Efficient Learning-based Image Enhancement: Application to Super-resolution and Compression Artifact Removal

Younghee Kwon[1]

Kwang In Kim[2]
http://www.mpi-inf.mpg.de/~kkim

Jin Hyung Kim[3]
http://ai.kaist.ac.kr/~jkim/

Christian Theobalt[2]
http://www.mpi-inf.mpg.de/~theobalt

[1] Google Inc.
Mountain View, CA, USA

[2] Max-Planck-Institut für Informatik
Saarbrücken, Germany

[3] KAIST
Daejeon, Korea

Many widely used imaging operations lead to specific degradations of images with respect to the ground truth. The removal of these degradations is one of the most important tasks in computer vision, image processing, and computational photography. For instance, image encoding deficiencies such as block artifacts have to be removed frequently. Deterioration and information loss due to the limitations of the optical system, such as limited sensor resolution or defocusing, should also be erased.

This paper presents an algorithm for learning-based image enhancement. At each pixel in the given degraded image, a small sub-window encompassing that pixel (patch) is extracted and the corresponding desired patch is estimated based on Gaussian process (GP) regression. As the output patches (i.e., the predictive means) overlap with their neighbors, the result of the regression step constitutes a set of candidates for each pixel location. The final pixel-valued output is synthesized by combining the candidates based on the corresponding predictive variances and trading the consistency with them with a global image prior as a regularizer [1].

While GP regression has been shown to be competitive on a wide range of small-scale applications, its application to large-scale problems is limited due to its unfavorable scaling behavior. A standard approximate approach to overcome this limitation is to introduce a small set of *inducing variables* $\mathbf{f}_{\mathcal{U}} = \{f(\mathbf{u}_1),\dots,f(\mathbf{u}_m)\}$ (corresponding to *inducing inputs* $\mathcal{U} = \{\mathbf{u}_1,\dots,\mathbf{u}_m\}$) through which the conditional independence of the training ($\mathbf{f}$) and target ($\mathbf{f}_*$) latent variables is assumed in the approximation of the joint prior (cf. the unified framework of [2]):

$$p(\mathbf{f}_*,\mathbf{f}) \approx q(\mathbf{f}_*,\mathbf{f}) = \int q(\mathbf{f}_*|\mathbf{f}_{\mathcal{U}})q(\mathbf{f}|\mathbf{f}_{\mathcal{U}})p(\mathbf{f}_{\mathcal{U}})d\mathbf{f}_{\mathcal{U}}. \tag{1}$$

The *training conditional* $q(\mathbf{f}|\mathbf{f}_{\mathcal{U}})$ is approximated subsequently. This leads to a set of approximations which are referred to as *sparse* GPs where the inference is carried out through $\mathbf{f}_{\mathcal{U}}$ summarizing l training data points.

In existing sparse GP algorithms, once identified, the inducing inputs $\mathcal{U}$ are fixed throughout the entire test set. The problem is then cast into an optimization where one constructs $\mathcal{U}$ based on a certain measure of approximation quality (e.g., marginal likelihood and information gain). The performance of a sparse approximation depends heavily on the inducing inputs $\mathcal{U}$. However, usually the corresponding optimization problem is non-convex and accordingly is not easy to solve.

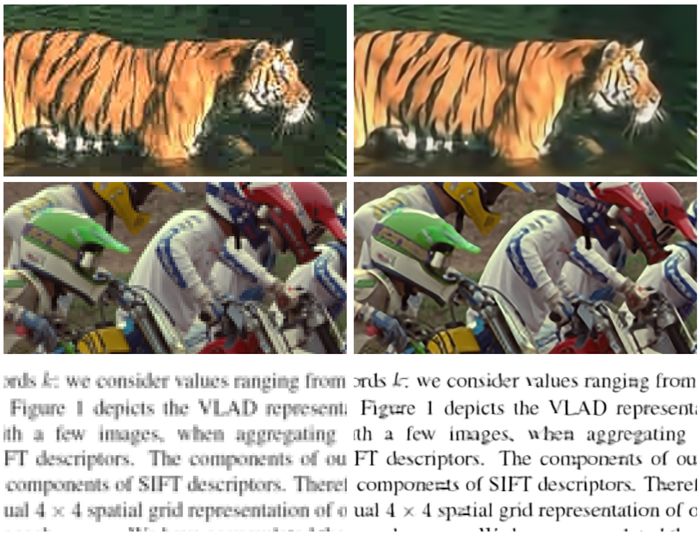
In this paper, we present a simple alternative to these *off-line* approaches: We build a sparse GP which is specially tailored for a given test input $\mathbf{x}_*$ (i.e., $\mathcal{U} \equiv \mathcal{U}_*$ is chosen depending on $\mathbf{x}_*$; The corresponding GP model is constructed only when it is presented with a test point $\mathbf{x}_*$). An important advantage of this *on-line* approach is that it naturally leads to an extremely simple strategy for identifying $\mathcal{U}_*$: If we introduce a spatial Markov assumption on $\{f_*,\mathbf{f}\}$

$$p(f_*|\mathbf{f},\mathcal{N}(f_*)) \approx q(f_*|\mathcal{N}(f_*)), \tag{2}$$

where $\mathcal{N}(f_*)$ denotes the values of the latent function f for the inputs in the spatial neighborhoods $\mathcal{N}(\mathbf{x}_*)$ (of $\mathbf{x}_*$), the decomposition (1) becomes exact once we use $\mathcal{N}(\mathbf{x}_*)$ for $\mathcal{U}_*$.

This approximation dramatically reduces the computation time during training. Actually, the only training component is building a data structure for nearest neighbor (NN)-search, which facilitates identifying $\mathcal{N}(f_*)$. However, for large scale problems ($l \approx 2*10^5$ in the current applications), this approximation might be still impractical. The second step of our approximation is to introduce an additional Markov-like assumption directly on the observations:

$$p(f_*|\mathcal{Y},\mathcal{N}_1(\mathbf{y}_*)) \approx q(f_*|\mathcal{N}_1(\mathbf{y}_*)), \tag{3}$$



Input Output

Figure 1: Examples of image enhancement: (top) JPEG artifact removal, (middle) (generic) single-image super-resolution, and (bottom) (document-specific) single-image super-resolution.

where $\mathcal{Y}$ is the set of training labels and $\mathcal{N}_1(\mathbf{y}_*)$ denotes the observed training target values in the spatial neighborhood $\mathcal{N}_1(\mathbf{x}_*)$ of $\mathbf{x}_*$. To guarantee that the resulting GPs are non-locally regularized, we set $\mathcal{N}(\mathbf{x}_*) \subset\subset \mathcal{N}_1(\mathbf{x}_*)$. The spatial Markov assumption (2) is fairly natural and has proven to be effective in many different applications while the second approximation step (3) is motivated by the large-scale behavior of full GPs: For large l, the predictive distribution $p(f_*|\mathcal{Y})$ of a full GP is not affected by the data points which are sufficiently distinct from $\mathbf{x}_*$ [3].

Since the only training component of the new approximation is building a data structure for NN-search, the off-line processing is very fast. Therefore, the resulting image enhancement system is very flexible as the it can be easily adapted to the distribution of a specific (non-generic) class of images. This is important especially when *a priori* knowledge of the problem is available in terms of a class-specific set of example images. For instance, if it is known that the image of interest to be processed is representing documents (whose statistical properties might be distinct from those of general images), one could quickly generate examples from this specific class of images on which the system is trained. While this leads to much better results (see the last low of Fig. 1), it is infeasible in conventional sparse GPs due to their high complexity in training (which includes the identification of inducing inputs).

We demonstrate the utility of our algorithm in two example image enhancement applications that can benefit from the high efficiency of our approximation (both in training and in testing): suppression of compression artifacts in JPEG images and single-image super-resolution (Fig. 1).

[1] P. V. Gehler and M. Welling. Product of "edge-perts". In *NIPS*, Cambridge, MA, 2005. MIT Press.

[2] J. Quiñonero-Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *JMLR*, 2005.

[3] P. Sollich and C. K. I. Williams. Using the equivalent kernel to understand Gaussian process regression. In *NIPS*, 2005.