BAYESIAN IMAGE RECONSTRUCTION WITH LOCAL LINEAR REGRESSORS

Colas Schretter

Vrije Universiteit Brussel, Belgium

ABSTRACT

I revisit the classical inverse problem of reconstructing a continuous restored image from a small set of blurred and noisy measurements. A Bayesian estimation framework using mixtures of local linear regressors addresses the strong numerical ill-conditioning that arises when estimating many collaborative local models from few samples. The method models the classical blur-and-sampling data formation and performs maximum a posteriori (MAP) imputation to regularize estimation under missing data. I also propose a progressive model selection strategy, where components are adaptively split during an expectation–maximization procedure so that the mixture complexity follows the local spatial structure of the latent image. This yields a compact, continuous, and in-core representation of the intensity field from which raster images of arbitrary definition can be generated on demand.

Index Terms— Image reconstruction, Bayesian statistics, Mixture models, Linear regression, Regularization

1. INTRODUCTION

The digital image restoration problem aims at approximating an unobserved continuous image from a few blurred, noisy, measured data samples. The data formation process involves a cascade of degradations of various severities. First, incident photons are refracted and scattered by an imperfect photographic lens stack, then the average of photoelectric stimulations over the pixel surface of the detector is converted to a weak electrical signal. This signal is perturbed by thermal noise and dark currents, then clipped and quantized.

Many authors have explored alternatives to linear transforms for image representation and coding. Kaup and Aach [1] proposed shape-adaptive basis functions. Takeda [2] used non-parametric local kernel regression for image reconstruction and denoising [3], deblurring [4] and super-resolution [5]. Thereafter, Seo [6] introduced locally adaptive regression kernel (LARK) descriptors for image analyses and object detection in computer vision. Weickert [7] opened the field of modeling anisotropic image diffusion in images with partial differential equations and applied this to image compression. Demaret [8] explored triangulations and adaptive splines. A subsequent work by Marwood [9] applied the idea to extreme image compression with a budget below 200 bytes per image.

In 2023, Gaussian splatting [10–12] revived Gaussian mixture density representations [13] where emissive image regions are represented by local kernels. In another field, Spatial Mixture-of-Experts (SMoE) [14] demonstrated that adaptive gated linear regression can be incorporated into advanced deep learning methods for localized prediction.

I adopt a fully statistical estimation of local linear models [15–17] that are compressed. Compact adaptive mixture models have proven universal approximation property and I can rely on a solid Bayesian framework for data-driven parametric estimation with missing data [18]. The parametric estimation only depends on the input data realization and a data formation model. Expanding further a prior work [19], I address the joint deblurring, denoising, upscaling, compression problem for image restoration with a sparse representation.

My contribution starts in Section 2 by using adaptive mixtures of local linear regressions as continuous latent image models. In Section 3, I describe a Bayesian image reconstruction method accounting jointly for data denoising, image deblurring and upscaling. The reconstruction problem is regularized by compressing the semi-parametric image representation. Results which are discussed in Section 4 outperform the DRCT-L [20] model for image upscaling and the KBNet [21] model for denoising. We conclude in Section 5.

2. IMAGE MODEL

Natural image features are characterized by smoothly varying intensity gradients separated by sharp edges between connected regions. In my adaptive image model, the uniform marginal sampling domain is covered by a small set of marginal receptor (gating) fields that is represented by a mixture density of K bivariate Gaussian kernels parameterized by a mean position $\boldsymbol{\mu}$ and a variance-covariance matrix $\boldsymbol{\Sigma}$:

$$d(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \quad \text{with} \quad \sum_{k=1}^{K} \pi_k = 1\,, \quad (1)$$

where in the 2-dimensional case, I set $\pi_k \propto 2\pi\sqrt{\det(\boldsymbol{\Sigma}_k)}$.

These prior weights are proportional to the area of the first standard deviation ellipse that is implicitly defined by

This work received funding from the Research Foundation - Flanders (FWO) project UNIMODE (G0B3521N).

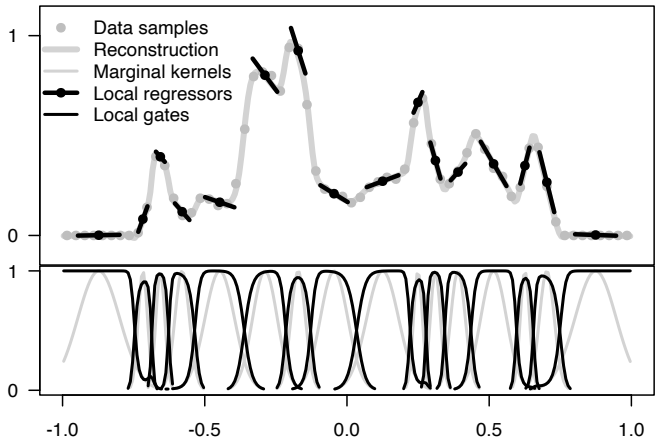


Fig. 1. Continuous reconstruction of a PET line profile with 16 local linear regressors estimated from 64 data samples.

the variance-covariance matrix $\boldsymbol{\Sigma}_k$. These priors can be interpreted as a peak-normalization of all kernels such that

$$d(\boldsymbol{x}) \propto \sum_{k=1}^{K} \exp\left(-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2_{\boldsymbol{\Sigma}_k}\right), \tag{2}$$

with the squared Mahalanobis distance to the kernel's position

$$\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2_{\boldsymbol{\Sigma}_k} = (\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\,. \tag{3}$$

The marginal partition of unity defining the relative importance (gate) for each of the $k \in [1,\ldots,K]$ component is

$$p(\boldsymbol{x}\,|\,k) = \frac{1}{d(\boldsymbol{x})}\exp\left(-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2_{\boldsymbol{\Sigma}_k}\right) \tag{4}$$

$$\text{with} \quad \sum_{k=1}^{K} p(\boldsymbol{x}\,|\,k) = 1\,. \tag{5}$$

I associate to each marginal kernel a linear regressor, parameterized by a mean intensity λ and a vector of spatial gradients $\boldsymbol{\delta}$, for capturing locally the smooth intensity variations. Linear regression planes can be evaluated with

$$\lambda(\boldsymbol{x}\,|\,k) = \lambda_k + \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu}_k)\,, \tag{6}$$

such that mixing regressors according to their relative spatial importance yields the average consensus from all regressors:

$$\lambda(\boldsymbol{x}) = \sum_{k=1}^{K} p(\boldsymbol{x}\,|\,k)\,\lambda(\boldsymbol{x}\,|\,k)\,. \tag{7}$$

Moreover, an estimate of the regression error may be computed from the discrepancy among regressor's responses at any query location $\boldsymbol{x}$. The explained variance from the model is the expected conditional squared error from the mean:

$$\sigma^2(\boldsymbol{x}) = \sum_{k=1}^{K} p(\boldsymbol{x}\,|\,k)\,(\lambda(\boldsymbol{x}\,|\,k)-\lambda(\boldsymbol{x}))^2\,, \tag{8}$$

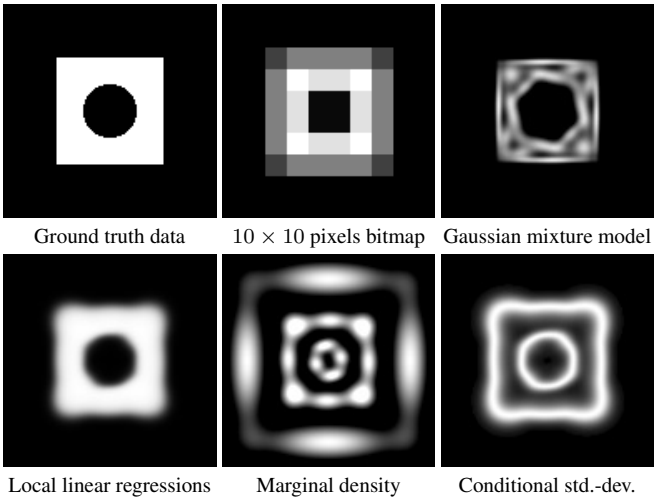


Fig. 2. Bitmap, Gaussian mixture model and mixture of local constant regression for representing a binary indicator image.

and this quantity will be useful for adaptive data-driven model selection in a Bayesian estimation by weighting the importance of regressors according to their radiometric proximity.

2.1. Examples

Fig. 1 shows an example of continuous signal reconstruction from a sparse mixture of local linear regressions. The marginal receptor field from 16 kernels defines the spatial extent (gate) of each regressor. My regularized parametric estimation converges to a solution such that convolving the mixture's marginal approximates the 64 blurred data samples, yielding a deblurred denoised continuous signal model.

For illustrating the rationale behind my local linear regression mixture model, I compare raster, density and regression representations in Fig. 2 for approximating a simple indicator function partitioning the image locations into background/empty (black) and foreground/solid (white). The ground truth data is a binary bitmap of 128×128 samples.

An alternative Gaussian mixture model (GMM) with 16 components is fitted to the data with a maximum-likelihood expectation-maximization (EM) method, such that 6 parameters are estimated per marginal kernel: the prior weight, two position coordinates, the vertical and horizontal variance for capturing the spread and a last covariance parameter describing the orientation of the kernel. The border of the square is better defined, in comparison to the grid representation.

Thinner kernels are fitting the transition between empty and solid regions. The central hole is approximated by six kernels defining an hexagonal shape while the center appears square in the bitmap. However, unwanted variations arise in the solid region. Furthermore, such density estimation models are not symmetric because an another solution will be found if I wanted to model the empty (black) regions instead.

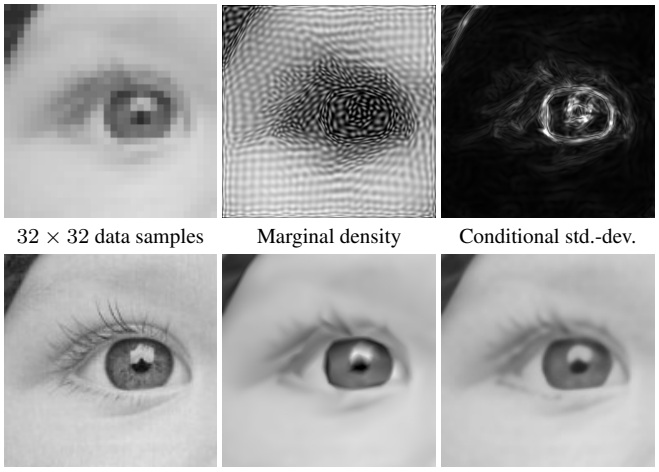


Fig. 3. Comparison between mixture of local linear regressions and a supervised learning model for image upscaling.

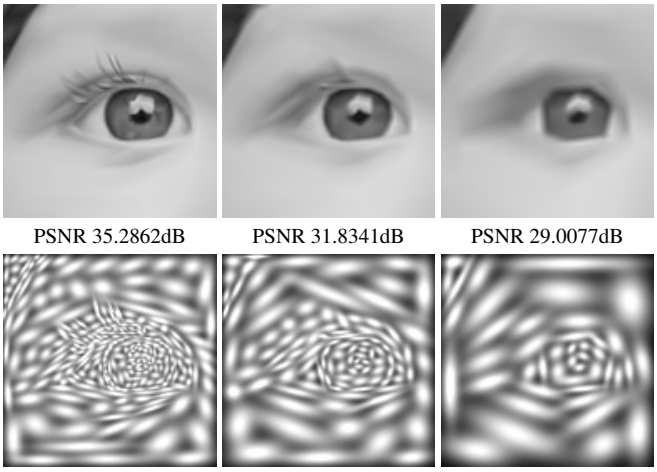


Fig. 4. Progressive images approximations with decreasing model sizes in function of the splitting threshold parameter T.

In the collaborative local regression model, the prior is not estimated since marginal kernels are peak-normalized, but a constant regressor is computed for capturing the local mean intensity. The number of parameters remains $16\times6 = 96$, that is exactly the same amount than in the GMM case. The regression function from mixing local linear regressions reconstructs a smooth approximation of the shape, without "holes".

3. IMAGE RECONSTRUCTION

Given a limited set of blurry and noisy input data samples, I should estimate K kernel's positions $\boldsymbol{\mu}_k$ and shape $\boldsymbol{\Sigma}_k$, along with the mean intensity λ_k and vertical and horizontal gradients $\boldsymbol{\delta}_k$ of local linear regressors. Such an inverse problem is ill-posed and difficult to solve with numerical (gradient descent) optimization methods, as many different latent images may explain the limited input data realization at hand.

I propose here to leverage knowledge from the data formation model for regularizing the solution. Moreover, I stabilize numerically an expectation-maximization algorithm by using an explicit numerical data imputation strategy. Finally, I introduce a progressive model selection strategy for setting the number K of model components and initializing marginal kernels at suitable locations on the image domain.

3.1. Data Formation

Let's assume I measure a grid D of $N_i \times N_j$ data samples

$$\tilde{d}_{i,j} = \left\lfloor \left(Y * \mathcal{N}\left(0, \boldsymbol{I}\,\sigma_b^2\right)\right)(\boldsymbol{x}_{i,j}) + \mathcal{N}\left(0, \sigma_n^2\right) \right\rceil_0^{255} \quad (9)$$

where the sample positions $\boldsymbol{x}_{i,j}$ are exact locations on the image space, aligned on a Cartesian grid.

The continuous field of image intensities Y is blurred by a sensor's point spread function, that is modeled by an isotropic Gaussian kernel, parameterized by the diagonal variance matrix $\boldsymbol{I}\,\sigma_b^2$, then the function is sampled at positions $\boldsymbol{x}_{i,j}$ and additive Gaussian white noise is added with variance σ_n^2. Finally the signal is quantized by a rounding operator and clipped to the classical 8-bit range. Heteroskedastic noise estimates can also be specified by a scalar variance σ_n^2 for each individual measurement, but I conform to uniform Additive White Gaussian Noise (AWGN) for the sake of clarity.

3.2. Data Imputation

Let $Y(\boldsymbol{x}) : [0, N_i] \times [0, N_j] \to \mathbb{R}$ be the pristine continuous deblurred noise-free intensity image that I want to reconstruct from the data sample grid D. I want to estimate a set of parameters for the mixture of regressions model:

$$\Theta = \{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \lambda_1, \boldsymbol{\delta}_1), \ldots, (\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K, \lambda_K, \boldsymbol{\delta}_K)\} \quad (10)$$

such that the log-likelihood of observing image λ is maximized, given the model's parameters.

This problem is different than estimating parameters for maximizing the log-likelihood of observing input data D:

$$L(\Theta \,|\, D) = \frac{1}{N_i \times N_j} \sum_{i=1}^{N_i} \sum_{i=j}^{N_j} \log \left(\sum_{k=1}^{K} p(\boldsymbol{x}_{i,j},\, \tilde{d}_{i,j} \,|\, k) \right) \quad (11)$$

while the log-likelihood from the continuous latent image is

$$L(\Theta \,|\, Y) = \int_0^{N_i} \int_0^{N_j} \log \left(\sum_{k=1}^{K} p([x\ y]^\top,\, Y([x\ y]^\top) \,|\, k) \right) \mathrm{d}x\, \mathrm{d}y \quad (12)$$

In the above formulations, $L(\Theta \,|\, D) \approx L(\Theta \,|\, Y)$ only if a very large number of samples is captured, such that the total amount of parameters $K \times 8$ is smaller than $N_i \times N_j$.

It is apparent from the full likelihood equation that some sort of numerical integration will be required for approximating the double integral over the spatial domain. I oversample the data grid by a factor 16, although a factor 8 provided only very slightly more blurred results. I perform mean imputation by computing the current upscaled regression at the subpixel locations and add the adjoint backprojected blurred residual from low-definition data residuals.

3.3. Expectation-Maximization

E-step For every imputed data point at position $\boldsymbol{x}_{i,j}$ and intensity $\tilde{d}_{i,j}$, I evaluate the joint spatio-radiometric likelihood of observing the intensity at that position. According to the Bayes theorem, the probability of each component is

$$p(k \,|\, i,j) = \frac{\exp\left(-\frac{1}{2}\left(\|\boldsymbol{x}_{i,j} - \boldsymbol{\mu}_k\|^2_{\boldsymbol{\Sigma}_k} + \|\tilde{d}_{i,j} - \lambda_k)\|^2\right)\right)}{\sum_{z=1}^{K} \exp\left(-\frac{1}{2}\left(\|\boldsymbol{x}_{i,j} - \boldsymbol{\mu}_z\|^2_{\boldsymbol{\Sigma}_k} + \|\tilde{d}_{i,j} - \lambda_z\|^2\right)\right)} \tag{13}$$

with the squared radiometric distance to the linear regressor

$$\|\tilde{d}_{i,j} - \lambda_k)\|^2 = \left(\tilde{d}_{i,j} - \lambda(\boldsymbol{x}_{i,j} \,|\, k)\right)^2 \Big/ \left(\bar{\sigma}^2(\boldsymbol{x}_{i,j}) + \sigma_n^2\right), \tag{14}$$

where the expected variance (squared conditional error) sums the expected variance from the model and the squared bias:

$$\bar{\sigma}^2(\boldsymbol{x}_{i,j}) = \sigma^2(\boldsymbol{x}_{i,j}) + \left(\tilde{d}_{i,j} - \lambda(\boldsymbol{x}_{i,j})\right)^2. \tag{15}$$

M-step For every mixture component $k \in [1 \dots K]$, I update the spatial means $\boldsymbol{\mu}_k$ and variances-covariances $\boldsymbol{\Sigma}_k$ representing the support of local regressors, as well as the mean intensities λ_k and the two spatial gradients $\boldsymbol{\delta}_k$ with

$$\boldsymbol{\mu}_k = \sum_{i=1}^{N_i}\sum_{j=1}^{N_j} p(k \,|\, i,j)\, \boldsymbol{x}_{i,j}\,, \tag{16}$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^{N_i}\sum_{j=1}^{N_j} p(k \,|\, i,j)\, \left[(\boldsymbol{x}_{i,j} - \boldsymbol{\mu}_k)\,(\boldsymbol{x}_{i,j} - \boldsymbol{\mu}_k)^\top\right], \tag{17}$$

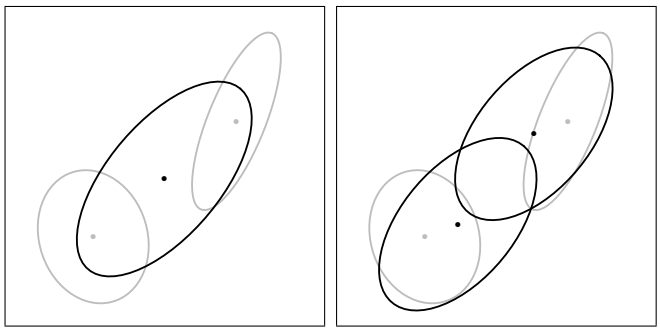
$$\lambda_k = \sum_{i=1}^{N_i}\sum_{j=1}^{N_j} p(k \,|\, i,j)\, \hat{d}_{i,j\,|\,k}\,, \tag{18}$$

$$\boldsymbol{\delta}_k = \sum_{i=1}^{N_i}\sum_{j=1}^{N_j} p(k \,|\, i,j)\, \left[(\boldsymbol{x}_{i,j} - \boldsymbol{\mu}_k)\,(\hat{d}_{i,j\,|\,k} - \lambda_k]\right), \tag{19}$$

then, "inflating" kernels with an isotropic "balloon" variance:

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k + \boldsymbol{I}\sigma_s^2 \quad \text{with} \quad \sigma_s^2 = \frac{1}{12} = \int_0^1 \left(x - \frac{1}{2}\right)^2 \mathrm{d}x\,. \tag{20}$$

This regularization is essential for modeling the gap between (imputed) samples, and therefore the amount of smoothing vanishes with the oversampling factor for data imputation.



Maximum-likelihood merge Binary splitting predictions

Fig. 5. Maximum-likelihood merge (black) of two hidden children components (gray) and a symmetric binary splitting.

For correcting noise, I computed the maximum a posteriori (MAP) estimate of the denoised intensity measurement given each local regressor. The MAP estimates have a closed form with Gaussian noise prior of standard deviation σ_n:

$$\hat{d}_{i,j\,|\,k} = \lambda(\boldsymbol{x}_{i,j} \,|\, k) + \frac{\bar{\sigma}^2\,(\boldsymbol{x}_{i,j})}{\bar{\sigma}^2\,(\boldsymbol{x}_{i,j}) + \sigma_n^2}\left(\tilde{d}_{i,j} - \lambda(\boldsymbol{x}_{i,j} \,|\, k)\right). \tag{21}$$

Then, I substitute actual data samples $\tilde{d}_{i,j}$ with their denoised estimates. Note that $\hat{d}_{i,j\,|\,k} = \tilde{d}_{i,j}$ when $\sigma_n^2 = 0$.

3.4. Model Selection

For adaptive model selection, I initialize the solution with a single kernel whose parameters are estimated by an ordinary least square regression. Then, mixture components are split, as illustrated in Fig. 5, by squashing their shape and displacing their means along the dominant eigenvector. The determinant of the generated kernels is halved by squashing their shape and their positions are shifted along the dominant eigenvector such that a maximum-likelihood merge recovers exactly the original variance-covariance matrix of the kernel.

During early iterations, components are split and positioned adaptively over the image plane, forming an initialization phase. At late convergence, the iterative parameter estimation always improves data likelihood and therefore diminishes the mean squared error. I split if the integrated data residual exceeds a set threshold that represents the expected noise level. Therefore, the number of mixture components is driven by the expected approximation accuracy.

The progressive model selection ensures that the space-variant image resolution is adaptively defined according to image details. I avoid that way the problem of model initialization for which only a data-driven criterion makes sense. In example, if the latent image is a simple ramp gradient, then the initial single component model will optimally approximate the solution and no splitting will occur if I set the threshold to the actual level of noise.

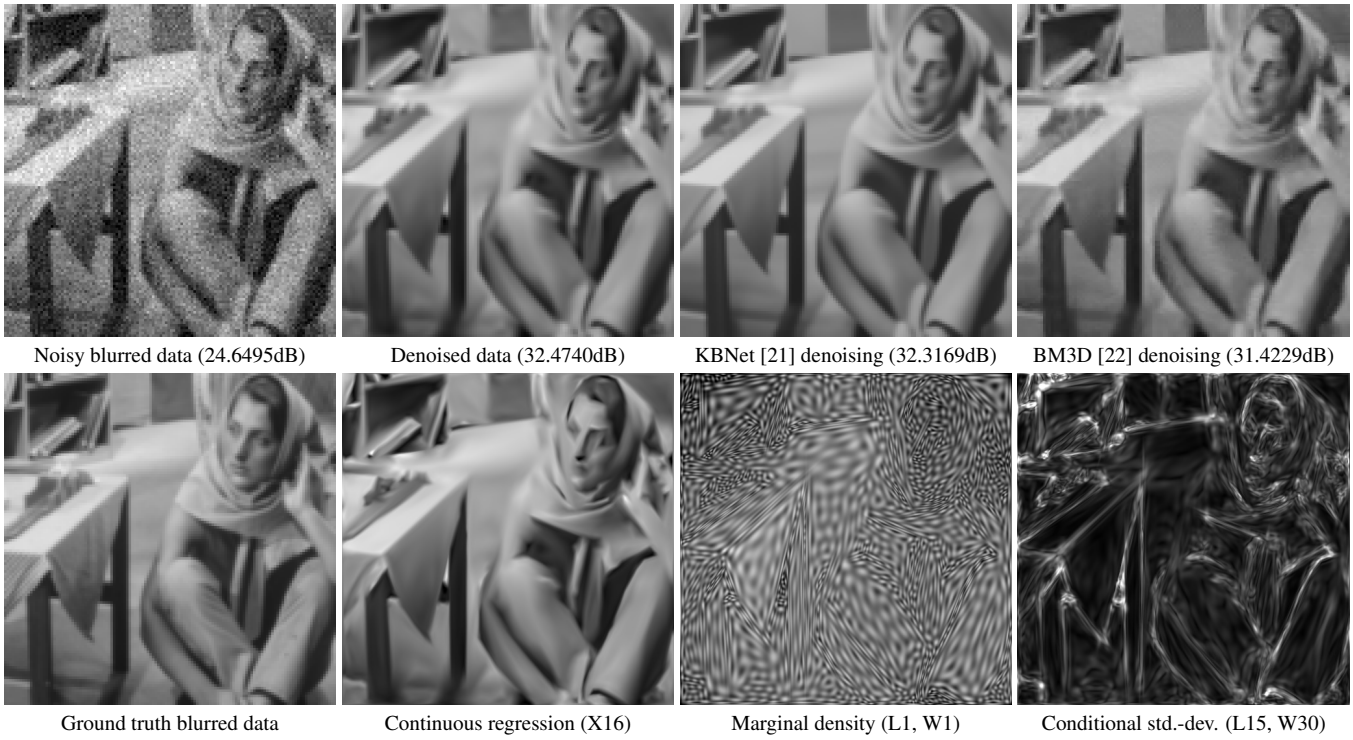


Fig. 6. Restoration with 1619 local regressors, estimated from 128×128 blurred $\left(\sigma_b^2 = 4/12\right)$ and noisy $\left(\sigma_n^2 = 15^2\right)$ samples.

4. RESULTS

Fig. 3 shows an example of image reconstruction in the noise-free case. Components have been initialized on a grid and a numerical missing data imputation of a factor 16 allows to estimate robustly model's parameters. A coarser spatial integration may be used, at the cost of blurring the solution. The ring around the iris is more contrasted and the shape of the reflection is sharper than with the larger DRCT-L model.

Fig. 4 shows several image approximations with decreasing model sizes that are estimated from the original 128×128 input samples. A fair quality of 35.29dB is reached from only $257 \times 8 = 2056$ parameters, for an effective compression factor of 8 in computer memory. These adaptive resolution images set fewer bigger model components in regions of low details. One could even edit the image by inserting, deleting, moving and setting the shape of components.

Fig. 6 shows the denoising effect while a sharp latent image is represented with a modest amount of components. Since data is corrupted, it makes sense to compress the solution by enforcing sparsity. Starting from a single component, a sequence of binary splittings generated 1619 components, so that the model adapts to the local visual complexity.

I implemented the method on GPU and it took 1.82s for computing 1000 iterations with a X4 upscaling factor, on a laptop Nvidia RTX 4070. Gaussian kernels are clipped after four standard deviations, for a constant time complexity per data sample, regardless of the number K of mixture elements.

Inference with the DRCT-L model required 1.70s for 4X upscaling. Even if the quantitative improvement between upsampling factors 16 and 4 was faint, this is, to my knowledge, the first 16X image upscaling method where the latent model size does not depend on the set upscaling factor.

5. CONCLUSION

This paper introduces early results with a Bayesian method for estimating parameters of mixture of local linear regression models. From blurry noisy and quantized data samples, a continuous image representation is estimated by accounting for deblurring in the system model and compensating for missing data with a statistical data imputation technique. This semi-parametric statistical model suits well for representing natural images with even fewer parameters than the initial data size.

Sparse models may be stored as an in-core compressed representation in computer memory and final images are rasterized on demand, depending on the display size. Such a decorrelation between encoding the source visual content and computing the regression function during rasterization is beneficial in today's world of heterogeneous display sizes, ranging from stamp-sized smartwatch displays to wide television screens, and even, ultra-wide coverage of the entire human's field of view on head-up displays for virtual reality.

Future works will aim at compressing model's parameters on file for streaming progressively very large image models.

6. REFERENCES

[1] Andre Kaup and Til Aach, "Coding of segmented images using shape-independent basis functions," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 937–947, 1998.

[2] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar, "Robust kernel regression for restoration and reconstruction of images from sparse noisy data," in *Proc. of the IEEE International Conference on Image Processing*, 2006, pp. 1257–1260.

[3] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[4] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar, "Deblurring using regularized locally adaptive kernel regression," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 550–563, 2008.

[5] Hiroyuki Takeda, Peyman Milanfar, Matan Protter, and Michael Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1958–1975, 2009.

[6] Hae Jong Seo and Peyman Milanfar, "Training-free, generic object detection using locally adaptive regression kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1688–1704, 2010.

[7] Irena Galić, Joachim Weickert, Martin Welk, Andrés Bruhn, Alexander Belyaev, and Hans-Peter Seidel, "Towards pde-based image compression," in *Variational, Geometric, and Level Set Methods in Computer Vision*, Nikos Paragios, Olivier Faugeras, Tony Chan, and Christoph Schnörr, Eds., Berlin, Heidelberg, 2005, pp. 37–48, Springer Berlin Heidelberg.

[8] Laurent Demaret, Nira Dyn, and Armin Iske, "Image compression by linear splines over adaptive triangulations," *Signal Processing*, vol. 86, no. 7, pp. 1604–1616, 2006.

[9] David Marwood, Pascal Massimino, Michele Covell, and Shumeet Baluja, "Representing images in 200 bytes: Compression via triangulation," in *IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 405–409.

[10] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, "3D Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1–14, 2023.

[11] Adam Celarek, Georgios Kopanas, George Drettakis, Michael Wimmer, and Bernhard Kerbl, "Does 3D Gaussian splatting need accurate volumetric rendering?," *Computer Graphics Forum*, vol. 44, no. 2, 2025.

[12] Yunxiang Zhang, Bingxuan Li, Alexandr Kuznetsov, Akshay Jindal, Stavros Diolatzis, Kenneth Chen, Anton Sochenov, Anton Kaplanyan, and Qi Sun, "Image-GS: Content-adaptive image representation via 2D Gaussians," *ACM Transactions on Graphics (TOG)*, vol. 44, no. 4, pp. 1–11, 2025.

[13] Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross, "EWA volume splatting," in *Proceedings of the Conference on Visualization '01*, USA, 2001, VIS '01, pp. 29–36, IEEE Computer Society.

[14] Nikoli Dryden and Torsten Hoefler, "Spatial mixture-of-experts," *Advances in Neural Information Processing Systems*, vol. 35, pp. 697–713, 2022.

[15] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[16] Trevor Hastie and Clive Loader, "Local regression: Automatic kernel carpentry," *Statistical Science*, vol. 8, no. 2, pp. 120–129, 1993.

[17] Guillaume Bouchard, "Localised mixtures of experts for mixture of regressions," in *Between Data Science and Applied Data Analysis*, pp. 155–164. Springer, 2003.

[18] Matthew Blackwell, James Honaker, and Gary King, "A unified approach to measurement error and missing data: Overview and applications," *Sociological Methods & Research*, vol. 46, no. 3, pp. 303–341, 2017.

[19] Colas Schretter, Jianyong Sun, and Peter Schelkens, "Image reconstruction with smoothed mixtures of regressions," in *International Conference on Image Processing (ICIP)*, 2018, pp. 400–404.

[20] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou, "DRCT: Saving image super-resolution away from information bottleneck," in *IEEE/CVF Computer Vision and Pattern Recognition*, 2024, pp. 6133–6142.

[21] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Honwei Qin, and Hongsheng Li, "KBNet: Kernel basis network for image restoration," *arXiv:2303.02881*, 2023.

[22] Ymir Mäkinen, Lucio Azzari, and Alessandro Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Transactions on Image Processing*, vol. 29, pp. 8339–8354, 2020.