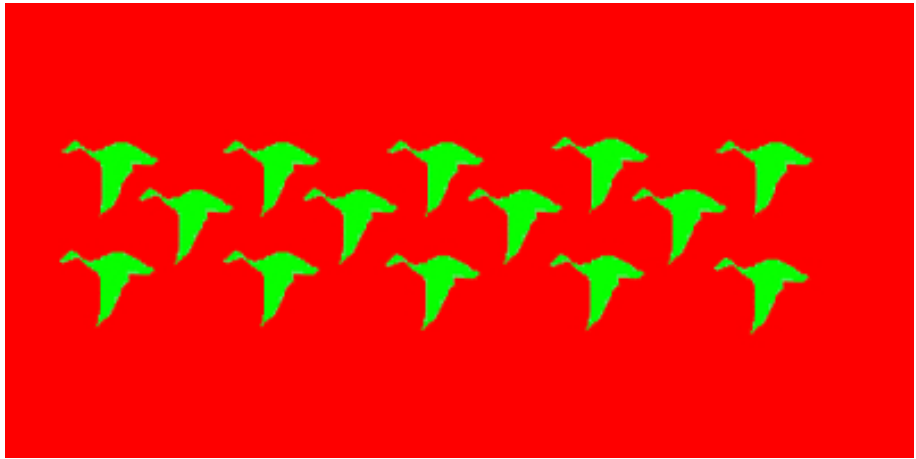
- PCA on pixel values
  (after conversion to a (L, M, S) space similar to CIE-RGB, see lecture from Arthur Leclaire)
  yields (roughly) the new color space :

$$(L + M + S, L + M - 2S, L - M)$$

- Same color opposition with PCA on 3×3 patches



- Similar to neural pre-processing made by the retina
- Similar to the so-called *opponent color spaces* :
  Luminance, red-green, blue-yellow

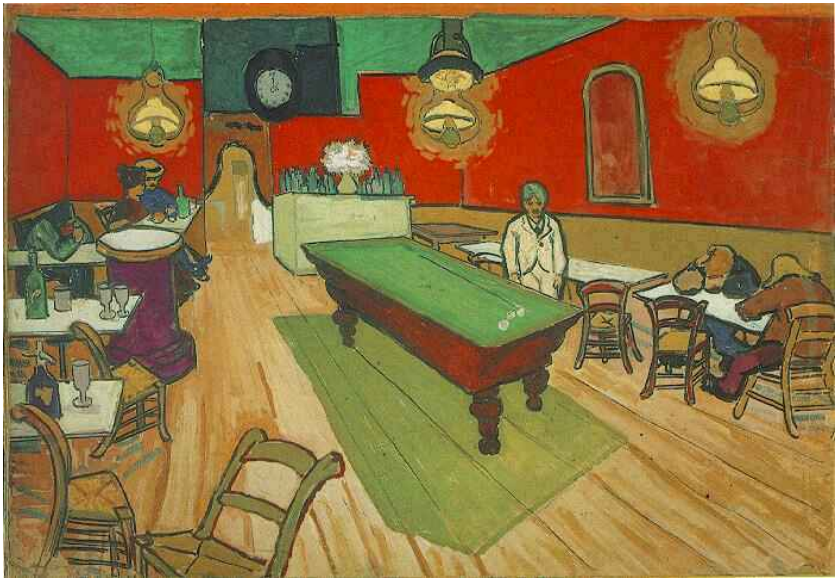


Stare at this image for about 30s, then go to the next slide

A practical consequence

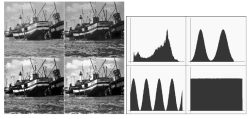


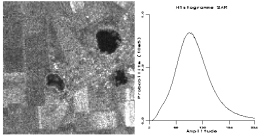
Opponent colors in art



First order statistics

- **Grey level histograms**, not informative for natural images
  different for other image modality, e.g. radar (Rayleigh $p(x) = \sigma^{-2} x \exp\left(-\frac{x^2}{\sigma^2}\right)$)





- **Color histograms** : exhibit 1D[1] or 2D[2] structures, depending on the complexity of the scene (1D for single Lambertian objects, 2D with specularity, texture, blur, etc.)

Top : RAW images, bottom JPEG images (figure from Omer 2004)

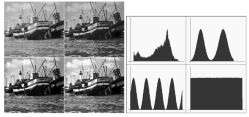
2D manifolds in the RGB space

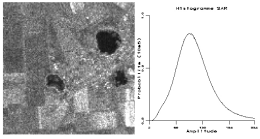
[1] Ido Omer and Michael Werman. "Color lines: image specific color representation". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2004.

[2] Antoni Buades, Jose-Luis Lisani, and Jean-Michel Morel. "Dimensionality of color space in natural images". In: *Journal of the Optical Society of America A* (2011).

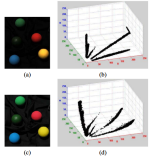
First order statistics

- **Grey level histograms**, not informative for natural images
  different for other image modality, e.g. radar (Rayleigh $p(x) = \sigma^{-2} x \exp\left(-\frac{x^2}{\sigma^2}\right)$)



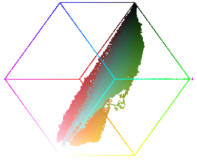


- **Color histograms** : exhibit 1D[1] or 2D[2] structures, depending on the complexity of the scene (1D for single Lambertian objects, 2D with specularity, texture, blur, etc.)



Top : RAW images, bottom JPEG images (figure from Omer 2004)





2D manifolds in the RGB space

[1] Ido Omer and Michael Werman. "Color lines: image specific color representation". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2004.

[2] Antoni Buades, Jose-Luis Lisani, and Jean-Michel Morel. "Dimensionality of color space in natural images". In: *Journal of the Optical Society of America A* (2011).

Power spectra of natural images

- For an image I, let $\hat{I}$ be its Fourier transform and $P = |\hat{I}|^2$ its power spectrum.
- Let $f(t) = \frac{1}{2\pi} \int_0^{2\pi} P(te^{i\theta}) d\theta$
- Empirically,

$$f(t) \simeq \frac{C}{t^\beta},$$

  with β close to 2.
- A fact known since the 1950's (TV signals).
- Due to the *scale invariance* of natural images :
  $I(\mathbf{x})$ and $I(k\mathbf{x})$ have the same distribution.

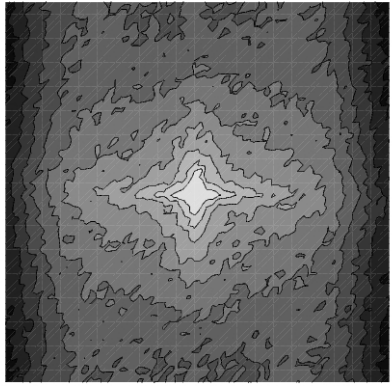


Figure 2: Contour plot of ensemble power spectrum of 45 images taken at focal length of 15mm. Center of figure is $k = 0$. Contours are placed at equal intervals in the logarithm of power, and spatial frequency is plotted on linear axes.

Figure from [Ruderman 1994][1], see also[2].

[1] Daniel L. Ruderman. "The statistics of natural images". In: *Network: Computation in Neural Systems* (1994).

[2] Daniel L. Ruderman. "Origins of Scaling in Natural Images". In: *Vision Research* (1997).

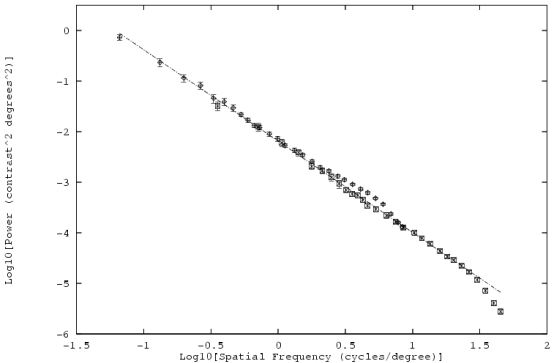
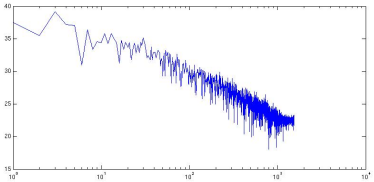


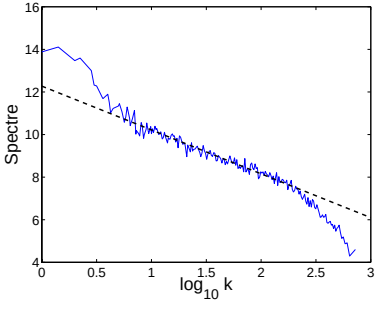
Figure 3: Orientationally averaged power spectrum with standard error bars for 15mm and 80mm focal length data (overlapping), along with regression line fitted as discussed in the text.

Figure [Ruderman 1994]

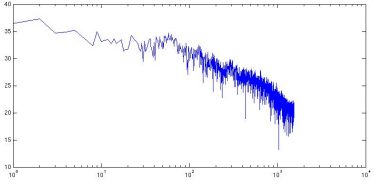



40
35
30
25
20
15
10^0
10^1
10^2
10^3
10^4

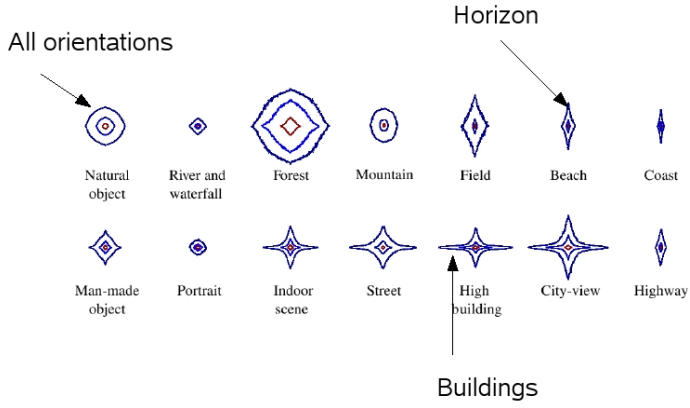



16
14
12
10
8
6
4
Spectre
0
0.5
1
1.5
2
2.5
3
log10 k




40
35
30
25
20
15
10
10^0
10^1
10^2
10^3
10^4

The shape of power spectra is related to image category
→ very quick and rough image classification



From [Torralba-Oliva 2003][1]

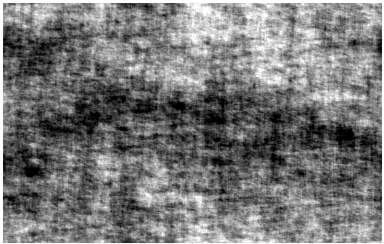
[1] Antonio Torralba and Aude Oliva. "Statistics of natural image categories". In: *Network: computation in neural systems* (2003).

Limitations

Not very informative on the structure of images:



A natural image



A synthetic image with the same spectrum

Except for some textures (as detailed later in this lecture), images are not Gaussian random fields

Covariances

- The covariance of zero-mean stationnary random field Z is defined as $\mathcal{C}(z) = E(Z(0)Z(z))$.
- The covariance of a random image I defined on a finite-domain Ω is estimated, for $z \in \Omega$, as

$$C(z) = \frac{1}{N^2} \sum_{z' \in \Omega} I(z')I(z' + z).$$

  where it is assumed that I is periodic.

- Wiener-Khintchin theorem (discrete version): let $\hat{I}$ be the discrete Fourier transform of I, then

$$|\hat{I}(\omega)|^2 = \hat{C}(\omega).$$

- Therefore, power spectrum and covariance are equivalent.
  Formal computations in the continuous case yield :
  $P(t) = Ct^{-\alpha} \leftrightarrow C(\mathbf{x}) = A + B|\mathbf{x}|^{2-\alpha}$, for $\alpha \neq 2$,
  $P(t) = Ct^{-2} \leftrightarrow C(\mathbf{x}) = A + B\log(\mathbf{x})$.

Covariances

- The covariance of zero-mean stationnary random field Z is defined as $\mathcal{C}(z) = E(Z(0)Z(z))$.
- The covariance of a random image I defined on a finite-domain Ω is estimated, for $z \in \Omega$, as

$$C(z) = \frac{1}{N^2} \sum_{z' \in \Omega} I(z')I(z' + z).$$

  where it is assumed that I is periodic.
- Wiener-Khintchin theorem (discrete version): let $\hat{I}$ be the discrete Fourier transform of I, then

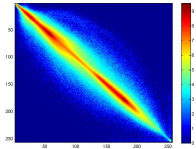
$$|\hat{I}(\omega)|^2 = \hat{C}(\omega).$$

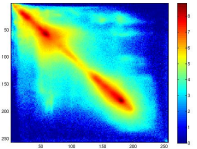
- Therefore, power spectrum and covariance are equivalent.
  Formal computations in the continuous case yield :
  $P(t) = Ct^{-\alpha} \leftrightarrow C(\mathbf{x}) = A + B|\mathbf{x}|^{2-\alpha}$, for $\alpha \neq 2$,
  $P(t) = Ct^{-2} \leftrightarrow C(\mathbf{x}) = A + B\log(\mathbf{x})$.

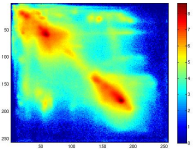
Other second order statistics : coocurrences

- Coocurrences : $B(\mathbf{x}, i, j) = Pr(I(0) = i \ \ \& \ \ I(\mathbf{x}) = j)$.



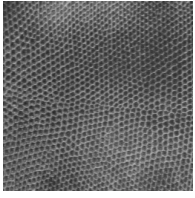


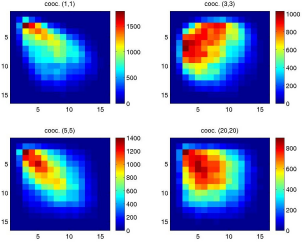




no simple parametric model,

- Useful non parametric description for some textures[1]





[1] Robert M. Haralick, Karthikeyan Shanmugam, and Its'hak Dinstein. "Textural Features for Image Classification". In: *IEEE Trans. on Systems, Man, and Cybernetics* (1973).

Other second order statistics : differences between nearby pixels

- Distribution of $I(\mathbf{x}+\mathbf{y}) - I(\mathbf{x})$ at fixed $\mathbf{y}$. Very good fit with generalized Gaussian distribution (GDD) :

$$p(x) = C_{\alpha,\eta} \exp\left(-|\eta x|^{\alpha}\right).$$



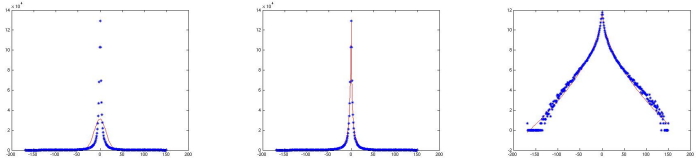
FIG. 1 – $\alpha = 0.5, \eta = 10$ FIG. 2 – $\alpha = 1, \eta = 10$ FIG. 3 – $\alpha = 2, \eta = 10$ FIG. 4 – $\alpha = 4, \eta = 10$



FIG. 5 – $\alpha = 0.5, \eta = 2$ FIG. 6 – $\alpha = 1, \eta = 2$ FIG. 7 – $\alpha = 2, \eta = 2$ FIG. 8 – $\alpha = 4, \eta = 2$

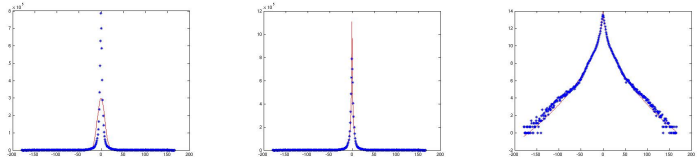
- Usual values for the ”shape” parameter : $1 \leq \alpha \leq 1$
  larger values for some textures
- Non-Gaussian distribution with sharp peaks at 0 and heavy tails



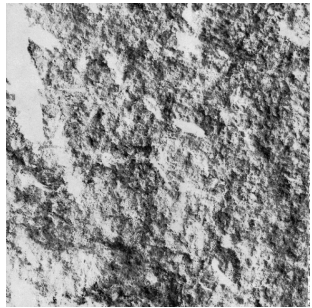


Left : best Gaussian fit
Middle : best GGD, $\alpha = 0.57$, $\eta = 0.47$ (method of moments)
Right : same on the log of the histogram





Left : best Gaussian fit
Middle : best GGD, $\alpha = 0.48$, $\eta = 1.59$ (method of moments)
Right : same on the log of the histogram

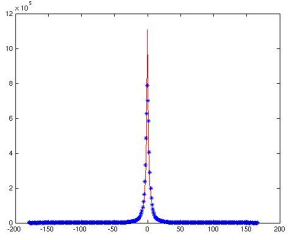


Left : best Gaussian fit
Middle : best GGD, $\alpha = 1.17$, $\eta = 0.04$ (method of moments)
Right : same on the log of the histogram

Generalized Gaussian distribution for pixel differences

- Highly non Gaussian distributions
- This behavior is mostly due to :
  - homogeneous regions (strong peak at 0), a ”blue sky” effect
  - edges (heavy tails of the distribution)
- Direct consequence of the occlusion phenomenon (most objects are opaque and hide parts of other objects).





Generalized Guassian distributions

- GDD is not only an excellent fit for pairwise pixel differences
  the same is true[1] for any filter with 0-mean !
- This is in particular true for wavelet coefficient marginals
- Applications
  - Bayesian denoising and compression[2]
  - edge detection, segmentation, object detection
  - texture classification[3]
  - variational restoration methods[4][5]
  - Modeling of heavy-tailed distribution of structures in latent spaces of deep generative models

---

[1] Christoph Zetzsche et al. “Neural network models and the visual cortex: the missing link between orientation selectivity and the natural environment”. In: *Neuroscience Letters* (1997).

[2] Eero P. Simoncelli. “Statistical Models for Images: Compression, Restoration and Synthesis”. In: *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, CA, USA: IEEE, Nov. 1997.

[3] Minh N. Do and Martin Vetterli. “Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback–Leibler Distance”. In: *IEEE Transactions on Image Processing* (Feb. 2002).

[4] Guillaume Tartavel, Gabriel Peyré, and Yann Gousseau. “Wasserstein Loss for Image Synthesis and Restoration”. In: *SIAM Journal on Imaging Sciences* (2016).

[5] Émilie Chouzenoux et al. “A Variational Approach for Joint Image Recovery and Feature Extraction Based on Spatially Varying Generalised Gaussian Models”. In: *Journal of Mathematical Imaging and Vision* (2024).

Non-Gaussianity as a quality measure of a signal decomposition

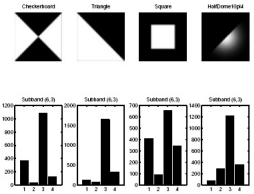


Figure 6. Kurtosis of Fourier(1), wavelet(2), ridgelet(3) and curvelet(4) coefficients; objects having only perfect straight edges, subband $2s = 6$, $j = i = 3$.

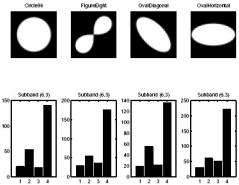


Figure 7. Kurtosis of Fourier (1), wavelet (2), ridgelet (3) and curvelet (4) coefficients; objects with curved edges, subband $2s = 6$, $j = i = 3$.

Figure from [Donoho-Flesia 2001][1]
Top : simple images Bottom : kurtosis of their decompositions in different dictionnary

More generally, parcimony[2] is a fundamental principle to guide the design of efficient image coding/reconstruction/representation strategy :
wavelet decomposition[3], sparse dictionnary[4], design of VAE[5], etc.

1 David L. Donoho and Ana Georgina Flesia. "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?" In: *Network: Computation in Neural Systems* (2001).

2 Rémi Gribonval and Elisa Riccietti. "Une brève histoire de la parcimonie: du traitement de signal à l'apprentissage profond". In: *GRETSI 2025*. 2025.

3 Stéphane G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". In: *IEEE Trans. on PAMI* (1989).

4 Michal Aharon, Michael Elad, and Alfred Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *IEEE Transactions on Signal Processing* (2006).

5 Irina Higgins et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *International Conference on Learning Representations (ICLR)*. 2017.

Non-Gaussianity as a quality measure of a signal decomposition

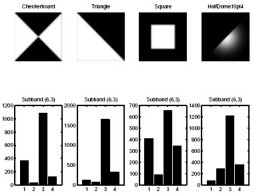


Figure 6. Kurtosis of Fourier(1), wavelet(2), ridgelet(3) and curvelet(4) coefficients; objects having only perfect straight edges, subband $2s = 6$, $j = i = 3$.

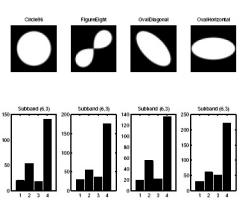


Figure 7. Kurtosis of Fourier (1), wavelet (2), ridgelet (3) and curvelet (4) coefficients; objects with curved edges, subband $2s = 6$, $j = i = 3$.

Figure from [Donoho-Flesia 2001][1]
Top : simple images Bottom : kurtosis of their decompositions in different dictionnary

More generally, parcimony[2] is a fundamental principle to guide the design of efficient image coding/reconstruction/representation strategy :
wavelet decomposition[3], sparse dictionnary[4], design of VAE[5], etc.

---

1 David L. Donoho and Ana Georgina Flesia. "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?" In: *Network: Computation in Neural Systems* (2001).

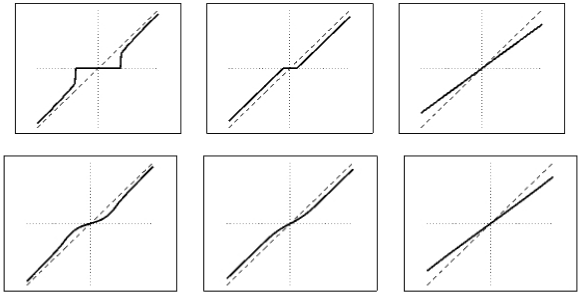
2 Rémi Gribonval and Elisa Riccietti. "Une brève histoire de la parcimonie: du traitement de signal à l'apprentissage profond". In: *GRETSI 2025.* 2025.

3 Stéphane G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". In: *IEEE Trans. on PAMI* (1989).

4 Michal Aharon, Michael Elad, and Alfred Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *IEEE Transactions on Signal Processing* (2006).

5 Irina Higgins et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *International Conference on Learning Representations (ICLR).* 2017.

Application 1 : Bayesian wavelet denoising



Transformation of the wavelet coefficients
Bayesian denoising with a generalized Gaussian model for the coefficients.
Top: maximum a posteriori (MAP); bottom: Bayesian least squares
From left to right: $\alpha = 0.5$; $\alpha = 1$; $\alpha = 2$
Noise variance = 1/3 of the signal variance
Excerpt from [Simoncelli 2005][1]

[1] Eero P. Simoncelli. "Statistical Modeling of Photographic Images". In: *Handbook of Image and Video Processing*. Ed. by Alan C. Bovik. 2nd. Burlington, MA: Academic Press, May 2005.

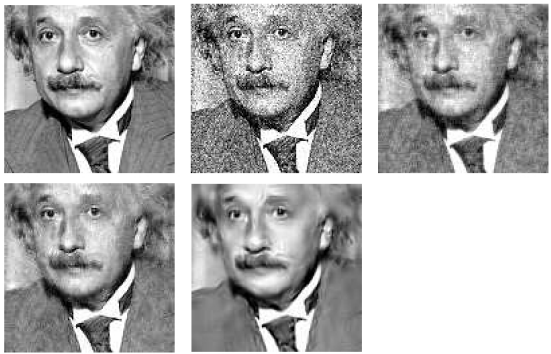


Fig. 10. Denoising examples. **Upper left:** Original 8-bit image (cropped for visibility). **Upper middle:** Contaminated by adding simulated Gaussian white noise, $\sigma = 21.4$, peak-signal-to-noise-ratio (PSNR)=22.06. **Upper right:** Denoised with a Gaussian marginal model (PSNR=27.87). **Lower left:** Denoised with generalized-Gaussian marginal model (PSNR=29.24). **Lower right:** Denoised with a GSM model (PSNR=30.86). All methods were implemented in an overcomplete wavelet domain (see [47, 39]). In each case, the noise variance was assumed known, and the denoising procedure was Bayes least-squares (i.e., the mean of the posterior distribution). Model parameters were estimated by maximizing the likelihood of the data.

Excerpt from [Simoncelli 2005]

Bayesian denoising and coefficient thresholding

We consider the denoising problem:

$$y = x + n,$$

with $n \sim N(0, \sigma_n)$, which can be written in the wavelet domain (orthonormal basis) as:

$$\tilde{w} = w + n$$

Bayesian framework:

$$p(w|\tilde{w}) \propto p(\tilde{w}|w).p(w)$$

Maximum a posteriori (MAP) estimation:

$$\begin{aligned}\hat{w} &= \arg\max_w p(w|\tilde{w}) \\ &= \arg\min_w \left[-\log p(\tilde{w}|w) - \log p(w)\right]\end{aligned}$$

Assume that:

- The coefficients follow a Gaussian distribution $w \sim N(0, \sigma_w)$, then

$$\hat{w} = \frac{\tilde{w}}{1 + \frac{\sigma_n^2}{\sigma_w^2}}$$

- The coefficients follow a generalized Gaussian distribution
  $\rightarrow$ no closed-form expression
  numerical approximation: thresholding

In practice, wavelet coefficients are not independent $\rightarrow$ need a finer model accounting for dependency.

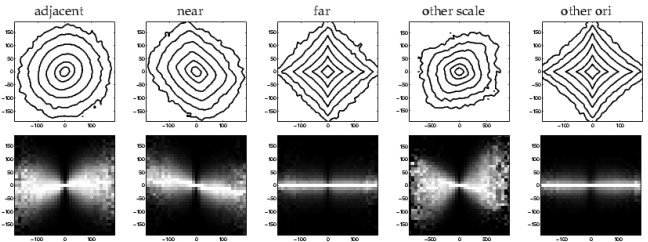


Fig. 8. Empirical joint distributions of wavelet coefficients associated with different pairs of basis functions, for a single image of a New York City street scene (see Fig. 1 for image description). The top row shows joint distributions as contour plots, with lines drawn at equal intervals of log probability. The three leftmost examples correspond to pairs of basis functions at the same scale and orientation, but separated by different spatial offsets. The next corresponds to a pair at adjacent scales (but the same orientation, and nearly the same position), and the rightmost corresponds to a pair at orthogonal orientations (but the same scale and nearly the same position). The bottom row shows corresponding conditional distributions: brightness corresponds to frequency of occurance, except that each column has been independently rescaled to fill the full range of intensities.

Excerpt from [Simoncelli 2005]

A priori on the gradient distribution and total variation

We consider the image restoration problem:

$$y = A.x + n$$

Bayesian framework:

$$p(x|y) \propto p(y|x).p(x)$$

Maximum a posteriori (MAP) estimation:

$$\begin{aligned}\hat{x} &= \arg\max_x p(x|y) \\ &= \arg\min_x \left[-\log p(y|x) - \log p(x)\right] \\ &= \arg\min_x \left[F(x,y) + R(x)\right]\end{aligned}$$

with
F: data fidelity term
R: prior $\rightarrow$ regularity
If we assume that n is i.i.d. Gaussian and a Laplacian distribution for the gradient, then

$$R(x) = \sum_i |\nabla x(i)| = TV(x),$$

then

$$\hat{x} = \arg\min_x ||A.x - y||^2 + TV(x),$$

the TV-L2 model (see A. Leclaire's lecture on restoration): .

Application 2 : Texture preserving variational restoration (Tartavel et al. 2016)

The problem:

- Recover an image u from $v = \Phi u + n$
- The degradation Φ is known (missing pixels, blur)
- Gaussian white noise n with known variance

Texture preservation:

- Visual quality
- Avoid degrading textures
- Common practice: adding noise a posteriori

Texture vs. PSNR:

Which ones are the most similar?

Application 2 : Texture preserving variational restoration (Tartavel et al. 2016)

The problem:

- Recover an image u from $v = \Phi u + n$
- The degradation Φ is known (missing pixels, blur)
- Gaussian white noise n with known variance

Texture preservation:

- Visual quality
- Avoid degrading textures
- Common practice: adding noise a posteriori

Texture vs. PSNR:

Which ones are the most similar?

Application 2 : Texture preserving variational restoration (Tartavel et al. 2016)

The problem:

- Recover an image u from $v = \Phi u + n$
- The degradation Φ is known (missing pixels, blur)
- Gaussian white noise n with known variance

Texture preservation:

- Visual quality
- Avoid degrading textures
- Common practice: adding noise a posteriori

Texture vs. PSNR:



Which ones are the most similar?

Variational restoration

Variational models:

Minimization of the cost function $E_v(u) = \underbrace{R(u)}_{\text{Regularization}} + \underbrace{F_v(u)}_{\text{Fidelity}}$

Regularization term:

- Total variation, $R(u) \propto \sum_x \|\nabla u_x\|_2$ (Rudin-Osher 1992)
- Non-local gradient, $R(u) \propto \sum_{x,x'} w_{x,x'} |u_x - u_{x'}|^2$ (Gilboa et al. 2008)
- Patch sparsity, $R(u) \propto \min_W \|P_u - DW\|^2 + \|W\|_1$ (Elad et al. 2006)

Fidelity term:

- White Gaussian noise, $F_v(u) \propto \|v - \Phi u\|^2$
- Indicator function $F_v = \mathbb{1}_C, C = \left\{u : \|v - \Phi u\|^2 \leq C\sigma^2\right\}$

Variational restoration

Variational models:

Minimization of the cost function $E_v(u) = \underbrace{R(u)}_{\text{Regularization}} + \underbrace{F_v(u)}_{\text{Fidelity}}$

Regularization term:

- Total variation, $R(u) \propto \sum_x \|\nabla u_x\|_2$ (Rudin-Osher 1992)
- Non-local gradient, $R(u) \propto \sum_{x,x'} w_{x,x'} |u_x - u_{x'}|^2$ (Gilboa et al. 2008)
- Patch sparsity, $R(u) \propto \min_W \|P_u - DW\|^2 + \|W\|_1$ (Elad et al. 2006)

Fidelity term:

- White Gaussian noise, $F_v(u) \propto \|v - \Phi u\|^2$
- Indicator function $F_v = \mathbb{1}_C, C = \left\{u : \|v - \Phi u\|^2 \leq C\sigma^2\right\}$

Variational restoration

Variational models:

Minimization of the cost function $E_v(u) = \underbrace{R(u)}_{\text{Regularization}} + \underbrace{F_v(u)}_{\text{Fidelity}}$

Regularization term:

- Total variation, $R(u) \propto \sum_x \|\nabla u_x\|_2$ (Rudin-Osher 1992)
- Non-local gradient, $R(u) \propto \sum_{x,x'} w_{x,x'} |u_x - u_{x'}|^2$ (Gilboa et al. 2008)
- Patch sparsity, $R(u) \propto \min_W \|P_u - DW\|^2 + \|W\|_1$ (Elad et al. 2006)

Fidelity term:

- White Gaussian noise, $F_v(u) \propto \|v - \Phi u\|^2$
- Indicator function $F_v = \mathbb{1}_{\mathcal{C}}$, $\mathcal{C} = \left\{u : \|v - \Phi u\|^2 \leq C\sigma^2\right\}$

Texture fidelity term

Formulation:

$$E_v(u) = R(u) + F_v(u) + \widetilde{F}_v(u)$$

With a gradient fidelity term:

$$\widetilde{F}_v(u) \propto \mathcal{W}(\mu_{\nabla u}, \mu_0)^2$$

Gradient model:

- ∇u modeled by a generalized Gaussian, $\mathrm{d}\mu_0(t) \propto e^{-\left|\frac{t}{s}\right|^p}$
- Parameters p and s estimated on the noisy image v

Texture fidelity term

Formulation:

$$E_v(u) = R(u) + F_v(u) + \widetilde{F}_v(u)$$

With a gradient fidelity term:

$$\widetilde{F}_v(u) \propto \mathcal{W}(\mu_{\nabla u}, \mu_0)^2$$

Gradient model:

- ∇u modeled by a generalized Gaussian, $\mathrm{d}\mu_0(t) \propto e^{-\left|\frac{t}{s}\right|^p}$
- Parameters p and s estimated on the noisy image v



Optimization in the TV case

Composite function:

Cost function $E = F + G + H \circ L$ with $L = \nabla$,

- $F = \widetilde{F}_v$ differentiable and non-convex
- $G = \mathbb{1}_{\mathcal{C}}$ proximable ($\mathrm{prox}_G = \mathrm{proj}_{\mathcal{C}}$)
- $H = \|\cdot\|_1$ proximable, with

$$\mathrm{prox}_{\lambda H}(x) = \arg\min_y \ \lambda H(y) + \frac{1}{2}\|y - x\|^2$$

Minimization algorithm (Condat 2013):

Iterative primal-dual algorithm:

$$\begin{cases} u_{\ell+1} = & \mathrm{prox}_{\tau G}\left(u_\ell - \tau \nabla F(u_\ell) - \tau L^* u^*_\ell\right) \\ u^*_{\ell+1} = & \mathrm{prox}_{\sigma^* H^*}\left(u^*_\ell + \sigma^* L(2u_{\ell+1} - u_\ell)\right) \end{cases}$$

No convergence guarantee: E is non-convex

Optimization in the TV case

Composite function:

Cost function $E = F + G + H \circ L$ with $L = \nabla$,

- $F = \widetilde{F}_v$ differentiable and non-convex
- $G = \mathbb{1}_{\mathcal{C}}$ proximable ($\mathrm{prox}_G = \mathrm{proj}_{\mathcal{C}}$)
- $H = \|\cdot\|_1$ proximable, with

$$\mathrm{prox}_{\lambda H}(x) = \arg\min_y \ \lambda H(y) + \frac{1}{2}\|y - x\|^2$$

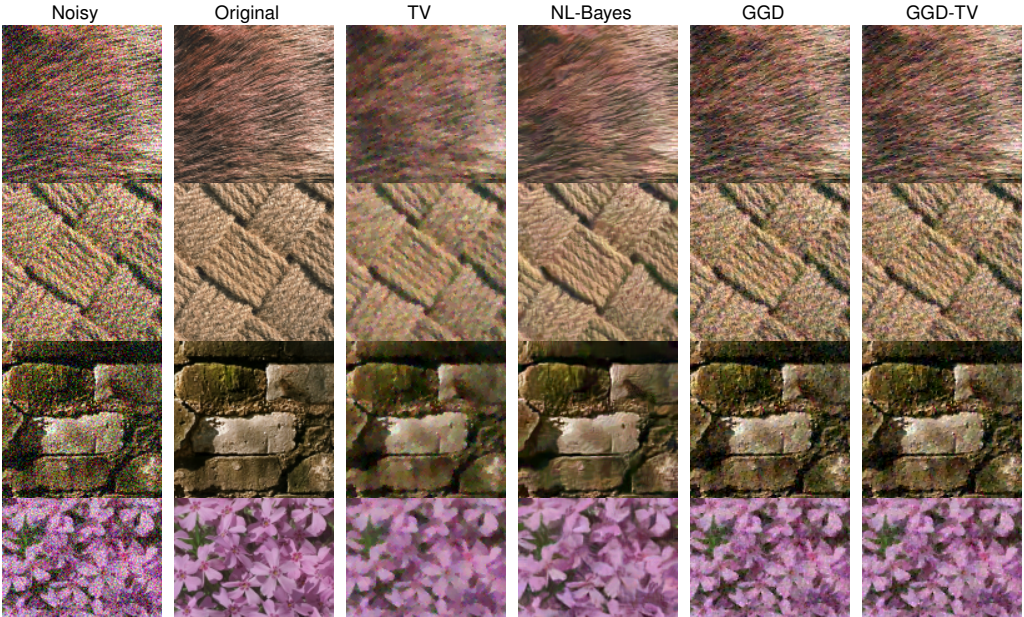
Minimization algorithm (Condat 2013):

Iterative primal-dual algorithm:

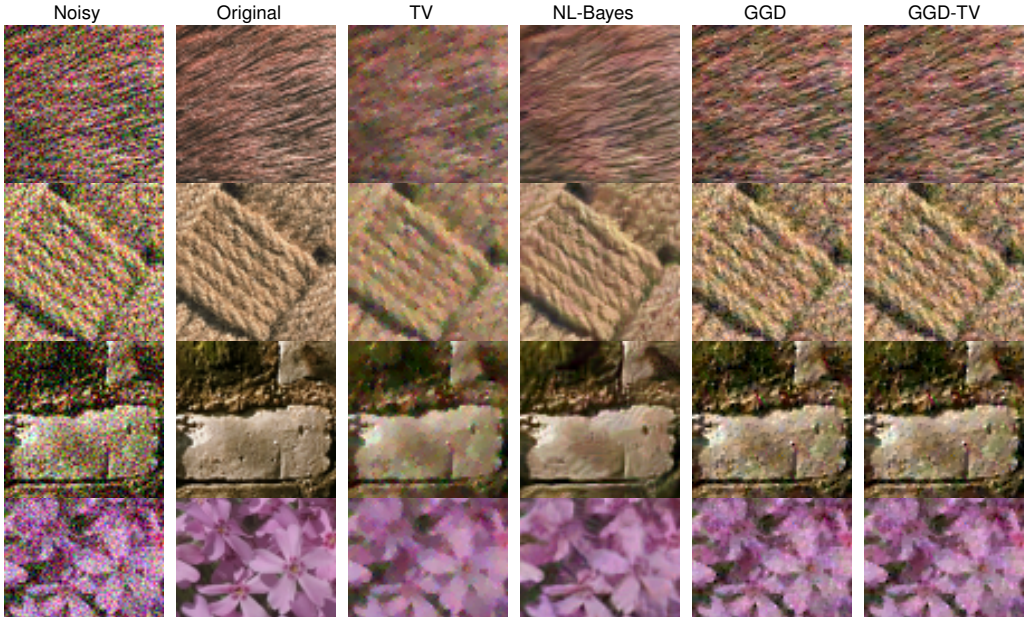
$$\begin{cases} u_{\ell+1} = & \mathrm{prox}_{\tau G}\left(u_\ell - \tau \nabla F(u_\ell) - \tau L^* u_\ell^*\right) \\ u_{\ell+1}^* = & \mathrm{prox}_{\sigma^* H^*}\left(u_\ell^* + \sigma^* L(2u_{\ell+1} - u_\ell)\right) \end{cases}$$

No convergence guarantee: E is non-convex

Denoising results



Denoising results (Zoom)



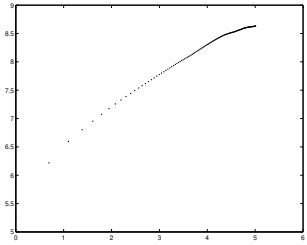
Back to power laws

- For the responses $c(y)$ of many filters at scale y, we observe behaviors of the form

$$\mathrm{var}(|c(y)|) \approx Cy^a.$$

  Examples of filters: jumps, wavelets, etc.

- Case of horizontal jumps at different distances, $S(y) = I(\mathbf{x} + (y, 0)) - I(\mathbf{x})$. We observe
  - Once again, generalized Laplacian distributions
  - A scaling law $var(S(y)) \approx Cy^a$



  - Consistent with the previous formulas for the covariance (since $var(I(0) - I(y)) = 2E(I(0))^2 - 2E(I(0)I(y))$).

Multifractal

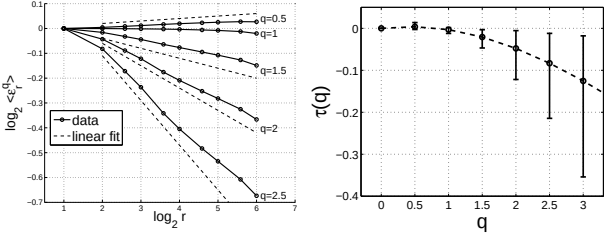
Multifractal behavior: in some cases we also observe[1][2]

$$E(|c(y)|^q) \approx C y^{\xi(q)}$$

with ξ a nonlinear function.

$$c(y) = \frac{1}{|B(0,y)|} \int_{B(0,y)} |I|$$

$$c(y) = \frac{1}{|B(0,y)|} \int_{B(0,y)} |\nabla I|$$



[Chainais 2007]

1 Antonio Turiel et al. "Self-similarity properties of natural images and self-organized criticality". In: *Physical Review Letters* (1998).

2 Pierre Chainais. "Infinitely divisible cascades to model the statistics of natural images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2007).

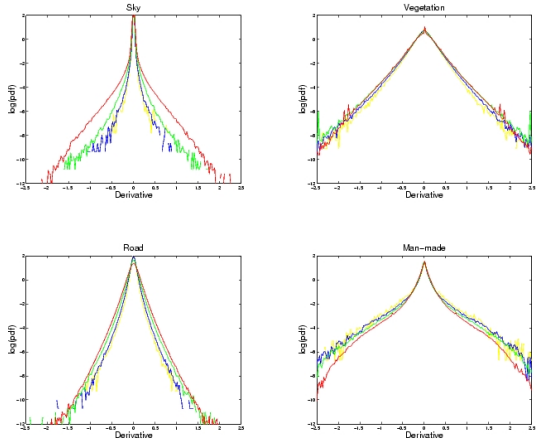


Figure 2.10: $log(pdf)$ of D at different scales and categories: red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8

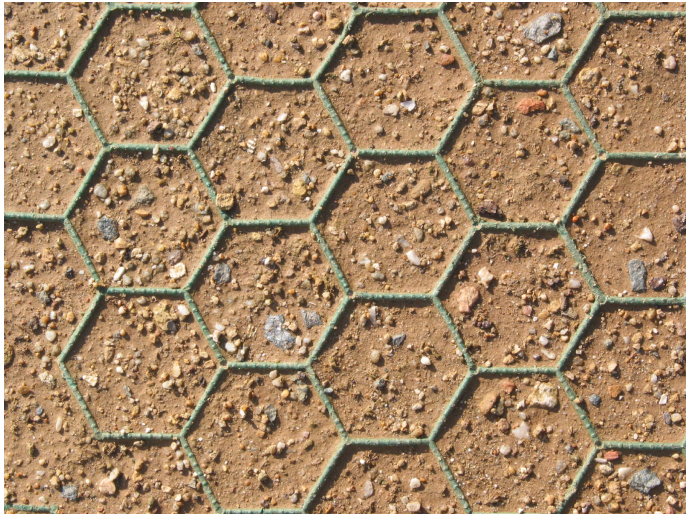
Excerpt from [Huang 2000]

Part II

Texture synthesis



Some textures



Texture perception

Our visual system can group elements into homogeneous regions within a few tenths of a second (pre-attentive vision)



From Rosenholtz 2014[1]

[1] Ruth Rosenholtz. “Texture perception”. In: *The Oxford Handbook of Perceptual Organization.* Ed. by Johan Wagemans. Oxford, UK: Oxford University Press, 2014.

Problem: which statistics are relevant to explain discrimination? [1]

- Begining of the 1970's (J. Beck) : first-order statistics on "properties" (contrast, orientation, colors)
- Same period, B. Julesz[2] : first- and second-order statistics between pointwise gray-level values

  Examples of indiscriminable pairs having the same second-order stats and different third- or fourth-order stats
- [Julesz 1981][3] : first- and second-order statistics between "textons" (structured elements, terminations, "T" junctions,...)

  Counter-example of discriminable textures having the same second-order stats (e.g. next slide)

Problem: which mathematical objects do textons correspond to?

As we are going to see, possible answers are : wavelets, sparse dictionnary atoms, learned deep features.

[1] Remark: it is the use of computers that made it possible to truly tackle this problem. In 1954, F. Attneave studied the perception of white noise ... drawn by hand with a random number generator!

[2] Bela Julesz et al. "Inability of humans to discriminate between visual textures that agree in second-order statisticsârevisited". In: *Perception* (1973).

[3] Bela Julesz. "Textons, the elements of texture perception, and their interactions". In: *Nature* (1981).

Problem: which statistics are relevant to explain discrimination? [1]

- Begining of the 1970's (J. Beck) : first-order statistics on "properties" (contrast, orientation, colors)
- Same period, B. Julesz[2] : first- and second-order statistics between pointwise gray-level values

  Examples of indiscriminable pairs having the same second-order stats and different third- or fourth-order stats
- [Julesz 1981][3] : first- and second-order statistics between "textons" (structured elements, terminations, "T" junctions,...)

  Counter-example of discriminable textures having the same second-order stats (e.g. next slide)

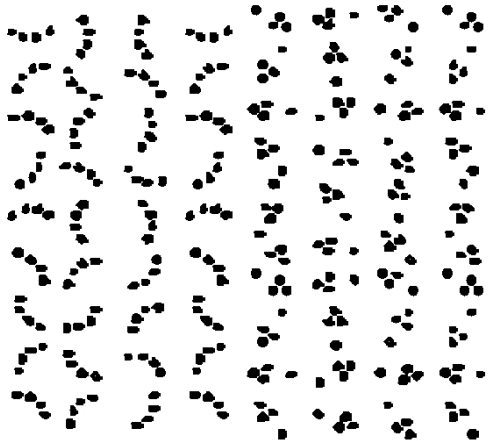
Problem: which mathematical objects do textons correspond to?

As we are going to see, possible answers are : wavelets, sparse dictionnary atoms, learned deep features.

[1] Remark: it is the use of computers that made it possible to truly tackle this problem. In 1954, F. Attneave studied the perception of white noise ... drawn by hand with a random number generator!

[2] Bela Julesz et al. "Inability of humans to discriminate between visual textures that agree in second-order statisticsârevisited". In: *Perception* (1973).

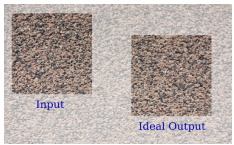
[3] Bela Julesz. "Textons, the elements of texture perception, and their interactions". In: *Nature* (1981).



Counter-example to Julesz's first conjecture

Texture synthesis

- Goal: starting from a texture image A, synthesize images that differ pointwise from A, but produce the same visual impression.
- Applications:
  - Computer graphics
  - Image enhancement and editing
  - Psychophysics
  - Testing the relevance of models
- Difficulties:
  - No completely satisfying mathematical model of textures
  - No quantitative criterion for evaluating the results



Ill. B. Galerne

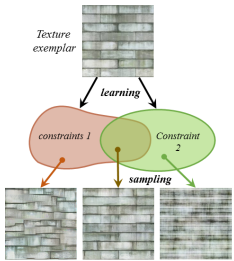


The commissar Yezhov vanishes (image www.newseum.org)

Generic models

Main families of methods:

- Synthesis by non-parametric patch-based methods
  Images obtained by resampling local neighborhoods (patches) / often copy-paste approaches
- Synthesis by statistical constraints
  Realizations of random models with statistical characteristics inspired by natural textures
  $\rightarrow$ projections onto sets of images defined by constraints



- Deep generative models

Non-parametric patch-based methods

General principle

Synthesis by patch copy-and-paste

- Consider the set of all patches of an exemplar image: D,
- The elements of D are sampled to synthesize a new texture image
- The synthesis is generally performed in a greedy manner

Efros-Leung 1999[1]

First approach (Efros-Leung 1999)

Idea:

- consider the texture as a realization of a Markov random field
- instead of sampling from a parametric model (whose parameters are learned from the texture), sample directly from an exemplar image A
- the synthesis is performed pixel by pixel, in a greedy manner
- for each new pixel of the region to synthesize, we assume that part of its neighborhood has already been synthesized, and we search A for the pixel whose neighborhood is closest.

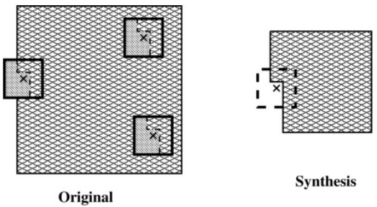


FIG. 1. *The nonparametric resampling algorithm.*

(fig: Levina-Bickel 2006)

[1] Alexei A. Efros and Thomas K. Leung. "Texture Synthesis by Non-Parametric Sampling". In: *ICCV'99*. 1999.

- From an exemplar A, we want to synthesize B
- We choose a small square window $F = \{a_1, \ldots, a_K\}$ (a patch).
- For p a pixel of A, let $N_A(p) = \{A(p + a_1), ..., A(p + a_K)\}$ be the patch placed at p
- We define a distance between patches, e.g.:

$$d_v(N_A(p), N_B(p')) = \frac{1}{K} \sum (A(p + a_i) - B(p' + a_i))^2$$

- to synthesize the value at a pixel p' of B we consider a set of **candidate pixels in** A

$$O_\epsilon(p') = \{p | d_v(N_A(p), N_B(p')) < \epsilon\}$$

  for $\epsilon > 0$.
- we draw $p \in O_\epsilon$ and assign the value $B(p') = A(p)$

To synthesize a new texture B: for each $p' \in \Omega_2$

- We compute $O_\epsilon(p')$,
- we draw a pixel p from $O_\epsilon(p')$ uniformly,
- we assign the value $B(p') = A(p)$

... but of course, we cannot compute $O_\epsilon(p')$, since the values taken by B in the neighborhood of p' are not known.
The distance d is therefore modified to take into account only the pixels of B where the synthesis has already been performed.
The synthesis then proceeds sequentially, starting from a small piece of the original image A (a few pixels).
Advantages: Excellent results, even with highly structured textures.
Drawbacks: Very slow, very sensitive to the choice of B, unstable.
$\rightarrow$ Online IPOL demo

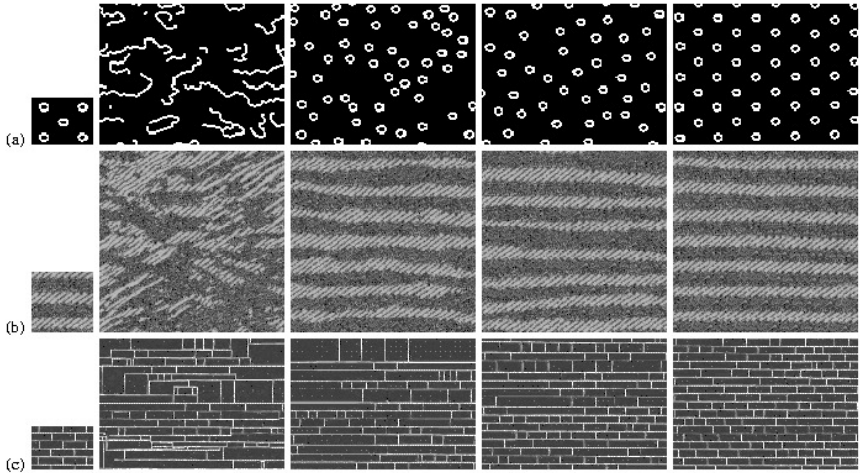
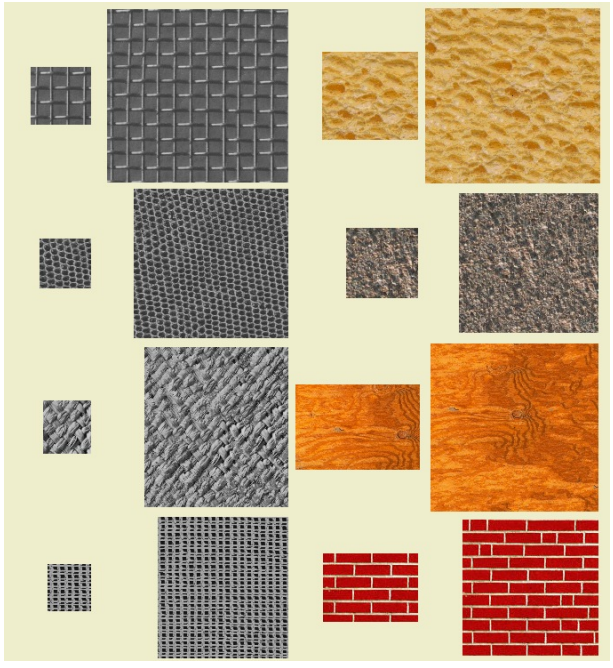
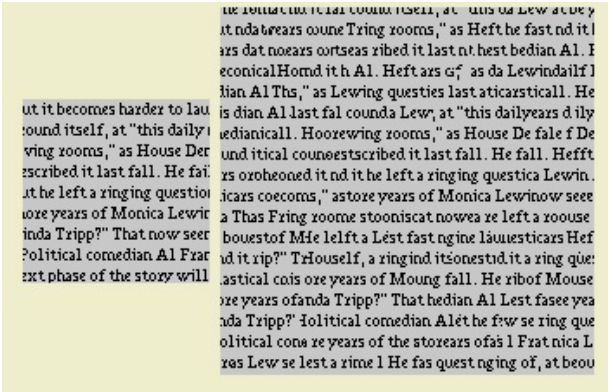


Figure 2. Results: given a sample image (left), the algorithm synthesized four new images with neighborhood windows of width 5, 11, 15, and 23 pixels respectively. Notice how perceptually intuitively the window size corresponds to the degree of randomness in the resulting textures. Input images are: (a) synthetic rings, (b) Brodatz texture D11, (c) brick wall.

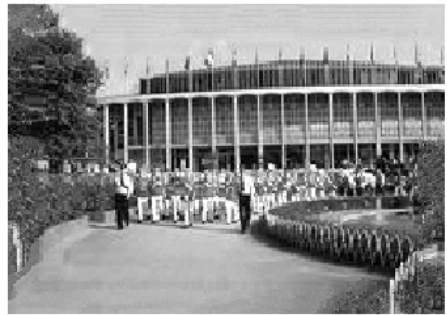
Excerpt from Efros-Leung 1999





Excerpt from Efros-Leung 1999





Excerpt from Efros-Leung 1999



(a) Input image



(b) $w = 9$



(c) $w = 15$



(d) $w = 21$

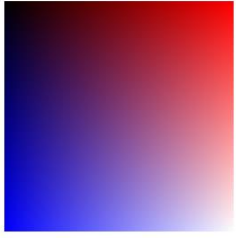
Input Image



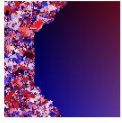
Output Image

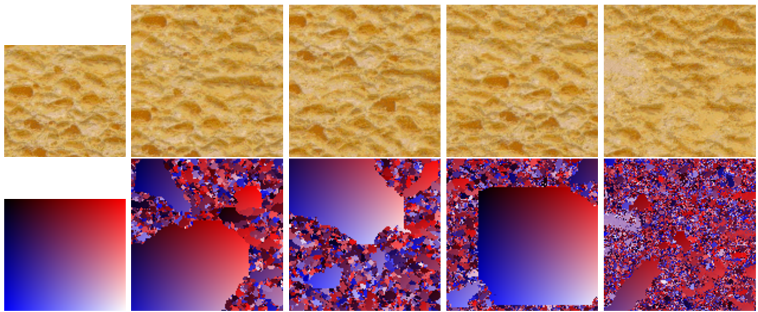


Position Map

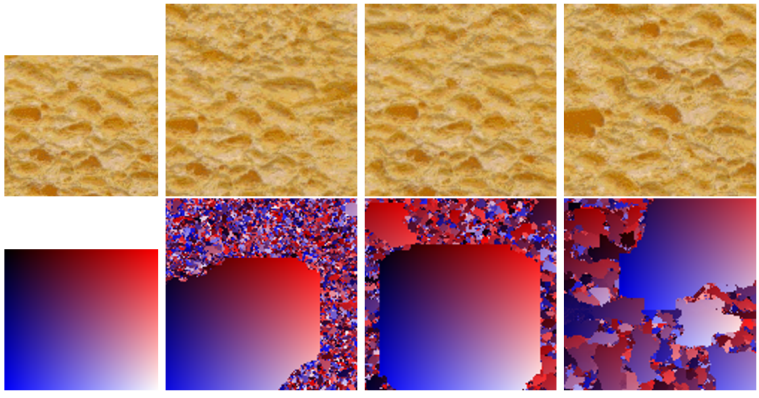


Copy Map





ϵ 0.05; 0.1; 0.3; 0.5



Patch size: 5, 7, 9

Other resampling methods

Many variants have been proposed for non-parametric texture synthesis by *patches*

- Vector quantization[1]
- Synthesis directly by patches (instead of pixels)[2]
  $\rightarrow$ IPOL demo (and next slide)
- Parallel implementations; GPU[3]
- Patch generation by Gaussian mixture modeling[4]
- Transport distance on patches[5]

[1] Li-Yi Wei and Marc Levoy. "Fast Texture Synthesis using Tree-structured Vector Quantization". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. 2000.

[2] Alexei A. Efros and William T. Freeman. "Image quilting for texture synthesis and transfer". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*. New York, NY, USA: ACM, 2001.

[3] Li-Yi Wei et al. "State of the Art in Example-based Texture Synthesis". In: *Eurographics 2009, State of the Art Report, EG-STAR*. Eurographics Association, 2009.

[4] Lara Raad, Agnès Desolneux, and Jean-Michel Morel. "A Conditional Multiscale Locally Gaussian Texture Synthesis Algorithm". In: *Journal of Mathematical Imaging and Vision* (2016).

[5] Bruno Galerne, Arthur Leclaire, and Julien Rabin. "A Texture Synthesis Model Based on Semi-discrete Optimal Transport in Patch Space". In: *SIAM Journal on Imaging Sciences* (2018).

Image quilting

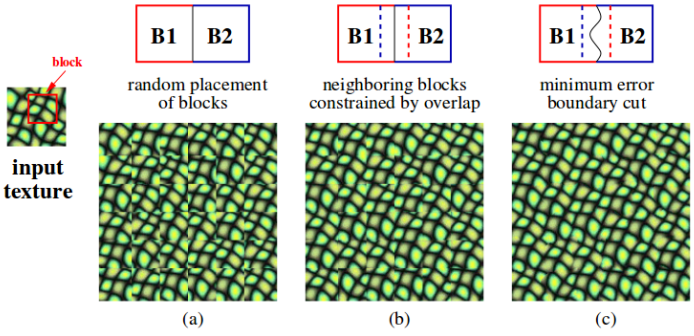
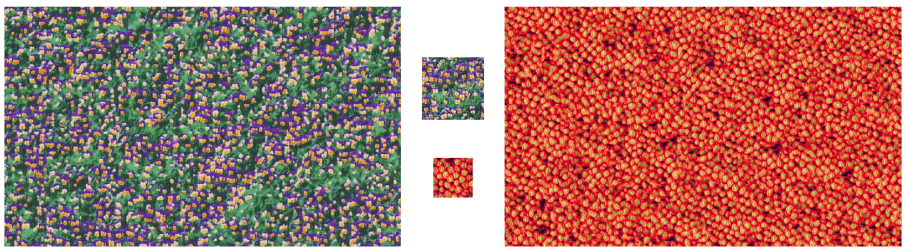


Illustration from Efros-Freeman 2001

Patches are stitched together sequentially with a stitching procedure (boundary between patches minimizing discontinuities)

Optimal transport on patches (Galerne-Leclaire-Rabin 2018)

- Initialization: Gaussian field
- Then: local transformations to force the patches of the image to follow the patch distribution learned from the exemplar
  (optimal assignment in the patch space)
- Involves solving a semi-discrete optimal transport problem on patches
- Synthesis is performed in a multiscale fashion



Illus. Galerne-Leclaire-Rabin 2018

Synthesis by statistical constraints

Synthesis by random phases (Galerne et al. 2011)[1]

- Synthesis by constraining the power spectrum
- Starting from an examplar image h_0, a sample is obtained as the image Z such that

$$\hat{Z} = |h_0| \exp(i\phi),$$

  with ϕ a random phase uniformly distributed between 0 and 2π
- This is equivalent to projecting a white noise image onto the set :

$$\{\text{images } h \text{ such that } |h| = |h_0|\}$$

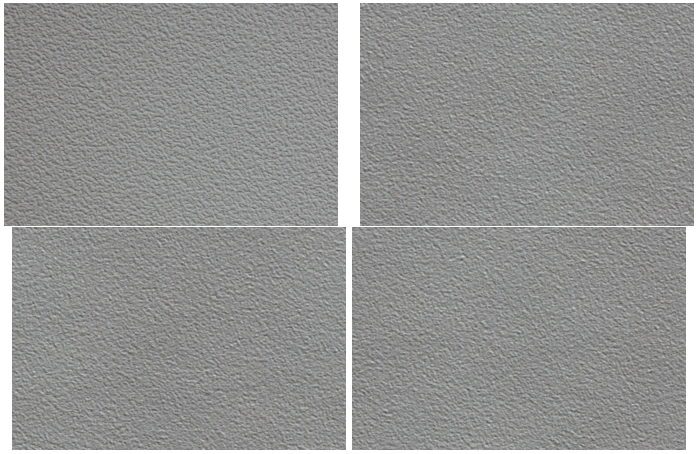
- Phase coherence between color channels is enforced by phase coherency (see next slides)

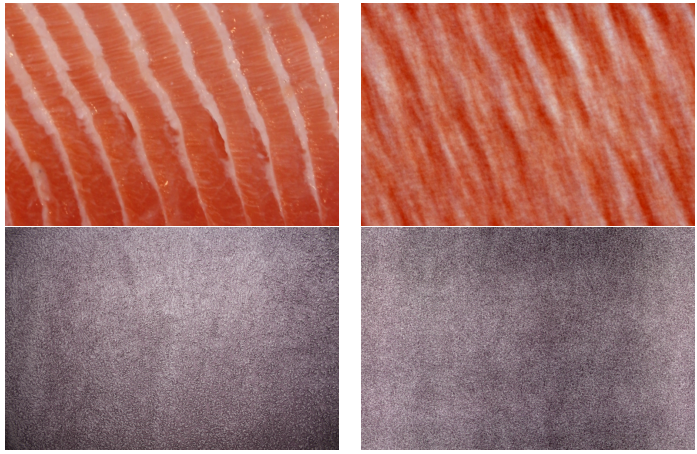
$\rightarrow$ Online IPOL demo[2]

[1] Bruno Galerne, Yann Gousseau, and Jean-Michel Morel. "Random Phase Textures: Theory and Synthesis". In: *IEEE Transactions on Image Processing* (2011).

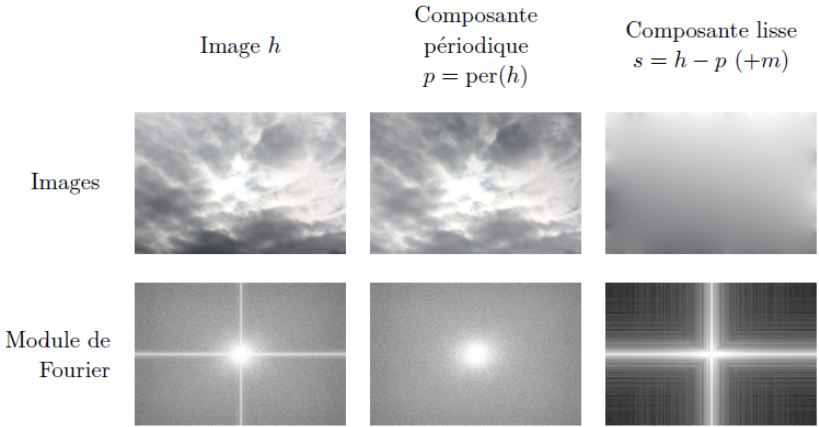
[2] Bruno Galerne, Yann Gousseau, and Jean-Michel Morel. "Micro-Texture Synthesis by Phase Randomization". In: *Image Processing On Line* (2011).





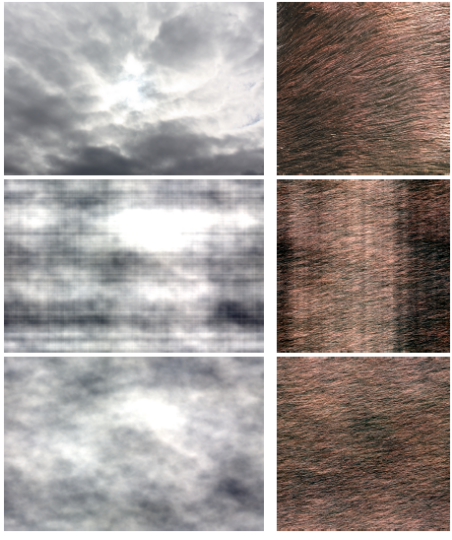


Removal of non-periodic components



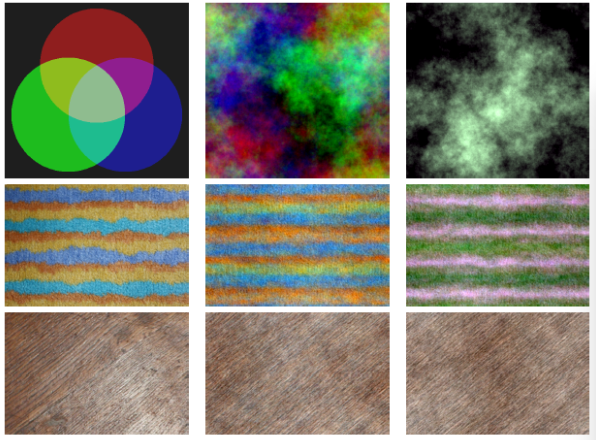
”Periodic + smooth” decomposition following [Moisan 2011]
The synthesis is obtained from the periodic component

Removal of non-periodic components



Importance of removing non-stationarities for synthesis

Color correlations



Spot h | BPA | "Mauvais" BPA

$$\hat{\boldsymbol{h}} = \begin{pmatrix} |\hat{h}_R| e^{i\varphi_R} \\ |\hat{h}_G| e^{i\varphi_G} \\ |\hat{h}_B| e^{i\varphi_B} \end{pmatrix} \qquad \hat{Z} = \begin{pmatrix} |\hat{h}_R| e^{i(\varphi_R+\theta)} \\ |\hat{h}_G| e^{i(\varphi_G+\theta)} \\ |\hat{h}_B| e^{i(\varphi_B+\theta)} \end{pmatrix} \qquad \hat{Z}_M = \begin{pmatrix} |\hat{h}_R| e^{i\theta} \\ |\hat{h}_G| e^{i\theta} \\ |\hat{h}_B| e^{i\theta} \end{pmatrix}$$

The same random phase is added to each channel
$\rightarrow$ preserves the correlations between channels

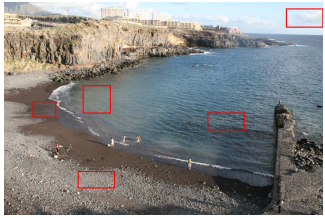
Extensions and applications

- Definition of a compact representative (texton)[1]
- Fast procedural synthesis[2]
- Initialization or constraints for more complex texture synthesis models (see previous and next slides)
- Texture mixing using optimal transport and Wasserstein barycenter[3]
- More generally, Gaussian fields play a central role for the modeling of natural 2D distributions (geosciences, remote sensing, climate studies, etc.)

1 Bruno Galerne, Arthur Leclaire, and Lionel Moisan. "Texton Noise". In: *Computer Graphics Forum* (2017).

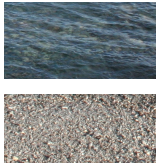
2 Ares Lagae et al. "Procedural Noise using Sparse Gabor Convolution". In: *ACM Transactions on Graphics* (2009).

3 Nicolas Bonneel et al. "Sliced and radon wasserstein barycenters of measures". In: *Journal of Mathematical Imaging and Vision* (2015).










Several regimes for the same physical texture











Another simple statistical constraint : first order statistics on wavelet coefficients (Heeger-Bergen 1995)[1]

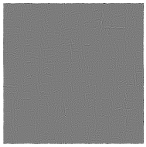
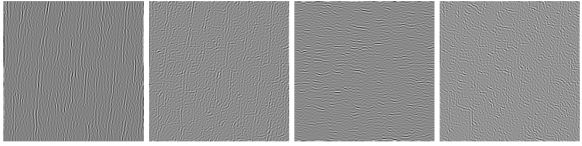
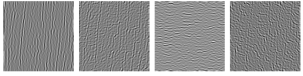
To synthesize a texture B from a sample A:

- Compute a wavelet decomposition $\{a_{i,j,k}\}$ of A: i orientation, j scale, k position,
- Start from B a white noise image, and decompose it into $\{b_{i,j,k}\}$,
- For each i, j, modify the histogram of $\{b_{i,j,k}\}_{k=1,\dots,N_{i,j}}$ to make it equal to that of $\{a_{i,j,k}\}_{k=1,\dots,N_{i,j}}$ (histogram specification, cf 2nd lecture of a. Leclaire),
- Reconstruct the image B,
- Modify the (color) histogram of B to make it equal to that of A,
- Iterate the three previous steps.

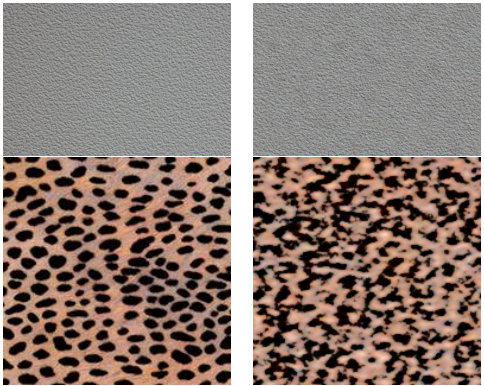
$\rightarrow$ Online IPOL demo[2]

[1] David J. Heeger and James R. Bergen. "Pyramid-based texture analysis/synthesis". In: *Proceedings of the International Conference on Image Processing (ICIP)*. 1995.

[2] Thibaud Briand et al. "The Heeger Bergen Pyramid Based Texture Synthesis Algorithm". In: *Image Processing On Line* (2014).

Original image	Associated steerable pyramid
	 High frequency residual  1st scale of oriented subbands with angle $\theta = 0,\ \frac{\pi}{4},\ \frac{\pi}{2}$, and $\frac{3\pi}{4}$  2nd scale of oriented subbands with angle $\theta = 0,\ \frac{\pi}{4},\ \frac{\pi}{2}$, and $\frac{3\pi}{4}$  Low frequency residual

In practice images are decomposed in a “steerable” pyramid basis
(Illustration Briand et al., Ipol)



Synthesis results (IPOL demo)

Method is limited to very simple micro-textures
More complex structures are characterized by dependency between responses at different positions, scales and orientations

Natural follow-up : second-order statistics on wavelet coefficents (Simoncelli-Portilla 1998, 2000)[1]

To synthesize a new image B from a sample A:

- Compute a wavelet decomposition $\{a_{i,j,k}\}$ of A (i: orientation, j: scale, k: position).
- Start from B a white noise image, and decompose it into $\{b_{i,j,k}\}$.
- Project the $\{b_{i,j,k}\}$ onto subspaces of elements having the same statistics (to be specified) as $\{a_{i,j,k}\}$.
- Reconstruct the image B.

Typically, these statistics include:

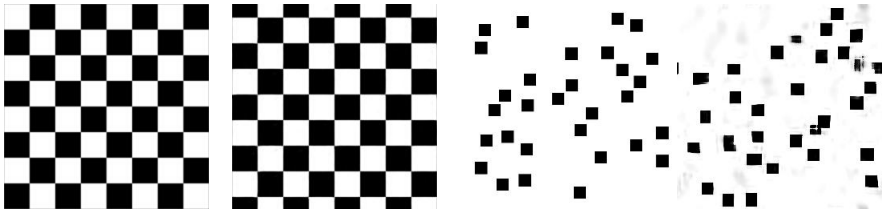
- Mean, variance, kurtosis, etc. of B.
- Covariances of $\{|b_{i,j,k}|\}$ and $\{|b_{i,j,k+p}|\}$, for given i, j, p.
- Covariances of $\{|b_{i_1,j_1,k}|\}$ and $\{|b_{i_2,j_2,k}|\}$, for given i_1, i_2, j_1, j_2.

Orders of magnitude: a few orientations (typically 4 to 6); p on the order of ten, which gives on the order of 1000 parameters.

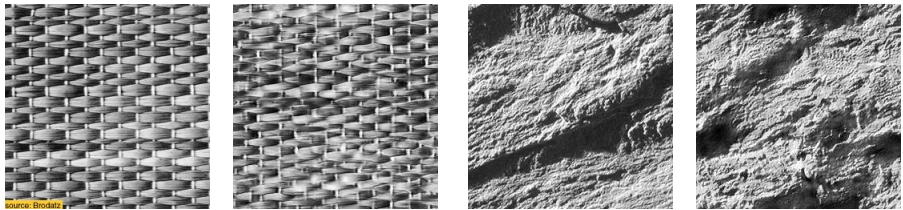
$\rightarrow$ IPOL Demo[2]

[1] Javier Portilla and Eero P. Simoncelli. "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients". In: *International Journal of Computer Vision* (2000).

[2] Jonathan Vacher and Thibaud Briand. "The Portilla-Simoncelli Texture Model: towards Understanding the Early Visual Cortex". In: *Image Processing On Line* (2021).



Matching of higher-order statistics of wavelet decompositions. Excerpt from Portilla-Simoncelli 2000



Matching of higher-order statistics of wavelet decompositions. Excerpt from Portilla-Simoncelli 2000



Matching of higher-order statistics of wavelet decompositions. Excerpt from Portilla-Simoncelli 2000

Two alternatives to wavelets

- Decomposition in a sparse dictionary
- Deep convolutional networks

Sparse decompositions

(Tartavel et al 2015)[1]

- Wavelets are replaced by a dictionary of sparse patches
- Constraints:
  - First-order statistics (marginals) on the decomposition coefficients
  - First-order statistics on the pixel color values
  - Fourier spectrum

[1] Guillaume Tartavel, Yann Gousseau, and Gabriel Peyré. “Variational Texture Synthesis with Sparsity and Spectrum Constraints”. In: *Journal of Mathematical Imaging and Vision* (2015).

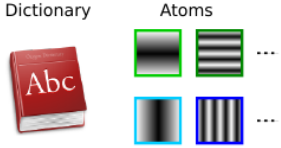
Sparse decompositions





$$P = D \times W$$

$$\mathbf{p}_k = \sum_n w_k^n \mathbf{d}^n$$



Illus. G. Tartavel

Sparse decompositions

Learning stage: non-convex, [Elad and Aharon, 2006]

Input image $v \Rightarrow$ dictionary D and weights $\widetilde{W}$

$$\min_{D,W} \|P_v - DW\| + \iota_{\mathcal{C}_s}(W)$$

Sparsity constraints:

$$\mathcal{C}_s = \{W : \|w_k\|_0 \leq S \quad \forall k\}$$

Decomposition stage:

$$\mathcal{D}(u, W) = \frac{1}{Z}\|P_u - DW\| + \iota_{\mathcal{C}_s}(W) + \iota_{\mathcal{C}_f}(W)$$

Distribution constraints:

$$\mathcal{C}_f = \{W : \|w^n\|_0 \leq \|\widetilde{w}^n\|_0 \quad \forall n\}$$



$$\mathbf{p}_k = \sum_n w_k^n \mathbf{d}^n$$

$$P = D \times W$$

Cost function:

Function to be minimized:

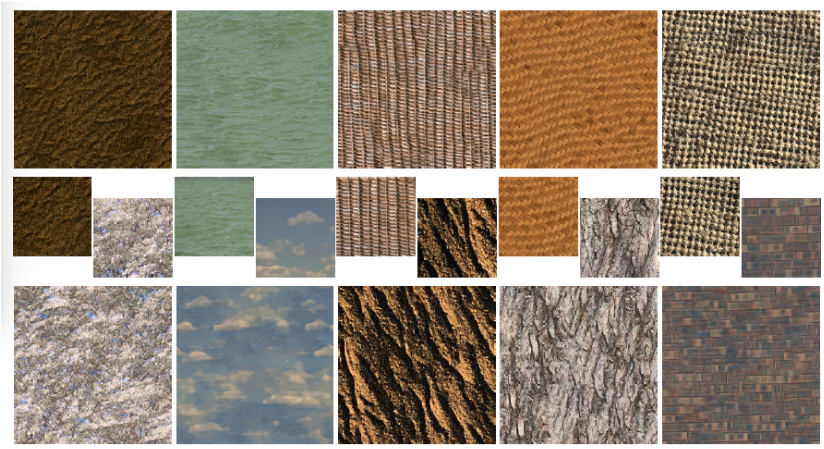
$$E_v(u, W) = \alpha \underbrace{\mathcal{W}^2(\mu_u, \mu_v)}_{\text{Histogram}} + \beta \underbrace{\mathcal{S}^2(u, v)}_{\text{Spectrum}} + \underbrace{\mathcal{D}^2(u, W)}_{\text{Decomposition}}$$

$\Rightarrow$ highly non-convex!

Minimization scheme:

- alternate descent
- w.r.t. u: gradient descent (using u_{hist} and u_{spec})
- w.r.t. W: greedy approximation derived from [Mallat and Zhang, 1993]

Sparse decompositions



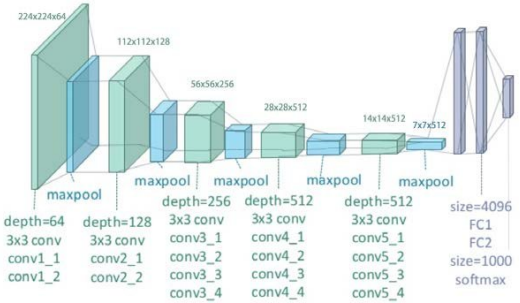
Sparse decompositions



From left to right: example, sparse method, Simoncelli-Portilla, random phase.

Synthesis by CNN

- Same general framework of statisticaly constrained synthesis
- Filters (wavelets) or sparse dictionaries are replaced by the learned filters of a deep neural network
- The retained statistics are the correlations between different features within a single layer of a neural network
- The network used is VGG19 (Simonyan-Zisserman 2014), trained for image classification



VGG architecture, from C.K. Yang et al. 2018

Deep convolutional networks (Gatys-Ecker-Bethge 2015)[1]

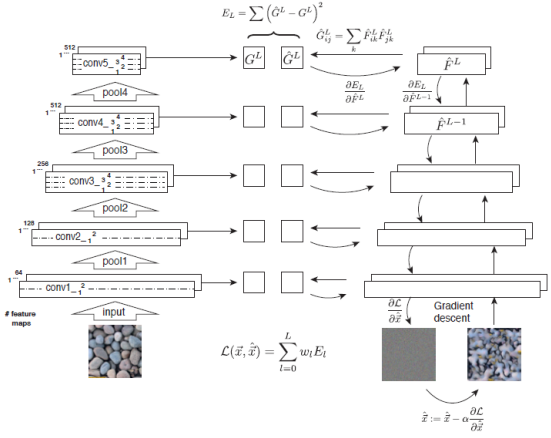


fig. Gatys et al.

Idea: impose the correlations between the different filter responses at the same position

[1] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. "Texture Synthesis Using Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems (NIPS)*. 2015.

Synthesis by CNN

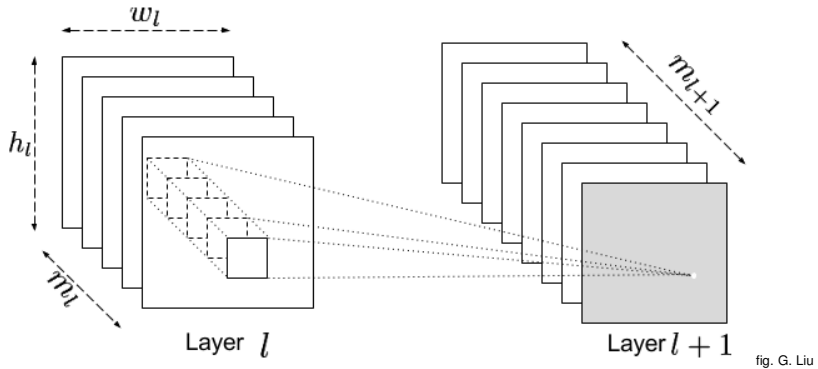


fig. G. Liu

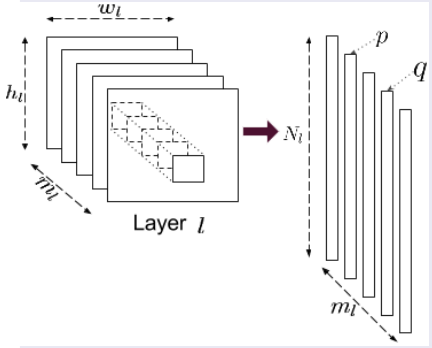
Layer l of the CNN

Texture features

For an example texture I, the feature map at layer l is $f^l \in \mathbb{R}^{m_l \times N_l}$.
m_l: number of filters, $N_l = w_l \cdot h_l$,
w_l, h_l size of the feature map.

Texture model

Inspired by Simoncelli et al.
We retain the correlation matrices $G^l \in \mathbb{R}^{m_l \times m_l}$,



$$G^l_{p,q} = f^l_p \times (f^l_q)^T = \sum_{i=1}^{N_l} f^l_p(i) \cdot f^l_q(i), \quad (1)$$

where p, q are the filter indices $p, q \in \{1, \cdots, m_l\}$.

Synthesis

We seek to synthesize an image having the same correlation matrices as the example

We start from white noise and define a cost function at layer l by

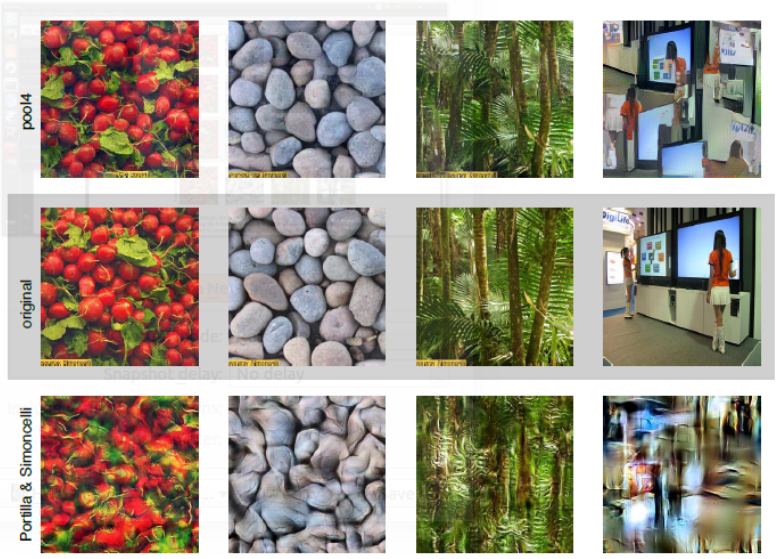
$$E_l = \frac{1}{4(N_l \cdot m_l)^2} \sum_{p=1}^{m_l} \sum_{q=1}^{m_l} \|G^l_{p,q} - \tilde{G}^l_{p,q}\|^2. \tag{2}$$

We combine several layers

$$\mathcal{L}_{cnn}(I, \tilde{I}) = \sum_{l=0}^{L} w_l E_l, \tag{3}$$

where w_l are weights
Then we minimize the cost function (backpropagation and the iterative quasi-Newton method L-BFGS)

Deep convolutional networks



Synthesis by CNNs

Synthesised



Source



Synthesis by CNNs

Synthesised



Source



Synthesis by CNNs

Synthesised



Source



Synthesis by CNNs



Synthesis artifacts

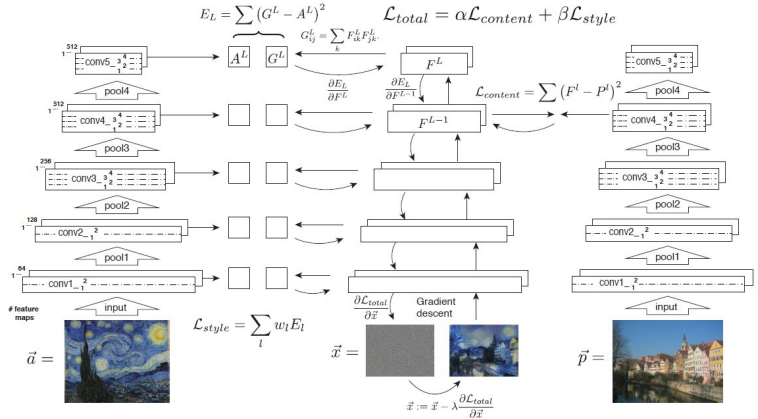
Style transfer

And also (from very similar ideas)





Style transfer



From Gatys et al. 2015

Style transfer

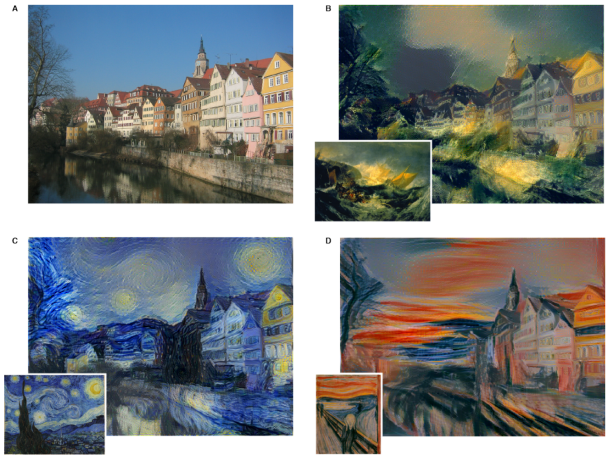


Illustration Gatys et al. 2015

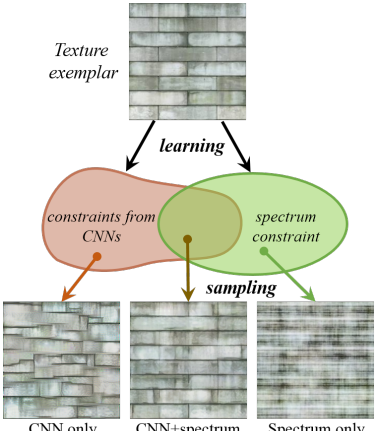
Limitation

Does not preserve long-range dependencies





A solution



Matching both the CNN covariances and the spectrum (Fourier)[1]

[1] Gang Liu, Yann Gousseau, and Gui-Song Xia. "Texture synthesis through convolutional neural networks and spectrum constraints". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE. 2016.

CNN + spectrum synthesis

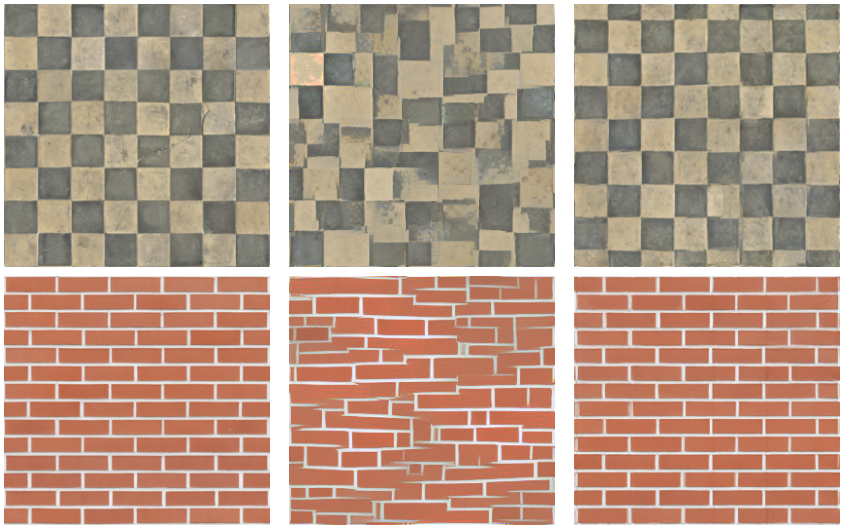


Figure: Left: original, middle: CNN, right: CNN + spectrum

CNN + spectrum synthesis

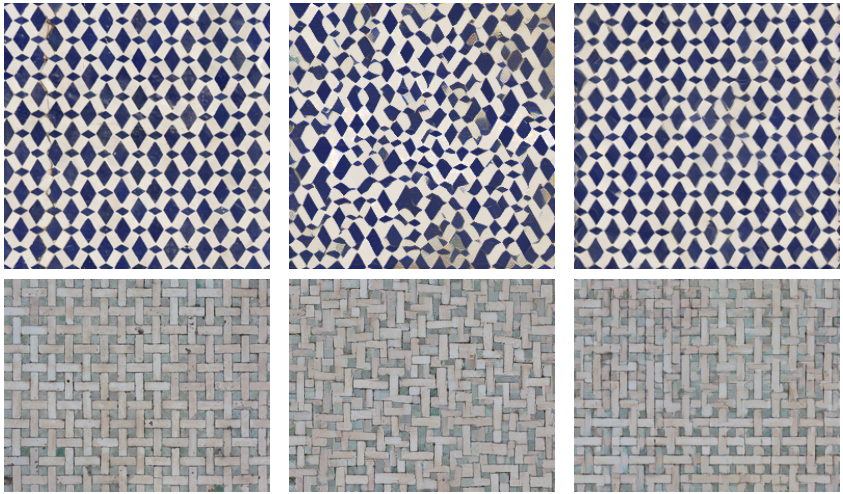
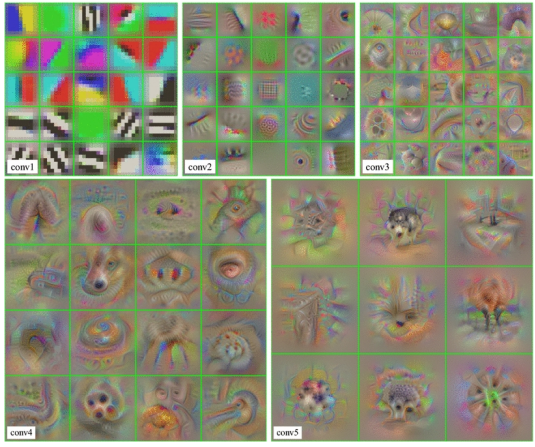


Figure: Left: original, middle: CNN, right: CNN + spectrum

Other optimization-based approaches

- Alternatives to pre-trained VGG: random weights (He et al 2016, Ustyuzhaninov et al. 2017), alternative architectures (Wang et al 2019)
- Additional or alternative statistical constraints:
  - Spectrum (Liu et al 2016, Gonthier et al 2022), intra-feature covariances (Sendik et al. 2017)
  - First-order statistics of the features (Risser 2017), joint distributions of the features (Heitz et al 2021)
- Multiscale representation (Snelgrove 2017)

And also, using similar optimization procedures,



Visualization by activation maximisation (excerpt from Mahendran et al. 2016)[1]

[1] Aravindh Mahendran and Andrea Vedaldi. "Visualizing deep convolutional neural networks using natural pre-images". In: *International Journal of Computer Vision* (2016).

Maximum entropy models

- Previous constraint-based methods: rely on **heuristics for sampling** (e.g. project, or start from a white noise image)
- Some methods optimize the diversity of samples by sampling from a distribution **maximizing the entropy**:
  FRAME Zhu et al. 1998[1], de Bortoli et al. 2021[2]
- Sampling from a Gibbs measure so that the constraints are satisfied on average (macrocanonical models)

[1] Song Chun Zhu, Yingnian Wu, and David Mumford. "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling". In: *International Journal of Computer Vision* (1998).

[2] Valentin De Bortoli et al. "Maximum entropy methods for texture synthesis: theory and practice". In: *SIAM Journal on Mathematics of Data Science* (2021).