Estimation of priors in natural images

Tomoyuki Obuchi$^{1,2}$, Hirokazu Koma$^2$, and Muneki Yasuda$^3$

$^1$ Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama, 226-8502, Japan.
$^2$ Cybermedia Center, Osaka University, Osaka 560-0043, Japan.
$^3$ Graduate School of Science and Engineering, Yamagata University, Yonezawa, 992-8510, Japan.

Abstract. We investigate prior distributions in natural images by using Boltzmann machine, to find some possible universal properties and individual characteristics of natural images. For simplicity, we specifically focus on binary pictures. We find that in most cases there emerges a structure with two sublattices, and the nearest-neighbor and next-nearest-neighbor interactions correspondingly take two discriminative values, which reflects individual characteristics of each set of pictures. On the other hand, in a longer spacial scale, a longer-range (though still rapidly-decaying) ferromagnetic interaction commonly appear in all the cases. The characteristic length scale of this longer-range interaction is universally about up to four lattice spacing $\xi \approx 4$. These results are derived by using the mean-field method which effectively reduces the computational time required in Boltzmann machine. An improved mean-field method called the Bethe approximation also gives the same result, which reinforces the validity of our analysis and findings.
1. Introduction

In any learning in biological systems or machines, the existence of a certain kind of widely-applicable rules or discriminative characteristics is assumed in objects of learning, otherwise the learning will lose its meaning. Extracted characteristics through learning are expressed by a particular way in each learning system, though each way is sometimes unclear even in machines designed by humans. In this paper, we would like to focus on this expressing way in machine learning. In particular, we consider Boltzmann machine \[1\] in the context of image processing.

Bayesian framework of image processing was initiated in \[2\], and now is an active research field including several disciplines \[3, 4, 5, 6\]. To state our problem clearly, let us briefly see Bayesian framework of image restoration. Suppose we have a degraded picture consisting from \(N\) pixels or bits \(\tau\) with bipolar representation \(\tau_i = \pm 1\), and this degraded picture is generated by adding some noise to a true one \(S\). The noise flips bits in the true picture. This flipping is assumed to independently occur at each bit with a probability \(p\). Consequently, the probability of \(\tau\) is given by

\[
P(\tau|S) = \frac{e^{\beta_p \sum_i \tau_i S_i}}{(2 \cosh \beta_p)^N},
\]

where \(\beta_p \equiv (1/2) \log(1-p)/p\). The problem is to infer \(S\) for given \(\tau\). For that purpose, we need some a priori knowledge about the true picture. Let us introduce this knowledge as a probability distribution \(P_{\text{prior}}(S)\). Based on Bayes’ formula, we can write down the probability \(S\) for given \(\tau\)

\[
P(S|\tau) = \frac{e^{\beta_p \sum_i \tau_i S_i} P_{\text{prior}}(S)}{\sum_S e^{\beta_p \sum_i \tau_i S_i} P_{\text{prior}}(S)}.
\]

Eq. (2) enables us to infer \(S\) by some reasonable ways. Two typical ways exist: one is to maximize the probability (2) itself (MAP estimation), and the other is to maximize the one-site marginals derived from eq. (2) (MPM estimation). In either case, the performance of restoration is strongly influenced by \(P_{\text{prior}}(S)\).

The prior distribution is an important ingredient not only for image restoration but also for other tasks in image processing. However in many earlier researches, the prior distribution \(P_{\text{prior}}\) is designed in rather heuristic ways. One typical choice is the distribution with a uniform nearest-neighboring interaction \(\omega\)

\[
P_{\text{prior}}(S) = P_{\text{unif}}(S|w) = \frac{1}{Z} e^{w \sum_{(i,j)} S_i S_j},
\]

where \(\sum_{(i,j)}\) denotes the summation over all the neighboring bits. This rather simple prior is known to work, but of course it can be generalized to get better performance. Actually, the reference \[7\] introduces longer-range interactions and gives notable refinement in image segmentation, though the theoretical reasoning of this good performance was absent. Naively thinking, we expect better performance if our prior \(P_{\text{prior}}\) more-and-more-precisely expresses characteristics of “true” pictures in “realistic” problems. This naive consideration thus implies that the better performance of \[7\] means the longer-range interactions actually exist in “true” pictures.
Motivated by the above consideration and focusing on a certain characteristic range of interactions $\xi$, in this paper we estimate the prior distribution $P_{\text{prior}}$ by Boltzmann machine learning through several sets of natural images. We also discern between some universal properties and individual characteristics in natural images. One good candidate of universal properties is the range of the interactions $\xi$ stated above, and actually we find that $\xi$ takes a common value $\xi \approx 4$ among the different sets of pictures with different natures, which is one of main results in this paper. As other possible characteristics, we also examine frustration and criticality [4, 5, 6]. We find that in all the sets of pictures frustration is absent or quite weak, hence natures of ground states do not change by frustration. One possible consequence of the very weakness of frustration is the smooth phase space with a few minima, implying that similar pictures in the sense of hamming distance give similar values of energy. This potentially provides some intuitive interpretation of energy introduced in the Boltzmann machine. The criticality suggested in [4, 5, 6] is not observed in our study. We speculate that the criticality in [4, 5, 6] might be accidental due to the set of pictures they have used, or due to using small-size patches in the analysis. On the other hand, concerning to individual characteristics, we find that the values of the nearest-neighboring (NN) and next-nearest-neighboring (NNN) interactions can strongly depend on the set of images used in the learning. This implies that much better performance in image processing can be obtained by tuning the NN and NNN interactions in the prior distribution depending on specific problems and pictures.

The above results are obtained by treating the fully-connected Boltzmann machine in the framework of the maximum likelihood estimation. The full connectivity is needed to extract the interaction range in a fair way. Maximization of the likelihood is known to be difficult in general, and hence we use two variants of the mean-field approximations, the naive mean-field approximation and the improved one called Bethe approximation, in this paper. The consistency between these two variants supports the validity of our approximation.

The remaining part of this paper is consisting as follows. In the next section, we introduce the Boltzmann machine and state the setup of our problem. The Boltzmann machine is known to be computationally infeasible, thus we use two approximations, which will be briefly explained in the section too. In sec. 3 we display the results. Reasonability of the concepts mentioned above is also examined in this section. In sec. 4 we give discussions about the criticality and proposition of a model of the prior distribution of natural images based on our observations. The last section is devoted to a conclusion.

2. Model, Task and Methods

2.1. Boltzmann Machine

Here we explain the definition of the so-called Boltzmann machine to state our objectives. The Boltzmann machine is defined by the following energy function or the Hamiltonian

$$\mathcal{H}(S|w, h) = -\sum_{\langle i,j \rangle} w_{ij}S_i S_j - \sum_i h_i S_i,$$  \hspace{1cm} (4)
where we call $S_i = \pm 1$ bit or spin, $w_{ij}$ interaction, and $h_i$ field. The probability distribution of spins is given by

$$p(S|w, h) = \frac{1}{Z(w, h)} e^{-\mathcal{H}(S|w, h)}.$$  

The average of an observable $\hat{O}$ over this distribution is hereafter denoted as

$$\sum_S \hat{O} p(S|w, h) \equiv \langle \hat{O} \rangle_{w, h}.$$  

2.2. Learning of effective interactions in images

Let us consider the Boltzmann machine in image processing. Suppose we have $B$ pictures $\{S^{(\mu)}\}_{\mu=1,\ldots,B}$ and assume that they are generated from a certain distribution. This distribution is approximated by the empirical distribution

$$p_D(S|\{S^{(\mu)}\}) = \frac{1}{B} \sum_{\mu=1}^{B} \delta(S - S^{(\mu)}),$$  

where $\delta(\cdots)$ denotes the delta function. We write the average by this distribution as

$$\sum_S p_D \hat{O} = \langle \hat{O} \rangle_D.$$  

A typical learning scheme is formulated to reproduce the average of several observables over the empirical distribution by the one over the Boltzmann distribution

$$\langle \hat{O} \rangle_{w, h} = \langle \hat{O} \rangle_D.$$  

We seek $w, h$ to satisfy this moment-matching condition for certain several observables. The choice of observables clearly influences on the result, and hereafter we choose the first and second moments of spins, or magnetization and pairwise correlations respectively, as observables. Let us fix the notations of the relevant quantities

$$m_i = \langle S_i \rangle_{w, h}, \quad \mu_i = \langle S_i \rangle_D,$$  

$$C_{ij} = \langle S_i S_j \rangle_{w, h} - m_i m_j, \quad \Gamma^D_{ij} = \langle S_i S_j \rangle_D - \mu_i \mu_j.$$  

This choice of observables is natural since they are in a conjugate relation with $h$ and $w$. In particular, the solution of these equations, $p(S|w^*, h^*)$, can be written as the maximizer of the log likelihood

$$p(S|w^*, h^*) = \arg \max_{p, w, h} \left\{ \sum_S p_D(S|\{S^{(\mu)}\}) \ln p(S|w, h) \right\}.$$  

From this equation the moment-matching conditions with respect to magnetizations and pairwise correlations are naturally derived. The task to be solved in this paper is to infer $w$ and $h$ from several sets of natural images, and to find some characteristic features.

2.3. Mean-field Methods

To find the optimal values of $w$ and $h$, some steepest-descent-type algorithms with respect to $w$ and $h$ are typically used. At each step of changing $w$ and $h$, we need to evaluate $\langle \hat{O} \rangle_{w, h}$,
which is infeasible for large size systems since the evaluation of the partition function $Z(w, h)$ generally requires the exponentially-growing time as the system size increases. To overcome this difficulty, we employ two variants of the mean-field methods. Namely the first is the naive mean-field method (NMF) \cite{8, 9, 10} and the other is the improved mean-field method called Bethe approximation (BA) \cite{11, 12}.

The original model of Boltzmann machine \cite{5} has interactions among variables which makes difficult to calculate the partition function. To overcome this, the basic strategy of the mean-field methods is decomposing this multi-body probability distribution into a batch of effective probability distributions consisting from a few number of variables. The NMF breaks eq. (5) into a batch of single-spin distributions, while the BA approximates eq. (5) by a appropriate combination of single-spin and two-spin distributions. Here we omit the detailed descriptions and quote only the results. Readers interested in the details should refer to \cite{8, 9, 10, 11, 12, 13, 14}.

The NMF result is

\[
\begin{align*}
    w_{ij} &= -\frac{(\Gamma - 1)}{2} + \frac{\delta_{ij}}{1 - \mu^2_i}, \\
    h_i &= \tanh^{-1} \mu_i - \sum_j w_{ij} \mu_j.
\end{align*}
\]

(12)

Note that in the original model \cite{5} the self-interaction terms $w_{ii}$ have no meaning but in the inverse problem we need those terms to match the dimension of the given data $\mu$ and $\Gamma$. Besides, the BA provides the formulas

\[
\begin{align*}
    w_{ij} &= \tanh^{-1} \left\{ \mu_i \mu_j - \frac{1}{2 (\Gamma - 1)} D_{ij} \right\}, \\
    h_i &= \tanh^{-1}(\mu_i) - \sum_j \tanh^{-1}(t_{ij} f(\mu_j, \mu_i, t_{ij})),
\end{align*}
\]

(14)

where $t_{ij} = \tanh w_{ij}$ and

\[
\begin{align*}
    D_{ij} &= \sqrt{1 + 4(1 - \mu_{ij}^2)(1 - \mu_{ij}^2)} (\Gamma - 1)_{ij}^2, \\
    f(\mu_1, \mu_2, t) &= \frac{1 - t^2 - \sqrt{(1 - t^2)^2 - 4t(\mu_1 - \mu_2 t)(\mu_2 - \mu_1 t)}}{2t(\mu_2 - \mu_1 t)}.
\end{align*}
\]

(16)

The BA is expected to give more accurate results since it includes higher body correlations than the NMF, and still the computational time of eqs. (14,15) is comparable with the one of NMF (12,13). The above formula of the BA is from \cite{12}.

3. Results

In this section, we estimate $w$ and $h$ of three sets of natural images. One is of aerial photos, the second is of human faces, the other is of forests. We choose these three sets in a rather artificial way, meaning that each set looks to have some discriminative properties, such as sizes
of clusters and shapes of edges, from other two sets. However, the arbitrariness of the choice of pictures is expected not to affect the results in this paper, since we focus on some possible universal aspects in a wide range of natural images.

Each set of pictures, aerial photos, faces, and forests, are downloaded from [15], [16], and [17], respectively. The original pictures are multi-colored and are binarized by ImageMagick [18], and we assume black dot is represented by +1 and the white one is by −1. Typical examples of each set, aerial photos, faces, and forests are given in Figs. 1 -3, respectively. The empirical distribution of pictures with a fixed size is generated by cutting many binarized pictures into several square-shaped patches with an appropriate size $L$ and mixing them equally.

![Figure 1. Some typical binarized aerial photos used in learning. From [15].](image1)

![Figure 2. Some typical binarized face pictures used in learning. From [16].](image2)

For simplicity in notation, hereafter let us represent each site $i$ by the orthogonal coordinates $r_i = (x_i, y_i)$ with integers $x_i, y_i = 1, 2, \cdots$. We allocate these coordinates $(x_i, y_i)$ to pixels as Fig. 4. The interaction between site $i$ and $j$ is rewritten as $w_{ij} = w(r_i, r_j)$. We assume the distance between two sites $i$ and $j$ is defined by the Euclid manner $r_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2$. 

3.1. Inferred interactions

3.1.1. Aerial photos  Here we see the inferred interactions of aerial photos. As clarified later by comparing with the other two sets of pictures, this case is the simplest case whose interactions only take positive values.

First of all, let us see some basic behavior of $w(r_i, r_j)$ inferred by the NMF in eqs. (12,13), especially focusing on the dependency on the distance $r = |r_j - r_i|$ and on the origin $r_i$. In this context, we rewrite $w(r_i, r_j) = w(r | r_i)$. In Fig. 5, we plot $w(r | r_i)$ of common $y$-coordinate $y_i = y_j$ against $r = |x_j - x_i|$ with $x_j = x_i + 1, x_i + 2, \cdots x_i + 8$. Namely, we plot the interactions inside row as changing the column from left to right. Different several curves are shown by changing the origin $r_i = (x_i, y_i)$. The size and number of patches are $L = 16$ and $B = 100,000$, respectively. We confirm that $w(r | r_i)$ along the opposite direction, namely from right to left, shows the quite similar behavior, and the one along the vertical direction inside column does as well. Thus, Fig. 5 is a good representative. Crucial observations from Fig. 5 are:

- Most of the interactions are positive for $r \leq 4 \equiv \xi$ and almost vanish for $r > 4$.
- A clear boundary effect exists for $y_i = 1$ (blue curves), namely the absolute values of $w(r | y_i = 1)$ are relatively larger than the other values of $y_i$.
- There is a periodic behavior in the interactions, especially for the NN ones at $r = 1$. 

![Figure 3](image1.png)  Some typical binarized forest pictures used in learning. From [17].

![Figure 4](image2.png)  Correspondence between coordinates and pixels. Black and white squares correspond to black and white pixels, respectively. The origin locates at the upper left edge of the patch.
Estimation of priors in natural images

The first three characters listed above are clearer than the fourth one, and we have checked that they are common among other rows and columns and similar counterparts exist for other sets of pictures of faces and forests. Hence, analyses below are based on these three findings. The last fourth one, a weaker periodic behavior, also seems to hold for other sets of pictures, but we avoid the analysis since it is hard to treat in a systematic manner due to the weakness of tendency and the complicacy of the periodicity.

Next, we see some orientation dependency of \( w(r_i, r_j) \). We plot \( w(r_i, r_j) \) when we vary \( r_j \) as \( r_j = (x_i + s, y_i - s) \) with \( s = 1, 2, \cdots, 6 \), namely along the downward slope of the 45-degree angle from fixed \( r_i \). The distance then becomes \( r_{ij} = \sqrt{2}s \). Fig. 5 again tells that the interactions are ferromagnetic and are not long ranged: They almost vanish for \( r \geq 3 \). This length scale is shorter than the one found in Fig. 5 implying that the interactions along the horizontal or vertical lines are stronger than the ones along other inclined directions. Here, it is hard to find any clear dependency on the choice of origins \( \{r_i\} \). Thus the periodicity becomes

\[
\begin{align*}
\text{Figure 5.} & \quad \text{Plots of interactions } w(r|r_i) \text{ against distance } r = |x_j - x_i| \text{ inside a common row } y_i = y_j \text{ as changing the origin of the plot } r_i = (x_i, y_i) \text{ (the top-right inset represents the moving direction of } r_j). \text{ The interactions are inferred by the NMF using } B = 100,000 \text{ patches of the size } L = 16 \text{ created from aerial photos.}
\end{align*}
\]

Namely, the absolute values of the NN interactions for \( x_i = 2 \) and 4 are similar, as well as the ones for \( x_i = 1 \) and 3, but the former ones for \( x_i = 2 \) and 4 are larger than the latter ones for \( x_i = 1 \) and 3.

- A weaker periodic behavior seems to exist among different \( y_i \) for even \( x_i = 2 \) and 4. For example for \( x_i = 2 \), curves with \( y_i = 2, 3 \) and 6 behave similarly, and other curves of \( y_i = 4 \) and 5 do as well.

The first three characters listed above are clearer than the fourth one, and we have checked that they are common among other rows and columns and similar counterparts exist for other sets of pictures of faces and forests. Hence, analyses below are based on these three findings. The last fourth one, a weaker periodic behavior, also seems to hold for other sets of pictures, but we avoid the analysis since it is hard to treat in a systematic manner due to the weakness of tendency and the complicacy of the periodicity.
of the NN interactions of the aerial photos. This sublattice structure in the NN links is well captured by the histogram in the left panel of Fig. 7. For later convenience, in this panel we also add another sublattice of the 45-degree angle. Weaker if we go along inclined directions. The absence of direction dependency is again checked by examining the interactions along opposite directions and the ones along the upward slope of the 45-degree angle.

The above findings of periodicity implies the existence of sublattice structure represented in the left panel of Fig. 7. For later convenience, in this panel we also add another sublattice structure in the NNN interactions though such a structure is not observed in the present case of the aerial photos. This sublattice structure in the NN links is well captured by the histogram of the NN interactions \( P(w_{ij} | r_{ij} = 1) \) with discriminating the links represented by the empty and filled squares, which is given as the right panel of Fig. 7. The histograms of these two different types are differently distributing in a clear way. This evidence strongly supports the presence of the sublattice structure.

The system-size dependence of the above results, especially the characteristic length scale of the interaction range, should be examined. To this end, we define the following averaged interactions

\[
\overline{w}(r | (\hat{x}_i, \hat{y}_i)) = \frac{1}{2(L - 2)} \left\{ \sum_{x_j = x_i+1}^{L} \sum_{y_j = y_i+1}^{L-1} w(r_{ij} = |x_j - \hat{x}_i| r_i = (\hat{x}_i, y_i)) \delta (r - r_{ij}) \right. \\
+ \left. \sum_{y_j = y_i+1}^{L-1} \sum_{x_j = x_i+1}^{L} w(r_{ij} = |y_j - \hat{y}_i| r_i = (x_i, \hat{y}_i)) \delta (r - r_{ij}) \right\}. \tag{18}
\]
As we have seen so far, the behavior of $w(r|\mathbf{r}_i)$ in $r \geq 2$ is stable and does not fluctuate so much in different directions and origins, which justifies to take this average. The first sum in eq. (18) is the contributions of $w(r|\mathbf{r}_i)$ along the horizontal direction going right from $\hat{x}_i$, and the second one is the ones along the vertical direction going down from $\hat{y}_i$. Contributions from the boundaries (blue curves in Fig. 5) are excluded in eq. (18). We can see sublattices $A$ and $B$ shown in Fig. 7 by appropriately choosing the coordinates $(\hat{x}_i, \hat{y}_i)$. We plot $w_A(r) \equiv w(r|(1,1))$ and $w_B(r) \equiv w(r|(2,2))$ of different patch sizes $L = 8, 16$ and 32 in Fig. 8. The numbers of patches used in Fig. 8 are $B = 400,000$ and 23400 for $L = 8$ and 32, respectively. The error bar of each data point in Fig. 8 are defined by $\sigma(r)/\sqrt{2(L-2)}$ where $\sigma(r)$ is the standard deviation of the terms in eq. (18) from $w(r|\hat{x}_i, \hat{y}_i)$. The result clearly demonstrates the absence of the size effect, which is consistent with that the characteristic length scale of $w(r|\mathbf{r}_i)$ is $\xi \approx 4$. Patches larger than $L = 8$, which is twice of $\xi \approx 4$, are enough to reproduce our findings.
Next, we compare the results by the NMF and BA in Fig. 9. We can see the difference between the NMF and the BA is negligibly small. Thus, in our purpose to find characteristics of the interactions, the NMF seems to be enough at least for aerial pictures we have treated.

3.1.2. Face pictures  
Next we show the interactions of face pictures. Roughly speaking, we can identify three regions in each picture in this category: The background, the hair and cloth, and the face itself, as seen in Fig. 2. The face region is expressed by patterns where both black and white pixels emerge frequently and alternately, which are considered to be produced by a dither process to discriminate this region from the other ones. Presumably due to this dither process, some antiferromagnetic interactions, which are absent in the aerial photos, are observed in this case as seen below.

We first see the NMF result as the case of aerial photos. In Fig. 10, we display the interactions $w(r|\mathbf{r}_i)$ against the distance $r$ inside a column, which are corresponding to Fig. 5. Again, we see that the inferred interactions drop off up to $r \approx 4$. The boundary effect at $y_i = 1$ is also present and the NN interactions seem to have similar periodicity to the one in Fig. 7. On the other hand, a new observation in Fig. 10 is:

- Some NN interactions take negative values.

As noted in the beginning of sec. 3.1.2 these antiferromagnetic interactions are considered to emerge for expressing patterns where black and white pixels alternatively appear, which are presumably produced by a dither process to discriminate the face region from other ones.

Orientation dependency of the inferred interactions is examined by seeing $w(r|\mathbf{r}_i)$ along the downward slope as Fig. 6. The result is in Fig. 11. New observations are:

- Some NNN interactions also take negative values.
- A periodicity in the NNN interactions is present. Namely, the NNN interactions at $y_i = 1$ and 3 behave similarly while they are different from the ones at $y_i = 2$ and 4 with similar values.

To visualize the periodicities of the NN and NNN interactions, we show graph representations of those interactions in Fig. 12 with employing the fact that the signs of the interactions are
different among the sublattices, namely the positive and negative interactions are colored by blue and red lines, respectively. Fig. 12 clearly exhibits the periodicity of the NN and NNN interactions and we can see there emerges a checker-board-like structure in the interacting network, which is consistent with the sublattice structure written in the left panel of Fig. 7. Some deviations from the checker-board structure are also observed, which introduces frustration into the system. However, the number of frustrated plaquettes is not large and we expect the effect of frustration is small, meaning the nature of the ground state is simple. If we take into account the NNN interactions, frustration can be enhanced by the antiferromagnetic NNN interactions in plaquettes consisting of four antiferromagnetic NN ones, but the effect to the ground state is again expected to be weak since those antiferromagnetic NNN interactions are small in the absolute value compared to other NN and NNN interactions, which can be confirmed in the histogram of the NN and NNN interactions, \( P(w_{ij}|r_{ij} = 1) \) and \( P(w_{ij}|r_{ij} = \sqrt{2}) \), given in Fig. 13. The histograms clearly reflect the sublattice structure shown in Fig. 7. Multiple peaks observed in the interactions represented by the filled square in \( P(w_{ij}|r_{ij} = 1) \) implies another additional periodicity in the NN interactions, but we do not pursue this point to avoid complicacy as we declared in sec. 3.1.1.

Even though the frustration is weak and possibly does not affect the ground state, some metastable states can emerge due to the frustration and can influence the nature of the system. In [4], the role of those metastable states are discussed in connection with biological visual
Figure 11. Plots of interactions $w(r|r_i)$ against distance $r_{ij} = \sqrt{2} s$ with $s = 1, 2, \cdots 6$ along the downward slope of the 45-degree angle from fixed $r_i$ (the top-right inset represents the moving direction of $r_j$). By the NMF using patches with the number $B = 576,096$ and the size $L = 16$ created from face pictures.

Figure 12. Graph representations of the NN (left) and NNN (right) interactions inferred from the face pictures of the size $L = 16$ plotted in the coordinate space $r = (x, y)$. Each site corresponds to each-pixel location and the links correspond to the interactions between the sites. Blue and red links denote positive and negative interactions, respectively. Clear periodicity and a checker-board-like structure are observed.
systems based on observations that the patterns of the metastable states can be interpreted as filters selecting certain directions of edges discriminating two uniform regions. This possibility is really interesting but we do not pursue this point since enumeration of the metastable states are not easy in our case because the system size is significantly larger than $[4]$, though some additional remarks will be presented in sec. 4.1 later.

As the aerial-photos case, we examine the finite-size effects and the difference between the NMF and BA. We plot $\bar{w}_A(r)$ and $\bar{w}_B(r)$, whose definitions are the same as the case of aerial photos, for different patch sizes $L = 8, 16$ and $32$ in Fig. 14. This figure tells that the size effect is again absent and the characteristic length scale is about $\xi \approx 4$ or 5. Similar plots to compare the NMF and BA are given in Fig. 15. We again see their is almost no difference between the NMF and the BA. Thus, the findings by the NMF for the patch size $L = 16$ are expected to be widely held.

3.1.3. Forest pictures Forest pictures are consisting of a variety of clusters of different sizes and of fractal edges, which are different from aerial photos with sharp edges and from face pictures with three clearly-discriminated domains. We will see the inferred interactions in this
be improved by the refined mean-field method such as the BA.

A slight difference between the NMF and BA is observed in this case, which might be related to the face-pictures case but the interactions tend to be positive like the aerial-photos case. A

![Figure 15](image.png)

**Figure 15.** Comparison in $\overline{w_A}(r)$ (left) and $\overline{w_B}(r)$ (right) between the NMF and BA of the size $L = 16$. The difference is negligible.

case behave in-between the precious two cases, in a sense that the periodicity is the same as the face-pictures case but the interactions tend to be positive like the aerial-photos case. A slight difference between the NMF and BA is observed in this case, which might be related to the fractal nature, since the fractal nature is connected to the criticality at which the NMF can be improved by the refined mean-field method such as the BA.

The interactions inside a row inferred by the NMF are plotted in Fig. 16 as Fig. 5 and Fig. 10. The interaction range, boundary effect, and periodicities are common as the previous

![Figure 16](image.png)

**Figure 16.** Plots of interactions $w(r|r_i)$ against distance $r = |x_j - x_i|$ inside a common row $y_i = y_j$ as changing the origin of the plot $r_i = (x_i, y_i)$ (the top-right inset represents the moving direction of $r_j$). The interactions are inferred by the NMF using $B = 576,096$ patches of the size $L = 16$ created from forest pictures.
two cases. As Figs. 6 and 11 the interactions along the downward slope of the forest pictures are displayed in Fig. 17. The behavior is similar to the aerial-photos case and the periodicity

![Plots of interactions](image)

**Figure 17.** Plots of interactions $w(r|r_i)$ against distance $r_{ij} = \sqrt{2s}$ with $s = 1, 2, \ldots, 6$ along the downward slope of the 45-degree angle from fixed $r_j$ (the top-right inset represents the moving direction of $r_j$). By the NMF using patches with the number $B = 576,096$ and the size $L = 16$ created from forest pictures.

is not clearly seen in the NNN interactions. The graph representations of the NN and NNN interactions are in Fig. 18 corresponding to Fig. 12. The checker-board-like structure is again observed in the NN-interaction network but not in the NNN-interaction one, meaning that the absence of frustration. Quantitative information about the NN and NNN interactions are obtained from the histograms in Fig. 19. The sublattice structure given in Fig. 7 is again observed in the NN interactions, including the other additional periodicity in common with the face-pictures case signaled by the multiple peaks in the histogram.

The finite-size effect is examined in Fig. 20. Comparing with the previous two cases, we can see that the curves of $\overline{w_A}(r)$ and $\overline{w_B}(r)$ do not drop off completely around $r \approx 4$, and seem to be longer ranged. A finite-size effect is observed in $L = 8$ where the values of interactions at $r = 1$ and 7 are larger than the other two sizes $L = 16$ and 32. These facts imply the range of interactions for the forest pictures is relatively longer than the cases of aerial photos and face pictures, which is expected to lead the finite-size effect at $L = 8$. To examine this point, we try the following exponential fit with two parameters $a$ and $b$:

$$
\overline{w}(r) = ae^{-(r-2)/b}. 
$$

(19)
The fitting is done against the region $2 \leq r \leq 6$ for $L = 16$ and 32, the resultant curves are given
Figure 18. Graph representations of the NN (left) and NNN (right) interactions inferred from the forest pictures of the size $L = 16$ plotted in the coordinate space $r = (x, y)$. Checker-board-like structures are observed in the NN interaction network except for the boundaries.

Figure 19. Histograms of the NN (left) and NNN (right) interactions derived by the NMF from forest pictures of the patch size $L = 16$. Categorization of interactions is done based on Fig. 7, but no clear difference is observed in the NNN interactions.

Figure 20. Plots of $\overline{w}_A(r)$ (left) and $\overline{w}_B(r)$ (right) for different patch sizes $L = 8, 16$, and 32. The numbers of used patches are $B = 1, 536, 384$ and $174, 096$ for $L = 8$ and 32, respectively. The interactions tend to be longer range than the cases of aerial and face pictures, which seems to lead a finite-size effect at $L = 8$. The curves of exponential fit to the region $2 \leq r \leq 6$ are given for $L = 16$ and 32.
in Fig. 20. The estimated parameters are summarized in Table 1. At least for the eyes, the

| Sublattice | Size L | a       | b       |
|------------|--------|---------|---------|
| A          | 16     | 0.16 ± 0.02 | 1.52 ± 0.25 |
| B          | 16     | 0.15 ± 0.03 | 1.11 ± 0.28 |
| A          | 32     | 0.17 ± 0.01 | 1.56 ± 0.17 |
| B          | 32     | 0.15 ± 0.03 | 1.23 ± 0.28 |

eponential fit is quite good, and the estimated parameter \( b \) takes values around 1.1-1.6. These facts imply the interactions in this case is seemingly longer-reach but still rapidly decaying, and it is enough to consider the range \( r \leq \xi = 2 + b \approx 4 \), which is in accordance with the cases of aerial and face pictures.

The comparison between the NMF and BA is given in Fig. 21. We observe that the NMF gives larger values of \( w(r) \), which is especially clear in the NN interaction of \( w_B \). This tendency was absent for the previous two sets of pictures, and may be related to the fractal nature of the forest pictures which can be connected to the criticality leading a bad performance of the NMF.

3.2. Inferred fields

Here we report histograms of local magnetizations, \( P(m_i) \), and of local fields, \( P(h_i) \), inferred by the NMF and BA, in each set of pictures. We show the results only for \( L = 16 \) since we do not find any meaningful size effect.

Fig. 22 shows the histograms of local magnetizations of aerial photos (left), face pictures (center), and forest pictures (right). All the histograms have a simple peak structure in the positive \( m_i > 0 \) region. This implies black pixels appear more frequently in all the sets of pictures.

In Fig. 23 we show the histograms of local fields inferred by the NMF for aerial photos (left), face pictures (center), and forest pictures (right). These figures exhibit a crucial failure of
Figure 22. Histograms of the local magnetizations of aerial photos (left), face pictures (center), and forest pictures (right) for the patch size $L = 16$, normalized as probability distribution. In all the cases, the magnetizations are positive, meaning that black pixels appear more frequently.

Figure 23. Histograms of local fields of aerial photos (left), face pictures (center), and forest pictures (right), inferred by the NMF and normalized as probability distribution. Even though the magnetizations are positive, the inferred local fields by the NMF tend to be negative thus are unphysical.

the NMF: The inferred fields tend to be negative even though the corresponding magnetizations are positive. We guess this failure is due to overestimation of the effect of the interaction terms in the NMF. As well known, the NMF has a bias for orders and tends to overestimate transition temperatures, implying the NMF tends to overestimate the effect of interactions. Fortunately, the sign problem in the NMF is resolved in the BA, the corresponding histograms are given in Fig. 24.

4. Discussion

4.1. Criticality

Here we examine the possible criticality of the inferred Boltzmann machine. The set up of the investigation is as follows. We have the interactions and fields derived by the BA, $\mathbf{w} = \mathbf{w}_{\text{BA}}$ and $\mathbf{h} = \mathbf{h}_{\text{BA}}$, for each set of pictures. Using these quantities, we define a new Ising model with temperature whose probability distribution is

$$p(S|\mathbf{w}_{\text{BA}}, \mathbf{h}_{\text{BA}}, T) = \frac{1}{Z(\mathbf{w}_{\text{BA}}, \mathbf{h}_{\text{BA}}, T)} e^{-\mathcal{H}(S|\mathbf{w}_{\text{BA}}, \mathbf{h}_{\text{BA}})},$$

(20)
Figure 24. Histograms of local fields of aerial photos (left), face pictures (center), and forest pictures (right), inferred by the BA and normalized as probability distribution. The signs of the local fields and the magnetizations coincide thus the failure of the NMF is resolved in the BA.

hence the original Boltzmann machine corresponds to $T = 1$. We employ a standard Monte-Carlo technique to simulate this Ising model, which enables us to calculate physical quantities with changing the temperature. If the original Boltzmann machine is critical, some characteristic features in certain physical quantities will appear around $T = 1$.

To make the point clearer, we calculate the specific heat ($N = L^2$ is the total number of spins)

$$C = \frac{1}{NT^2} \left( \langle H^2(S|w_{BA}, h_{BA}) \rangle_{w_{BA}, h_{BA}, T} - \langle H(S|w_{BA}, h_{BA}) \rangle_{w_{BA}, h_{BA}, T}^2 \right),$$

and identify the peak location of the specific heat as the “critical” point. Here the brackets $\langle \cdots \rangle_{w_{BA}, h_{BA}, T}$ denote the average over the Ising model. This is a natural choice since the specific heat is connected to the variation ratio of the entropy as the temperature changes, and is known to actually show a characteristic divergence at the critical point in many systems. The data of the specific heat of the Ising models is shown in Fig. 25. For comparison, on the upper row we display the data with the fields set to be the values inferred by the BA, $h = h_{BA}$, while on the lower row the data without the fields, $h = 0$, is shown. As seen clearly, the peaks of specific heat locate far above $T = 1$ in all the figures, implying that the inferred Boltzmann machines are rather in lower temperature regions than at the critical point. The data without the fields shows sharp peaks and thus those Ising models really enjoy phase transitions. In addition, the Ising model for face pictures shows another complicated structure: we see not only a peak at $T \approx 3.0$ but also a shoulder-like structure at $T \approx 2.6$. The shoulder might be an indication of another transition, which can be connected to multiple transitions in the Ising model on the Union-Jack lattice [19]. This seems to be reasonable since the inferred interactions of the face pictures are similar to the ones of the Union-Jack lattice (see Fig. 12).

Our observation implies that the criticality suggested in [4] might be an accidental thing. We speculate that this is perhaps due to the choice of types of natural images (woods of Hacklebarney State Park) or due to the smallness of used patch sizes (up to $L = 4$).

Relation between the criticality and the models of images has also been reported in some other earlier researches in image processing [3]. In most of those studies, the prior distribution is assumed to have only the uniform NN ferromagnetic interaction. In that simple model, a
criticality might be observed. Our interpretation about this is as follows. Natural images have a lot of complex structures of a vast variety of sizes and they can neither be expressed in paramagnetic nor ferromagnetic phases. As a result, the inferred prior distribution becomes critical to adapt to such complex structures in natural images.

In any case, our opinion is that the criticality is a fascinating concept in explaining properties of natural images or of biological functions [4] and can be true in certain cases, but it is not always applicable. At least, to achieve efficient signal processing systems, we think the criticality would not be so useful.

4.2. Summary and model proposition

We summarize our observations so far:

- The range of interactions is about $\xi \approx 4$, and in the region $r \geq 2$ the interaction is positive and rapidly decays as $r$ grows.
- The sublattice structure shown in Fig. 7 is widely present. Correspondingly, the NN and NNN interactions have each (at least) two types.
- Boundary effects exist and tend to give larger values of interactions than the bulk ones.
- Frustration is absent, or quite weak even if it exists.
- The NMF strongly underestimates the local fields, which is resolved by the BA.
- The criticality is absent for all the sets of images.
Based on these findings, we propose a model of the prior distribution of natural images with six parameters:

\[
    w_{ij} = w(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} 
        w^1_{NN}, & ((\mathbf{r}_i, \mathbf{r}_j) \leftrightarrow \text{blank square}) \\
        w^2_{NN}, & ((\mathbf{r}_i, \mathbf{r}_j) \leftrightarrow \text{filled square}) \\
        w^1_{NNN}, & ((\mathbf{r}_i, \mathbf{r}_j) \leftrightarrow \text{blank circle}) \\
        w^2_{NNN}, & ((\mathbf{r}_i, \mathbf{r}_j) \leftrightarrow \text{filled circle}) \\
        ae^{-|\mathbf{r}_i - \mathbf{r}_j|^2/b}, & \text{(otherwise)}
    \end{cases}
\]  

These six parameters can be adaptively determined depending on specific problems, but we admit it is not always easy. For convenience in such situations, we display example values of these parameters estimated through our Boltzmann machine learning for \(L = 16\) in Table 2. In particular, the NN and NNN interactions, \(w_{NN}\) and \(w_{NNN}\), are the average values of \(w_{ij}\) over the corresponding histograms shown in Figs. 7, 13, and 19. The parameters \(a\) and \(b\) are estimated through the fitting based on eq. (19), although the data of aerial photos and face pictures seems to drop off more rapidly than the exponential decay and the fitting quality is not so good. Two different values of \(a\), and also of \(b\), corresponding to two sublattices are averaged to give the values in Table 2. As observed, \(w^1_{NNN}\) and \(w^2_{NNN}\) are similar for aerial and forest pictures, in which it is hard to see clear periodicity in the NNN interactions. This prior with estimation of the parameters is the main result of the present paper.

|        | Aerial | Face | Forest |
|--------|--------|------|--------|
| \(w^1_{NN}\) | 0.07   | -0.85 | -0.03  |
| \(w^2_{NN}\) | 0.32   | 0.2  | 0.43   |
| \(w^1_{NNN}\) | 0.24   | -0.14 | 0.3    |
| \(w^2_{NNN}\) | 0.22   | 0.4  | 0.37   |
| \(a\)       | 0.1    | 0.3  | 0.16   |
| \(b\)       | 0.7    | 1.1  | 1.3    |

Table 2. Model parameters estimated by our Boltzmann machine learning of natural images of \(L = 16\) by the NMF.

over the corresponding histograms shown in Figs. 7, 13, and 19. The parameters \(a\) and \(b\) are estimated through the fitting based on eq. (19), although the data of aerial photos and face pictures seems to drop off more rapidly than the exponential decay and the fitting quality is not so good. Two different values of \(a\), and also of \(b\), corresponding to two sublattices are averaged to give the values in Table 2. As observed, \(w^1_{NNN}\) and \(w^2_{NNN}\) are similar for aerial and forest pictures, in which it is hard to see clear periodicity in the NNN interactions. This prior with estimation of the parameters is the main result of the present paper.

5. Conclusion

In this paper, we investigated the prior distributions of natural images by employing the Boltzmann machine. We prepared three sets of different pictures: Namely, aerial photos, face pictures, and forest pictures. To reduce the huge computational time of learning process, we used the NMF which enables us to treat relatively large patch sizes up to \(L = 32\). The result is stable against the change of system size if the linear size \(L\) is larger than or equal to \(L = 16\). The refined mean-field method, the BA, is also employed to check the validity of the naive mean-field method, and we found that the NMF result is reliable for the interactions among the sets of pictures we studied.
As individual characters to each set of pictures, we found that the NN and NNN interactions strongly depend on the set of pictures. Both negative and positive values can appear for those NN and NNN interactions. On the other hand, as universal aspects, we observed that the inferred interactions are essentially short-range. For longer distance than $r \geq 2$, the interactions basically are positive and decay rapidly among all the sets of pictures. The characteristic length scale is about $\xi \approx 4$. Some simple periodic behavior are also observed in all the cases. Summarizing these properties, we propose a model prior distribution with six parameters at most. It will be an interesting future work to examine the performance of this model distribution in image processing tasks such as image restoration.

As another topic, we also examined the concepts of frustration and criticality. The frustration are present for the interactions inferred from the face pictures but absent for other two sets of pictures. The criticality is absent for all the cases, which is contrast to some earlier works [4, 5, 6]. Based on this observation, we speculated that the criticality can appear for small patches of images or for the Boltzmann machine with a few parameter, but is not always applicable to processing of natural images.

Image processing is not a simple theme, and can be treated from different motivations such as biological modeling, efficiency in information processing, or getting new physical ideas [20]. Among those motivations, in a sense we analyzed how natural images are seen by the Boltzmann machine in this paper. Such study, treating a theoretical model as a really-existing experimental system and investigating it, will give some hints for all the above motivations, and will increase more and more in future. We hope our present study provides some perspectives to those researches.
[1] Ackley D H, Hinton G E and Sejnowski T J 1985 A Learning Algorithm for Boltzmann Machines Cognitive Science 9 147-169
[2] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images IEEE Trans. PAMI 6 721-741
[3] Tanaka K 2002 Statistical-mechanical approach to image processing (Topical review) J. Phys. A: Math. Gen. 35 81-150
[4] Stephens G J, Mora T, Tkačik G and Bialek W 2013 Statistical Thermodynamics of Natural Images Phys. Rev. Lett. 110 018701
[5] Ruderman D L and Bialek W 1994 Statistics of natural images: Scaling in the woods Phys. Rev. Lett. 73 814-817
[6] Ruderman D L 1994 The statistics of natural images Network: Computation in Neural Systems 5 517-548
[7] Krähenbühl P and Koltun V 2011 Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials In Proc. NIPS 109-117
[8] Peterson C and Anderson J R 1987 A Mean Field Theory Learning Algorithm for Neural Networks Complex Systems 1 995-1019
[9] Hinton G E 1989 Deterministic Boltzmann Learning Performs Steepest Descent in Weight-Space Neural Computation 1 143-150
[10] Kappen H J and Rodríguez F B 1997 Efficient learning in Boltzmann Machines using linear response theory Neural Computation 10 1137
[11] Yasuda M and Tanaka K 2012 Approximate Learning Algorithm in Boltzmann Machines, Neural Computation 21(11) 3130-3178
[12] Ricci-Tersenghi F 2012 The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods J. Stat. Mech. P08015
[13] Opper M and Saad D 2001 Advanced mean field methods: theory and practice (The MIT press)
[14] Mezard M and Montanari A 2009 Information, Physics and Computation (Oxford University Press)
[15] Database of Geospatial Information Authority of Japan: http://www.gsi.go.jp/tizu-kutyu.html
[16] Color FERET: http://www.nist.gov/itl/iad/ig/colorferet.cfm
[17] SUN Database: http://groups.csail.mit.edu/vision/SUN/
[18] ImageMagick: http://www.imagemagick.org/script/index.php
[19] Baxter R J 2007 Exactly Solved Models in Statistical Physics (Dover Publications)
[20] Lee C H, Yamada Y, Kumamoto T and Matsueda H 2014 arXiv:1403.0163