Towards the entropy of images

Pictures in less than a thousand words

Alex Striff
Advisor: Vincent Meunier

Reed College

July 10, 2020

Outline

- What are we after?
- Information theory
- The maximum entropy method
- Toy models of image fluctuation
- The Wang-Landau algorithm
- Next steps

Information from a random variable X being $x \in \mathcal{X}$?

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$
- If $P(x) < P(x')$, $I(x) > I(x')$

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$
- If $P(x) < P(x')$, $I(x) > I(x')$
- For independent events x and y, $I(x, y) = I(x) + I(y)$

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$
- If $P(x) < P(x')$, $I(x) > I(x')$
- For independent events x and y, $I(x, y) = I(x) + I(y)$

$$I(x) = -\log P(x)$$

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$
- If $P(x) < P(x')$, $I(x) > I(x')$
- For independent events x and y, $I(x, y) = I(x) + I(y)$

$$I(x) = -\log P(x)$$

$$S(X) = \langle I_X \rangle_X$$

Information from a random variable X being $x \in \mathcal{X}$?

- If $P(x) = 1$, $I(x) = 0$
- If $P(x) < P(x')$, $I(x) > I(x')$
- For independent events x and y, $I(x, y) = I(x) + I(y)$

$$I(x) = -\log P(x)$$

$$\begin{aligned} S(X) &= \langle I_X \rangle_X \\ &= -\sum_{x \in \mathcal{X}} P(x) \log P(x) \end{aligned}$$

Inference

- A noisy measurement of an image (I) produces data (D).

Inference

- A noisy measurement of an image (I) produces data (D).

Bayes' rule

$$P(I \mid D) = \frac{P(D \mid I)P(I)}{P(D)}$$

- We infer the posterior from the likelihood and prior, and normalize by the evidence.

Inference

- A noisy measurement of an image (I) produces data (D).

Bayes' rule

$$P(I \mid D) = \frac{P(D \mid I)P(I)}{P(D)}$$

- We infer the posterior from the likelihood and prior, and normalize by the evidence.
- MAP estimate: maximize $\ln P(D \mid I) + \ln P(I)$

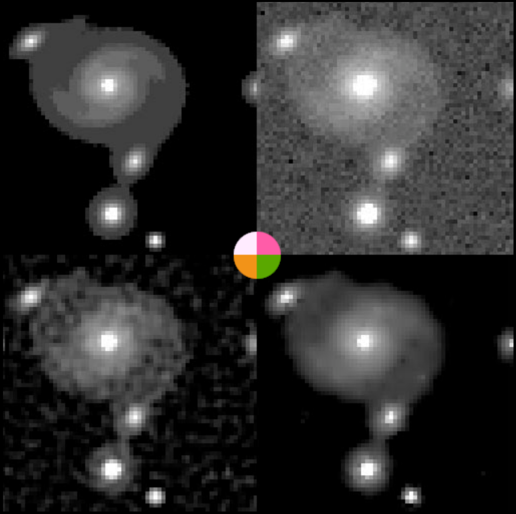
Inference

- A noisy measurement of an image (I) produces data (D).

Bayes' rule

$$P(I \mid D) = \frac{P(D \mid I) P(I)}{P(D)}$$

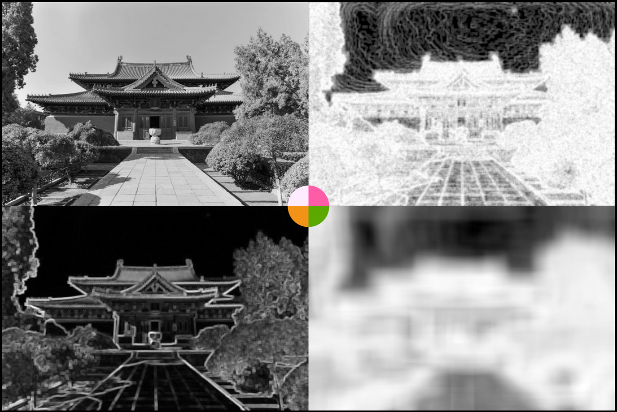
- We infer the posterior from the likelihood and prior, and normalize by the evidence.
- MAP estimate: maximize $\ln P(D \mid I) + \ln P(I)$
- Maximum entropy method: $P(I) = e^{\lambda S(I)}$



Original "Measured" MEM Multiscale MEM

E Pantin and J-L Starck. "Deconvolution of astronomical images using the multiscale maximum entropy method". In: *Astronomy and Astrophysics Supplement Series* 118.3 (1996), pp. 575–585

Intensity “Entropy”

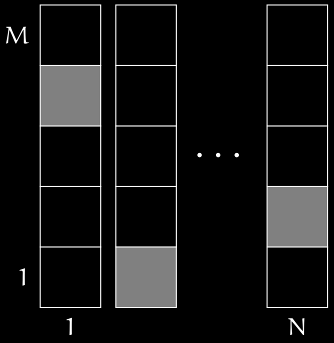


Original 5×5 IE 5×5 SD 41×41 IE

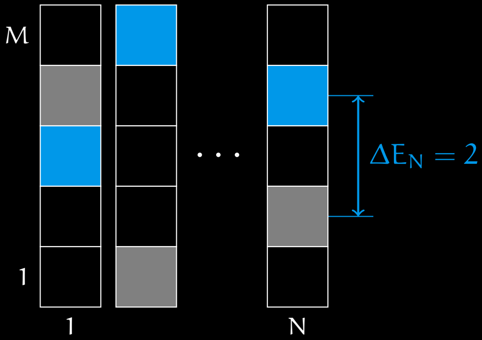
IE: Intensity entropy

SD: Standard deviation

Images as discrete-level systems

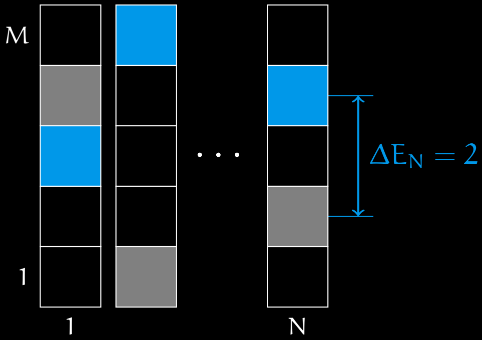


Images as discrete-level systems

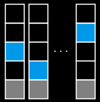


$$E = \sum_i \Delta E_i$$

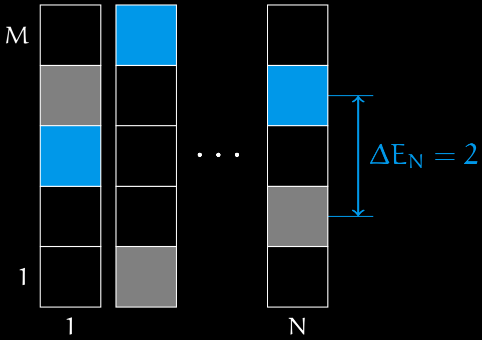
Images as discrete-level systems



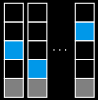
$$E = \sum_i \Delta E_i$$



Images as discrete-level systems

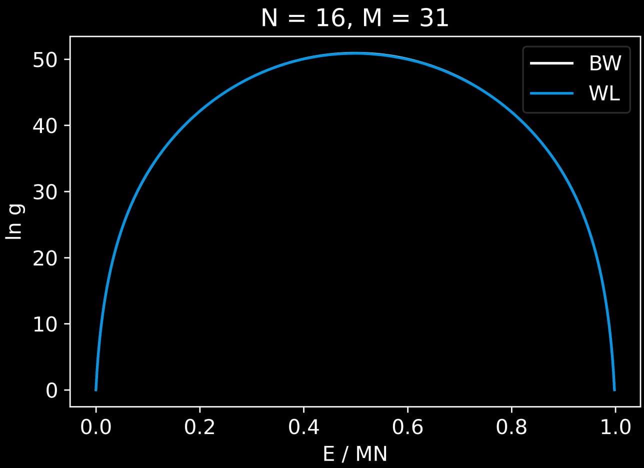


$$E = \sum_i \Delta E_i$$

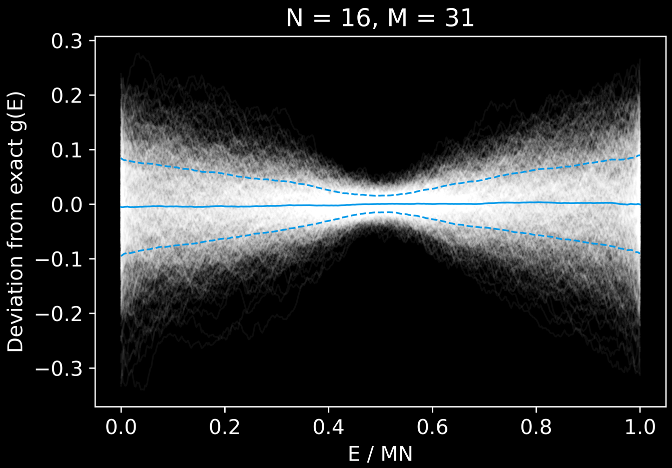


$$g(E) = \sum_k (-1)^k \binom{N}{k} \binom{N+E-kM-1}{E-kM}$$

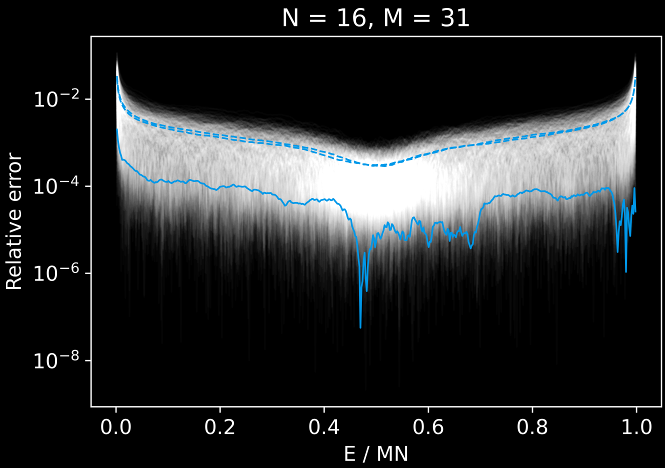
The Wang-Landau Algorithm



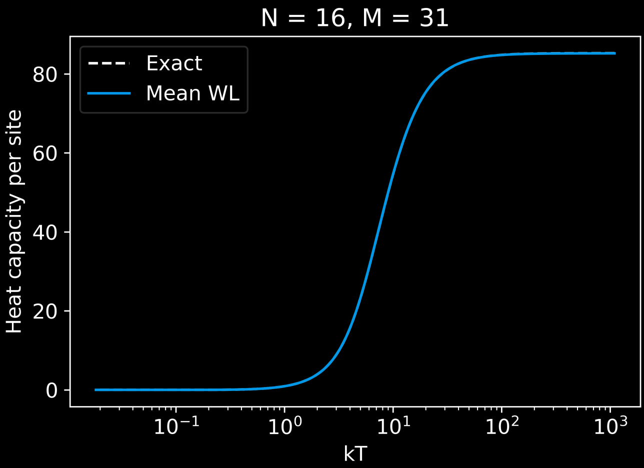
The Wang-Landau Algorithm



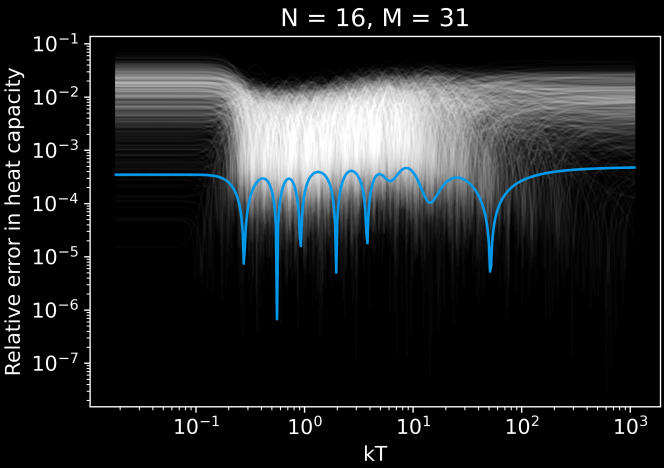
The Wang-Landau Algorithm



The Wang-Landau Algorithm



The Wang-Landau Algorithm



Next steps

- Connecting continuous and discrete
- Natural images
  - Spatial information
  - Color
  - Stereopsis
  - Motion
- High-level features like faces
- Generalization
- Inference

Acknowledgements

How to assign an *information dimension* to the random variable X?

How to assign an *information dimension* to the random variable X?

$$X_m = \frac{\lfloor mX \rfloor}{m}$$

How to assign an *information dimension* to the random variable X?

$$X_m = \frac{\lfloor mX \rfloor}{m}$$

$$d(X) = \lim_{m \to \infty} \frac{S(X_m)}{\log m}$$

How to assign an *information dimension* to the random variable X?

$$X_m = \frac{\lfloor mX \rfloor}{m}$$
$$d(X) = \lim_{m \to \infty} \frac{S(X_m)}{\log m}$$

- *Lebesgue decomposition theorem:* $d(X)$ is the fraction of P_X that is discrete.

How to assign an *information dimension* to the random variable X?

$$X_m = \frac{\lfloor mX \rfloor}{m}$$
$$d(X) = \lim_{m \to \infty} \frac{S(X_m)}{\log m}$$

- *Lebesgue decomposition theorem:* $d(X)$ is the fraction of P_X that is discrete.
- For a n-vector X with finite $H(\lfloor X \rfloor)$, $0 \leqslant d(X) \leqslant n$.

References

- Edwin T Jaynes. “Information theory and statistical mechanics”. In: *Physical review* 106.4 (1957), p. 620.
- E Pantin and J-L Starck. “Deconvolution of astronomical images using the multiscale maximum entropy method”. In: *Astronomy and Astrophysics Supplement Series* 118.3 (1996), pp. 575–585.
- Claude E Shannon. “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.