Joint bayesian convolutional sparse coding for image super-resolution

Qi Ge\textsuperscript{1,2}*, Wenze Shao\textsuperscript{1}, Liqian Wang\textsuperscript{1}

\textsuperscript{1} College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, \textsuperscript{2} National Engineering Research Center of Communications and Networking, Nanjing University of Posts and Telecommunications, Nanjing, China

* geqi@njupt.edu.cn

Abstract

We propose a convolutional sparse coding (CSC) for super resolution (CSC-SR) algorithm with a joint Bayesian learning strategy. Due to the unknown parameters in solving CSC-SR, the performance of the algorithm depends on the choice of the parameter. To this end, a coupled Beta-Bernoulli process is employed to infer appropriate filters and sparse coding maps (SCM) for both low resolution (LR) image and high resolution (HR) image. The filters and the SCMs are learned in a joint inference. The experimental results validate the advantages of the proposed approach over the previous CSC-SR and other state-of-the-art SR methods.

Introduction

Image super-resolution (SR) arms to reconstruct a high resolution (HR) image from a single or several low resolution (LR) image of the scene together [1–5]. The resolution limitations of low-cost imaging sensors is overcome by SR methods, and the degradation in the LR images caused by blur and the motion of camera or scene are utilized to reconstruct the HR image by SR methods. Since the motion parameters are estimated along with HR image solely from the LR images, SR is a difficult inverse problem, especially the single image SR.

To reconstruct a HR image from the LR image, information usually was lost during the down-sampling procedure, and more prior knowledge should be exploited. The errors are easily caused in estimating the missing pixels of HR image based on some simple assumptions that does not hold in many practical systems. For better reconstructing complicated structures in natural images, local priors are exploited in image patches by the single image super resolution (SISR) for estimating the HR image [6–9]. Example-based methods are one of the most important SISR methods. The existing example-based methods can be categorized into sparse coding-based and mapping-based methods [10–12]. Sparse coding-based methods train a couple of dictionaries for LR and HR image patches, and there are many approaches are proposed to build mapping functions between the LR and HR patches [13–16]. The mapping functions are learned using LR and HR patch pairs [6], [17–18]. For obtaining the final results, the pixels in the overlapped patches are averaged by the previous example-based methods. However, the pixels in the overlapped patches usually are not consistent, and the averaging process may
destroy the inconsistency. Some approaches are proposed to overcome this problem and show the improved performance in image reconstruction [19].

Recently, the convolutional sparse coding based super-resolution (CSC-SR) method [19] utilizes a global sparse coding method for preserving the consistency better. CSC has been applied in unsupervised learning for visual features [20], [21]. It represents the input signal by the linear combination of $N$ sparse feature maps corresponding to $N$ filters. The number of filters in decomposing LR image is different from the number of filters in reconstructing HR image in CSC-SR method. By processing the global image rather than in each local patch, the previous CSC-SR has shown its outperformance over the sparse coding (SC) based ones. Although CSC-SR adopts a more adaptive decomposition-reconstruction strategy, this model still utilizes the fixed number of filters. Therefore, these parameters should be assigned a priori and a structure of latent space representing the input data should be applied to solve CSC problem. In that case, the sparse coding maps and the filters should be learned optimal. Otherwise, the errors in estimating the parameters may lead to instability in reconstructing the HR image.

To address the above problems, we present a novel convolutional sparse coding based super-resolution method. There are two stages in the proposed method. Firstly, we learn the filters and the sparse coding maps adaptively by modeling them using a Beta-Bernoulli distribution in decomposing the LR image. Secondly, we use the same distributions associated with the sparse coding maps when reconstructing the HR image. In each stage, the sparse coding maps are learned by minimizing the CSC problem with a variance Bayesian inference process. The Bernoulli distributions have an ability of controlling the frequency with the factors. Since learning processes on LR image and HR image are based on the same distribution, the filters and the sparse coding maps in each stage are inferred simultaneously under a joint inference process. The experimental results of the proposed method validate the competitive results with the state-of-the-art methods.

**Related works**

Due to the consistency between the overlapped patches, the existing sparse coding-based methods [13–16] may smooth out the consistency including the high frequency edges and structures of the image. To preserve the consistency, a convolutional sparse coding is proposed to encode the whole image in [22]. The CSC based super-resolution (CSC-SR) method learns the sparse coding maps of the LR and HR image separately by solving the CSC problem. We use $3 \times 3$ low pass filter with all coefficients being 1/9 to extract the smooth component of the LR image, which is the same as the work [19]. Denote by $y$ the LR image, by $Y_s$ the smooth component of $y$, by $Y$ the residual component of $y$. $Y$ represents high frequency edges and texture structures. In CSC-SR, after the smooth component of LR image $Y_s$ extracted by a low pass filter, and a group of LR filters are learned to decompose the residual component $Y$:

$$
\min_{A,f} \left\| Y - \sum_{i=1}^{N_f} f_i^T \otimes A_i \right\|_F^2 + \lambda \sum_{i=1}^{N_f} \| A_i \|_1, \text{s.t. } \| f_i \|_F^2 \leq 1
$$

(1)

where $Y$ is the residual component including high frequency edges and texture structures, $\{f_i\}_{i=1..N_f}$ are the filters, $\{A_i\}_{i=1..N_f}$ are the sparse coding maps. $\| \cdot \|_F$ represents the Frobenius norm and $\| \cdot \|_1$ represents l-1 norm. $\otimes$ is convolution operator. And we use the initial sparse coding map provided on the website, http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm. Gu et al. [19] use a stochastic average based alternating direction multiplier algorithm (SA-ADMM) to efficiently solve the CSC model (1) with a large number of filters. The HR image is also decomposed into a smooth component and residual component. $X$ denotes the residual component which represents high frequency edges and structures of HR image. The
$N_1$-dimensional sparse coding space of $(A_1^l, A_2^l, \ldots, A_{N_1}^l)$ can be transformed to $N_2$-dimensional sparse coding space by multiply a trained mapping function $M \in R^{N_1 \times N_2}$ [19]. By amplifying the transformed matrix $(A_1^l, A_2^l, \ldots, A_{N_1}^l) M$ with $k$ factor, the column vectors of the transformed matrix are the initial the sparse coding map set $(A^h_j)_{j=1\ldots N_2}$. The filters and the mapping functions can be learned by solving another CSC minimization problem,

$$\min_{x,f} \left\| X - \sum_{j=1}^{N_2} f^h_j \otimes A^h_j \right\|_F^2 + \beta \sum_{j=1}^{N_2} \| A^h_j \|_1, \text{s.t.} \| f^h_j \|_2^2 \leq 1, \quad m_j \geq 0, \quad |m_j|_1 = 1, \quad (2)$$

where $A^h_j = A^l(x,y)m_j$ represents the $j$-th sparse coding map of HR image, $M$ represents the mapping function matrix of size $N_1 \times N_2$, $(f^h_j)_{j=1\ldots N_2}$ is the set of filters of HR image. After solving the minimization problem Eq (2) by SA-ADMM algorithm, the mapping function $M$ and the filters $f^h_j, j = 1 \ldots N_2$ can be learned and finally the residual component with high frequency texture structure of the HR image can be reconstruct by the summation of the convolution results of HR filters and the sparse coding maps $X = \sum_{j=1}^{N_2} f^h_j \otimes A^h_j$. By amplifying the smooth component of LR image $Y$, with $k$ factor, the smooth component of HR image $X_i$ is obtained. The reconstructed HR image $x$ is represented by $x = X_i + \hat{X}$.

**Joint bayesian convolutional sparse coding for image super-resolution**

The performance of convolutional-sparse-coding (CSC)-based super resolution still depends on the appropriate choice of the unknown parameters [19]. To address the limitation, we present a joint Bayesian convolutional sparse coding framework for image super resolution (JB-CSC SR). Firstly, rather than the non-parameter Bayesian based sparse coding, we develop a new beta process to build the Bayesian based CSC model. Secondly, since we use different numbers of the filters and the corresponding sparse coding maps to decompose/reconstruct the LR and HR images, two CSC problems are solved by jointly learning a mapping function between the sparse coding maps in the two feature spaces.

**Bayesian based convolutional sparse coding model for decomposing LR image**

In this paper, we propose a new CSC model for super resolution with Bayesian prior. As shown in S1 Fig, the SR model is a linear combination of $N$ atoms with the corresponding coefficients. Likewise, the traditional CSC model is the linear combination of the convolutions with the corresponding linear coefficients, where all the linear coefficients equal to one. Therefore, adaptive linear coefficients should be added in the CSC model.

For conveniently describing our model, we transform the residual component of LR image $Y$, the filters $(f^l_{i-1} \ldots N_1)$ and the sparse coding maps $A^l_i$ in Eq (1) into 1-dimensional, thus, we have the residual component $Y \in R^P$, filters $(f^l_{i-1} \ldots N_1) \in R^P$, sparse coding maps $a^l_{i-1} \ldots N_1 \in R^{(m \times n) \times (n \times s)}$, where $P = m \times n, S = s \times s$. The number of filters for decomposing the LR image is $N_1$, $d^l_i \in R^P$ is the $i$-th 1-dimensional version of the convolution $d^l_i = f^l_i \otimes a^l_i$. With the base measure $h_0$ and the parameters $a, b > 0$, a representation of Beta Process is as follows,

$$\hat{h} = \sum_{i=1}^{N_1} \pi_i \delta_{d^l_i}, \quad \pi_i \sim \text{Beta}(a/N_1, b(N_1 - 1)/N_1) \quad (3)$$

where $\delta_{d^l_i}$ is unit point mass at $d^l_i$. A draw $\hat{h}$ from the process is a set of $N$ probabilities $\pi_{i=1\ldots N_1}$.
each associated with $d_i, i = 1 \ldots N_1$ that are drawn i.i.d. from the base measure $h_0$. Considering $\pi_i$ to be a Bernoulli distribution parameter, we use $h$ to draw a binary vector $z' = (z'_1, z'_2, \ldots, z'_N) \in \mathbb{R}^N$, $z'_i \in \{0, 1\}^N$, $z' \sim \prod_{i=1}^{N_1} \text{Bernoulli}(\pi_i)$. In the limit $N_1 \to \infty$, the number of the non-zero elements in $z'$ itself is derived from Poisson($\alpha$) that control the number of used convolutions $d_i, i = 1 \ldots N_1$. Concatenating the convolutions $d_i, i = 1 \ldots N_1$ together, we have the convolution matrix $D = (d_1, d_2, \ldots, d_{N_1})$. Since the residual component $Y$ can be expressed as the linear combination of the convolutions $d_1, d_2, \ldots, d_{N_1}$, it can be expressed in the following term,

$$Y = Dw' + e'$$

where $w' = (w_1', w_2', \ldots, w_{N_1}')^T \in \mathbb{R}^N$ is the coefficient vector for linearly representing $y$. $e' \in \mathbb{R}^D$ denotes the error term, $D' = (d_1', d_2', \ldots, d_{N_1'}) = (f_{1} \otimes a_1, f_{2} \otimes a_2, \ldots, f_{N_1} \otimes a_{N_1})$. With the binary vector $z' \in \mathbb{R}^N$, the parameters in Eq (5) can be expressed as,

$$w' = z' \otimes s', s' \sim N(0, \gamma_s^{-1}1_N), e' \sim N(0, \gamma_e^{-1}I_p),$$

$$f_{i'} \sim N(0, S^{-1}I_d), d_{i'} \sim N(0, P^{-1}I_p)$$

where $\otimes$ represents the Hadamard vector product, $I_{N_1}, I_S, I_P$ represent the identity matrix with size of $N_1 \times N_1, S \times S, P \times P$, respectively. $\gamma_s \sim \text{Gamma}(\epsilon, \delta), \gamma_e \sim \text{Gamma}(\epsilon, \delta)$ are drawn from the Gamma distributions, respectively. Following the model Eq (5) and the general structure of beta process described in [23], the convolutional sparse coding model with Bayesian prior can be expressed as,

$$\min_{a_i} \|y - (f_{1} \otimes a_1, f_{2} \otimes a_2, \ldots, f_{N_1} \otimes a_{N_1})w'\|_2^2 + \lambda \sum_{i=1}^{N_1} \|a_i\|_1, s.t. \|f_i\|_F^2 \leq 1$$

where $w'$ is defined as in Eq (5). We rewrite the formula Eq (6) as follows,

$$\min_{a_i} \|y - \sum_{i=1}^{N_1} f_{i} \otimes a_i'w'\|_F^2 + \lambda \sum_{i=1}^{N_1} \|a_i'w'\|_1, s.t. \|f_i\|_F^2 \leq 1$$

Eq (7) is a typical convolutional sparse coding minimization problem with Bayesian priors. The weighted linear combination of the convolutions, $\sum_{i=1}^{N_1} f_{i} \otimes a_i'w'$, is equivalent to the summation of the convolution of the filters $f_{i}, i = 1 \ldots N_1$ and the weighted sparse coding maps $a_i'w', i = 1 \ldots N_1$. Analogous to [19], we adopt the SA-ADMM algorithm to solve the minimization problem Eq (7) to train the filters $f_{i=1 \ldots N_1}$ and learn the new sparse coding maps $a_i'w', i = 1 \ldots N_1$ with the Bayesian prior.

**Bayesian based convolutional sparse coding model for reconstructing HR image**

Authors in [19] train the mapping function to obtain the sparse coding maps of the HR image from the sparse coding maps of the LR image. However, inappropriate choosing the unknown parameters, including the number of the convolutions for HR image and the parameters of the mapping function, may lead to instability of the performance of the algorithm. To address this problem, we build the convolutional sparse coding (CSC) model with Bayesian prior to learn the HR filters to reconstruct the HR image. For convenient expression, given the sparse coding
maps learned from Eq (7), we have $A' = (a'_1, w'_1, \ldots, a'_N, w'_N) \in \mathbb{R}^{(m + s - 1)(n + s - 1) \times N_1}$. $N_2$ is the number of filters for reconstructing the HR image. The authors in [19] transform $A'$ from $N_1$-dimension feature space to the $N_2$-dimension feature space by right-multiplying it with the mapping function $M \in \mathbb{R}^{N_2 \times N_1}$, $(m_j, m_j) = 1, m_j$ is the $j$-th column of $M$), they have $A'M \in \mathbb{R}^{(m + s - 1)(n + s - 1) \times N_2}$. And then by zooming each column of $A'M$ with $k$ factor, the sparse coding maps $A'' = (a''_1)_{j=1}^{N_2} \in \mathbb{R}^{(m + s - 1)(n + s - 1) \times N_2}$ for HR image is obtained. Different from [19], we apply Bayesian prior to the mapping function $M$, the HR filters, and the sparse coding maps together.

Let $k$ denote the zooming factor, thus we represent the 1-dimensional version of the residual component of HR image representing high frequency edges and structures as $X \in \mathbb{R}^{kP}$. We represent the mapping function as $M = (m_1, m_2, \ldots, m_N) \in \mathbb{R}^{N_1 \times N_2}$, the HR filters as $f_j^h \in \mathbb{R}^{N_2}, j = 1 \ldots N_2$, and the sparse coding maps as $a_j^h, j = 1 \ldots N_2 \in \mathbb{R}^{m_2 \times k}$. We initialize each column of mapping function $M$ as $m_j \sim N(0, N_1^{-1}I_{N_1})$. With the same base measure $h_0$ and the same parameters $a, b > 0$ in Eq (3), we have a representation of Beta Process as follows

$$h^h = \sum_{j=1}^{N_2} \pi_j \delta_{d_j^h}, \pi, \sim \text{Beta}(a/N_2, b(N_2 - 1)/N_2)$$

where $\delta(d^h_j)$ is unit point mass at $d^h_j$. Similar to the Beta Process in Sub-section III-A, we also have $h$ to draw a binary vector $z^h = (z^h_1, z^h_2, \ldots, z^h_N) \in \mathbb{R}^{N_2}$, $z^h_j \in \{0, 1\}^{N_2}$, $z^h \sim \prod_{j=1}^{N_2} \text{Bernoulli}(\pi_j)$. The number of the non-zero elements in $z^h$ itself is drawn from Poisson($a/b$) and it controls the number of used convolutions $d^h_j, j = 1 \ldots N_2$, when $N_2 \rightarrow \infty$. Given the binary vector $z^h$, we have the diagonal matrix $\Lambda = \text{diag}(z^h) \in \mathbb{R}^{N_2 \times N_2}$. By multiplying the mapping function $M$ with $\Lambda$, we have $MA = (z^h_1 m_1, z^h_2 m_2, \ldots, z^h_N m_N)$, thus the columns of the mapping function are weighted by the elements of $z^h$. We represent the sparse coding maps as $A^h = (a^h_1)_{j=1}^{N_2}$ for HR image. First, we transform $A'$ with $N_1$ sparse coding maps for LR image to $A'MA$ with $N_2$ sparse coding maps by the new mapping function $MA$. By zooming $A'MA$ with $k$ factor, we have the sparse coding maps $A^h(x, :) = A'MA(x/k, :)$, where $x$ represents the horizontal coordinate in the image domain.

Given the convolution matrix $D^h = (d^h_1, d^h_2, \ldots, d^h_N)$, the residual component $X$ of HR image can be represented as a linear combination of the convolutions, $d^h_1, d^h_2, \ldots, d^h_N$,

$$X = D^h w^h + e^h,$$

where $w^h = (w^h_1, w^h_2, \ldots, w^h_N)^T \in \mathbb{R}^{N_2}$ and $e^h \in \mathbb{R}^{kP}$ represent the coefficient vector and the error term, respectively. The coefficient vector $w^h$ can be decomposed by the binary vector $z^h \in \mathbb{R}^{N_2}$. Similar to Eq (5), we have,

$$w^h = z^h \odot s^h, s^h \sim N(0, \gamma_s^{-1}I_{N_2}), e^h \sim N(0, \gamma_e^{-1}I_{kP}), f^h \sim N(0, kS^{-1}I_{kS}), d^h \sim N(0, kP^{-1}I_{kP})$$

$I_{N_2}, I_{kS}, I_{kP}$ represent the identity matrix with size of $N_2 \times N_2, kS \times kS, kP \times kP$, respectively. $\gamma_s$ and $\gamma_e$ are the defined as in Eq (5). Following the parameter settings in Eq (10), the convolutional
sparse coding model with Bayesian prior for learning the HR filters can be expressed as,

\[
\{\mathbf{f}^i, \mathbf{a}^i_k\} = \min_{\mathbf{x}} \| \mathbf{X} - (\mathbf{f}^i_1 \otimes \mathbf{a}^i_{k_1}, \mathbf{f}^i_2 \otimes \mathbf{a}^i_{k_2}, \ldots \mathbf{f}^i_{N_2} \otimes \mathbf{a}^i_{k_{N_2}}) \mathbf{w}^i \|_F^2 + \lambda \sum_{j=1}^{N_2} \| \mathbf{a}^i_{k_j} \|_1, s.t. \| \mathbf{f}^i_1 \|_F^2 \leq 1
\]

where \( \mathbf{w}^i \) is defined as in Eq (10). \((\mathbf{f}^i_1 \otimes \mathbf{a}^i_{k_1}, \mathbf{f}^i_2 \otimes \mathbf{a}^i_{k_2}, \ldots \mathbf{f}^i_{N_2} \otimes \mathbf{a}^i_{k_{N_2}}) \mathbf{w}^i \) represents the weighted linear combination of the convolutions, we rewrite the formula Eq (11) as follows,

\[
\{\mathbf{f}^i, \mathbf{a}^i_k\} = \min_{\mathbf{x}} \| \mathbf{X} - \sum_{j=1}^{N_2} \mathbf{f}^i_j \otimes \mathbf{a}^i_{k_j} \mathbf{w}^i_j \|_F^2 + \lambda \sum_{j=1}^{N_2} \| \mathbf{a}^i_{k_j} \|_1, s.t. \| \mathbf{f}^i_1 \|_F^2 \leq 1
\]

Eq (12) is also a convolutional sparse coding minimization problem with Bayesian priors. We solve it by SA-ADMM algorithm. After training the HR filters \( \{\mathbf{f}^i\}_i \), the sparse coding maps \( \{\mathbf{a}^i_k\}_i \), and the mapping function \( \mathbf{MA} \), the residual component of HR image is formulated as the summation of the convolutions of HR filters and the sparse coding maps:

\[
\hat{X} = \sum_{i=1}^{N_2} \mathbf{f}^i_1 \otimes \mathbf{a}^i_{k_j} \mathbf{w}^i_j. \]

The estimation of HR image is reconstructed by combing the residual component \( \hat{X} \) and the smooth component. The training framework of the proposed JB-CSC super-resolution method is shown in S2 Fig.

**Variance bayesian inference process**

Gibbs sampling is usually used to perform Bayesian inference to sample the parameters from the conditional distribution of each hidden variable, given the other observations [24]. The authors in [25] update the parameters according to the posterior distributions over the model for Gibbs sampler. The inference process performs updating. It is also called sampling over authors in [25] update the parameters according to the posterior distributions over the model components [24].

Given Eq (15), the posterior distribution over a convolution to sampling the \( i \)-th convolution,

\[
p(\mathbf{d}^i|\mathbf{z}^i) \propto \int \prod_{j=1}^{N_2} N(\mathbf{y}_d^i|\mathbf{D}^i(\mathbf{z}^i \circ \mathbf{s}^i), \gamma^{-1}_x \mathbf{I}_{P})
\]

Let \( \mathbf{y}_d^i \) denote the contribution of the convolution \( \mathbf{d}^i \) to the residual component of LR image \( \mathbf{y} \), then we have,

\[
\mathbf{y}_d^i = \mathbf{y} - \mathbf{D}(\mathbf{z}^i \circ \mathbf{s}^i) + \mathbf{d}(\mathbf{z}^i \circ \mathbf{s}^i)
\]

With \( \mathbf{y}_d^i \), the posterior distribution in Eq (14) can be rewritten as,

\[
p(\mathbf{d}^i|\mathbf{z}^i) \propto \int \prod_{j=1}^{N_2} N(\mathbf{y}_d^i|\mathbf{D}^i(\mathbf{z}^i \circ \mathbf{s}^i), \gamma^{-1}_x \mathbf{I}_{P})N(\mathbf{d}^i|\mathbf{0}, \gamma^{-1}_x \mathbf{I}_{P})
\]

Given Eq (15), the posterior distribution over a convolution can be written as

\[
p(\mathbf{d}^i|\mathbf{z}^i) \propto N(\mathbf{d}^i|\mu_i, \gamma^{-1}_x \mathbf{I}_{P})
\]

where \( \mu_i = \frac{1}{\gamma_x} \sum_{j=1}^{N_2} (\mathbf{z}_j \circ \mathbf{s}_j) \mathbf{y}_d^i, P = \gamma_x (\mathbf{z}_j \circ \mathbf{s}_j)^2. \)

Once the convolutions have sampled, we sample the parameter \( \mathbf{z}_j \). By the distribution of the \( i \)-th convolution, the posterior probability distribution over \( \mathbf{z}_j \) can be expressed as

\[
p(\mathbf{z}_j | \mathbf{y}_d^i) \propto N(\mathbf{y}_d^i|\mathbf{d}^i(\mathbf{z}_j \circ \mathbf{s}^i), \gamma^{-1}_x \mathbf{I}_{P})Bernoulli(\mathbf{z}_j|\pi_j).
\]
With the prior probability of \( z_i' = 1 \) given by \( \pi_i \), we can write the following posterior probability,

\[
p(z_i' = 1|\cdot) \propto \pi_i \exp \left( -\frac{\gamma_e}{2} \| y_d - d_i s_i' \|_2^2 \right).
\]  

(18)

The prior probability of \( z_i' = 0 \) is given by \( 1-\pi_i \), the posterior probability can be derived as,

\[
p(z_i' = 0|\cdot) \propto (1-\pi_i) \exp \left( -\frac{\gamma_e}{2} \| y_d - d_i s_i' \|_2^2 \right)
\]  

(19)

Analogous to the K-SVD algorithm, we derive the expression of sampling \( z_i' \),

\[
z_i' \sim \text{Bernoulli} \left( z_i' \left| \frac{\pi_i \xi}{1 + \pi_i (\xi - 1)} \right. \right), \quad \xi = \exp \left( -\frac{\gamma_e s_i'}{2} \right. \left( \| d_i s_i' \|_2^2 - 2 (y_d)^T d_i \right) \)
\]  

(20)

Having sampled the convolution and the binary vector, we adopt the method of sampling of \( s_i' \), \( \gamma_n \), \( \pi_0 \), \( \gamma_e \) according to [25].

**Summary of algorithm**

The proposed model consists of two stages: first, we solve the CSC model problem to learn LR filters and the sparse coding coding maps; second, we solve the CSC problem to learn the HR filters, the sparse coding maps, and the mapping function for reconstructing the HR image. By incorporating the Bayesian prior and the inference process into the convolutional sparse coding model, we summarize the algorithm of the proposed method in **Algorithm 1**.

**Algorithm 1**

1. Input the train LR image \( Y \), decompose it into smooth component \( Y_s \) and residual component \( Y_r \);
2. Initialize the parameters \( z^i, s^i, \varepsilon^i, d^i, f_i^{1...N_I} \) as in Eq (5).
3. Solving Eq (6) by SA-ADMM algorithm to obtain \( f_i^{1...N_I} \) and the weighted sparse coding maps \( A^i = (a^i w^i_j)_{j=1...N_I} \) while updating the parameters for \( z^i \) by Eq (20), and updating \( s^i, \varepsilon^i, f_i^{1...N_I} \) accordingly. Then the smooth component \( Y_s \) is zoomed into \( X_s \) with k factor.
4. Initialize \( M = (m_1, m_2, \cdots, m_{N_I}) \in \mathbb{R}^{N_I x N_s}, m_i \sim N(0, S^{-1} I_{N_I}) \) and the parameters \( z^h, s^h, \varepsilon^h, d^h, f_i^{1...N_I} \) as in Eq (10);
5. Transform the sparse coding maps \( A^i = (a^i w^i_j)_{j=1...N_I} \) to \( A^i M \Lambda \), where \( \Lambda = \text{diag}(z^i) \in \mathbb{R}^{N_s x N_s} \), by zooming it with k factor, get the initial sparse coding maps \( A_s^h \) for HR image;
6. Solving Eq (12) by SA-ADMM algorithm to obtain the filters \( f_i^{1...N_I} \), sparse coding maps \( A_s^h \), and the mapping function \( M_s \) while updating the parameters for \( z^h \) analogous to Eq (20), and updating \( s^h, \varepsilon^h, d^h \) accordingly.
7. Reconstruct the residual component \( X \) of HR image by \( \hat{x} = \sum_{j=1}^{N_I} f_j^0 \otimes a_j^i w^j \).

The final HR image is obtained by \( X = X_s + \hat{x} \).

**Experimental results**

In this section, we compare the proposed joint Bayesian convolutional sparse coding (JBCSC) method with four state-of-the-art SR methods, including the Beta process joint dictionary learning method (BPJDL) [16], convolutional neural network based method (CNN-SR) [18], convolutional sparse coding based method (CSC-SR) [19], and adjusted ANR (A-ANR) [26]. We use the source code provided in the authors’ website, and use the parameters which they
recommend. The original ground truth images are downsized by bi-cubic interpolation to generate LR-HR image pairs for both training and evaluation.

Parameter setting

To have a fair comparison between CSC-SR [19] and our model, we adopt the same training set provided in [19], the same regularization parameters, and the same implementation on the image boundary. The beta-distribution parameters are set as \( a = N_1 \) for LR image and \( a = N_2 \) for HR image, and \( b = 1 \) for both HR and LR image. The hyper-parameters in Eq (5) and Eq (10) were set as \( c = d = e = f = 10^{-6} \), which is provided in [24]. All these parameters in the Bayesian inference process have not been tuned. We conduct all the comparison experiments on a PC with Intel Core i7 25.GHz CPU 4GB RAM on Matlab platform.

Comparison with state-of-the-arts

We compare the proposed JB-CSC method with other recent SR methods on the widely used test images from Set5 [27] and Set14 [28] for different zooming factors. The test images of Set5 and Set14 are available at https://sites.google.com/site/jbhuang0604/publications/struct_sr. These two test image datasets are both publicly released and popularly used for image super-resolution works [1–6], [8–26] without any consent or any copyright. We show the SR results of BPJDL [16], A-ANR [26], CNN-SR [18], convolutional sparse coding based method (CSC-SR) [19], and the proposed JB-CSC method in S3–S6 Figs. In S3 Fig, we show the SR results of image Bird with zooming factor \( k = 2 \). In S4, S5 and S6 Figs we show the SR results of image Foreman, Face Butterfly with zooming factor \( k = 3 \). And the PSNR values by the competing methods are shown in Table 1.

In S3 Fig, we can see from the small window in the right-bottom of the images, the SR results by CNN-SR over-smooth the edge. And very little ringing artifacts are generated by BPJDL, A-ANR, and CSC, since the zooming factor is low. The PSNR values of the results at this factor \( k = 2 \) show the outperformance of our method.

In S4 Fig, in the small window of the images, the SR results of the competing methods, such as BPJDL and A-ANR show that more edges are over-smoothed. Comparing CSC-SR and the proposed method, the edges of the image can be preserved better than CSC-SR.

| Zooming factor \( k = 2 \) | Zooming factor \( k = 3 \) |
|---------------------------|---------------------------|
| **BPJDL** | **CNN** | **A-ANR** | **CSC** | **JB-CSC** | **BPJDL** | **CNN** | **A-ANR** | **CSC** | **JB-CSC** |
| Butterfly | 31.43 | 32.20 | 31.94 | 31.96 | 32.02 | 26.42 | 27.58 | 27.22 | 27.11 | 27.60 |
| Face | 35.75 | 35.60 | 35.72 | 35.71 | 35.76 | 33.45 | 33.57 | 33.74 | 33.80 | 33.94 |
| Bird | 40.99 | 40.63 | 40.98 | 41.49 | 41.64 | 34.53 | 34.92 | 34.48 | 34.78 | 35.90 |
| Woman | 35.23 | 34.93 | 35.27 | 35.31 | 35.37 | 30.50 | 30.92 | 31.19 | 31.27 | 31.42 |
| Foreman | 36.49 | 36.19 | 36.91 | 36.64 | 36.79 | 32.91 | 33.34 | 34.22 | 34.24 | 34.33 |
| Coast | 30.60 | 30.49 | 30.55 | 30.65 | 30.71 | 30.07 | 27.19 | 27.27 | 27.27 | 27.38 |
| Flowers | 32.91 | 33.04 | 33.03 | 33.15 | 33.72 | 28.62 | 28.98 | 29.05 | 29.05 | 29.15 |
| Zebra | 33.62 | 33.30 | 33.67 | 33.77 | 33.83 | 28.73 | 28.90 | 29.06 | 29.30 | 29.43 |
| Lena | 35.58 | 36.48 | 36.57 | 36.66 | 36.76 | 33.73 | 33.39 | 33.50 | 33.43 | 33.75 |
| Bridge | 27.77 | 27.70 | 27.78 | 27.84 | 27.91 | 24.99 | 25.07 | 25.17 | 25.20 | 25.33 |
| Baby | 38.54 | 38.41 | 38.43 | 38.48 | 38.53 | 35.15 | 35.00 | 35.14 | 35.28 | 35.44 |
| Peppers | 36.71 | 36.75 | 37.06 | 36.90 | 36.97 | 34.02 | 34.35 | 34.71 | 34.72 | 34.86 |
| Man | 30.80 | 30.82 | 30.88 | 30.97 | 31.02 | 28.05 | 28.18 | 28.29 | 28.34 | 28.44 |
| Barbara | 28.68 | 28.59 | 28.70 | 28.77 | 28.81 | 26.82 | 26.65 | 26.47 | 26.67 | 26.78 |
| Ave | 34.01 | 33.94 | 34.11 | 34.16 | 34.24 | 30.31 | 30.57 | 30.68 | 30.83 | 30.97 |

https://doi.org/10.1371/journal.pone.0201463.t001
In S5 Fig, the results of the proposed method are compared with the results of BPJDL, A-ANR, CNN-SR, and CSC-SR. In the small windows, we can see that more textures of the hair are preserved much better in BPJDL. This comparison verifies that the models with Bayesian prior, such as BPJDL and the proposed method, have the outperformance in preserving the texture details of the images.

In S6 Fig, the proposed method is compared with the competing methods with the zooming factor \( k = 4 \). In (b), we can see that the result of BPJDL shows that the ringing artifacts increase as the zooming factor increases. As shown in (d), (c), (e), (f) of S6 Fig, the ringing artifacts remain in the results of A-ANR, CSC-SR and the proposed method while not in the result of CNN-SR. However, the edges are over-smoothed by CNN-SR. The PSNR value of the proposed method is higher than the competing methods.

**Conclusion**

In this paper, we present a convolutional sparse coding based super resolution method with a joint Bayesian learning strategy (JB-CSC). JB-CSC employs a coupled Beta-Bernoulli process to incorporate the Bayesian prior into the convolutional sparse coding model, which avoids instability caused by estimating the unknown parameter. Different from the CSC method, the filters and sparse feature maps for both low resolution (LR) image and high resolution (HR) are learned adaptively through the Bayesian learning strategy. The experimental results validate the advantages of the proposed approach, compared with the previous CSC-SR and other state-of-the-art SR methods.

**Supporting information**

S1 Fig. The difference between the sparse coding model and the convolutional sparse coding model.
(TIF)

S2 Fig. Pipeline of the proposed JB-CSC super-resolution training framework.
(TIF)

S3 Fig. Super resolution results on image Bird by different algorithms.
(TIF)

S4 Fig. Super resolution results on image Foreman by different algorithms.
(TIF)

S5 Fig. Super resolution results on image Face by different algorithms.
(TIF)

S6 Fig. Super resolution results on image Butterfly by different algorithms.
(TIF)

S1 Dataset.
(ZIP)

**Acknowledgments**

This work was supported in part by the Natural Science Foundation of China under Grant 61502244, Grant 61771250, Grant 61602257, Grant 61402239, the Science Foundation of Jiangsu Province under Grant BK20150859, BK20160904, China Postdoctoral Science Foundation 2015M580433.
Author Contributions

Data curation: Liqian Wang.
Formal analysis: Wenze Shao.
Funding acquisition: Qi Ge.
Investigation: Qi Ge.
Methodology: Qi Ge.
Project administration: Qi Ge.
Resources: Qi Ge.
Software: Qi Ge.
Supervision: Qi Ge.
Visualization: Qi Ge.
Writing – original draft: Qi Ge.
Writing – review & editing: Qi Ge.

References

1. Katsaggelos K, Molina R, Mateos J. Super Resolution of Images and Video. San Rafael, CA: Morgan and Claypool, 2007.
2. Song H, Huang B, Liu Q, Zhang K. Improving the Spatial Resolution of Landsat TM/ETM+ Through Fusion With SPOT5 Images via Learning-Based Super-Resolution. IEEE Transactions on Geoscience and Remote Sensing, 2015; 53(3): 1195–1204.
3. Glasner D, Bagon S, and Irani M. Super-resolution from a single image. in Proceedings of International conference on computer vision. 2009; 349–356.
4. Elad M, and Feuer A, Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. IEEE Transactions on Image Processing. 1997; 6(12): 1646–1658. https://doi.org/10.1109/83.650118 PMID: 18285235
5. Song H, Wang G, Cao A, Liu Q, Huang B B, Improving the Spatial Resolution of FY-3 Microwave Radiation Imager via Fusion With FY-3/MERSI. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2017; 10(7): 3055–3063.
6. Chen Y, Ranftl R, Pock T, Insights Into Analysis Operator Learning: From Patch-Based Sparse Models to Higher Order MRFs. IEEE Transactions on Image Processing. 2014; 23(3): 1060–1072. https://doi.org/10.1109/TIP.2014.2299065 PMID: 24474375
7. Fattal R. Image upsampling via imposed edge statistics. in Proceedings of international conference on computer graphics and interactive techniques. 2007; 26(3).
8. Tai Y.-W, Liu S, Brown M. S, Lin S. Super resolution using edge prior and single image detail synthesis. in Proceedings of computer vision and pattern recognition. 2010; 2400–2407.
9. Yang J, Wright J, Huang T, Ma Y. Image super-resolution as sparse representation of raw image patches. in Proceedings of International conference on computer vision. 2008; 1–8.
10. Gao X, Zhang K, Tao D, Li X. Joint learning for single-image super-resolution via a coupled constrain. IEEE Transactions on Image Processing. 2012; 21(2): 469–480. https://doi.org/10.1109/TIP.2011.2161482 PMID: 22928969
11. Zhang K, Tao D, Gao X, Li X, Li J. Single image super-resolution with multiscale similarity learning. IEEE Trans. Neural Netw. Learn. Syst. 2013; 24(10):1648–1659. https://doi.org/10.1109/TNNLS.2013.2262001 PMID: 24808601
12. Ham D K Sohn. Depth superresolution by transduction. IEEE Transactions on Image Processing. 2015; 24(5): 1524–1535. https://doi.org/10.1109/TIP.2015.2405342 PMID: 25769140
13. Yang J, Wright J, Huang T, S, Ma Y. Image super-resolution via sparse representation. IEEE Transactions on Image Processing. 2010; 19(11): 2861–2873. https://doi.org/10.1109/TIP.2010.2050625 PMID: 20489687
14. Yang J, Wang Z, Lin Z, Cohen S, Huang T. Coupled dictionary training for image super-resolution. IEEE Trans. Image Process. 2012; 21(8):3467–3478. https://doi.org/10.1109/TIP.2012.2192127 PMID: 22481818

15. Lu X, Yuan Y, Yan P. Alternatively constrained dictionary learning for image super-resolution," IEEE Transactions on Systems, Man, and Cybernetics. 2014; 44(3): 366–377.

16. He L, Qi H, Zaretzki R. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. in Proceedings of computer vision and pattern recognition. 2013; 345–352.

17. Cui Z, Chang H, Shan S, Zhong B, Chen X. Deep network cascade for image super-resolution. in Proceedings of European conference on computer vision ECCV 2014; 49–64.

18. Dong C.-C, He K, Tang X. Learning a deep convolutional network for image super-resolution. in Proceedings of European conference on computer vision, ECCV 2014; 184–199.

19. Gu S, Meng W, Xie Q, Meng D, Feng X, Zhang L. Convolutional Sparse Coding for Image Super-Resolution. in Proceedings of international conference on computer vision, ICCV 2015; 1823–1831.

20. M Sorel S Filip. Fast convolutional sparse coding using matrix inversion lemma. Digital Signal Processing, 2016; 44–51.

21. Zeiler M. D, Taylor G. W, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning. in Proceedings of international conference on computer vision. 2011; 2018–2025.

22. Zeiler M. D, Krishnan D, Taylor G. W, Fergus R. Deconvolutional networks. in Proceedings of international conference on computer vision, 2010;2528 -2535.

23. Polatkan G, Zhou M, Carin L, Blei D, Daubechies, I. A Bayesian Nonparametric Approach to Image Super-Resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015; 37(2): 346–358. https://doi.org/10.1109/TPAMI.2014.2321404 PMID: 26353246

24. Paisley J, Carin L. Nonparametric factor analysis with beta process prior. in Proceedings of international conference on machine learning, ICML 2009; 777–784.

25. Akhtar N, Shafait F, Mian S. A. Discriminative Bayesian Dictionary Learning for Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016; 38(12):2374–2388. https://doi.org/10.1109/TPAMI.2016.2527652 PMID: 26886965

26. Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution. in Proceedings of asian conference on computer vision, ACCV 2014; 111–126.

27. Bevilaqua M, Roumy A. Guillemot C, Morel M.-L. Low complexity single-image super-resolution based on nonnegative neighbor embedding. in Proc. 23rd Brit. Mach. Vis. Conf. 2012; 135.

28. Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. in Proc. Int. Conf. Curves Surf. 2010;711–730.