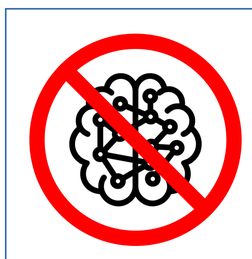
BAYESIAN IMAGE RECONSTRUCTION WITH LOCAL LINEAR REGRESSORS

Colas Schretter colas.schretter@{vub,ulb,imec}.be

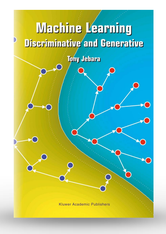
ICIP 2026 TAMPERE

HYBRID GENERATIVE / DISCRIMINATIVE EXPECTATION-MAXIMIZATION

- Sparse mixtures of local linear regressors are used for continuous image representation.
- Model parameters are estimated from limited data with Bayesian maximum *a posteriori* (MAP) imputation of denoised missing samples using the expected conditional variance from the model.
- progressive binary splitting on data likelihood for adaptive model selection.
- Approximate the pristine denoised and upscaled latent image, not overfitting measured data.
- Hybrid expectation-maximization (EM) combines generative (competitive) estimation for marginal Gaussian kernels and discriminative (collaborative) minimum MSE fitting for attached regressors.
- The positions and shapes of marginal kernels act as adaptive-resolution smooth pixels that are not constrained on a grid and may overlap. The model can be rasterized at any required image definition.



NO DEEP NETWORK
NO BLACK BOXING
NO TRAINING DATA
NO INVERSE CRIMES

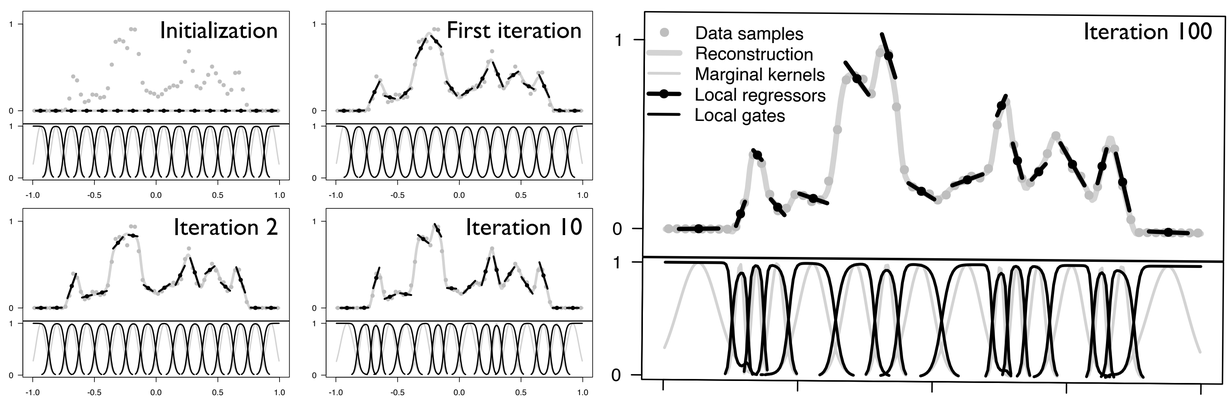


GENERATIVE
Competitive — P(X,Y)

DISCRIMINATIVE
Collaborative — P(Y|X)

EXAMPLE ON A ONE-DIMENSIONAL SIGNAL

Continuous reconstruction with 16 local linear regressors estimated from 64 noise-free data samples



MARGINAL KERNELS, LOCAL LINEAR REGRESSORS, MISSING DATA IMPUTATION

The structured image representation model allows for closed-form parametric updates

Marginal support kernels

The uniform sampling domain is covered by a mixture of marginal receptors (gating) that is represented by a mixture density of K bivariate Gaussian kernels parameterized by a mean position $\boldsymbol{\mu}$ and a symmetric covariance matrix $\boldsymbol{\Sigma}$:

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

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

These priors are proportional to the area of the first standard deviation ellipse such that kernels are peak-normalized:

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

with the squared Mahalanobis distance to the kernel's position

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

The marginal partition of unity defines the relative spatial responsibility for each mixture component $k \in \{1,\ldots,K\}$:

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

Local linear regressors

A local regressor is associated to each marginal kernel. Linear regressors are parameterized by a mean intensity λ and a vector of spatial gradients $\boldsymbol{\delta}$, for capturing locally the smooth intensity variations. Regression planes can be evaluated with

$$\lambda(\boldsymbol{x} \,|\, k) = \lambda_k + \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_k)\,, \quad (5)$$

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

$$\lambda(\boldsymbol{x}) = \sum_{k=1}^{K} p(k \,|\, \boldsymbol{x})\, \lambda(\boldsymbol{x} \,|\, k)\,. \quad (6)$$

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:

$$\sigma^2(\boldsymbol{x}) = \sum_{k=1}^{K} p(k \,|\, \boldsymbol{x})\, (\lambda(\boldsymbol{x} \,|\, k) - \lambda(\boldsymbol{x}))^2\,, \quad (7)$$

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.

Missing data imputation

Let $Y(\boldsymbol{x}) : [0, N_i) \times [0, N_j) \to \mathbb{R}$ be a pristine continuous intensity image to reconstruct from a grid D of $N_i \times N_j$ blurred noisy data samples $\tilde{d}_{i,j}$. One must estimate $K \times 8$ parameters:

$$\Theta = \{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \lambda_1, \boldsymbol{\delta}_1), \ldots, (\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K, \lambda_K, \boldsymbol{\delta}_K)\} \quad (8)$$

such that the log-likelihood of observing the ground truth latent image Y is maximized, given the model's parameters.

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

$$L(\Theta \,|\, D) = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \log\left(\sum_{k=1}^{K} p(\boldsymbol{x}_{i,j},\, \tilde{d}_{i,j} \,|\, k)\right) \quad (9)$$

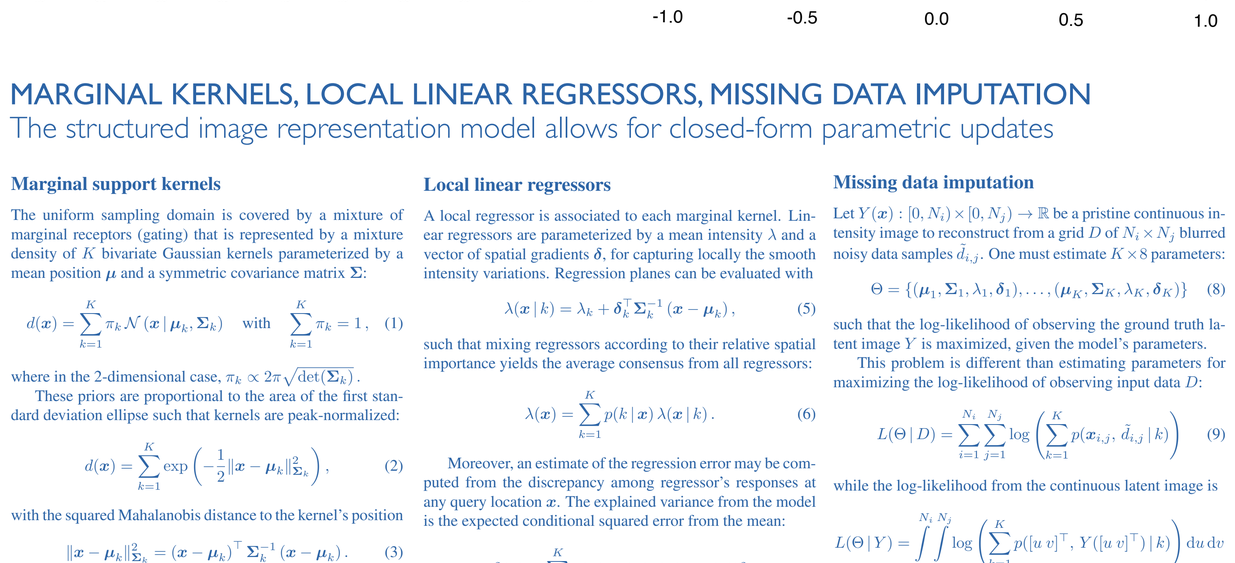
while the log-likelihood from the continuous latent image is

$$L(\Theta \,|\, Y) = \int_0^{N_i}\int_0^{N_j} \log\left(\sum_{k=1}^{K} p([u\ v]^\top,\, Y([u\ v]^\top) \,|\, k)\right) \mathrm{d}u\, \mathrm{d}v \quad (10)$$

In the above formulations, $L(\Theta \,|\, D) \approx L(\Theta \,|\, Y)$ only if a very large number of noise-free samples is captured. Imputation is achieved by computing the current upscaled regression at the subpixel locations and add the blurred data residual.

STALLED IMPROVEMENT IN NEURAL NETWORKS FOR IMAGE RESTORATION

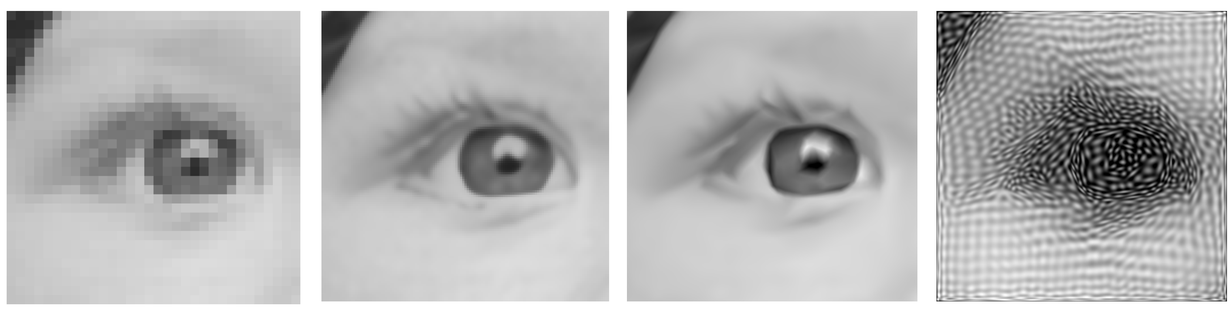
SGD training with ADAM overfits training sets, hence models are very large and not robust



State of the Art from www.paperswithcode.com (Acquired by Meta in 2019, shutdown in July 2025)

NAIVE IMAGE UPSCALING WITH DENSE MODELS — ONE KERNEL PER DATA SAMPLE

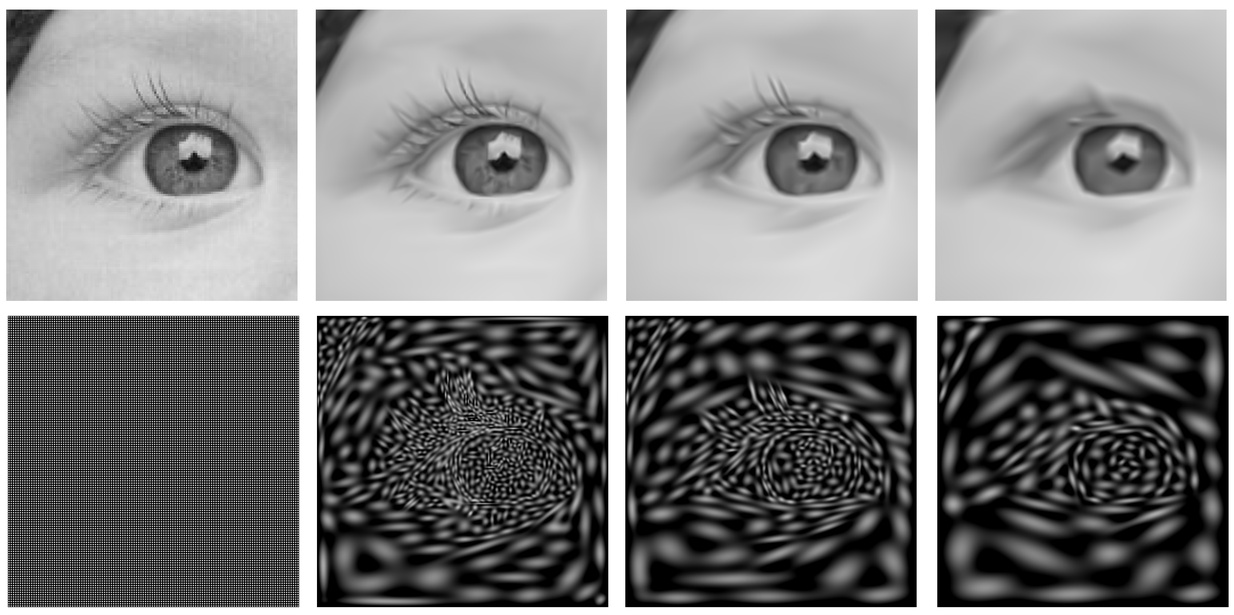
Comparison between local linear regressors and a supervised learning model for image upscaling



32 × 32 data samples | DRCT-L upscaling 4X, 31.39 dB | Local linear regression 16X, 31.43 dB | 1024 marginal kernels

PROGRESSIVE MODEL SELECTION FOR IN-CORE IMAGE COMPRESSION

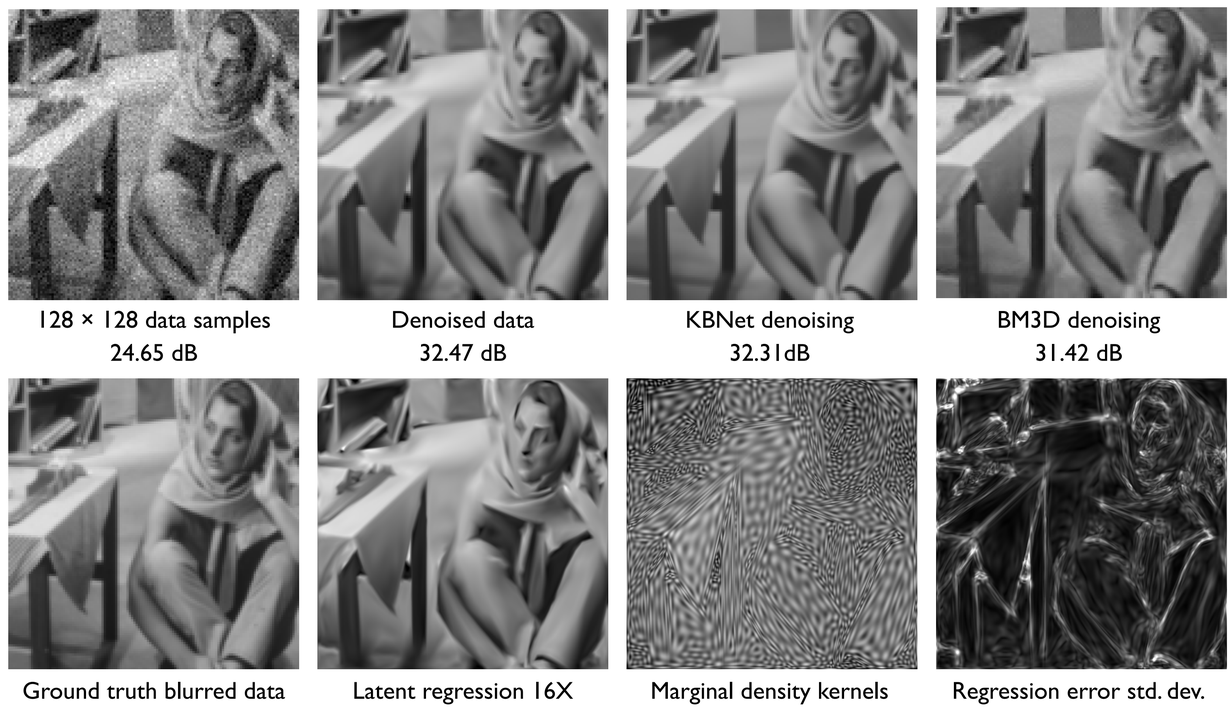
Approximations with decreasing model sizes in function of the splitting threshold parameter T



T = 0, 16384 components, ∞ dB, 256 kB | T = 5, 623 components, 38.87 dB, 9968 bytes | T = 10, 257 components, 35.29 dB, 4112 bytes | T = 20, 108 components, 31.83 dB, 1728 bytes

JOINT DATA DENOISING, IMAGE UPSCALING AND SPARSE MODEL SELECTION

Restoration with 1619 local regressors, estimated from 16384 blurred and noisy data samples

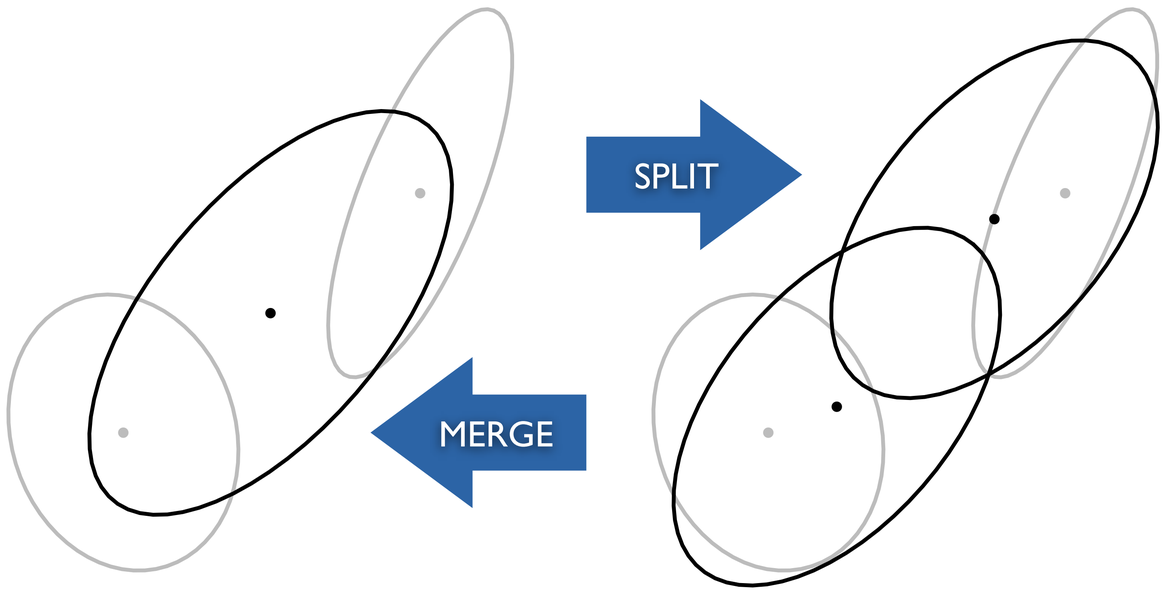


128 × 128 data samples, 24.65 dB | Denoised data, 32.47 dB | KBNet denoising, 32.31dB | BM3D denoising, 31.42 dB

Ground truth blurred data | Latent regression 16X | Marginal density kernels | Regression error std. dev.

ADAPTIVE MODELS WITH BINARY SPLITTING BASED ON LOCAL LIKELIHOOD

Components are split in two where local linear regression poorly explains the imputed data



A (parent) component is too large to explain the underlying density

Maximum-likelihood parametric merge of the two children recovers the parent

fwo This work received funding from the Research Foundation - Flanders (FWO) Project UNIMODE (G0B3521N)

Scan this QR code to read the paper

imec VUB ETRO ELECTRONICS & INFORMATICS