

Single Image Super-Resolution Using Gaussian Process Regression With Dictionary-Based Sampling and Student- t Likelihood

Haijun Wang, Xinbo Gao, *Senior Member, IEEE*, Kaibing Zhang, and Jie Li

Abstract—Gaussian process regression (GPR) is an effective statistical learning method for modeling non-linear mapping from an observed space to an expected latent space. When applying it to example learning-based super-resolution (SR), two outstanding issues remain. One is its high computational complexity restricts SR application when a large data set is available for learning task. The other is that the commonly used Gaussian likelihood in GPR is incompatible with the true observation model for SR reconstruction. To alleviate the above two issues, we propose a GPR-based SR method by using dictionary-based sampling (DbS) and student- t likelihood. Considering that dictionary atoms effectively span the original training sample space, we adopt a DbS strategy by combining all the neighborhood samples of each atom into a compact representative training subset so as to reduce the computational complexity. Based on statistical tests, we statistically validate that student- t likelihood is more suitable to build the observation model for the SR problem. Extensive experimental results show that the proposed method outperforms other competitors and produces more pleasing details in texture regions.

Index Terms—Super-resolution, Gaussian process regression, dictionary-based sampling, student- t likelihood.

Manuscript received May 28, 2016; revised November 17, 2016 and March 9, 2017; accepted April 24, 2017. Date of publication May 3, 2017; date of current version May 26, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61432014, Grant U1605252, Grant 61672402, and Grant 61471161, in part by the National Key Research and Development Program of China under Grant 2016QY01W0204, in part by the Key Industrial Innovation Chain in Industrial Domain under Grant 2016KTZDGY-02, in part by the Fundamental Research Funds for the Central Universities under Grant BDZ021403 and Grant JB149901, in part by National High-Level Talents Special Support Program of China under Grant CS11117200001, in part by the Program for Changjiang Scholars and Innovative Research Team in the University of China under Grant IRT13088, in part by Doctoral Startup Foundation of the Xi'an Polytechnic University under Grant BS1616, in part by the Key Scientific Research Project in Colleges and Universities of Henan Province under Grant 16A520063 and Grant 15A110020, and in part by the Doctoral Research Fund of Henan University of Science and Technology. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Feng Wu. (*Corresponding author: Xinbo Gao*)

H. Wang is with the School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471023, China (e-mail: hajion@msn.com).

X. Gao is with the State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbga@mail.xidian.edu.cn).

K. Zhang is with the College of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China (e-mail: zhangkaibing@xpu.edu.cn).

J. Li is with the Video and Image Processing System Laboratory, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: leejie@mail.xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2700725

I. INTRODUCTION

SINGLE image super-resolution (SR) reconstruction refers to a special image restoration technique that can produce a latent high-resolution (HR) image with more details from one low-resolution (LR) observation. This technique shows great potential in many practical applications such as criminal investigation, video surveillance, medical imaging, digital entertainment, and high definition television (HDTV).

In general, the existing single image SR approaches can be roughly divided into three categories: interpolation-based, reconstruction-based, and example learning-based methods.

Interpolation-based methods use interpolation kernels to predict the missing pixel values on the HR image grids. Typical methods include classical bilinear, bicubic, and nearest neighbor interpolation, and so on [1]. However, due to the isotropic property of applied kernel functions, these methods fail to approximate underlying image structures, leading to blurred results. In recent years, more structure-adaptive interpolation methods have been proposed [2]–[4] so as to adapt different image structures. The interpolation-based methods are fast and easy to implement, but they cannot effectively suppress blurring artifacts in the obtained HR images.

Reconstruction-based methods usually incorporate a certain prior knowledge (represented by one or several regularization terms) into the reconstruction process for solving ill-posed SR problem [5]. The popular priors include total variation [8], bilateral total variation [9], gradients prior [10], edge-specific priors [11], Zernike-moment [7], non-local similarity [12], steering kernel prior [6], and feedback-control framework [13]. Although reconstruction-based SR approaches are good with the preservation of sharp edges and suppression of unwanted noise, they are still clumsy at generating novel details, leading to watercolor-like artifacts or over-smooth textures in resultant images.

Example learning-based methods use a prepared training set to learn the mapping relationship between the LR and HR feature spaces by machine learning technologies. According to how the mappings are established, the existing example learning-based SR methods can be further divided into two subcategories, namely coding-based and regression-based methods [14]–[18].

Coding-based SR approaches assume that isometric manifolds exist in the coupled LR-HR feature spaces, which makes it possible to convey the coding coefficients of the LR space to the corresponding HR space to synthesize the HR image.

Early methods directly learn the coefficients based on the similar neighbors in the feature space, including locally linear embedding (LLE) [20]–[22], and sparse neighbor embedding [26]. Recently, sparse representation has been successfully applied to example learning-based SR. This particular type of methods seeks sparse coding coefficients regarding dictionaries learned from the LR and HR feature spaces to generate HR images. Based on the construction way of dictionaries, there are two mainlines, i.e., orthogonal dictionary-based methods such as PCA-based methods [30] and wavelet-based methods [28], and over-complete dictionary-based methods such as sparse coding [23]–[25], [29], Beta process [31], and K-SVD based methods [27].

The other kind of example learning-based SR approaches is referred to as regression-based methods. This subclass directly learns the mapping relationship from LR and HR feature spaces. Typical methods include Markov random fields [19], kernel regression [39], neural network [34]–[36], neighborhood regression [32], linear regression [33], [38], support vector regression [37], and Gaussian process regression (GPR) [40]–[42].

Gaussian process (GP) is a powerful and elegant probabilistic learning tool for learning the non-linear mapping relationship based on kernel function. Theoretically, the time complexity of GP is the order of $O(v^3)$ when the number of training samples is v , so it is prohibitive for many large-scale applications. Recently, several GPR-based SR methods have been proposed to alleviate the high complexity. For example, He and Siu [40] propose a self-learning framework which partitions the input LR image into several local tiles and separately learn a GP model for each tile. Li *et al.* [41] use K-means to partition the training dataset into a set of clusters and then learn one local GP model regarding each cluster. Wang *et al.* [42] adopt grid sampling to learn the non-local similarities by GP model, where an adjustable sampling interval is used to balance the reconstruction quality and computational complexity. Generally, all these existing GPR-based methods utilize the Gaussian likelihood to make inference. However, they do not examine whether the likelihood assumption is compatible with the actual observation model. Moreover, their computation is considerably intensive when learning from a large dataset for a regression task.

Aiming at the above issues of GPR-based SR, we propose a novel GPR-based SR framework by incorporating dictionary-based sampling (DbS) strategy and GPR model with Student-*t* likelihood, called DSGPR. The DbS is based upon sparse representation that the training samples can be well represented by a set of atoms in an over-complete dictionary, i.e., the sample space can be spanned by the atoms. Because these atoms are not real samples, it is insufficient to directly use these atoms to learn the mappings from the LR to HR spaces. As such, a natural alternative is to approximate each atom by its neighbors found in the original training dataset. By this way, the combination of the atoms' neighbors is suitable to approximately represent the whole samples and enables to train a GPR model with a relatively lower cost. In addition, we statistically find that a heavy-tailed Student-*t* likelihood

is more robust and suitable than commonly-used Gaussian likelihood for GPR-based SR.

The major contributions of the proposed method are three-fold: 1) In order to decrease the computational complexity while keeping the accurate mapping, we suggest using dictionary atoms to select a representative subset from the original large training set for GPR training. 2) Unlike the previous GPR-based SR methods that use Gaussian likelihood to train the model, we incorporate Student-*t* likelihood to learn the hyper-parameters based on extensive statistical analysis. The new model shows more robust and effective than these traditional methods. 3) It is experimentally shown that the newly proposed SR framework produces finer details than other existing competitors.

The rest of the paper is organized as follows. Section II provides an overview of the proposed framework. Section III introduces the details of the proposed method and statistically validates the rationality of Student-*t* likelihood assumption for GPR-based SR. Section IV provides the experimental results in comparison with other state-of-art methods. Finally, we conclude the paper in Section V.

II. OVERVIEW OF THE PROPOSED SR FRAMEWORK

For ease of description, the lowercase bold letters such as \mathbf{x} and \mathbf{y} denote column vectors, the lowercase letters such as x and y denote scalars, and the capital letters such as X and Y denote matrices or images. Fig. 1 illustrates the training flowchart the proposed SR framework. As demonstrated, there are two phases involved in the proposed SR framework, i.e., the training phase and the testing phase, respectively.

In the training phase, we first blur and down-sample each HR training image H_q to mimic the degradation of the observed LR image, and then upscale the degraded LR image with bicubic interpolation algorithm to produce the coupled interpolated LR image I_q . Next, we create the original training dataset L by extracting the patches from I_q and the center pixels of corresponding patches from the high frequency (HF) image via $F_q = H_q - I_q$. Then we use K-SVD algorithm [44] to train a dictionary D based on the prepared training dataset L . With the learned dictionary D , we find the neighborhoods of each atom from L and combine them together to build a compact training subset T for the learning of GPR. Finally, we train a GPR model with Student-*t* likelihood to learn the mappings from the interpolated image to the missing HF details.

In the testing phase, we first interpolate the input LR image I to the desired size with bicubic interpolation algorithm. Then all the patches in the interpolated image S_I are taken as inputs to predict the missing HF details in the input LR image via the learned GPR model. Next the predicted HF details are added into the interpolated test image to produce the initial SR estimate. Finally, we employ iterative back projection (IBP) algorithm to further improve the consistence between the resultant image and input LR image.

III. THE PROPOSED FRAMEWORK

In this section, we follow the previously defined notations and present the details of the proposed SR framework.

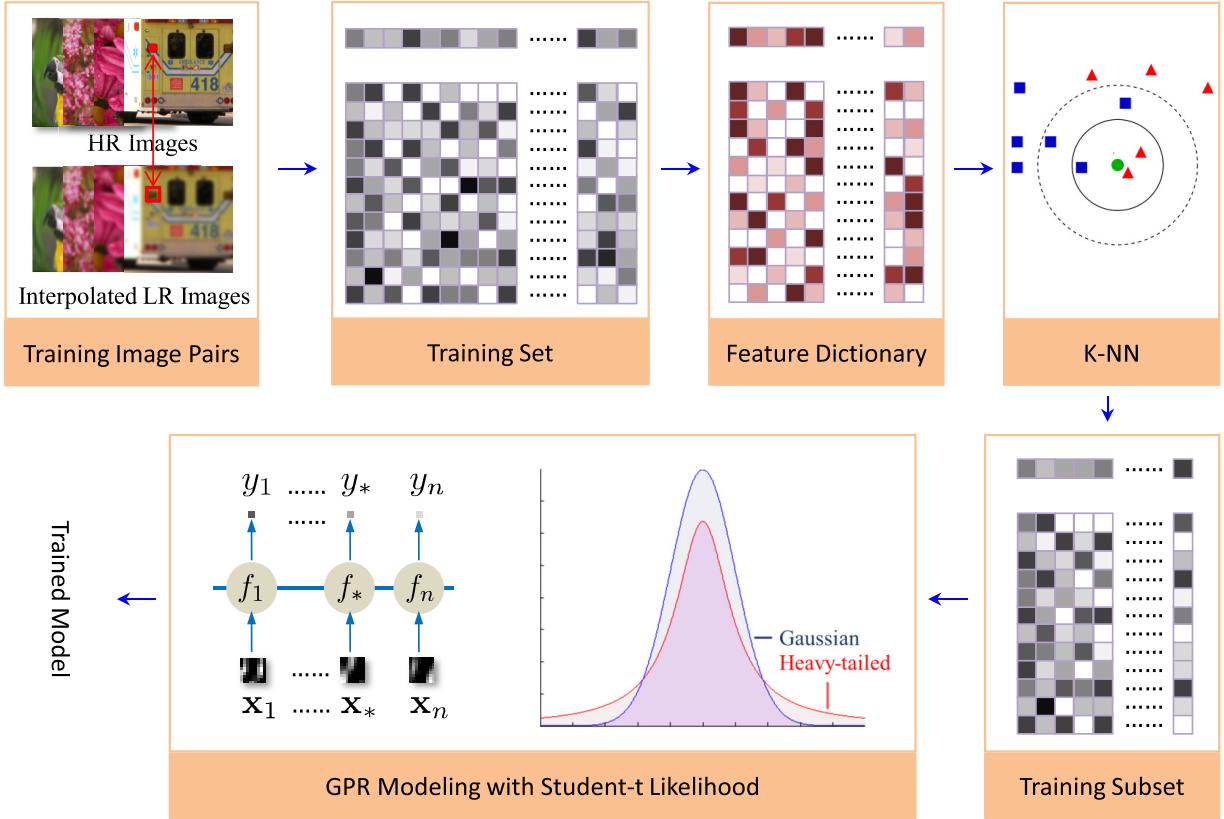


Fig. 1. Pipeline of the training phase of the proposed method.

A. Sampling From the Training Dataset

The computation of traditional GPR is expensive when a large original training set is directly applied to the learning task. To overcome this difficulty, we construct a representative subset from the original training dataset to sparsify the GP model. In this paper, we use a set of dictionary atoms in $D = [\mathbf{d}_1, \dots, \mathbf{d}_l]$ to achieve this purpose in that the atoms in a learned dictionary can be regarded as reasonable representatives of the original training dataset L . Considering that atoms are not real samples and insufficient to model the mapping relationship, we find q neighborhoods $\{\mathbf{n}_{i1}, \dots, \mathbf{n}_{iq}\} \subset L$ for each atom \mathbf{d}_i and combine all the neighborhood samples to produce the training subset $T = \cup_{i=1}^l \{\mathbf{n}_{i1}, \dots, \mathbf{n}_{iq}\}$ with the size of $n = l \times q$. As a result, the training space can also be approximately spanned by T . This mechanism can be roughly explained in the following reason. Given a sample $\mathbf{s}_i \triangleq <\mathbf{x}_i, y_i> \in L$, it can be represented as sparse linear combination of the atoms in a learned dictionary, i.e., $\mathbf{s}_i \approx D\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the corresponding coefficient vector. Since each atom \mathbf{d}_i can be approximately represented by its neighbors in T , namely $\mathbf{d}_i \approx T\boldsymbol{\mu}_i$, there exists a representation matrix A satisfying $D \approx TA$, where $A = [\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_l]$. Hence, T spans the training sample space by $\mathbf{s}_i \approx D\boldsymbol{\alpha} \approx TA\boldsymbol{\alpha} = T\boldsymbol{\eta}$.

B. GPR Using Gaussian Likelihood

GPR is one of most popular statistical learning method that enables to model an unknown non-linear mapping function $f(\cdot)$ hidden in a dataset [45]–[48].

Given the reduced training subset $T \triangleq \{<\mathbf{x}_i, y_i>\}_{i=1}^n$, GPR assumes that the training dataset satisfies the observation model:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{y} \triangleq [y_1, \dots, y_n]^T$, $\mathbf{f} \triangleq [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]^T$, $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n]^T$, and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma_n^2)$, i.e., the error or noise term is independent and identically distributed (i.i.d.) Gaussian. Our learning task is to estimate the latent noiseless functions $\mathbf{f}^* = [f(\mathbf{x}_1^*), \dots, f(\mathbf{x}_m^*)]^T$ for the prediction of the noiseless HF details missing in the interpolated LR image.

GPR uses a kernel $k(\mathbf{x}, \mathbf{x}')$ to evaluate the covariance between \mathbf{x} and \mathbf{x}' . Note that the targets $\{y_i\}_{i=1}^n$ are extracted from the HF images, so their statistical means are zeros. A joint Gaussian distribution of the output and estimated function is formulated as follows:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right), \quad (2)$$

where $X \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ is the input matrix of training samples, $X^* = [\mathbf{x}_1^*, \dots, \mathbf{x}_m^*]^T$ is the input matrix of test samples, and $K(X_1, X_2)$ is an $r_1 \times r_2$ -dimensional covariance matrix evaluated by all the samples in X_1 and those in X_2 , where r_i is the column (sample) number of X_i .

In essence, the learning of GPR is to obtain the optimal value of hyper-parameters in the kernel function $k(\mathbf{x}, \mathbf{x}')$, which can be achieved by maximizing the marginal likelihood $p(\mathbf{y}|X)$. The prior is Gaussian, i.e., $\mathbf{f}|X \sim \mathcal{N}(\mathbf{0}, K)$, and the

likelihood is a factorized Gaussian $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2 I)$, where K denotes $K(X, X)$. We can compute the log marginal likelihood $\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2}\log 2\pi$ because $p(\mathbf{y}|X) = \int p(\mathbf{y}|\mathbf{f}, X)p(\mathbf{f}|X)d\mathbf{f}$. Finally, we can utilize conjugate gradient method to learn the GPR model.

Based on the condition distribution of joint Gaussian, we have the predictive distribution as follows:

$$\mathbf{f}^*|X, \mathbf{y}, X^* \sim \mathcal{N}(\bar{\mathbf{f}}^*, \text{cov}(\mathbf{f}^*)), \quad (3)$$

where

$$\bar{\mathbf{f}}^* = K(X^*, X)K_y^{-1}\mathbf{y}, \quad (4)$$

$$\text{cov}(\mathbf{f}^*) = K(X^*, X^*) - K(X^*, X)K_y^{-1}K(X, X^*), \quad (5)$$

$$K_y \triangleq K_f + \sigma_n^2 I = K(X, X) + \sigma_n^2 I. \quad (6)$$

The posterior mean $\bar{\mathbf{f}}^*$ in (4) is an optimal Bayesian estimate of \mathbf{f}^* , and the posterior variance in $\text{cov}(\mathbf{f}^*)$ provides the predictive confidence.

As aforementioned, because the error terms $\epsilon_i \in \epsilon$ in an ordinary GPR are usually assumed as i.i.d. Gaussian, the observation model is formulated as

$$p(y_i|f_i, \sigma_n^2) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y_i - f_i)^2}{2\sigma_n^2}\right), \quad (7)$$

where $f_i = f(\mathbf{x}_i)$.

Gaussian likelihood is simple for the inference because the partial derivative of the marginal likelihood is easily computed. To be more specific, the GP prediction is reduced to compute the mean and covariance of a multivariate Gaussian, which can be done exactly by simple matrix algebra. For this reason, most existing GPR-based SR methods typically use Gaussian likelihood to learn the mapping relationships from the prepared training dataset. However, due to the complicated imaging procedure, Gaussian likelihood may not well characterize the true observation model. In the following, we will experimentally show that Gaussian likelihood is also sensitive to the outliers in the training set.

C. GPR With Student-*t* Likelihood

In this section, we first statistically analyze the reason why we do not use Gaussian likelihood, and then follow Student-*t* likelihood for GPR-based SR reconstruction.

1) Tests on Gaussian Likelihood: Gaussian likelihood gets popular in the existing GPR model for various predictive tasks. When it is applied to SR, the rationality is not verified at all in literatures. In this subsection, we make a statistical analysis on whether the Gaussian likelihood is suitable for SR or not.

Under the observation model $y_i = f(\mathbf{x}_i) + \epsilon_i$ and the assumption that ϵ_i is i.i.d. Gaussian, i.e., $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$, the predictive residuals should be Gaussian too. Here the residuals are referred to as the difference between the SR estimate and the observed HR image. In order to validate the assumption, we conduct a validation experiment by reconstructing all the 69 training images used in [23] under the proposed framework with Gaussian likelihood, and collect predictive residuals for each image, and part of the results are shown in Fig 2.

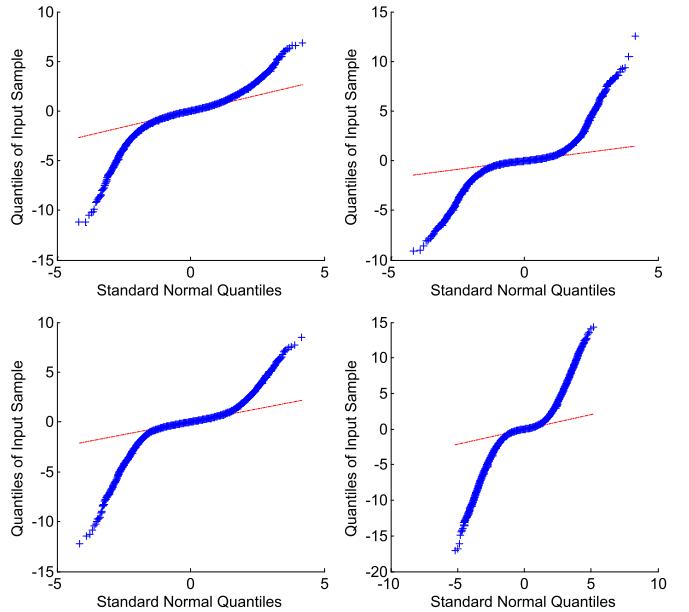


Fig. 2. QQ plot of sample data versus standard normal. The subfigures from top to bottom, from left to right are QQ plots of the residuals for the first, second, and third image among the 69 training images, and the QQ plot of the overall residual.

In Fig. 2, we use quantile-quantile (QQ) plot to illustrate whether the residual distribution is Gaussian or not. In statistics, QQ plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other. If the Gaussian assumption is true, the QQ plot should be along a straight line. The results show that all the 69 QQ plots are similar to each other. Due to the space limitations, Fig. 2 only shows the QQ plots of the residuals for the first, second, and third training image (TrImg) among the 69 training images, and the QQ plot of the average residual, respectively. However, the results are not as Gaussian as we think. Suppose ζ_q and ζ_q are the q -quantile of the sample residual distribution and the standard normal distribution, respectively. Then we can find that ζ_q satisfies $\zeta_q < \zeta_q$ for $\forall q < t_l$, and $\zeta_q > \zeta_q$ for $\forall q > t_u$, where t_l and t_u are two constants satisfying $t_l < 0.5$ and $t_u > 0.5$. These plots indicate that the likelihood is a zero-mean distribution with a heavier tail than Gaussian distribution.

Besides the QQ plot, we also statistically test the Gaussian assumption by “nortest” package of R. The results are shown in Table I, where AD, CVM, Lillie, and Pearson represent Anderson-Darling test, Cramer-von Mises test, Lilliefors (Kolmogorov-Smirnov) test, and Pearson chi-square test, respectively. Based on the results presented in Table I, all the p-values are much smaller than the specified significance level of 0.01, so it is reasonable to reject the null hypothesis that the errors are Gaussian.

Based on the above discussions, Gaussian likelihood is not the best representation of residual distribution from the probability perspective, and a heavy tailed distribution is more suitable for the likelihood of GPR.

In addition to statistical incompatibility, [49] also shows that the prediction accuracy of the GPR with Gaussian likelihood

TABLE I
THE DETAILS OF THE STATISTICAL TEST FOR NORMALITY

Method	Item	1st TrImg	2nd TrImg	3rd TrImg	Average
AD	test statistic	703.15	2275.3	1362.2	221710
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
CVM	test statistic	125.87	441.33	242.89	42638
	p-value	7.37e-10	7.37e-10	7.37e-10	7.37e-10
Lillie	test statistic	0.092366	0.17956	0.14426	0.15056
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
Pearson	test statistic	6745	26054	12260	3214700
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

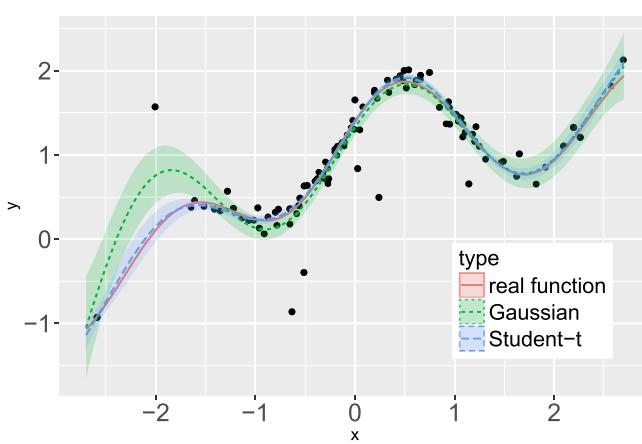


Fig. 3. A demo to demonstrate the GPR with outliers in [49]. There are 100 training cases are shown as dots, with the input on the horizontal axis, and the target on the vertical axis. The red solid line is the real function, and the blue dotted line gives the mean of the predictive distribution using GPR in which the likelihood was assumed to come from a Student- t distribution. The green dotted line gives the mean of the predictive distribution using GPR in which the likelihood was assumed to be Gaussian. The color shading areas are the corresponding confidence intervals of GPR prediction with different likelihoods. The leftmost outlier at around (-2, 1.5) pulls the posterior mean of the unknown function value predicted by GPR with Gaussian likelihood considerably far from the real function.

may be significantly reduced due to its sensitivity of outliers, which can be explained by Fig. 3.

In Fig. 3, obviously, the leftmost outlier pulls the posterior mean of the unknown function value considerably far from the real function. The result implies that GPR with Gaussian likelihood is not robust against outliers. In practice, the image patches in training dataset are in diverse structures, so the existence of outliers may affect the performance of GPR if Gaussian likelihood is used.

2) *Student- t Likelihood:* Assuming that the error (or noise) terms are i.i.d. Student- t distribution, i.e., $\epsilon_i \sim T(0, \sigma_n^2, \nu)$, the observation model becomes

$$p(y_i|f_i, \sigma_n^2, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi\sigma_n^2}} \left(1 + \frac{(y_i - f_i)^2}{\nu\sigma_n^2}\right)^{-(\nu+1)/2}, \quad (8)$$

where ν is the degree of freedom, and σ_n is the noise variance which is determined by hyper-parameter settings. We will detail the settings of them in Section III-E.

Comparing with the Gaussian distribution, Student- t distribution is heavy-tailed and more suitable to model the likelihood of the proposed SR framework. Fig. 3 compares the predictive

differences between the Gaussian likelihood and Student- t likelihood. Obviously, GPR with Student- t likelihood shows better robustness against outliers in the training dataset.

D. Kernel Selection

The key of GPR is how to select a suitable kernel function $k(\mathbf{x}, \mathbf{x}')$ to well evaluate the similarity between samples. Considering the anisotropic characteristic of image patches, we combine the anisotropic linear kernel and the noise kernel in the proposed model, i.e.,

$$k(\mathbf{x}, \mathbf{x}') = c\mathbf{x}^\top\mathbf{x}' + \sigma_n^2\delta(\mathbf{x} - \mathbf{x}'), \quad (9)$$

where c scales $\mathbf{x}^\top\mathbf{x}'$ to be closer to the true covariance, σ_n^2 is the noise variance, and δ is the Kronecker delta function. The noise kernel $\sigma_n^2\delta(\mathbf{x} - \mathbf{x}')$ is beneficial to keep the robustness against noise and guarantee the positive definite of the gram matrix K_y .

E. Hyper-Parameter Settings

There are several hyper-parameters such as c , σ_n^2 and ν to be determined, which can be optimized by maximizing the marginal likelihood, i.e., $\theta^* = \arg \max_{\theta} p(\mathbf{y}|X)$, where $\theta = \{c, \sigma_n, \nu\}$.

We initialize these hyper-parameters according to their physical meanings. σ_n is the noise variance. Since interpolation image I_q can be seen as an estimation of the latent HR image H_q , it is reasonable that $H_q - I_q$ estimates the noise. Therefore, σ_n^2 is estimated as below,

$$\sigma_n^2 \triangleq \frac{1}{n-1} \sum_{i=1}^n \left[y_i - \left(\frac{1}{n} \sum_{j=1}^n y_j \right) \right]^2, \quad (10)$$

where $y_i, y_j \in H_q - I_q$. c is a coefficient to scale the kernel function to be closer to the true covariance, which provides scalability and adaptation to the composite kernel, so we set c to 1 naturally. The freedom degree ν determines the degree of heavy-tail. The smaller ν is, the heavier the tail is. So ν is initialized to 1 to make the distribution of likelihood heavy-tailed.

When the hyper-parameters are initialized, we compute the marginal likelihood $p(\mathbf{y}|X)$ by

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|\mathbf{f}, X)p(\mathbf{f}|X)d\mathbf{f}. \quad (11)$$

Suppose that ϵ in Eqn. (1) is i.i.d., the likelihood $p(\mathbf{y}|\mathbf{f}, X)$ can be computed as

$$p(\mathbf{y}|\mathbf{f}, X) = \prod_i^n p(y_i|f_i, \sigma_n^2, \nu). \quad (12)$$

With the Student- t likelihood, the marginal likelihood and the conditional posterior $p(\mathbf{f}^*|X, \mathbf{y}, \mathbf{y}^*)$ are not analytically tractable anymore. In this paper, we resort to variational Bayes method [45], [49] to make an approximate inference for optimizing the hyper-parameters.

F. Iterative Back Projection

Fundamentally, an observed LR image is seen as the outcome of underlying scene via a typical degradation process including blurring, down-sampling and noising

Algorithm 1 DSGPR-Based SR

```

1 Input :  $u$  training interpolated LR and HR image pairs
       $\{I_q, H_q\}_{q=1}^u$ 
2 Output : Final SR result  $S_F$ 
   // Training Phase
3  $\{I_q, F_q\}_{q=1}^u = \{I_q, H_q - I_q\}_{q=1}^u;$ 
4 Randomly select  $v$  patch pairs  $\{P_l^{(I)}, P_l^{(F)}\}_{l=1}^v$  from  $\{I_q, F_q\}_{q=1}^u$ ;
5 Let  $L \triangleq \{\langle x_l, y_l \rangle\}_{l=1}^v = \{\langle P_l^{(I)}, \text{cen}(P_l^{(F)}) \rangle\}_{l=1}^v$ ;
   //  $\text{cen}(\cdot)$  returns the center pixel of a patch
6 Use K-SVD to train a dictionary  $D = [\mathbf{d}_1, \dots, \mathbf{d}_n]$ ;
7 Find the  $q$  neighborhoods  $\{\mathbf{n}_{i1}, \dots, \mathbf{n}_{iq}\}$  for each atom  $\mathbf{d}_i$  from the
   training dataset  $L$ ;
8 Combine all the neighborhood to produce the refined training subset  $T$ ,
   i.e.,  $T \triangleq \bigcup_{i=1}^q \{\mathbf{n}_{i1}, \dots, \mathbf{n}_{iq}\}$ ;
9 Train the GPR model  $f(\cdot)$  based on  $T$ ;
   // Testing Phase
10 Interpolate test LR image  $I$  to  $S_I$ ;
11 Initialize SR result image  $S_H = S_I$ ;
12 for each test input patch in  $\mathcal{T} \triangleq \{\mathbf{x}_k^*\}_{k=1}^m$  from  $S_I$  do
13    $y_k^* \leftarrow f(\mathbf{x}_k^*)$ ;
14    $y_k^* \leftarrow y_k^* + \text{cen}(\mathbf{x}_k^*)$ ;
15   Replace the corresponding pixel in  $S_H$  with  $y_k^*$ ;
16 end
17 Apply IBP on  $S_H$  to get  $S_F$ ;
18 return  $S_F$ ;

```

as follows,

$$Y = (X * B) \downarrow + \omega,$$

where X is an known HR image, Y is the observed LR image, B is a blur kernel, \downarrow is the down-sampling operation, and ω is the Gaussian white noise. The task of SR aims to estimate an HR image X from a given LR image Y .

To enhance the consistency between the input LR image and the resultant HR image, most previous example learning-based SR approaches often employ iterative back projection (IBP) algorithm to improve the quality of SR estimate. IBP assumes that the super-resolved result S_H should generate the same LR image Y under the same degradation model, which can be modeled by an L2-norm minimization problem as

$$\arg \min_{S_H} \|Y - (S_H * B) \downarrow\|_2^2. \quad (13)$$

The local optimal solution to (13) can be sought by a gradient descent as

$$S_H^{(t+1)} = S_H^{(t)} + \tau \left((Y - (S_H^{(t)} * B) \downarrow) \uparrow \right) * p, \quad (14)$$

where τ is the iterative step size and p is a “back-projection” filter.

G. Summary of the Proposed SR Algorithm

To provide a better view of the proposed DSGPR based SR method, Algorithm 1 provides the core steps of the proposed DSGPR-based SR method.

H. Complexity Analysis

In the training phase, the time complexity and space complexity of K-SVD are about $O((r+1) \times v^2)$ and $O((r+1) \times v)$, respectively, where r is the square image patch size and v is

TABLE II
THE AVERAGE PERFORMANCE COMPARISON OF $3 \times$ SR ON
THREE DATASETS FOR DSGPR FRAMEWORK
WITH DIFFERENT LIKELIHOODS

Dataset	Sech2	Laplace	Student- <i>t</i>	Gaussian
Kodak	27.82	27.82	27.88	27.73
	0.901	0.900	0.903	0.898
	0.949	0.948	0.950	0.947
Set5	29.96	30.02	30.07	28.90
	0.879	0.880	0.881	0.865
	0.900	0.901	0.902	0.891
Set14	27.03	27.05	27.11	26.38
	0.863	0.862	0.865	0.850
	0.914	0.914	0.916	0.903

the number of original training samples. The time complexity and space complexity of k -nearest neighbors (KNN) are $O((r+1) \times vq)$ and $O((r+1) \times v)$, respectively. In addition, the time complexity and space complexity of GPR are $O(n^3)$ and $O(n^2)$, respectively.

In the testing phase, the time complexity and space complexity of GPR prediction are $O((n \times m) \times r^2 + (n \times m))$ and $O(n \times m)$, respectively, where m is the number of image patches extracted from the magnified LR image by bicubic interpolation.

IV. EXPERIMENTAL RESULTS

In order to verify the effectiveness of the proposed SR method, six state-of-art methods, including BPJDL [31], SpReg [39], NARM [30], SCSR [23], UDF [4], and SRGPR [40], are used as competitors to evaluate the SR performance. We conduct $3 \times$ SR experiments on 24 test images from Kodak (<http://r0k.us/graphics/kodak/>), “set5” [43], and “set14” [27] dataset. All the test LR images are mimicked by first blurred with a 7×7 Gaussian kernel with a standard deviation 1.1 and then down-sampled with a factor of 3. We use K-SVD to train the dictionary by using 100,000 randomly selected patches, and the default dictionary size and neighborhood size are set to 512 and 1 experimentally. Note that we use the same 69 training images as in [23] to prepare our training set. Three quantitative assessment indices of PSNR, SSIM, and FSIM are employed to evaluate the objective quality of the obtained SR results [50]. We conduct all the comparison experiments on a PC with Intel Core i3-3220 3.30GHz CPU and 8GB RAM.

For convenience in demonstrating results, we denote the proposed framework with Gaussian likelihood as DSGPR-G, and the same framework with Student-*t* likelihood as DSGPR-S.

A. Determination of Parameters

Both the dictionary size and the neighborhood size are two paramount parameters that affect the final SR performance. In order to determine the reasonable parameters, we use DSGPR-G to conduct a set of validation experiments based on the same training set under different parameter settings. Fig. 4 shows the variations of the SR performance with different combinations of the two parameters.

Based on Fig. 4, we can see that, for each fixed neighborhood size, bigger dictionary size leads to higher values of

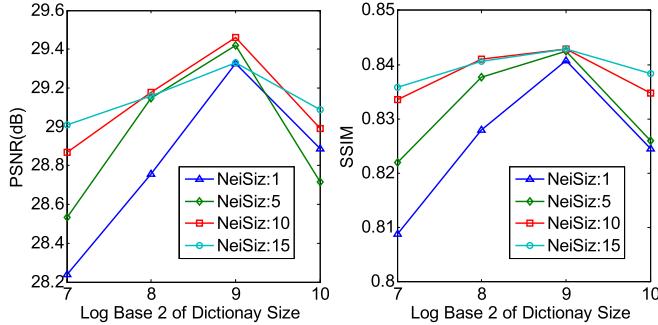


Fig. 4. Performance variations with different combination of dictionary size and neighborhood size (NeiSiz). The horizontal axis of each subfigure denotes the different candidate dictionary sizes such as $128(2^7)$, $256(2^8)$, $512(2^9)$, $1024(2^{10})$. The lines with different colors and shapes correspond to the performance at different neighborhood sizes.

TABLE III

THE 3 \times PERFORMANCE COMPARISON ON KODAK DATASET FOR DSGPR FRAMEWORK WITH GAUSSIAN AND STUDENT-*t* LIKELIHOODS

Img	Gau	Stu	Img	Gau	Stu	Img	Gau	Stu
1	24.21	24.26	9	29.36	29.47	17	29.83	30.02
	0.819	0.822		0.939	0.942		0.939	0.943
	0.921	0.922		0.952	0.955		0.960	0.963
2	31.20	31.27	10	29.16	29.33	18	26.18	26.35
	0.908	0.910		0.930	0.935		0.881	0.887
	0.951	0.953		0.948	0.952		0.941	0.945
3	31.51	31.70	11	27.27	27.40	19	26.06	26.19
	0.940	0.945		0.882	0.887		0.880	0.885
	0.967	0.970		0.939	0.943		0.936	0.939
4	30.69	30.89	12	30.45	30.52	20	28.15	28.34
	0.922	0.927		0.924	0.927		0.943	0.946
	0.969	0.971		0.960	0.962		0.959	0.963
5	24.07	24.26	13	22.49	22.64	21	26.40	26.56
	0.869	0.875		0.799	0.806		0.909	0.913
	0.928	0.935		0.923	0.927		0.933	0.938
6	25.93	26.03	14	26.83	27.04	22	28.57	28.67
	0.841	0.847		0.881	0.888		0.892	0.894
	0.925	0.928		0.946	0.952		0.953	0.955
7	29.82	30.04	15	29.31	29.44	23	31.41	31.72
	0.948	0.952		0.937	0.940		0.965	0.967
	0.952	0.957		0.968	0.970		0.981	0.983
8	21.95	22.02	16	29.42	29.58	24	25.21	25.38
	0.848	0.850		0.882	0.889		0.881	0.886
	0.920	0.922		0.954	0.957		0.938	0.941

PSNR and SSIM. When the dictionary size is 512, DSGPR-G achieves the best performance under different neighborhood sizes. To balance the SR quality and the computational efficiency, we suggest setting the dictionary size and the neighborhood size to 512 and 1 in that this configuration greatly reduces the complexity but very slightly sacrifices the reconstruction quality. In this case, it is easy to implement the proposed GPR-based SR on most resource-limited computation platforms.

B. Effects of Student-*t* Likelihood

In the framework of DSGPR, we have found that Student-*t* likelihood is more compatible with the true observation model than Gaussian likelihood. Since the heavy-tailed distributions available such as Weibull, Lognormal, and Lévy are only used

TABLE IV
THE 3 \times COMPARISON OF OBJECTIVE QUALITY ASSESSMENT ON 24 IMAGES OF KODAK DATASET

Image	SCSR	NARM	SpReg	SRGPR	BPJDL	UDF	DSGPR
1	23.26	23.87	22.99	22.32	23.17	23.08	24.26
	0.744	0.725	0.726	0.702	0.739	0.718	0.822
2	30.30	31.03	30.15	29.53	30.32	30.36	31.27
	0.879	0.866	0.875	0.861	0.879	0.880	0.910
3	30.65	31.81	30.46	29.07	30.74	30.61	31.70
	0.924	0.920	0.922	0.912	0.923	0.923	0.945
4	29.54	30.88	29.41	28.30	29.58	29.65	30.89
	0.884	0.877	0.880	0.868	0.884	0.887	0.927
5	22.83	23.91	22.62	20.96	22.88	22.59	24.26
	0.799	0.803	0.790	0.761	0.799	0.784	0.875
6	24.81	25.55	24.54	24.19	24.78	25.10	26.03
	0.786	0.780	0.770	0.761	0.778	0.781	0.847
7	27.46	29.75	27.11	25.08	27.50	27.22	30.04
	0.894	0.924	0.889	0.860	0.894	0.888	0.952
8	20.21	21.77	19.91	19.35	20.26	20.05	22.02
	0.754	0.802	0.737	0.718	0.755	0.737	0.850
9	27.53	29.56	27.17	26.50	27.61	27.26	29.47
	0.905	0.916	0.899	0.892	0.903	0.903	0.942
10	28.02	29.64	27.73	26.52	28.05	28.13	29.33
	0.887	0.904	0.880	0.863	0.887	0.894	0.935
11	26.10	27.06	25.87	25.04	26.11	26.01	27.40
	0.832	0.832	0.821	0.806	0.829	0.832	0.887
12	28.85	30.47	28.52	27.51	29.03	28.93	30.52
	0.892	0.895	0.886	0.878	0.889	0.897	0.927
13	21.60	22.15	21.40	21.03	21.55	21.61	22.64
	0.714	0.696	0.698	0.688	0.710	0.699	0.806
14	25.47	26.33	25.21	24.47	25.46	25.50	27.04
	0.823	0.812	0.813	0.804	0.820	0.820	0.888
15	27.50	29.52	26.96	26.11	27.61	27.23	29.44
	0.900	0.905	0.896	0.888	0.902	0.907	0.940
16	28.58	29.31	28.29	27.82	28.47	28.74	29.58
	0.847	0.839	0.834	0.827	0.840	0.848	0.889
17	28.19	28.70	28.03	26.47	28.27	28.32	30.02
	0.892	0.886	0.888	0.880	0.892	0.898	0.943
18	24.88	25.63	24.70	23.92	24.92	25.00	26.35
	0.809	0.796	0.800	0.783	0.810	0.810	0.887
19	24.67	25.87	24.19	23.29	24.75	24.43	26.19
	0.835	0.835	0.821	0.813	0.834	0.826	0.885
20	26.70	29.11	26.53	23.67	26.68	26.52	28.34
	0.917	0.920	0.912	0.904	0.916	0.915	0.946
21	25.06	26.13	24.78	24.14	24.98	25.05	26.56
	0.859	0.865	0.849	0.840	0.855	0.858	0.913
22	27.32	28.13	27.15	26.07	27.33	27.51	28.67
	0.832	0.821	0.824	0.807	0.832	0.837	0.894
23	30.21	31.79	29.96	28.16	30.28	30.09	31.72
	0.943	0.946	0.941	0.934	0.944	0.944	0.967
24	24.17	24.96	23.97	23.10	24.19	24.31	25.38
	0.814	0.822	0.805	0.784	0.816	0.821	0.886
Average	26.41	27.62	26.15	25.11	26.44	26.39	27.88
	0.849	0.849	0.840	0.827	0.847	0.846	0.903

for nonnegative regression but not suitable for the prediction of residuals. In order to verify the effectiveness of the heavy-tailed Student-*t* likelihood, here we conduct experiments using other three heavy-tailed distributions including Sech-square (Sech2), Laplace, Student-*t* to compare their average performance with Gaussian likelihood. The contrastive results with different likelihoods are listed in Table II. In the table, each dataset corresponds to three lines, which are PSNR, SSIM, and FSIM from top to bottom, respectively.

From Table II, we can see that: (1) The performance of three heavier-tailed likelihoods are all better than that of Gaussian. The fact tells us that that the heavier-tailed distributions are more suitable to model the likelihood of GPR. (2) Among the available likelihoods, Student-*t* achieves the best performance. Taking the average PSNR as an example, the student-*t* likelihood achieves the performance gain of 0.15dB, 1.17dB,

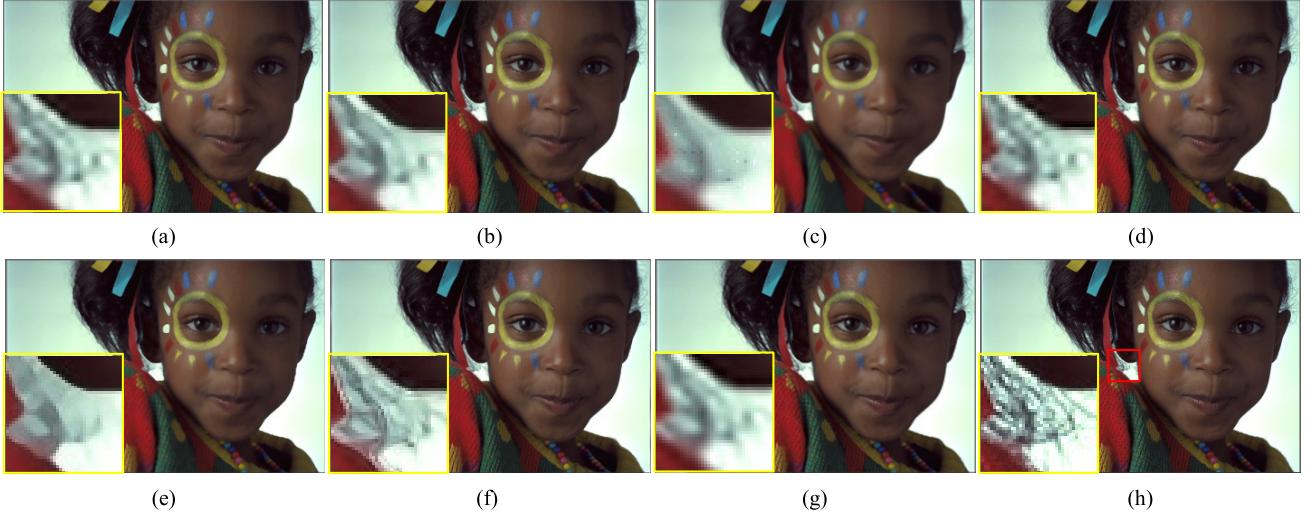


Fig. 5. Comparison of SR results ($3\times$) on *kodak15* image: (a) BPJDL [31] (PSNR: 27.61, SSIM: 0.902, FSIM: 0.937). (b) SpReg [39] (PSNR: 26.96, SSIM: 0.896, FSIM: 0.936). (c) NARM [30] (PSNR: 29.52, SSIM: 0.905, FSIM: 0.933). (d) SCSR [23] (PSNR: 27.50, SSIM: 0.900, FSIM: 0.937). (e) UDF [4] (PSNR: 27.23, SSIM: 0.907, FSIM: 0.933). (f) SRGPR [40] (PSNR: 26.11, SSIM: 0.888, FSIM: 0.928). (g) Proposed Method (PSNR: 29.44, SSIM: 0.940, FSIM: 0.970). (h) Original.

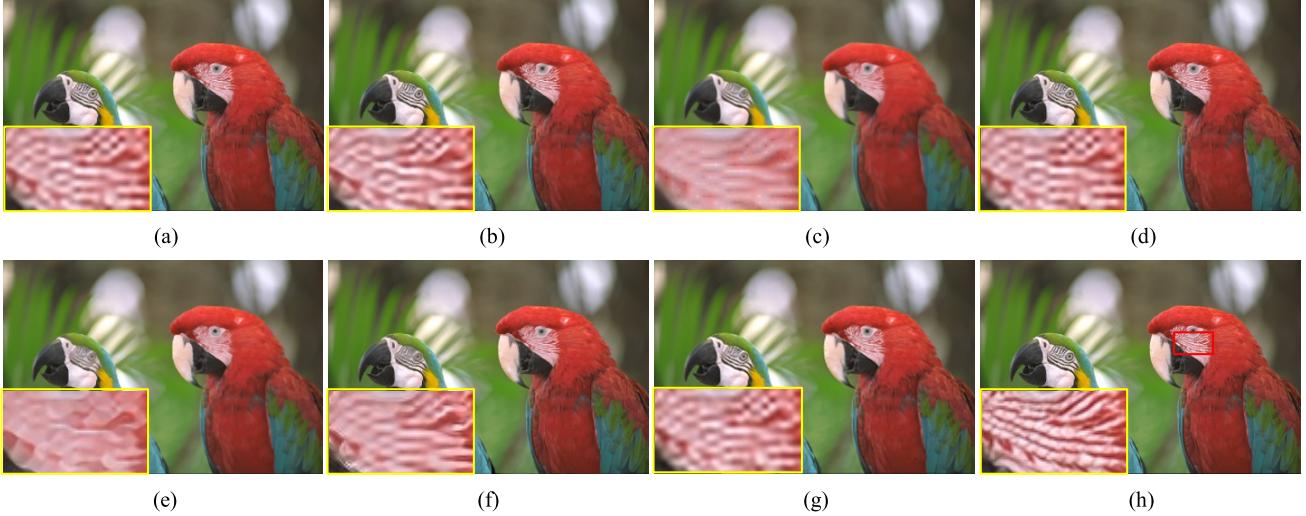


Fig. 6. Comparison of SR results ($3\times$) on *kodak23* image: (a) BPJDL [31] (PSNR: 30.28, SSIM: 0.944, FSIM: 0.965). (b) SpReg [39] (PSNR: 29.96, SSIM: 0.941, FSIM: 0.964). (c) NARM [30] (PSNR: 31.79, SSIM: 0.946, FSIM: 0.964). (d) SCSR [23] (PSNR: 30.21, SSIM: 0.943, FSIM: 0.965). (e) UDF [4] (PSNR: 30.09, SSIM: 0.944, FSIM: 0.962). (f) SRGPR [40] (PSNR: 28.16, SSIM: 0.934, FSIM: 0.959). (g) Proposed Method (PSNR: 31.72, SSIM: 0.967, FSIM: 0.983). (h) Original.

and 0.73dB over Gaussian likelihood in Kodak, Set5, and Set14 dataset, respectively. Therefore, it is reasonable to replace the Gaussian likelihood with Student-*t* likelihood for better reconstruction quality.

Since the performance gain in Kodak is the least significant among the three datasets, it is necessary to show the comparative details in Table III. By analyzing the structures of the images in each dataset, we notice that the images in the Kodak dataset are long shot relatively and contain large smooth regions. Therefore, the residuals (the subtraction between the HR image and the corresponding interpolated image) are closer to the mean. This means that there are less residuals in the tail of the distribution. So the departure of residual distribution from the Gaussian is not significant.

As a result, the performance gain in Kodak is not comparable to that in Set5 and Set14. However, in Table III, the maximum gain of PSNR is 0.31 for the 23th image, the minimum PSNR gain is 0.05 for the 1st image, and the average PSNR gain is 0.152. Therefore, DSGPR-S is still superior to DSGPR-G in terms of three assessment indices.

C. Quality Assessment

Among all the 24 test images in Kodak dataset, except that the PSNR values of 6 images are at the second rank, the quantitative assessment results of most test images obtained by the proposed method outperforms those of other competitors. Table IV lists the detail comparison of objective quality assessment between the proposed GPR-based SR method and



Fig. 7. Comparison of SR results ($3\times$) on *kodak18* image: (a) BPJDL [31] (PSNR: 24.92, SSIM: 0.810, FSIM: 0.897). (b) SpReg [39] (PSNR: 24.70, SSIM: 0.800, FSIM: 0.893). (c) NARM [30] (PSNR: 25.63, SSIM: 0.796, FSIM: 0.872). (d) SCSR [23] (PSNR: 24.88, SSIM: 0.809, FSIM: 0.896). (e) UDF [4] (PSNR: 25.00, SSIM: 0.810, FSIM: 0.881). (f) SRGPR [40] (PSNR: 23.92, SSIM: 0.783, FSIM: 0.884). (g) Proposed Method (PSNR: 26.35, SSIM: 0.887, FSIM: 0.945). (h) Original.

other competitors. Each method corresponds to two lines, which are PSNR and SSIM from top to bottom, respectively. The overall quality achieved by the proposed method is better than that by other competitors.

Besides the objective quality assessments, we further compare the subjective quality of different methods. Although the objective quality of *Kodak15* and *Kodak23* ranks at the second level, the subjective quality of the proposed method is comparable to other competitors. This can be seen from the magnified interest areas in the girl's hair in Fig. 5 and the bird's face in Fig. 6. Moreover, we also visually compare the results of *Kodak18* and *Kodak24* in Figs. 7 and 8, respectively. One can find that the folds and stripe in *Kodak18*, and beams in *Kodak24* are visually more compelling than those in the SR results obtained from other competitors.

D. Effectiveness of the Sampling Strategy

In order to validate the effectiveness of the dictionary-based sampling strategy, we compare the proposed DbS strategy with

the traditional random sampling (RS) strategy under the same GPR-based SR framework. In order to show the stability of the comparison, we repeat RS 50 times and compare the average results with DbS as illustrated in Fig. 9. Based on the Fig. 9, the proposed DbS strategy is superior to RS strategy in terms of PSNR, SSIM, and FSIM.

E. Acceleration

In Eqn. (4), if we compute the projection matrix $P \triangleq K_y^{-1}y$ in advance, we can avoid computing the inverse of K_y online for each prediction. In this way, the original prediction is reduced to a simple matrix multiplication, i.e., $\bar{\mathbf{f}}^* = K(X^*, X)P$, which can greatly accelerate the computational efficiency of DSGPR-based SR method. In Table V, we compare the average CPU-time of the seminal GPR-based method [40] with that of the proposed method. As shown, the running time of our method is significantly less than that of [40].

It is worthy of emphasizing that the complexity of the proposed method can be further reduced by parallel programming.

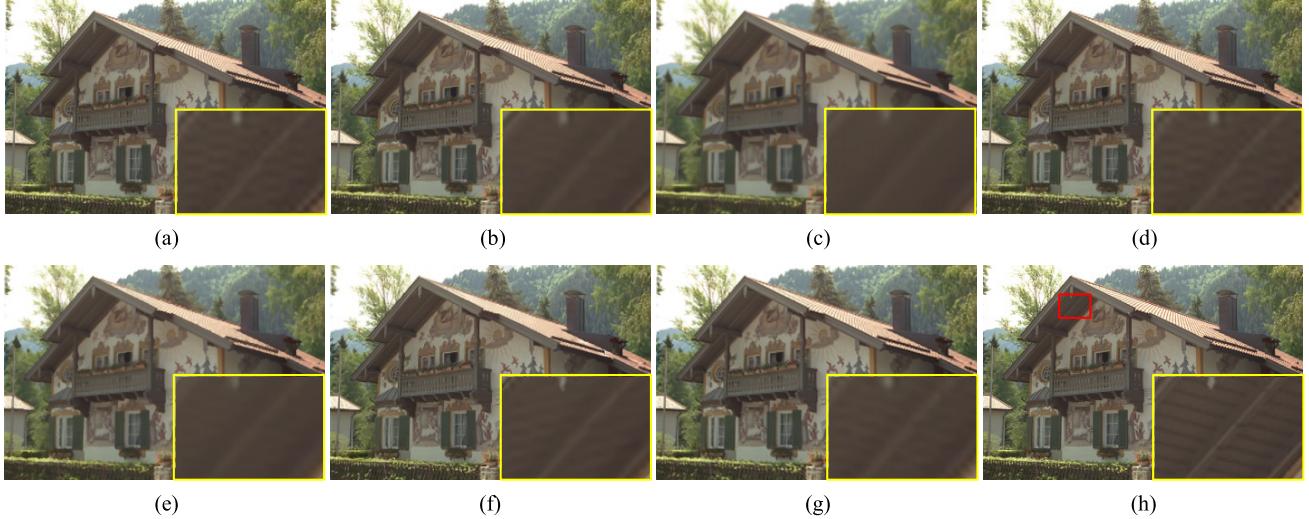


Fig. 8. Comparison of SR results ($3\times$) on *kodak24* image: (a) BPJDL [31] (PSNR: 24.19, SSIM: 0.816, FSIM: 0.897). (b) SpReg [39] (PSNR: 23.97, SSIM: 0.805, FSIM: 0.892). (c) NARM [30] (PSNR: 24.96, SSIM: 0.822, FSIM: 0.895). (d) SCSR [23] (PSNR: 24.17, SSIM: 0.814, FSIM: 0.896). (e) UDF [4] (PSNR: 24.31, SSIM: 0.821, FSIM: 0.886). (f) SRGPR [40] (PSNR: 23.10, SSIM: 0.784, FSIM: 0.873). (g) Proposed Method (PSNR: 25.38, SSIM: 0.886, FSIM: 0.941). (h) Original.

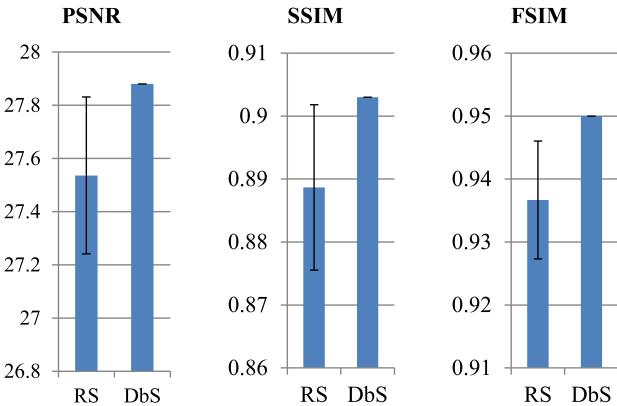


Fig. 9. Performance comparisons between DbS and RS strategies with the same training sample number of 512. The bars in the histogram represent the standard deviation.

TABLE V
THE COMPARISON OF AVERAGE PREDICTION CPU-TIME

Method	CPU Time (s)
Proposed	9.83
SRGPR in [33]	2624.13

Specifically, we can divide the test set $T^* = \{x_1^*, \dots, x_m^*\}$ into several independent groups $\{T_u^*\}$ s.t. $\cup_u T_u^* = T^*$ and $T_i^* \cap T_j^* = \emptyset$ for $i \neq j$, and then perform the prediction of each group T_u^* in parallel.

F. Performance on Other Upsampling Rate

Besides the performance evaluation on $3\times$ SR results, we also conduct $2\times$ SR to further verify the effectiveness of our proposed DSGPR. Table VI provides the compared results under the same experimental configuration as Sec. IV-A. In the table, each image corresponds to three lines, which are PSNR, SSIM, and FSIM from top to bottom, respectively. Taking the PSNR as example, the maximum gain is 0.54,

TABLE VI
THE $2\times$ PERFORMANCE COMPARISON ON KODAK DATASET FOR DSGPR FRAMEWORK WITH GAUSSIAN AND STUDENT-*t* LIKELIHOODS

Img	Gau	Stu	Img	Gau	Stu	Img	Gau	Stu
1	26.51	26.60	9	32.27	32.54	17	0.979	0.983
	0.933	0.939					0.988	0.992
	0.971	0.974						
2	32.99	33.16	10	31.94	32.25	18	28.65	29.01
	0.962	0.965		0.977	0.978		0.961	0.966
	0.985	0.986			0.985	0.989		0.981
3	33.62	33.84	11	29.52	29.75	19	28.96	29.09
	0.977	0.979		0.956	0.961		0.956	0.959
	0.989	0.990			0.979	0.982		0.975
4	33.31	33.70	12	32.37	32.53	20	30.71	30.88
	0.973	0.976		0.969	0.971		0.979	0.981
	0.990	0.993			0.986	0.987		0.986
5	26.85	27.35	13	24.58	24.86	21	28.89	29.17
	0.958	0.965		0.926	0.936		0.967	0.971
	0.976	0.983			0.973	0.977		0.977
6	28.01	28.15	14	29.43	29.78	22	30.92	31.20
	0.940	0.947		0.961	0.966		0.961	0.966
	0.973	0.974			0.984	0.987		0.984
7	32.92	33.46	15	31.48	31.70	23	34.28	34.67
	0.985	0.988		0.976	0.979		0.987	0.988
	0.986	0.991			0.988	0.990		0.994
8	24.31	24.46	16	31.31	31.39	24	27.54	27.89
	0.940	0.945		0.954	0.958		0.959	0.965
	0.967	0.971			0.983	0.985		0.978

the minimum gain is 0.08, and the average gain is 0.27. Table VI indicates that the Student-*t* likelihood is superior to the Gaussian likelihood when applied in the DSGPR method.

Table VII lists the details of $2\times$ performance comparisons between DSGPR and other competitors, and each image corresponds to two lines, which are PSNR and FSIM from top to bottom, respectively. Obviously, the proposed DSGPR method outperforms other competitors in terms of two quality assessment indices.

G. Performance on Other Benchmark Test Sets

In order to make the comparison more robust, other benchmark test datasets including “set5” and “set14” are applied to

TABLE VII

THE 2 \times COMPARISON OF OBJECTIVE QUALITY ASSESSMENT
ON 24 IMAGES OF KODAK DATASET

Image	SCSR	NARM	SpReg	SRGPR	BPJDL	UDF	DSGPR
1	25.18	25.21	25.20	24.69	25.29	24.90	26.60
	0.874	0.839	0.875	0.893	0.874	0.841	0.939
2	32.00	32.18	32.02	31.37	32.12	31.88	33.16
	0.939	0.913	0.937	0.942	0.939	0.925	0.965
3	32.56	33.01	32.67	32.23	32.69	32.62	33.84
	0.962	0.950	0.962	0.967	0.962	0.955	0.979
4	31.79	32.37	31.86	31.03	31.95	31.82	33.70
	0.943	0.928	0.943	0.949	0.943	0.935	0.976
5	25.08	25.30	25.22	24.50	25.29	25.09	27.35
	0.902	0.889	0.904	0.923	0.902	0.888	0.965
6	26.75	26.88	26.69	26.24	26.81	26.65	28.15
	0.891	0.865	0.891	0.910	0.890	0.867	0.947
7	30.46	31.56	30.59	29.63	30.64	30.90	33.46
	0.955	0.958	0.956	0.960	0.955	0.958	0.988
8	22.60	23.10	22.56	22.05	22.75	22.56	24.46
	0.883	0.875	0.884	0.892	0.883	0.865	0.945
9	30.33	31.11	30.36	29.89	30.48	30.31	32.54
	0.955	0.948	0.956	0.962	0.955	0.952	0.979
10	30.36	31.18	30.40	29.85	30.54	30.51	32.25
	0.947	0.944	0.947	0.952	0.947	0.946	0.978
11	28.20	28.39	28.25	27.79	28.35	28.14	29.75
	0.919	0.898	0.918	0.929	0.919	0.906	0.961
12	31.36	31.95	31.25	30.19	31.32	31.31	32.53
	0.948	0.933	0.946	0.951	0.948	0.940	0.971
13	23.39	23.41	23.39	23.22	23.47	23.21	24.86
	0.854	0.821	0.856	0.885	0.854	0.822	0.936
14	27.73	27.90	27.76	27.29	27.88	27.77	29.78
	0.915	0.894	0.915	0.931	0.915	0.903	0.966
15	29.98	30.55	30.09	29.44	30.08	30.05	31.70
	0.951	0.939	0.951	0.956	0.951	0.949	0.979
16	30.26	30.40	30.26	29.93	30.38	30.22	31.39
	0.923	0.899	0.922	0.933	0.923	0.907	0.958
17	30.79	30.93	30.81	29.52	30.96	30.97	33.11
	0.951	0.945	0.951	0.960	0.951	0.951	0.983
18	27.06	27.07	27.10	26.78	27.19	27.00	29.01
	0.909	0.889	0.910	0.927	0.909	0.896	0.966
19	27.18	27.41	27.16	25.53	27.34	26.91	29.09
	0.919	0.901	0.920	0.926	0.919	0.898	0.959
20	29.23	30.29	29.53	25.88	29.43	29.68	30.88
	0.959	0.949	0.959	0.965	0.959	0.955	0.981
21	27.31	27.57	27.35	27.09	27.48	27.37	29.17
	0.933	0.921	0.933	0.944	0.932	0.925	0.971
22	29.50	29.57	29.50	28.87	29.62	29.36	31.20
	0.919	0.896	0.919	0.931	0.919	0.906	0.966
23	32.50	33.43	32.69	31.51	32.66	32.46	34.67
	0.973	0.968	0.973	0.977	0.973	0.971	0.988
24	26.17	26.23	26.20	24.55	26.29	26.06	27.89
	0.911	0.898	0.912	0.919	0.911	0.900	0.965
Average	28.66	29.04	28.70	27.88	28.79	28.66	30.44
	0.926	0.911	0.927	0.937	0.926	0.915	0.967

TABLE VIII

THE 3 \times COMPARISON OF AVERAGE OBJECTIVE QUALITY ASSESSMENT ON SET5 AND SET14 DATASET

Dataset	SCSR	NARM	SpReg	SRGPR	BPJDL	UDF	DSGPR
Set5	28.09	28.42	26.57	25.19	26.89	26.43	30.07
	0.845	0.809	0.825	0.797	0.831	0.816	0.881
Set14	26.19	26.50	24.63	23.77	24.87	24.70	27.11
	0.841	0.721	0.789	0.772	0.798	0.789	0.865

evaluate the performance of newly proposed method. Due to the limited space, we only show their average performance of 3 \times SR in Table VIII, and each dataset corresponds to two lines, which are PSNR and FSIM from top to bottom, respectively. Based on the results in Table VIII, we can find that the proposed DSGPR method outperforms other competitors in all the cases for the two datasets.

V. CONCLUSION

In this paper, we have proposed a novel GPR-based SR method by integrating a dictionary-based sampling scheme and Student- t likelihood. To lessen the issue of the high complexity of GPR, the dictionary-based sampling is proposed to sparsify the original large training dataset by combining the atoms' neighborhoods into a compact training subset. Besides, we statistically verify that it is not commonly-used Gaussian likelihood but a heavy-tailed distribution compatible with the true observation model. Therefore, in order to improve the compatibility, we suggest using Student- t distribution to define the likelihood of the proposed GPR-based SR model. Experimental results validate the effectiveness of the proposed DSGPR-based SR framework and show the superiority objectively and subjectively in comparing with other competitors.

However, as a pixel-wise method, DSGPR-based SR method cannot fully utilize the structure information in the neighbor pixels of an image patch. For this issue, a multi-task learning-based GPR algorithm can be presented to design a patch-wise SR method for reducing artifacts and improving the consistency of image structures in the future research. Besides, deep learning-based methods have endlessly broken the records in many machine learning tasks and become the state-of-the-art [51]–[53], and how to incorporate deep learning into the GPR framework is also a challenge in our future work.

REFERENCES

- [1] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
- [2] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.
- [3] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.
- [4] L. Wang, H. Wu, and C. Pan, "Fast image upsampling via the displacement field," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5123–5135, Sep. 2014.
- [5] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324–335, Dec. 1993.
- [6] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4544–4556, Nov. 2012.
- [7] X. Gao, Q. Wang, X. Li, D. Tao, and K. Zhang, "Zernike-moment-based image super resolution," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2738–2747, Oct. 2011.
- [8] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, Dec. 2008.
- [9] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Process.*, vol. 90, no. 2, pp. 405–414, Feb. 2010.
- [10] J. Sun, Z. B. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
- [11] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph.*, vol. 26, no. 3, p. 95, Jul. 2007.
- [12] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 36–51, Jan. 2009.
- [13] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, "Fast image/video upsampling," *ACM Trans. Graph.*, vol. 27, no. 5, p. 153, Dec. 2008.
- [14] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.

- [15] K. Zhang, D. Tao, X. Gao, X. Li, and J. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1648–1659, Oct. 2013.
- [16] K. Zhang, X. Gao, D. Tao, J. Li, and X. Li, "Coarse-to-fine learning for single-image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1109–1122, May 2017.
- [17] X. Lu and X. Li, "Multiresolution imaging," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 149–160, Jan. 2014.
- [18] B. Ham, D. Min, and K. Sohn, "Depth superresolution by transduction," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1524–1535, May 2015.
- [19] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.
- [20] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2004, pp. 275–282.
- [21] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Oct. 2009, pp. 349–356.
- [22] X. Lu, Y. Yuan, and P. Yan, "Image super-resolution via double sparsity regularized manifold learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2022–2033, Dec. 2013.
- [23] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [24] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
- [25] X. Lu, Y. Yuan, and P. Yan, "Alternatively constrained dictionary learning for image superresolution," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 366–377, Mar. 2014.
- [26] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3194–3205, Jul. 2012.
- [27] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, Jun. 2010, pp. 711–730.
- [28] S. Mallat and Y. Guoshen, "Super-resolution with sparse mixing estimators," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2889–2900, Nov. 2010.
- [29] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2216–2223.
- [30] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.
- [31] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 345–352.
- [32] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1920–1927.
- [33] C.-Y. Yang and M.-Y. Yang, "Fast direct super-resolution by simple functions," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 561–568.
- [34] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 184–199.
- [35] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [36] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
- [37] K. S. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1596–1610, Jun. 2007.
- [38] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 846–861, Mar. 2015.
- [39] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.
- [40] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 449–456.
- [41] J. Li, Y. Qu, C. Li, Y. Xie, Y. Wu, and J. Fan, "Learning local Gaussian process regression for image super-resolution," *Neurocomputing*, vol. 154, no. 4, pp. 284–295, Apr. 2015.
- [42] H. Wang, X. Gao, K. Zhang, and J. Li, "Image super-resolution using non-local Gaussian process regression," *Neurocomputing*, vol. 194, no. 6, pp. 95–106, Jun. 2016.
- [43] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. 23rd Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2012, p. 135.
- [44] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [45] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006, pp. 7–30.
- [46] S. T. Ounpraseuth, "Gaussian processes for machine learning," *J. Amer. Statist. Assoc.*, vol. 103, no. 481, p. 429, Mar. 2008.
- [47] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2006, pp. 1257–1265.
- [48] C. E. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (GPML) toolbox," *J. Mach. Learn. Res.*, vol. 11, pp. 3011–3015, Nov. 2010.
- [49] R. M. Neal, "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification," Dept. Statist., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-9702, Jan. 1997.
- [50] F. Gao and J. Yu, "Biologically inspired image quality assessment," *Signal Process.*, vol. 124, no. 7, pp. 210–219, Jul. 2016.
- [51] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, Nov. 2016.
- [52] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.
- [53] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.



Haijun Wang received the B.Eng. degree in computer software and the M.Eng. degree in computer software and theory from Hunan University, Changsha, China, in 2002 and 2007, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2016. Since 2002, he has been with the School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, China. His research interests include machine learning and computer vision.



Xinbo Gao (M'02–SM'07) received the B.Eng., M.Sc., and Ph.D. degrees from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively, all in signal and information processing. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of the Ministry of Education, a Professor of Pattern Recognition and Intelligent System, and the Director of the State Key Laboratory of Integrated Services Networks, Xi'an. He has authored six books and around 200 technical articles in refereed journals and proceedings. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He is a fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He served as the General Chair/Co-Chair, the Program Committee Chair/Co-Chair, or the PC Member for around 30 major international conferences. He is on the Editorial Boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier).



Kaibing Zhang received the M.Sc. degree in computer software and theory from Xihua University, Chengdu, China, in 2005, and the Ph.D. degree from Xidian University, Xi'an, China, in 2012. From 2013 to 2015, he was a Post-Doctoral Research Fellow with the School of Electronic Engineering, Xidian University. He is currently an Associate Professor with the College of Electrics and Information, Xi'an Polytechnic University, Xi'an. His main research interests include pattern recognition and computer vision. In these areas, he has authored over 20 technical articles in refereed journals and proceedings including the IEEE TIP, TNNLS, CVPR, *Signal Processing* (Elsevier), Neurocomputing, and ICIP.



Jie Li received the B.Sc. degree in electronic engineering, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuit and systems from Xidian University, Xi'an, China, in 1995, 1998, and 2004, respectively. She is currently a Professor with the School of Electronic Engineering, Xidian University. Her research interests include image processing and machine learning. In these areas, she has authored around 50 technical articles in refereed journals and proceedings including the IEEE T-IP, T-CSVT, and *Information Sciences*.