Learning how to combine internal and external denoising methods

Harold Christopher Burger, Christian Schuler, and Stefan Harmeling

Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract. Different methods for image denoising have complementary strengths and can be combined to improve image denoising performance, as has been noted by several authors [11, 7]. Mosseri et al. [11] distinguish between internal and external methods depending whether they exploit internal or external statistics [13]. They also propose a rule-based scheme (PatchSNR) to combine these two classes of algorithms. In this paper, we test the underlying assumptions and show that many images might not be easily split into regions where internal methods or external methods are preferable. Instead we propose a learning based approach using a neural network, that automatically combines denoising results from an internal and from an external method. This approach outperforms both other combination methods and state-of-the-art stand-alone image denoising methods, hereby further closing the gap to the theoretically achievable performance limits of denoising [9]. Our denoising results can be replicated with a publicly available toolbox[1].

1 Introduction

Image denoising is the long-standing problem of finding a clean image, given a noisy one. Usually, one seeks to denoise images corrupted with additive white Gaussian (AWG) noise, where it is often assumed that the variance of the noise is known. Most often, the images one wishes to denoise are so-called natural images (i.e. every-day scenes). The quality measure of interest is the peak signal-to-noise ratio (PSNR), which is monotonically related to the mean squared error.

Denoising methods can be divided into *internal* and *external* methods [11]: (i) internal methods denoise image patches using only other noisy image patches from the same image. In contrast, (ii) external methods denoise image patches using external clean image patches (i.e. patches coming from a database of clean images). For instance:

Internal denoising methods:

- NLM (non-local means) [2] denoises a noisy image using only patches from the same image: No explicit assumptions are made regarding all natural images.
- BM3D [6] is conceptually similar to NLM, but uses a more effective noise-reduction strategy than NLM, which averages similar-looking patches.

[1] `http://webdav.is.mpg.de/pixel/prj/neural_denoising/gcpr2013.html`

External denoising methods:

– EPLL [14] denoises image patches using a probabilistic prior for the image patches learned on a database of clean image patches.
– MLP is the currently best performing method, see [4, 3]. It uses a multi-layer perceptron to automatically learn a denoising method.

Other denoising methods can be less clearly classified, e.g. LSSC [10] learns a dictionary on the noisy image at hand and exploits this dictionary in a manner reminiscent of BM3D (speaking for an internal method), but the initialization of the dictionary is also important. Therefore it seems that external information also plays a role in LSSC, similarly for KSVD [1].

Recent denoising methods (such as BM3D [6], LSSC [10], EPLL [14]) perform on average equally well. This is surprising, considering that the methods rely on fundamentally different approaches. This has naturally led to the question if there are inherent limits to how well it is possible to denoise, and if so, whether current methods are approaching these limits. Even though the approaches taken are different [5, 9], the consensus is that current methods are indeed not far away from theoretical limits, especially at lower noise levels.

Contributions: In this paper, we will study the performance of internal and external methods across an image database and patch-wise across single images. Furthermore, we propose a method that automatically combines the advantages of external and internal approaches using learning. In particular our contributions are:

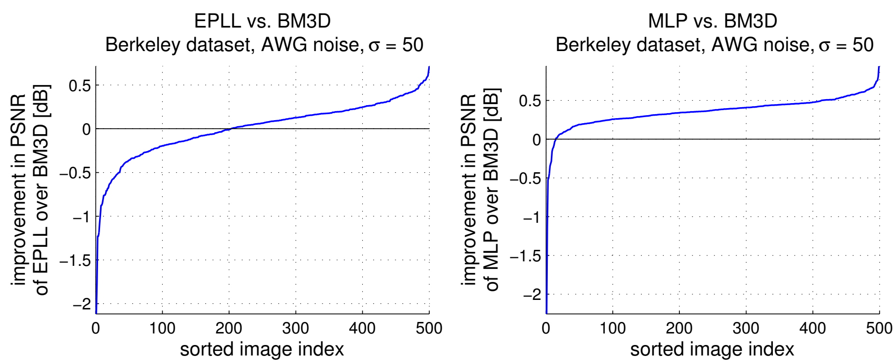
1. We show that internal denoising methods tend to be better for images depicting artificial objects, whereas external denoising methods are better for images of natural scenes.
2. We show that there is no trivial rule to decide whether to use external or internal denoising on a patch-by-patch basis.
3. We show that a combining strategy can be learned by an MLP that outperforms both internal and external approaches across a wide range of images.
4. We show that the new combined approach gets close to theoretical bounds.

2 Related work

Work on image denoising is extensive and we already mentioned some of the best performing methods in the introduction. In the following we limit our discussion on publications that also try to combine different denoising methods:

RTF. Jancsary et al. [7] observe that there is no single best denoising method, but that even in a single image depending on the image content one method might be preferable over others (see Fig. 5 in [7]). For that reason, they not only consider regression tree fields (RTFs) based on some filterbank ($\mathrm{RTF}_{\mathrm{plain}}$), but they also study a version that additionally exploits the output of BM3D ($\mathrm{RTF}_{\mathrm{BM3D}}$) and a version that additionally uses the output of four denoising methods simultaneously ($\mathrm{RTF}_{\mathrm{all}}$). Their finding is that the approach combining several methods is the best. In general their approach is based on learning RTFs on a large dataset of images, thus automatically determining how image features and different denoising methods can be combined. However, they do not discuss the distinction between internal and external methods.

PatchSNR. Zontak and Irani [13] study the merits of internal vs. external statistics for the task of super-resolution and also for denoising, where they observe that NLM works better with internal noisy patches (internal-NLM) than with noise-free patches from external images (external-NLM). Following up on this work, Mosseri et al. [11] introduce the corresponding distinction between internal and external denoising algorithms. To combine the advantages of these two paradigms, they propose a patch-wise signal-to-noise-ratio called PatchSNR, which as they claim indicates whether an internal (low PatchSNR) or an external (high PatchSNR) denoising method should be applied. The resulting denoised patches are blended together to form a single image, and they show that their results are slightly better than the stand-alone methods.



(a) EPLL vs. BM3D (b) MLP vs. BM3D

Fig. 1. No method is *always* the best. (a) Performance profile of EPLL vs. BM3D. (b) Performance profile of MLP vs. BM3D.

3 Internal vs. external denoising

3.1 Comparison on a large dataset of images

To compare the performance of two denoising algorithms, we plot the sorted differences between PSNRs achieved on a large set of noisy images. We call such a plot a *performance profile*. Fig. 1 (a) shows such a performance profile for EPLL [14], an external method, against BM3D [6], an internal method. We see that EPLL is worse than BM3D on 40% of the image images (blue line below zero) and better than BM3D on 60% of the images (blue line above zero). On some images EPLL is much better (about 0.5dB), while on other images BM3D is much better (more than 1dB in the extreme case).

Fig. 1 (b) shows a similar comparison for MLP [4, 3] (also an external method) vs. BM3D. We see that MLP is the clear winner, being superior on almost all images (blue line above zero). However, there are also some images where BM3D wins (close to image index zero). Even though Fig. 1 (b) shows that MLP is good over a large range of images, we can *not* conclude that one algorithm is the best on *all* images.

MLP vs. BM3D: +0.84dB | MLP vs. BM3D: +0.82dB | MLP vs. BM3D: +0.82dB

MLP vs. BM3D: -1.09dB | MLP vs. BM3D: -0.66dB | MLP vs. BM3D: -0.54dB

Fig. 2. MLP vs. BM3D for $\sigma = 25$: MLP wins (top row), BM3D wins (bottom row).

Is there some underlying principle that would allow us to predict whether an internal (such as BM3D) or an external algorithm (such as MLP) is better on a given image? To answer this question, we show in the first row of Fig. 2 images where MLP excels and in the second row images where BM3D is better. We notice that MLP tends to outperform BM3D on images containing smooth areas or irregular textures, whereas BM3D outperforms MLP mainly on images with regular, repeating textures (many more images supporting this in the supplementary material). Put differently, MLP is better for images of nature, while BM3D is better for images of man-made objects. This also makes sense intuitively, since an internal method like BM3D exploits the self-similarity of images which is much higher in images showing highly regular structures, while common images of nature show many irregular patterns, which are not easily matched to each other by an internal method.

Conclusion. We hypothesize that current internal methods are good at repetitive image structure, while external methods are good at irregular image content. In order to combine the strength of both paradigms, can we easily decide on a patch-by-patch level whether to apply internal or external denoising? To answer this question we compare internal and external denoising methods pixel-wise on single images.

3.2 Pixel-wise comparison on single images

Having denoised a given image with two methods, we can create a so-called *preference image* that shows a white pixel if the first method is closer to the truth or a black pixel if the second method is better. Such a visualization to compare methods pixel-wise has been previously used in [7] to compare four methods simultaneously (with four colors, Fig. 5 in [7]) and in [11].

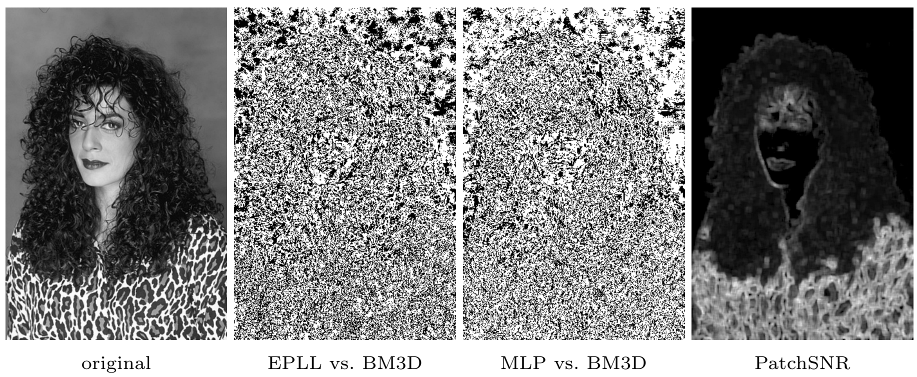


Fig. 3. Preference images and PatchSNR for image "woman" ($\sigma = 25$) where external methods (EPLL, MLP) perform better than internal methods (BM3D). EPLL: 24.86dB, MLP: **25**.**43**dB, BM3D 24.52dB.

Fig. 3 shows the image "woman" used by Mosseri et al. [11], who compare the performance of internal-NLM against external-NLM on that image. They conclude that smooth image patches should preferentially be denoised with an internal denoising method, whereas patches with details should rather be denoised with an external denoising method. Furthermore, they conclude that the higher the noise, the higher the preference for internal denoising. To exploit these insights, they apply the PatchSNR (briefly introduced in Sec. 2). The higher the PatchSNR, the higher the preference for external denoising, and vice-versa for internal denoising. [11] shows that this approach is effective for combining internal-NLM and external-NLM. Their approach also yields better results when combining BM3D and EPLL. However, the two preference images (two middle images in Fig. 3) show that the preference of EPLL or MLP over BM3D is much less clear-cut. This is somewhat surprising since the image "woman" is an example where the external methods EPLL and MLP outperform the internal method BM3D. Also, we used the ground truth to decide which method is better and still do not see a clear pattern for which pixels should use which method.

As a second example, we consider image "Barbara" (see Fig. 4) which is an example of an image where internal methods such as BM3D are better than external methods like EPLL and MLP. The reason for this is that there is a repetitive pattern on the table cloth and the trousers which are hard to recover

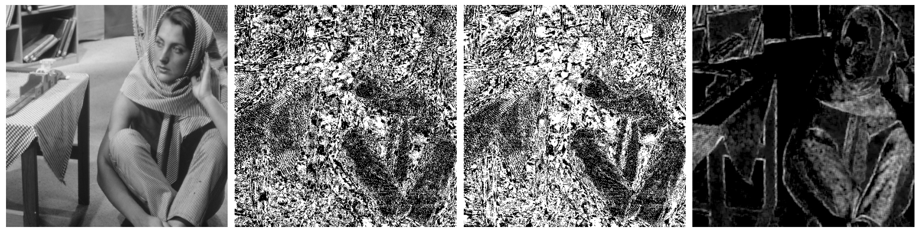


Fig. 4. Preference images and PatchSNR for image "Barbara" ($\sigma = 50$), where external methods (EPLL, MLP) perform worse than internal methods (BM3D). EPLL: 24.83dB, MLP: 25.28dB, BM3D: **27.22**dB.

for external methods. This is supported by the two preference images in the middle of Fig. 4. We clearly see dark areas for the trousers and the table cloth, indicating that BM3D is preferred for those regions. However, across large areas of the image, there is not a clearly preferred approach. On the other hand, PatchSNR gives a strong preference to BM3D on the smooth regions of the image and low preference for the trousers and the table cloth. This is opposite to our findings.

Conclusion. For the analysis of these two images we draw the following conclusions (there are more images in the supplementary material supporting our findings):

1. External denoising is usually better on irregular and smooth regions.
2. Internal denoising is usually better on regular, repeating structures.
3. Overall there is no easy way to determine which is better. In particular, our findings contradict those of the PatchSNR criterion [11].

Further notes on PatchSNR: Why are our conclusions different from Mosseri et al.'s [11]? The reason lies in the patch size of the methods: [11] compares external- vs. internal-NLM, both of which use small (7×7) patches. External methods using small patch sizes tend to overfit the noise, especially for smooth patches, which has been also noted by Mosseri et al. [11], as well as by Zontak and Irani [13]. In contrast, we consider MLP and EPLL. MLP uses 39×39 patches. Even though EPLL uses small (8×8) patches, it requires several iterations, which spreads out image information and therefore effectively increases the patch size. On the other hand, internal methods (like BM3D) using small patches are less prone to this effect due to the fact that similar patches are likely to be found in the vicinity of a given patch [13].

4 Learning to combine internal and external methods

Since we have seen that there is no trivial criterion to decide whether an internal or an external method should be applied to a given patch, we propose to use a learning approach based on neural networks to combine the complementary strengths of internal and external denoising methods (see Sec. 3.1).

Technically, we could combine any number of denoising methods with a neural network (similar to the learning-based approach proposed by [7], who combine the results of four denoising algorithms). However, we will show that it is sufficient to combine one internal method (BM3D) with one external method (MLP). Our method uses the original noisy image patch together with the denoised patches of MLP and BM3D as input. We choose do so because applying a denoising algorithm inevitably removes information contained in the noisy image. However, exactly that information might be missing in the denoised patches by BM3D or MLP.

Multi-layer perceptrons. The neural network we employ is a multi-layer perceptron that non-linearly transforms a vector-valued input into a vector-valued output. It is composed of a sequence of differentiable functions whose parameters can be trained efficiently given labeled training data with a combination of the back-propagation algorithm and stochastic gradient descent [8]. Usually, layers performing an affine transformation and layers performing an element-wise non-linearity (such as tanh) are applied in sequence. For example, $f(x) = W_2 \tanh(W_1 x + b_1) + b_2$ is a multi-layer perceptron with a single hidden layer whose parameters $\theta = \{W_1, W_2, b_1, b_2\}$ can be learned.

Training. Our neural network takes as input x the concatenation of three input patches, one from the noisy image, and one from each of the denoising results (BM3D and MLP), extracted from the same image location. The output of the neural network is a clean image patch. As a pre-processing step, we approximately de-correlate the three input patches using a pre-learned 3×3 matrix (one for each noise level). De-correlating the inputs of a neural network is considered good practice [8]. In our case the use of a 3×3 whitening matrix can be intuitively justified by the fact that two of the inputs (BM3D and MLP) look very similar (see supplementary material).

To generate training data, we add noise to images from a large image data set, and apply both BM3D and MLP. This provides us the input/output pairs required for training. Note that BM3D and MLP are computationally relatively inexpensive, allowing us to generate plentiful training data (we denoised approximately 9×10^4 images).

Hyper-parameters. We use four hidden layers with 2047 hidden units each. The input patches are each of size 25×25, the output patch are of size 17×17. We also experimented with smaller architectures, leading to worse results, see supplementary material. We used a constant learning rate of 0.1, as suggested in [4].

Training and running times. We train six neural networks, one for each of the noise levels $\sigma = \{10, 25, 35, 50, 75, 170\}$. Training each neural network is computationally intensive: About 4×10^8 training samples are needed before the results converge. This requires roughly one month of training time on a GPU. However, the running time of applying a trained neural network to a noisy image is relatively short: Approximately one minute on a CPU for an image of size 512×512 (such as "Lena"). Running times on GPU are less than six seconds. This compares favorably to other denoising methods such as EPLL (approximately five minutes on CPU) or KSVD [1] and LSSC [10] (approximately one hour on CPU), but unfavorably to BM3D (about five seconds on CPU). The total computation time of our method on an image of size 512×512 is therefore about two minutes on CPU: One minute for MLP, a few seconds for BM3D and one minute for our neural network combining MLP and BM3D.

5 Results

In the following, we show that combining MLP and BM3D with a neural network (as explained in the previous section) outperforms the current state-of-the-art stand-alone methods as well as previous attempts to combine denoising methods. We can also show that the proposed approach further closes the gap to the theoretical limits of denoising.

Comparison against competing methods. Tab. 1 compares our method to the combination approach of [11] as well as to stand-alone denoising methods on a (held-out) test set of 100 images, for six different noise levels. In all tested settings, our approach outperforms the existing methods.

σ	Mosseri et al. [11] BM3D and EPLL	our results BM3D and MLP	BM3D [14]	EPLL [6]	MLP [3]
170	20.14	**21.96**	19.85	21.21	21.87
75	24.16	**24.53**	23.96	24.16	24.42
50	25.64	**25.95**	25.45	25.50	25.83
35	27.07	**27.36**	26.89	26.98	27.29
25	28.54	**28.79**	28.35	28.47	28.75
10	33.17	**33.34**	33.11	33.17	33.31

Table 1. Average PSNR values [dB] on 100 test images from the BSDS300 dataset. Note that MLP [3] is better than the blend of BM3D and EPLL proposed by Mosseri et al [11] at every noise level.

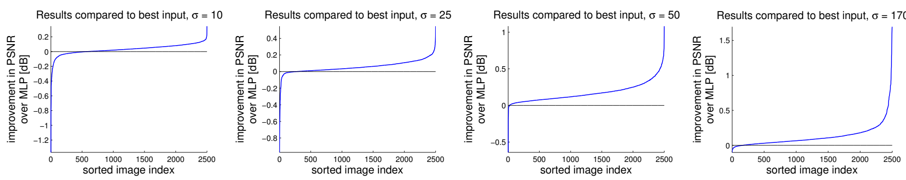
Table 2 compares our method to RTF-based methods [7] and to [11] on the dataset used in [7]. Our method achieves the highest PSNR also on these images. Note that it outperforms also RTF_{all} even though RTF_{all} combines four denoising methods, whereas we combine only two.

single methods	FoE [12]	BM3D [6]	EPLL [14]	LSSC [10]	$\text{RTF}_{\text{Plain}}$ [7]	MLP
PSNR	24.47dB	25.09dB	25.18dB	25.09dB	24.76dB	25.45dB

combining m.	RTF_{BM3D} [7]	RTF_{All} [7]	Mosseri [11]	our result
PSNR	25.38dB	25.51dB	25.30dB	**25.58**dB

Table 2. Results obtained with our approach and other methods on the dataset of images used in [7], with $\sigma = 50$. Top: Stand-alone methods, bottom: methods combining the results of other methods. Note that MLP outperforms both RTF_{BM3D} [7] and Mosseri et al. [11]. Our approach outperforms all competitors.

Performance profiles against the best input method. We now compare the results achieved with our method against the input methods (BM3D and MLP). For each image in a dataset of 2500 test images (that have not been used for training), we compare our method against the best of the two input methods (BM3D and MLP) for that image. Our method outperforms both BM3D and MLP on 76.92%, 89.12%, 96.92%, 99.12%, 98.8% and 93.48% of the images on the noise levels $\sigma = 10, 25, 35, 50, 75$, and 170, respectively. Figure 5 plots these results as performance profiles for four noise levels (more results in the supplementary material). Our method usually achieves results that are better than the best of the two inputs methods.



$\sigma = 10$ $\sigma = 25$ $\sigma = 50$ $\sigma = 170$

Fig. 5. Our results are better than the best of the two inputs in almost all cases.

Comparison against bounds. Tab. 3 compares our results against recently estimated bounds for image denoising [9]. Our proposed method that combines BM3D and MLP gets much closer to the bounds (last row). For $\sigma = 50$, half of the remaining possible gain over MLP is achieved. Note, that for noise levels $\sigma = 10$ and $\sigma = 25$, the bounds are difficult to estimate (the lower the noise, the more difficult). On these noise levels, our method achieves still better results than MLP, proving by examples that the limits are not yet reached.

	$\sigma = 10$	$\sigma = 25$	$\sigma = 35$	$\sigma = 50$	$\sigma = 75$	$\sigma = 170$
gain over BM3D by MLP [3]	0.07	0.3	0.33	0.34	0.38	2.19
gain over BM3D by our results	0.15	0.38	0.45	0.52	0.53	2.32
possible gain over BM3D [9]	–	–	0.6	0.7	1	–

Table 3. Improvements in dB over BM3D on 2500 test images.

6 Conclusion

Internal and external denoising approaches have complementary strengths and weaknesses. It has been previously claimed that external methods are preferred for patches with details, whereas internal methods are better for smooth patches. Our conclusions contradict previous findings: Internal methods are better on regions with regular, repeating structures. For irregular patterns, external methods are better. We have presented a simple patch-based method using neural networks that effectively combines the results of two denoising algorithms. The results surpass those of any previously published method. Bayesian patch-based bounds on image denoising have been estimated for medium to high noise levels, but are difficult to estimate at low noise levels. It was therefore not known if further improvements over BM3D at low noise levels were possible, but we have shown by example that improvements over BM3D were indeed possible at low noise levels.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing (TIP) 54(11), 4311–4322 (2006)
2. Buades, A., Coll, B., Morel, J.: A non-local algorithm for image denoising. In: International Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2, pp. 60–65. IEEE (2005)
3. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising with multi-layer perceptrons, part 1: comparison with existing algorithms and with bounds. arXiv:1211.1544 (2012)
4. Burger, H., Schuler, C., Harmeling, S.: Image denoising: Can plain neural networks compete with BM3D? Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2392–2399 (2012)
5. Chatterjee, P., Milanfar, P.: Is denoising dead? IEEE Transactions on Image Processing (TIP) 19(4), 895–911 (2010)
6. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing (TIP) 16(8), 2080–2095 (2007)
7. Jancsary, J., Nowozin, S., Rother, C.: Loss-specific training of non-parametric image restoration models: A new state of the art. In: European Conference of Computer Vision (ECCV). IEEE (2012)
8. LeCun, Y., Bottou, L., Orr, G., Müller, K.: Efficient backprop. Neural networks: Tricks of the trade pp. 546–546 (1998)
9. Levin, A., Nadler, B., Durand, F., Freeman, W.: Patch complexity, finite pixel correlations and optimal denoising. In: European Conference on Computer Vision (ECCV) (2012)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: IEEE 12th International Conference on Computer Vision (ICCV). pp. 2272–2279 (2010)
11. Mosseri, I., Zontak, M., Irani, M.: Combining the power of internal and external denoising. In: International Conference on Computational Photography (ICCP). pp. 1–9. IEEE (2013)
12. Schmidt, U., Gao, Q., Roth, S.: A generative perspective on mrfs in low-level vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1751–1758. IEEE (2010)
13. Zontak, M., Irani, M.: Internal statistics of a single natural image. In: International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 977–984. IEEE (2011)
14. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: International Conference on Computer Vision (ICCV). pp. 479–486. IEEE (2011)