Dear editor and referees,

Thank you very much for your comments and criticism. We found that we could answer all your questions and could also explain better why our claim was not straightforward.

As a result, we have rewritten the paper and hope that you will accept to review it again.

The authors

Review 1

This paper investigates the performance of a specific deep neural network for denoising. It is shown that normalizing the input to the neural network affects the results only slightly, which means it could be possible to learn a smaller network on a smaller, normalized data set.

Thank you, we fully agree with your summary of our paper.

The paper is well written and the results are interesting, especially for having a single neural network which works well across all noise levels. It is certainly worth trying to train a new neural network including the mentioned invariances.

However, judging from Figure 2, I don't think it's clear that a mean normalized network is guaranteed to be better, as stated in line 52 on page 3. While this may be the case, it could also be that the patch distributions vary slightly with different means. After normalization, the neural network could not learn this dependence anymore and would throw away prior information. I would recommend to weaken this statement or show that the mentioned variance dependent patch mean shift actually improves the performance.

We agree with your analysis and thus make it clear now that training with mean normalized patches was *a priori* risky because the conditional law that governs the patch variation given its mean varies with the mean.

In this case, mean quantification might be a viable alternative because one can reasonably expect the conditional laws being continuous with respect to their indexing means. To further confirm it, we added an experiment along with the following discussion:

> *Put differently, what we have shown so far is that the trained deep network is not very different from its most informative section. Hence, one may instead train a network on mean normalized patches*

> *to improve performance, thanks to a denser data distribution. However, in so doing, one implicitly trains on a marginal distribution and thus risks losing information. For lack of a complete probabilistic description of the natural patch distribution, the exact behavior of the patch law conditioned on the patch mean is elusive. Yet it can still be argued that they are continuous with respect to the patch means. The fact that the second moments of the laws indexed by the patch means not far away from each other are similar indicate just that. Note that here the test is not intended as a modeling attempt since the first and second moments are the sufficient statistics of the Gaussian law, known to be rather inadequate for describing the natural patches as a whole.*

Also, I would recommend to add labels to all axes in the figures.

All the figures are now labeled.

For future research the translational invariance could also be interesting (train a convolutional neural network instead of predicting overlapping patches).

Thank you for the suggestion, although we think that it may lead to an even larger network. Certainly, we may first try it on a smaller image to see how it fares.

Review 2

It is not well organized, so that the three algorithms presented appear unexpectedly, instead of be introduced from the very Introduction. There is not a neat justification of the three algorithms. Also, the experiments are not as well described as desirable. This lack of organization makes it difficult to evaluate the paper's degree of novelty.

We now make it clear that our main claim is Algorithm 2, which makes a single existing network work well across all levels of Gaussian noise, while Algorithm 1 is a test that validates our preliminary analysis on the natural patch distribution. Algorithm 3 is now removed, which was intended to support the other minor claim that the neural network has acquired a certain symmetry that may explain its enormous size.

The wording and organisation of the experiments have also been reviewed. In addition, since Fig.2 used in the first submission might be misleading on the actual RMSE gap between our proposed generic network and the dedicated one, we rescaled them so as to emphasize that this difference is almost negligible, even more so in view of the noise levels applied. This difference is less than 0.1 for noise standard deviation as high as 75.

Although introduced as a way of reducing redundancy in deep networks, the method described in the paper is basically a patch normalization. I don't think this is contribution worth publishing in SPL.

Our claim, which is now reformulated, should have been that, through an investigation of the distribution invariance of the natural image patches, we succeeded in making a single existing neural network work well across all levels of Gaussian noise. As we now highlight in the paper, the present contribution solves what was explicitly stated as a major difficulty in the commented paper *Image denoising: Can plain neural networks compete with BM3D?* (Burger, Schuler, and Harmeling 2012) Here is the quote:

> *Our most competitive MLP is tailored to a single level of noise and does not generalize well to other noise levels compared to other denoising methods. This is a serious limitation which we already tried to overcome with an MLP trained on several noise levels. However, the latter does not yet achieve the same performance for $\sigma = 25$ as the specialized MLP.*

The proposed normalization may look simple *a posteriori*, but it was far from obvious that it could work. To the best of our knowledge, this has not been proposed elsewhere and thus constitutes a novel solution to the posed problem. We believe that the result is very relevant for future research in image denoising with neural networks. The method is easy to explain, hence its submission as a short letter.

Finally, it is not clear how the computational advantages claimed in the introduction are finally achieved. The reader expects a set of experiments in which large networks are scaled down without losing accuracy thanks to the presented normalization steps, but I don't find such experiments.

For the image denoising task, we reduced the number of neural networks required of a general purpose denoising algorithm down to just one. This is an effective redundancy reduction. As a result, it suffices to train just one neural network, which we did not do because there was a well trained one available. Algorithmically speaking, our result also means that the training time for a generic neural network based algorithm can be divided by at least a factor of ten (if we consider noise standard deviations ranging from 5 to 50 for instance), which is certainly not negligible in view of what is stated in (Burger 2013)

> *Each experiment is the result of many days and sometimes even weeks of computation time on a modern GPU (we used nVidia's C2050).*

> *We used nVidia's C2050 GPU and achieved a speed-up factor of more than one order of magnitude compared to an implementation on a quad-core CPU.*

For the same reason, we decided to postpone the experiments of possible neural network domain reduction for a later work.

I don't think a paper dealing with such a hot topic as image denoising can have only 9 references.

We have added six more references, including those related to denoising algorithms that use both global and local learning, and some theoretical background of vector quantification and sufficient statistics.

References

Burger, H. C. 2013. Modelling and Learning Approaches to Image Denoising.

Burger, H., C. Schuler, and S. Harmeling. 2012. Image denoising: Can plain Neural Networks compete with BM3D?. *Computer Vision and Pattern Recognition.*