

Learning Adaptive Parameter Tuning for Image Processing

Jingming Dong^a, Iuri Frosio^b, and Jan Kautz^b

^aUCLA Vision Lab, University of California, Los Angeles, CA, USA

^bNVIDIA, Santa Clara, CA, USA

Abstract

The non-stationary nature of image characteristics calls for adaptive processing, based on the local image content. We propose a simple and flexible method to learn local tuning of parameters in adaptive image processing: we extract simple local features from an image and learn the relation between these features and the optimal filtering parameters. Learning is performed by optimizing a user defined cost function (any image quality metric) on a training set. We apply our method to three classical problems (denoising, demosaicing and deblurring) and we show the effectiveness of the learned parameter modulation strategies. We also show that these strategies are consistent with theoretical results from the literature.

Introduction

The effectiveness of most image processing¹ algorithms depends on a careful parameter choice. For instance, denoising methods commonly require a *denoising strength* or a *patch size* to be set. These parameters can be adjusted per image, but neglecting the local image characteristics leads to sub-optimal results. Setting the filtering parameters adaptively has obvious benefits; e.g., the denoising strength can be higher in smooth areas where the risk of blurring out details is low, and in turn it can be lower in highly textured areas where noise is less visible. Adaptiveness can also be easily achieved by mixing the output of different algorithms, each operating at best in a different part of an image.

Adaptiveness, however, does not come without a cost. It requires to establish a rule to process the pixels, based on the local image properties. In most cases, this means establishing a mapping from a collection of features, describing the image at a local scale, to a set of parameters that determines the behaviour of the processing algorithm. The rule to modulate the parameters across an image can be derived heuristically, or based on a statistical model, but better results are obtained by a learning approach. The learning procedure can in fact be designed to optimize a specific cost function, related to the problem in hand. Nonetheless, previous learning methods have proposed adaptive filtering only for specific image processing problems [2, 3, 4].

In contrast, we propose a general approach to learn parameter tuning for adaptive image processing. We learn the mapping from local features to parameters on a training set, which then generalizes to unseen images. We demonstrate the generality and effectiveness of our approach on three classical image processing problems: denoising, demosaicing and deblurring. We show in the case of denoising that the proposed method is capable of

learning a parameter modulation strategy, consistent with previously derived theoretical results [5]. For demosaicing, we blend the output of three demosaicing methods with adaptive mixing parameters; this does not only produce a better demosaiced image, but also highlights untold strengths and weaknesses in the demosaicing algorithms. In the case of deblurring, we learn how to adaptively set the regularization hyper-parameter without resorting to any global, discrepancy principle [6]. Finally, we show how the objective function affects the parameter modulation strategy, coherently with [7].

Related work

We mainly identify three approaches² for the development of adaptive algorithms: 1) heuristic-driven, 2) based on local statistics analysis, and 3) learning-based.

Heuristic-driven · The rules to modulate the parameters across an image are derived from empirical observation, experience or intuition. This is the case of the adaptive unsharp masking technique in [8]. The filter was designed to enhance high detail areas, while leaving unaltered the smooth regions. The enhancement strength is guided by the classification of the pixels into low, medium and high dynamic classes. In the denoising context, Chen *et al.* [9] proposed an adaptive Total Variation (TV) regularization method to avoid the staircase artifact introduced by TV. The idea is to identify edges and ramps in an image and apply ℓ_1 or ℓ_2 regularization respectively. The blending rule for the two regularization terms is determined heuristically. Although the heuristic approach generally leads to an improvement of the filter performances, it is not suitable to reveal complex or counter-intuitive interactions between the local image characteristics and the filtering parameters [2].

Analysis of local statistics · Thaipanich *et al.* [10] used SVD and K-Means to group similar patches in an image and consequently derived an adaptive Non Local Means (NLM) denoising method. Lebrun *et al.* described the non-local Bayes denoising algorithm in [11], based on the local covariance matrix of a collection of similar patches; they also established the connection between their algorithm and the PCA-based algorithm in [12]. Duval *et al.* performed an extended analysis of the filter parameter of NLM denoising, for a simplified patch weight model [5]. They showed that large patches provide a precise but biased estimate of the noise-free image. The filter parameter has to be consequently modulated across the image to increase the accuracy. To that end, they proposed an iterative method to compute the filter parameter in each pixel. Generally speaking, a careful definition of the lo-

¹We refer *image processing algorithm* to any algorithm that operates on the image domain, ranging from classical signal processing to any pre-processing step for later computer vision algorithms.

²We are not considering here the class of linear adaptive filters (e.g. Wiener or Kalman filters), for which a wide, well established theory exists.

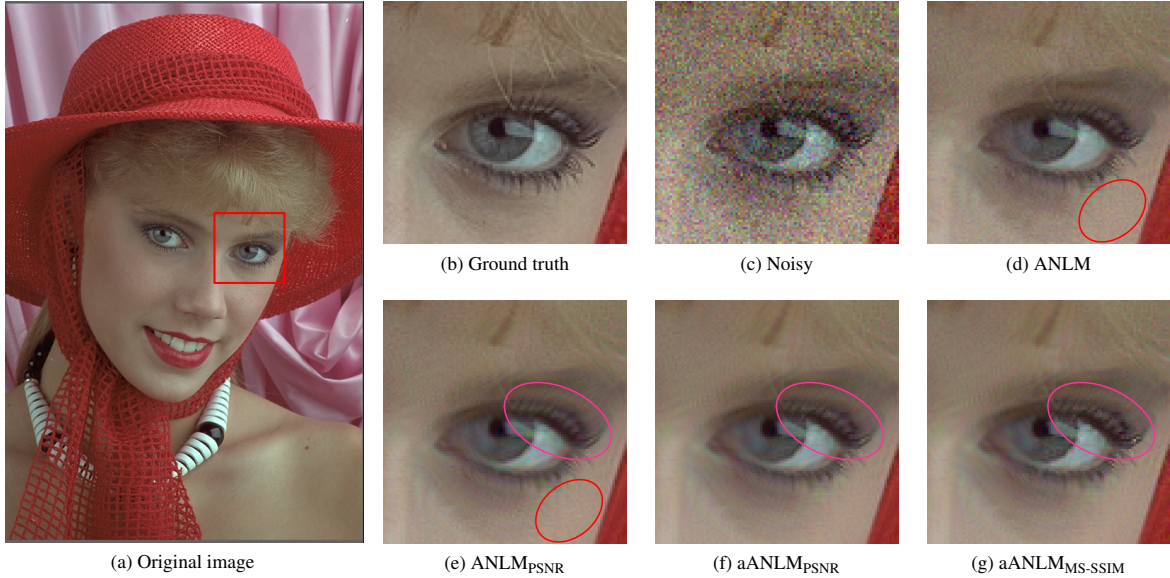


Figure 1: Result of Approximate Non Local Mean (ANLM) denoising. ANLM works like standard NLM, but using a limited number (here, 16) of nearest neighbor patches. Panel (d) shows the output of ANLM using the patch size and filtering parameter prescribed in [1]. Learning the *global* patch size and filtering parameter through optimization of PSNR on a training set of images, reduces the noise level on the skin (see the red ellipse for ANLM and ANLM_{PSNR}). Learning to *adaptively* modulate these parameters increases the contrast of small details (see the pink ellipse on the eyelashes when using *globally* learned parameters in (e) ANLM_{PSNR}, and *adaptively* learned parameters in (f) aANLM_{PSNR}). Learning adaptive parameters that optimize for MS-SSIM instead of PSNR further increases the contrast of small details, see (g). Better seen at 4x zoom.

cal statistics of the signal guarantees superior performances when compared to the heuristic approach, but the price to be paid is the high computational cost of the statistical analysis. Furthermore, complex interactions between the local image characteristics and the optimal filter parameters cannot be identified if not included a-priori in the statistical model. Finally, these algorithms generally optimize a statistical cost function, *e.g.*, ℓ_2 in [5], that may or may not be related to the application in hand.

Learning-based · Using machine learning to learn adaptive processing has the potential to overcome these issues. The main idea is to learn the relation between the image characteristics and the optimal processing procedure at a local scale. Zhang *et al.* [13] used this approach to learn the parameters of an adaptive bilateral filter aiming at denoising and deblurring by minimizing a least squares error over pixel clusters computed from Log of Gaussian responses. Frosio *et al.* [2] showed how to adaptively optimize the spatial and grey level range of the same filter, to maximize the PSNR on a set of noisy training images. The framework described in their paper is applied only to the case of image denoising and Gaussian noise; furthermore, the function putting in relation the local features and the filtering parameters is simpler than the one we propose in Sect. . The learning approach has also been adopted for Monte Carlo denoising in rendering, where a set of secondary features are extracted from local mean primary features including color, world positions, shading normals among others. These features are mapped to the filter parameters through a perceptron [4]. Fanello *et al.* [14] described how to learn optimal linear filters applied to clusters of patches, agglomerated by a random forest. The optimal filters are found by minimizing the ℓ_2 reconstruction error. Hammond and Simoncelli derived a closed form solution for blending two denoising filters based on wavelets [3], whereas

Liu *et al.* learned adaptive regularizations for patches based on the estimated distribution of features in transform domain [15]. It is remarkable to notice that each of the aforementioned approaches is targeted at a unique specific application with a pre-defined metric. On the other hand, our approach is not restricted to a specific application as long as one is able to define a task-specific objective function.

Method

To process an image, we first extract a set of features describing the characteristics of each pixel in the input image. We use a parametric model to put in relation the feature vector and the processing parameters for each pixel, and we estimate the model parameters by optimizing a user-defined cost function on a set of training images.

We describe the local image properties in position (x, y) by a feature vector $\mathbf{f}_{x,y} = [f_{x,y}^0 f_{x,y}^1 \dots f_{x,y}^{F-1}]^\top$ with F elements, where $f_{x,y}^0 = 1$ by definition. Typical features are the local image variance and entropies, as in [2]. The vector $\mathbf{p}_{x,y} = [p_{x,y}^0 p_{x,y}^1 \dots p_{x,y}^{P-1}]^\top$ contains the P processing parameters for the pixel (x, y) . For instance, in adaptive bilateral filter, the $P = 2$ parameters are the spatial and range sigmas. We relate $\mathbf{f}_{x,y}$ and $\mathbf{p}_{x,y}^k$ with a logistic function:

$$p_{x,y}^k = p_{\min}^k + \frac{p_{\max}^k - p_{\min}^k}{1 + e^{-h(\mathbf{f}_{x,y}; \theta_k)}}, \quad (1)$$

and

$$h(\mathbf{f}_{x,y}; \theta_k) = \theta_k^0 + \theta_k^{1\top} \mathbf{f}_{x,y} + \mathbf{f}_{x,y}^\top \Theta_k^2 \mathbf{f}_{x,y}, \quad (2)$$

where θ_k^0 is a scalar, θ_k^1 is a $F \times 1$ vector, Θ_k^2 is a $F \times F$ triangular matrix, θ_k is a vector containing θ_k^0 and the elements of θ_k^1 and Θ_k^2 , whereas p_{\min}^k and p_{\max}^k are the minimum and maximum values of $p_{x,y}^k$ (these are user-defined, reasonable values for a specific filter). We choose the logistic function since it builds a continuous map from \mathbb{R}^P to a bounded interval, while the model in Eq. (2) allows representing more complex interactions between features, compared to [2].

We denote the set of M input training images with $\{\mathbf{i}^j\}_{j=0}^{M-1}$, while $\{\mathbf{o}^j\}_{j=0}^{M-1}$ is the corresponding set of desired output images. For instance, in the case of denoising, \mathbf{i}^j and \mathbf{o}^j are respectively the noisy and noise-free images. Adaptive processing for the j -th image is described by:

$$\tilde{\mathbf{o}}^j = g\left(\mathbf{i}^j; \theta_0, \dots, \theta_{P-1}\right), \quad (3)$$

where g is the image processing algorithm of interest.

The learning procedure consists of the optimization of a cost function E with respect to $\{\theta_0, \dots, \theta_{P-1}\}$ on the set of training pairs $\{(\mathbf{i}^j, \mathbf{o}^j)\}_{j=0}^{M-1}$. We typically choose a cost $e = E(\{\tilde{\mathbf{o}}^j\}, \{\mathbf{o}^j\}; \theta_0, \dots, \theta_{P-1})$ related to the quality of the processed images (for instance, [2] uses the average PSNR of $\{\tilde{\mathbf{o}}^j\}$). Computing the derivatives of e with respect to the processing parameters may be difficult or even impossible in certain cases: for instance, when the parameters are discrete, as in the case of the patch size for NLM; or when the processing algorithm is iterative, as for TV denoising [6, 16]; or when the cost function is not differentiable, as in the case of FSIM[17]. Therefore, we resort to a derivative-free optimization algorithm, the Nelder-Mead simplex method [18]. After training, we use the optimal $\{\hat{\theta}_0, \dots, \hat{\theta}_{P-1}\}$ to compute $\mathbf{p}_{x,y}^k$ for each pixel of any image, out of the training set, through Eqs. (1) and (2). Adaptive filtering is then performed as in Eq. (3).

Applications

In the following we report results for the proposed learning procedure applied to three classical image processing problems: denoising, demosaicing and deblurring.

Image denoising

The first application we consider is denoising of color images through NLM [1]. This filter leverages the image self-similarity to first denoise single patches in the image. Then it averages the denoised patches to get the processed image in a collaborative way. Let's consider a $p^0 \times p^0$ patch, \mathbf{q}^n , corrupted by zero mean Gaussian noise with variance σ^2 . The denoised patch, \mathbf{q}^d , is [1]:

$$\mathbf{q}^d = \sum_{j=1}^N w_j \mathbf{r}_j, \quad (4)$$

$$w_j = \frac{e^{-\max\{d^2(\mathbf{q}^n, \mathbf{r}_j)/(p^0)^2 - 2\sigma^2, 0\}/(p^1\sigma)^2}}{\sum_{j=1}^N e^{-\max\{d^2(\mathbf{q}^n, \mathbf{r}_j)/(p^0)^2 - 2\sigma^2, 0\}/(p^1\sigma)^2}}, \quad (5)$$

where $\{\mathbf{r}_j\}_{j=1}^N$ is a set of N noisy patches, neighbors of \mathbf{q}^n , and $d(\mathbf{q}^n, \mathbf{r}_j)$ is the ℓ_2 distance between \mathbf{q}^n and \mathbf{r}_j . Global (non-adaptive) values for the patch size p^0 and the filtering parameter $p^1\sigma$ have been empirically investigated and reported in [1] as a function of σ . The choice of these parameters is critical when the Approximate NLM (ANLM) is used, *i.e.* when N is small to

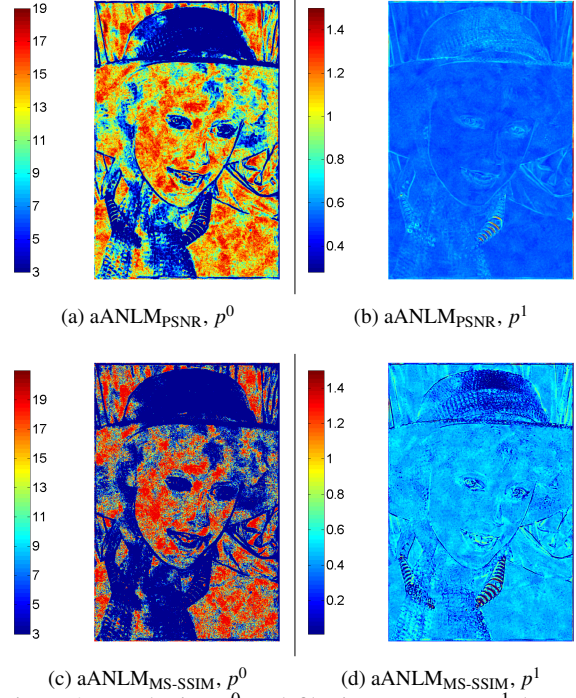


Figure 2: Patch size p^0 and filtering parameter p^1 learned by aANLM_{PSNR} and aANLM_{MS-SSIM} for the image in Fig. 1.

reduce the computational cost (*e.g.*, $N = 16$ in [19]) and preserve edges better [5].

We used the procedure in Sect. to learn adapting p^0 and p^1 across an image for ANLM, $N = 16$. The training set was composed of 12 of the 24, 512×768 images in the Kodak dataset [20], noise-free and corrupted by zero-mean, Gaussian noise with $\sigma = 20$. We used two different cost functions: the mean PSNR on the training set, $e = \sum_{j=1}^N \text{PSNR}(\tilde{\mathbf{o}}_j, \mathbf{o}_j)$; and the mean MS-SSIM [21], $e = \sum_{j=1}^N \text{MS-SSIM}(\tilde{\mathbf{o}}_j, \mathbf{o}_j)$. Optimizing for MS-SSIM is particularly interesting in the case of image denoising, as the final consumer of the image is a human observer and MS-SSIM correlates with the human judgement better than PSNR [21, 17]. We first performed the optimization using a single, unary feature ($F = 1$), thus not extracting in practice any local feature from the image. In this case $h(\mathbf{f}_{x,y}; \theta_k)$ in Eq. (2) is constant and our training procedure boils down to estimating the global, non-adaptive p^0 and p^1 parameters that maximize PSNR and MS-SSIM on the training set. We refer to these filters respectively as ANLM_{PSNR} and ANLM_{MS-SSIM}. Then we used $F = 7$ features to describe the local image characteristics. These are, apart from the first unary feature, the 3×3 and 5×5 local variance, the 3×3 and 7×7 entropy and the 3×3 and 7×7 gradient entropy. These features are sensitive to edges and textures [2]. We will refer to these adaptive filters respectively as aANLM_{PSNR} and aANLM_{MS-SSIM}. Notice that the extracted features are also expected to be noisy in this case. We therefore preprocessed the feature maps with NLM, using a 9×9 patch and a filtering parameter equal to 0.4σ , as suggested in [1]. Since the statistics of the noise in the feature maps is complicated or even unknown, we performed the nearest neighbor search and computed the patch weights using the original, noisy image.

	Noisy	ANLM [1]	ANLM _{PSNR}	ANLM _{MS-SSIM}	aANLM _{PSNR}	aANLM _{MS-SSIM}
p^0	-	5	9	19	Eq. (1)	Eq. (1)
$p^1 \cdot \sigma$	-	0.40σ	0.51σ	0.49σ	Eq. (1)	Eq. (1)
PSNR	22.11	30.06	30.27	30.00	30.80	30.48
SSIM	0.6750	0.8952	0.8992	0.9006	0.9056	0.9084
MS-SSIM	0.8192	0.9417	0.9445	0.9461	0.9480	0.9496

Table 1: Average image quality metrics measured on the testing dataset for ANLM. The values of the patch size p^0 and the filtering parameter p^1 are reported for the baseline and non-adaptive ANLM. The best result in each row is shown in bold.

We evaluated the performance of ANLM_{PSNR}, ANLM_{MS-SSIM}, aANLM_{PSNR} and aANLM_{MS-SSIM} on the 12 images of the Kodak dataset not used for training. For each filtered image we measured the PSNR, SSIM [22] and MS-SSIM [21]. We also compared to ANLM using the patch size $p^0 = 5$ and filtering parameter $p^1 \sigma = 0.4\sigma$ suggested in [1]. The average metrics are reported in Table 1, together with the values of the learned, global (non-adaptive) parameters. Training a non-adaptive filter to maximize the PSNR leads to an increase in the average PSNR on the testing data, when compared to the original ANLM (from 30.06dB to 30.27dB). This can be visually appreciated in Fig. 1, showing that ANLM_{PSNR} removes more noise than ANLM in the smooth areas. We observe a similar increase in MS-SSIM for ANLM_{MS-SSIM} compared to ANLM (from 0.9417 to 0.9461) but a slight decrease in PSNR (from 30.06dB to 30.00dB). This is due to the fact that different image quality metrics have different local maxima. In this specific case, optimizing MS-SSIM does not necessarily means optimizing PSNR at the same time.

Turning the global filtering procedure into an adaptive one (from ANLM_{PSNR} to aANLM_{PSNR}) results in an additional gain of approximately 0.5dB in PSNR, and an increase in SSIM and MS-SSIM. Fig. 1 visually demonstrates the advantages of using learning and adaptiveness together: compared to ANLM_{PSNR}, aANLM_{PSNR} produces slightly more contrasted small details (on the eyelashes in this case). Lastly, Fig. 1 highlights the differences between optimizing PSNR and MS-SSIM. MS-SSIM is only partially affected by small residual noise close to edges, which is on the other hand poorly visible by a human observer. This explains the slightly higher residual noise and, at the same time, the better visual appearance achieved by aANLM_{MS-SSIM}, consistent with what was reported in [7].

Fig. 2(a)-(b) shows the parameters learned by aANLM_{PSNR} for the image in Fig. 1. The learned strategy uses as-large-as-possible patches in the smooth areas (*e.g.*, the skin), and as-small-as-possible close to the edges (*e.g.*, at the face border) and on irregular textures (*e.g.*, the hat). The filtering parameter p^1 increases along edges, and it is at maximum in areas where saturation occurs (*e.g.*, on the necklace). This learned strategy is remarkably similar to the one derived by Duval *et al.* [5], based on a statistical analysis of NLM. Their main finding is that, close to the edges, we cannot find exact replicas of the reference patch. Consequently, averaging dissimilar patches introduces a bias. The bias can be reduced by using small patches, but this increases the uncertainty because of the poor statistics associated with the smaller number of pixels. To compensate for this, the filtering parameter p^1 is increased to include pixels from patches that are far from the reference one. Duval *et al.* [5] demonstrates analytically that such strategy is optimal when a least squares error (or, simi-

larly, PSNR) has to be optimized. Our method is capable of automatically learning the same strategy and even go beyond that. For instance, in the white area of the necklace, the filtering parameter is increased to its maximum since noise is reduced by saturation and the reliability of the measured pixels is consequently higher.

Even more interestingly, Fig. 2(c)-(d) shows that a different modulation strategy has to be adopted when a different metric (MS-SSIM in this case) is maximized. Since MS-SSIM is only slightly affected by noise close to edges, the filtering parameter p^1 is set to its minimum here, while the patch size p^0 is minimal. According to the analysis of Duval *et al.*, this reduces the bias (*i.e.*, the low pass filtering effect of NLM) and increases the variance (*i.e.*, leave more residual noise) in the filtered image, but it also produces more pleasant images as shown in Fig. 1(g).

Image demosaicing

The second application we consider is the mixture of demosaicing algorithms. Demosaicing is the process of recovering a full resolution color image from a subsampled (*e.g.*, Bayer) pattern. Several demosaicing approaches have been proposed in the literature [23], each with strengths and weaknesses. Adaptively blending the output of different demosaicing algorithms is expected to achieve a better overall image quality. For a mixture of P demosaicing algorithms, the output image at pixel (x,y) is defined as:

$$\tilde{\mathbf{d}}_{x,y} = \sum_{k=0}^{P-1} p_{x,y}^k \mathbf{d}_{x,y}^k / \sum_{k=0}^{P-1} p_{x,y}^k \quad (6)$$

where $\mathbf{d}_{x,y}^k$ is the output of the k -th algorithm, $p_{\min}^k = 0$ and $p_{\max}^k = 1$, $\forall k$. The learned parameters here are used to compute the blending factors, $p_{x,y}^k / \sum_{k=0}^{P-1} p_{x,y}^k$.

We consider a mixture of $P = 3$ recently published, state-of-the-art demosaicing algorithms: Adaptive Residual Interpolation (ARI, [26]), an algorithm based on the exploitation of inter-color correlation (ECC, [25]), and image demosaicing with contour stencils (CS, [27]). For training and testing we employed the images of the McMaster dataset [24], after splitting it into two separate sets. We first learned how to blend the output of the three algorithms globally (*i.e.*, $F = 1$), by maximizing the mean PSNR (Mix_{PSNR}). For adaptive blending, we used as features the local variance, intensity entropy and gradient entropy, computed separately on three channels of the Bayer pattern, for a total of $F = 10$ features. The local window size for extracting the features is 7×7 , close to the operating scale of most demosaicing algorithms. Learning was performed to maximize the PSNR, SSIM and MS-SSIM on the training data. The corresponding algorithms are indicated as aMix_{PSNR}, aMix_{SSIM}, and aMix_{MS-SSIM}. Results are reported in Table 2.

The learned, adaptive mixture, $\text{aMix}_{\text{MS-SSIM}}$, outperforms the best of the three original methods by 0.96dB on average in terms of PSNR, when MS-SSIM is maximized in training. The improvement is consistent for all the three metrics considered here (PSNR, SSIM and MS-SSIM) even if only one of them is optimized during training. Consistent with the case of denoising, training to maximize SSIM or MS-SSIM produces the best results in terms of SSIM or MS-SSIM. This also produces an improvement in terms of PSNR since these image quality metrics are correlated, even if they do not share the same local maxima. Remarkably, even without any adaptiveness, the learned mix (Mix_{PSNR}) gains 0.86dB on average in terms of PSNR, compared to the best of ECC, ARI and CS. The learned (non-adaptive) blending factors are in this case 0.50 for ECC, 0.44 for ARI and 0.06 for CS. The low importance given to CS on the average is somehow expected, given the lower performance achieved by this method and reported in Table 2. On the other hand, the fact that the blending factor for CS is non-zero demonstrates that CS can be coupled with the other two algorithms to increase the image quality. This advantage is maximized when turning to an adaptive approach, which results into an additional 0.1dB on the average compared to the non adaptive approach.

Fig. 3 shows the blending factors for three testing images. ECC and ARI are the best methods when used alone, and they are almost equally distributed in the mixture. Close to the edges, the learned modulation strategy select either one or the other of these methods. In homogeneous areas they appear to be equally important, with the exception of very dark or bright areas, where the weights associated to CS becomes predominant, even if this is the worst algorithm (among the three considered here) when used alone. The quality of the images obtained with the proposed method can be appreciated in Fig. 4, showing a better reconstruction of the border of the red pepper compared to state-of-the-art ECC and ARI.

Image deblurring

The last application we consider is image deblurring through TV regularization. This problem is of particular importance in the medical and astronomical fields, where imaging apparatuses with a known point spread function measure single channel images with a very limited number of photons [6, 16]. The problem can be stated as follow. Given the vectorial representation of a noise-free image, \mathbf{o} , the measured, noisy, and blurred image, \mathbf{i} , is:

$$\mathbf{i} = \mathbf{H} \cdot \mathbf{o} + \mathbf{n} \quad (7)$$

where \mathbf{H} is a matrix representing a linear blur and \mathbf{i} is corrupted by photon counting noise \mathbf{n}^3 with a Poisson distribution [6]. The inverse problem of estimating \mathbf{o} from \mathbf{i} is formulated in a Bayesian context as a *maximum a posteriori* problem. We estimate the noise free image, $\hat{\mathbf{o}}$, as:

$$\hat{\mathbf{o}} = \arg \min_{\mathbf{o}} -\log(L(\mathbf{i}|\hat{\mathbf{o}})) + p_{x,y}^0 R(\hat{\mathbf{o}}), \quad (8)$$

where $L(\mathbf{i}|\hat{\mathbf{o}})$ is the likelihood term (the Kullback-Leibler divergence for Poisson noise) and $R(\hat{\mathbf{o}})$ is the regularization term. TV regularization is often adopted for its edge preserving capability.

³With an abuse of notation, we indicate here the photon counting noise as additional noise \mathbf{n} .

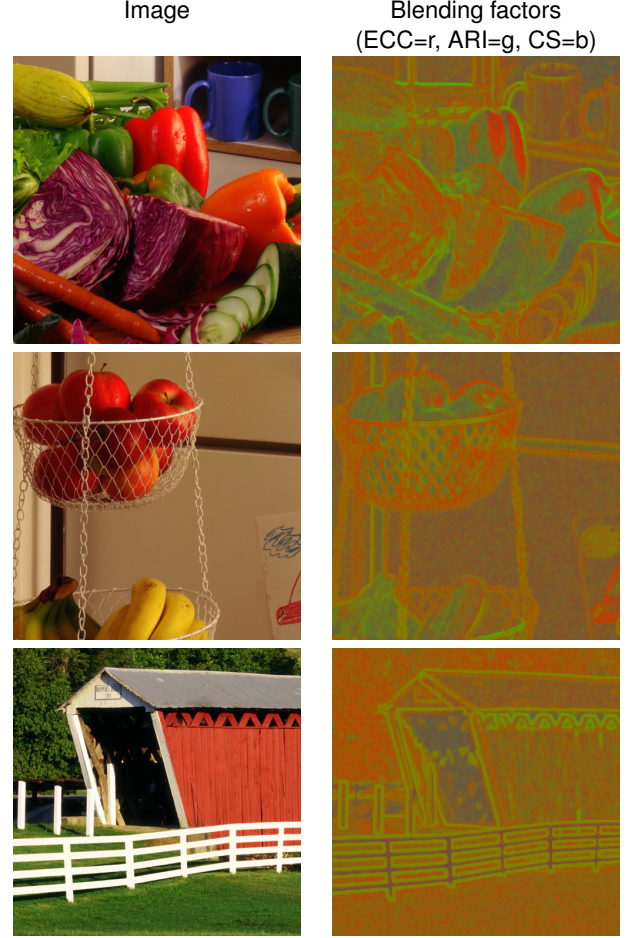


Figure 3: Blending factors learned by $\text{aMix}_{\text{MS-SSIM}}$ (right column) for three images of the McMaster dataset [24]. The blending factor for ECC, $p^0 / \sum p^k$, is associated with the red channel, whereas those of ARI ($p^1 / \sum p^k$) and CS ($p^2 / \sum p^k$) are associated with the green and blue channels, respectively. ECC and ARI are the preferred methods for edges and textured areas, whereas CS has more importance in very dark or bright areas of the images.

The regularization hyper-parameter $p_{x,y}^0$ is traditionally kept constant across the image and chosen, for instance, through a discrepancy principle [6]. For simplicity, we solve Eq. (8) by steepest descent, even if more efficient algorithms can be used.

We used the procedure from Sect. to learn how to adapt $p_{x,y}^0$ across the image. The training dataset was composed of 12 of the 24 images of the Kodak dataset, converted to grey levels. Each image was blurred with a Gaussian kernel of 7×7 pixels and a standard deviation of 2 pixels. White, photon counting noise was added to the blurred images, assuming a maximum number of 1024 photons per pixel. During learning we optimized the average PSNR and MS-SSIM on the training dataset. We first performed training for $F = 1$ (i.e. for a global, non-adaptive hyper-regularization parameter p^0). We refer to these filters as TV_{PSNR} and $\text{TV}_{\text{MS-SSIM}}$. We then performed training using $F = 7$ features: the unary feature and the 5×5 and 9×9 local grey level average, standard deviation and their ratio. The rationale for these features is that the optimal regularization hyper-parameter may be

	ECC [25]	ARI [26]	CS [27]	Mix _{PSNR}	aMix _{PSNR}	aMix _{SSIM}	aMix _{MS-SSIM}
PSNR	38.85	38.37	36.67	39.71	39.81	39.80	39.81
SSIM	0.9666	0.9633	0.9516	0.9717	0.9725	0.9725	0.9725
MS-SSIM	0.9951	0.9945	0.9913	0.9963	0.9963	0.9963	0.9964

Table 2: Average image quality metrics on the testing dataset for demosaicing. The best result in each row is shown in bold.

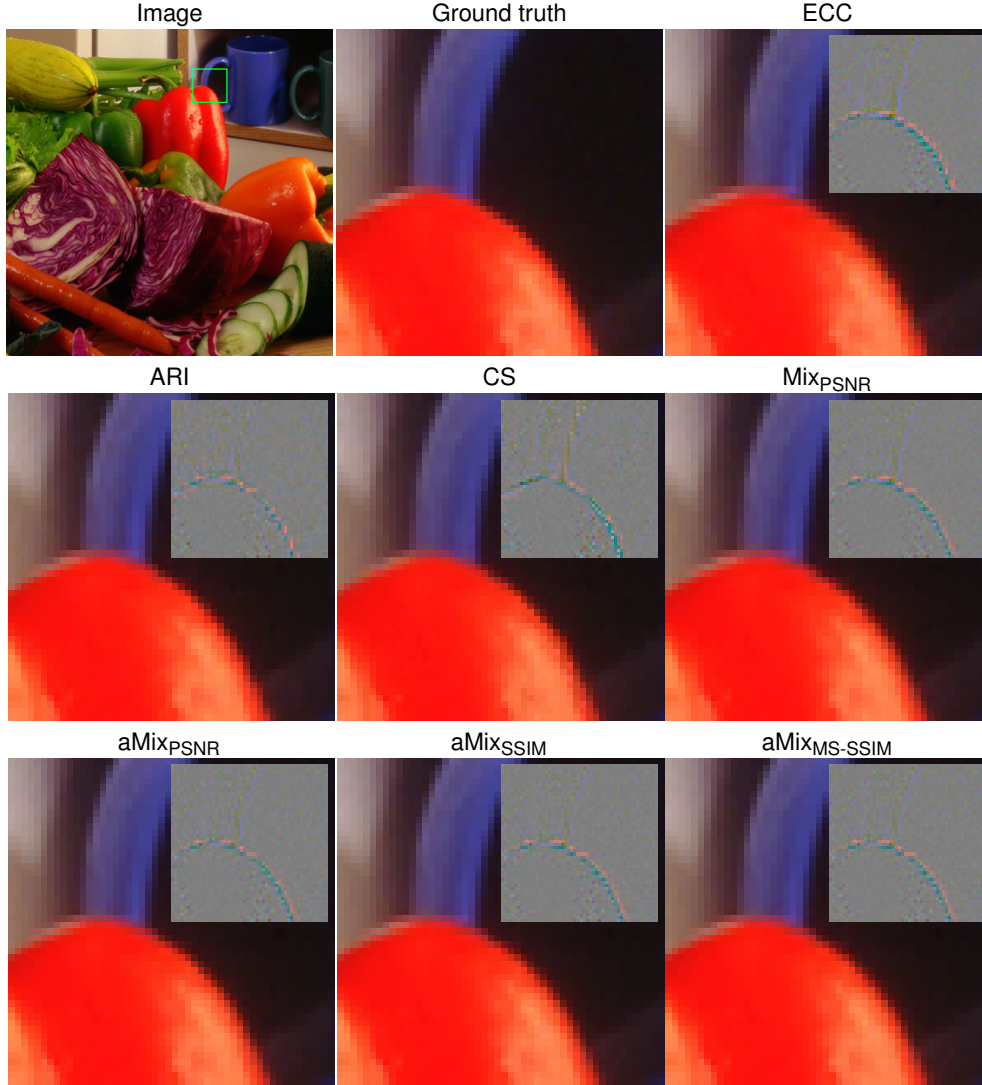


Figure 4: Output of different demosaicing algorithms for the green patch in the upper left panel. The inset panels shows the residual error. The learned blending strategy gains approximately 1dB on the average compared to previous state-of-the-art results. Better seen at $4\times$ zoom.

affected by the signal-to-noise ratio (which is proportional to the number of photons) and by the presence of structure in the local image (associated with the local variance). We refer to the corresponding adaptive filters as aTV_{PSNR} and $aTV_{MS-SSIM}$.

The average image quality metrics achieved on the testing dataset (the second half of the Kodak dataset) are reported in Table 3. This shows a slight advantage of using the adaptive strategy over the non adaptive one, both in the case of the maximization of PSNR and MS-SSIM. Fig. 5(f) and 5(h) show that the adaptive filter learned by maximizing the PSNR is capable of better reconstructing small details in the image, like the writing on the wall.

This is achieved using a larger regularization hyper-parameter close to the edges (Fig. 5(b) and 5(k)). Nonetheless, since most of the pixels in the smooth part of the image are still filtered using the global optimum hyper-parameter, the overall gain in terms of PSNR is small. Maximizing MS-SSIM instead of PSNR leads to a slightly different adaptation strategy and to images with more contrast on small details, that are also a bit more noisy (Fig. 5(g) and 5(i)).

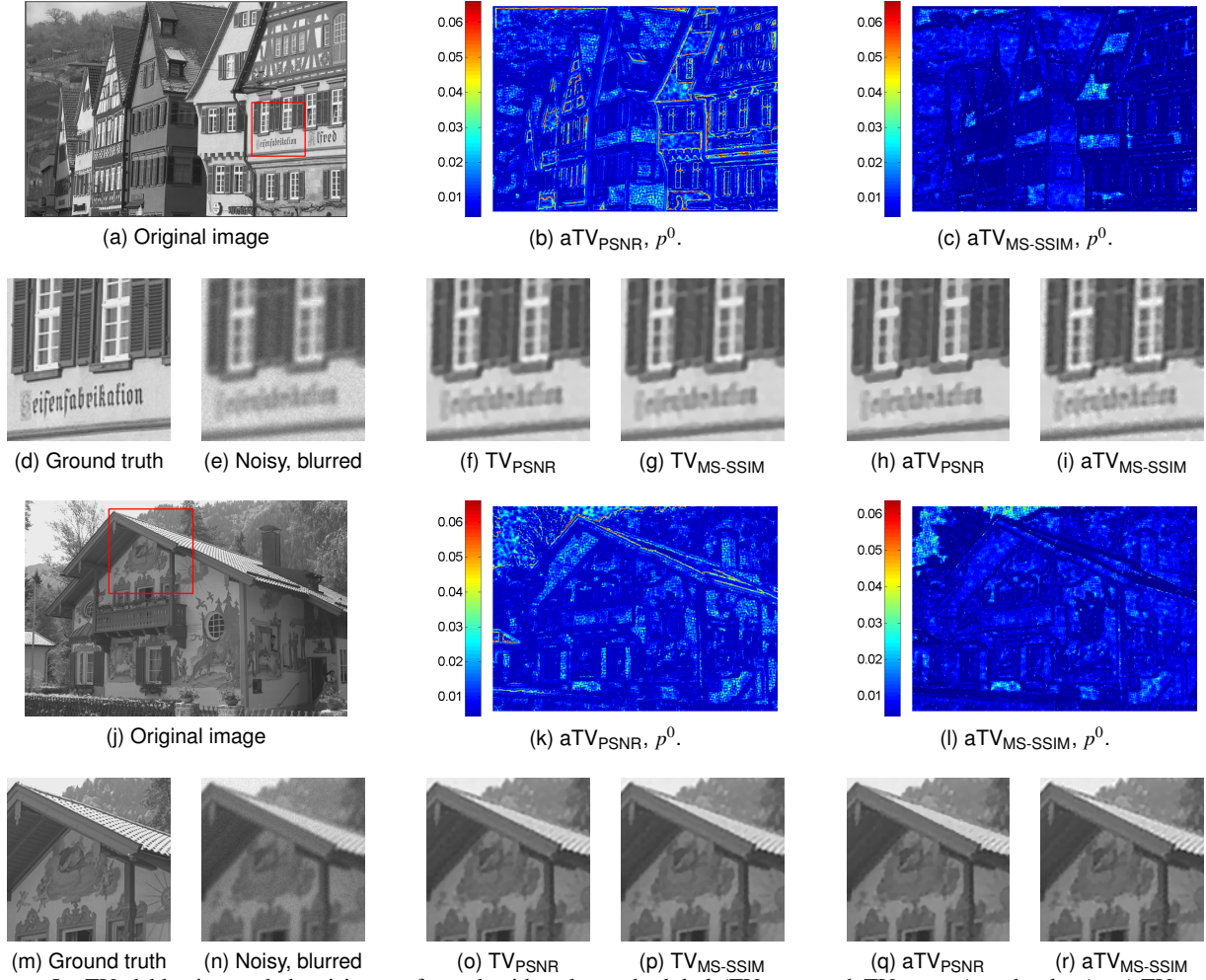


Figure 5: TV deblurring and denoising performed with a learned *global* (TV_{PSNR} and $\text{TV}_{\text{MS-SSIM}}$) and *adaptive* (aTV_{PSNR} and $\text{aTV}_{\text{MS-SSIM}}$) regularization hyper-parameter, p^0 . The writing on the wall (panels a, d) is better restored by the adaptive filters (panels h, i) and it is slightly more contrasted and noisy when MS-SSIM is maximized. The texture of the roof (panels j, m) is better restored by the adaptive filter maximizing MS-SSIM (panel r).

	TV_{PSNR}	$\text{TV}_{\text{MS-SSIM}}$	aTV_{PSNR}	$\text{aTV}_{\text{MS-SSIM}}$
p^0	0.0065	0.0069	Eq. (1)	Eq. (1)
PSNR	28.455	28.45	28.49	28.43
SSIM	0.8745	0.8745	0.8764	0.8774
MS-SSIM	0.9410	0.9411	0.9421	0.9425

Table 3: Average image quality metrics on the testing dataset for TV deblurring. The values of the hyper-regularization parameter p^0 are also reported. The best result in each row is shown in bold.

Conclusion

We have presented a general method to learn tuning of parameters in adaptive image processing. We applied it effectively to three classical problems, but the same procedure can be applied to other image processing problems (*e.g.*, segmentation, compression, and so on), once a proper cost function and set of features have been defined.

The proposed learning procedure always led to an improvement in the quality of the processed images. We demonstrated its

application in the case of TV deblurring, an iterative processing procedure for which it is not possible to write an analytical cost function to optimize; the improvement in this case is small as a single parameter is optimized and the space of the possible outputs is highly constrained by the filter flow-chart. For a complex filter like ANLM, where two adaptive parameters are considered, we obtained a gain similar to that reported in [5], where adaptiveness is driven by a statistical risk criterion. For demosaicing, we demonstrated how to apply our method to blend different filters with orthogonal properties; this has the potential to produce large improvements in the learned adaptive algorithm, and in fact we achieved state-of-the-art results by learning a global mixture, and a further slightly improvement by introducing adaptiveness.

We also observed, consistent with [7], that optimizing for different cost functions leads to different parameter modulation strategies. Maximizing for perceptual image quality metrics improves their score and likely also the actual perceived quality. Since we resorted to derivative-free optimization, optimizing for complex, state-of-the-art metrics like FSIM [17], that better corre-

lates with human judgement of image quality, is also doable with our method. We leave these aspects for future investigations.

References

- [1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model Sim.*, 2005.
- [2] I. Frosio, K. Egiazarian, and K. Pulli, "Machine learning for adaptive bilateral filtering," in *SPIE*, 2015.
- [3] D. K. Hammond and E. P. Simoncelli, "A machine learning framework for adaptive combination of signal denoising methods," *IEEE TIP*, 2007.
- [4] N. K. Kalantari, S. Bako, and P. Sen, "A machine learning approach for filtering monte carlo noise," *ACM TOG*, 2015.
- [5] V. Duval, J.-F. Aujol, and Y. Gousseau, "On the parameter choice for the non-local means," *SIAM J. Imag. Sci.*, 2011.
- [6] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, "A discrepancy principle for poisson data," *Inv. Prob.*, 2010.
- [7] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE TCI*, 2017.
- [8] A. Polesel, G. Ramponi, and V. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE TIP*, 2000.
- [9] Q. Chen, P. Montesinos, Q. S. Sun, P. A. Heng, and D. S. Xia, "Adaptive total variation denoising based on difference curvature," *Image Vision Comput.*, 2010.
- [10] T. Thaipanich, B. T. Oh, P.-H. Wu, and C.-C. Kuo, "Adaptive nonlocal means algorithm for image denoising," in *ICCE*, 2010.
- [11] M. Lebrun, M. Colom, A. Buades, and J. M. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, 2012.
- [12] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, 2010.
- [13] B. Zhang and J. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise removal," *IEEE TIP*, 2008.
- [14] S.-R. Fanello, C. Keskin, P. Kohli, S. Izadi, J. Shotton, A. Criminisi, U. Pattacini, and T. Paek, "Filter forests for learning data-dependent convolutional kernels," in *CVPR*, 2014.
- [15] H. Liu, R. Xiong, J. Zhang, and W. Gao, "Image denoising via adaptive soft-thresholding based on non-local samples," in *CVPR*, 2015.
- [16] I. Frosio, C. Olivieri, M. Lucchese, N. Borghese, and P. Boccacci, "Bayesian denoising in digital radiography: A comparison in the dental field," *Computerized Medical Imaging and Graphics*, 2013.
- [17] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE TIP*, 2011.
- [18] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, 1965.
- [19] Y.-T. Tsai, M. Steinberger, D. Pajk, and K. Pulli, "Fast ANN for high-quality collaborative filtering," *CGF*, 2015.
- [20] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process-Image*, 2015.
- [21] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *ACSSC*, 2003.
- [22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TIP*, 2004.
- [23] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," in *SPIE*, 2008.
- [24] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journ. Electron Imaging*, 2011.
- [25] S. P. Jaiswal, O. C. Au, V. Jakhetiya, Y. Yuan, and H. Yang, "Exploitation of inter-color correlation for color image demosaicking," in *ICIP*, 2014.
- [26] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, "Adaptive residual interpolation for color image demosaicking," in *ICIP*, 2015.
- [27] P. Getreuer, "Image demosaicking with contour stencils," *IPOL*, 2012.

Author Biography

Jingming Dong received his Ph.D. in Computer Science from University of California, Los Angeles in 2017. His research interests are Computer Vision, Machine Perception and Deep Learning. Currently he focuses on local representation learning with application in wide-baseline matching, general object recognition and semantic mapping in visual inertial navigation systems. Before he received his B.Sc. in Computer Science and Technology from Fudan University, Shanghai, China in 2011.

Iuri Frosio got his PhD in biomedical engineering at the Politecnico of Milan in 2006. He was a research fellow at the Computer Science Department of the University of Milan from 2003 and an assistant professor in the same department from 2006 to 2013. In the same period, he worked as a consultant for various companies in Italy and in the US. He joined NVIDIA in 2014 as senior research scientist. His research interests include image processing, computer vision, parallel programming and reinforcement learning.

Jan Kautz is a Senior Director of Research at NVIDIA, where he leads the Learning & Perception Research team. He is working predominantly on computer vision problems (from low-level vision through geometric vision to high-level vision), as well as machine learning problems (including deep reinforcement learning, generative models, and efficient deep learning). Before joining NVIDIA in 2013, Jan was a tenured faculty member at University College London. He is a member of ACM and IEEE.