Deep convolutional neural network for image denoising in the medical field
Deep Learning - Final Project Report

Arthur Calvi
École normale supérieure Paris-Saclay
4, avenue des Sciences, 91190 Gif-sur-Yvette
arthur.calvi@ens-paris-saclay.fr

Benoit Roussel
École normale supérieure Paris-Saclay
4, avenue des Sciences, 91190 Gif-sur-Yvette
benoit.roussel@ens-paris-saclay.fr

Abstract

Medical and biological images like any other form of imaging techniques are susceptible to noise and artefacts. The presence of noise makes the images unclear and may perplex the identification and analysis of diseases or entity. Hence, denoising of images is a mandatory and essential pre-processing technique for further image processing stages. Deep learning algorithms are very effective at denoising but conventionally needed both clean and noisy images. Even when possible, collecting clean samples remains a daunting task. However, recent advances in self supervised learning alleviate the use of deep learning algorithms by allowing to train them only on noisy data. In this project, these self-supervised technics are investigated to denoise medical or biological image data such as X-ray, MRI and Microscopy.

1. Introduction

The major drawback of using deep learning algorithm for denoising is the collection of clean samples. In a lot of imagery domains, it remains a difficult or even impossible task. One common technic to obtain decent clean samples is to average several images to reduce the noise level. However, for numerous experiments, either the acquisition of the image may damage the entity observed or the entity is moving or deforming itself, making the averaging ineffective. Hence, the traditional supervised training appears to be inadequate for those cases.

Nevertheless, self supervised technics aim to reach the same performance as supervised ones only by looking at noisy data. This training scheme removes the collection issue of clean samples and widen the application scope of deep learning technics. In this project, we investigate the capability of these technics to denoise MRI, X-Rays and microscopy images.

2. Problem definition

Nowadays, self-supervised technics have not been yet largely tested nor trained on real noisy data. The most performant algorithms have been trained on artificial noises added to clean data. Popular noises are Gaussian or Poissonian as they can model the principle noise component for numerous type of acquisition. However, real noise is always more complex. One other weak point of self supervised technics is often relying on some strong assumptions such as i.i.d noise or even constraints such as explicitly describing the noise model to compute the loss and the estimation of the clean image [8]. These points are limiting the use of this kind of technics for real data.

In this report: x and y respectively denote the clean and noisy pixel value, Ω_y denotes the noisy context which is the receptive field without the central pixel. The challenge with self supervised learning is to learn to output the clean pixel x just at looking at the noisy pixel y and its surrounding Ω_y. For some algorithms, the estimation of the clean pixel is help by specifying a noise model $p(y|x)$. This means that one trained model is limited to a specific noise.

3. Related work

Firstly, we partially reproduced the ablation study of the Bayesian inference of the SSDN algorithm [8] and extend it to two other datasets containing MRI and X-rays images. This study shows the advantages of using the information of the noisy pixel y at test time. Secondly, we evaluated several configurations of the SSDN algorithms on real noisy data. This experiment illustrates the capacity of the algorithm to generalize to more complex noise. In addition, this experiment also points out the weaknesses of the algorithm to treat real and complex noises.

4. Methodology

The purpose of this project is to investigate the usage of self supervised technics to denoise real data. The complex-

ity of the project is to firstly find medical/biological datasets with clean and noisy pairs to compare self-supervised and supervised technics. Secondly, those datasets should differ from their noise complexity in order to assess the generalization of models to more complex noises. Finally, all self-supervised algorithms have to be modified to take into account already noisy images during evaluation and perhaps training. In addition, a more complex noise will be eventually implemented.

4.1. Self-supervised denoising

In this project, we will mainly focus on the SSDN algorithm, but we will also compare it with other self supervised technics such as noise2noise and noise2void.

Self-supervised deep denoising (ssdn) : This algorithm uses the same noisy image as input and target. To avoid learning the identity during the training, the central pixel of the receptive field is masked; this technique is called the blindspot network. The network learn to generate a Multivariate Gaussian $p(x|\Omega_y) \sim \mathcal{N}(\mu_x, C_x)$ for each pixel by minimizing the negative log likelihood of the training data $p(y|\Omega_y)$ which depends on both the prior $p(x|\Omega_y)$ and the noise model $(p(y|x)$.

$$\underbrace{p(y|\Omega_y)}_{\text{Training data}} = \int \underbrace{p(y|x)}_{\text{Noise model}} \underbrace{p(x|\Omega_y)}_{\text{Prior}} dx \tag{1}$$

Namely, the network is learning to map a context Ω_y to a clean pixel parametrized as a multivariate Gaussian (μ_x, C_x). In addition, to use the noisy measurement y of pixel x, a Bayesian inference is performed. In practice, it is done by using the posterior mean $\mathbb{E}_x[p(x|y, \Omega_y)]$ which maximizes PSNR [8].

For instance, in the case of a Gaussian noise model : $p(y|x) \sim \mathcal{N}(0, \sigma^2 I)$. We want to describe $p(y|\Omega_y) \sim \mathcal{N}(\mu_y, \Sigma_y)$ with eq 4.1. A convolution between two independant Guassian distributions gives the following relationships : $\mu_y = \mu_x$ and $\Sigma_y = \Sigma_x + \sigma^2 I$. The loss which is the negative log likelihood of $p(y|\Omega_y)$ is then given by :

$$L(y, \mu_y, \Sigma_y) = \frac{1}{2}[(y - \mu_y)_y^{T-1}(y - \mu_y)] + \frac{1}{2}\log|\Sigma_y| + C$$

Finally at test time the post estimate has to be computed :

$$\mathbb{E}_x[p(x|y, \Omega_y) = (\Sigma_x^{-1} +^{-2} I)^{-1}(\Sigma_x^{-1}\mu_x +^{-2} y)^{-1}$$

The two critical steps of this algorithm heavily depend on the noise model description.

Noise-2-noise (N2N) [9] also allows to cope without clean ground truth training data. It starts out with noisy image pairs.

$$x^j = s^j + n^j \quad and \quad x^{'j} = s^j + n^{'j} \tag{1}$$

that is, the two training images are identical up to their noise components n^j and $n^{'j}$, which are, in the image generation model, just two independent samples from the following same distribution :

$$p(n|s) = \prod_i p(n_i|s_i) \tag{2}$$

Hence, a convolutional neural network can be trained and fine-tuned with a pair of noisy images, by selecting one as the groundtruth and the other as in the input image. Even though, it attempts to learn a mapping from a noisy input to a noisy target, the training will still converge to the correct solution. The key to this phenomenon lies in the fact that the expected value of the noisy input is equal to the clean signal.

Noise-2-Void (N2V) [6] go a step further by proposing to derive both parts of our training sample, the input and the target, from a single noisy training image x^j. If it were to simply extract a patch as input and use its center pixel as target, the network would just learn the identity, by directly mapping the value at the center of the input patch to the output. In practice, the blind spot corresponding to the center pixel of a patch is replaced with a random selected value from the surrounding area.

4.2. Dataset description

This NIH Chest X-ray Dataset [10, 1] contains 112,120 X-ray images with disease labels from 30,805 unique patients. To create these labels, the authors used Natural Language Processing to text-mine disease classifications from the associated radiological reports. The labels are expected to be over 90% accurate and suitable for weakly-supervised learning. It is advantageous to use Chest X-ray exams, because they are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging and the lack of large publicly available datasets with annotations means it is still very difficult, if not impossible, to achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites with chest X-rays.

The energy deposited at a pixel location (the image intensity) does not correspond exactly to the number of captured monochromatic photons, since the X-rays follow a compound Poisson noise process. Whereas the number of X-rays follows a Poisson noise distribution, the X-ray energy converted counts follow a compound Poisson noise distribution due to the wide spectrum of the energy.

The Radiological Society of North America (RSNA) BraTS dataset [3, 2] describes a retrospective collection of brain tumor mpMRI scans acquired from multiple different institutions under standard clinical conditions, but with different equipment and imaging protocols, resulting in a heterogeneous image quality reflecting diverse clinical practice across different institutions. Inclusion criteria comprised pathologically confirmed diagnosis and available MGMT promoter methylation status. These data contains a total number of cases from 660 to 2,000.

The image intensity in Magnetic Resonance magnitude Images (MRI) in the presence of noise is shown to be governed by a Rician distribution. Low signal intensities (Signal-to-Noise Ratio (SNR) ¡ 2) are therefore biased due to the noise. The noise characteristics in phase images are also studied and shown to be very different from those of the magnitude images. Common to both, however, is that the noise distributions are nearly Gaussian for SNR larger than two.

The Fluorescence Microscopy Dataset [13] consists of 12,000 real fluorescence microscopy images obtained with commercial confocal, two-photon, and wide-field microscopes and representative biological samples such as cells, zebrafish, and mouse brain tissues. By averaging raw data, authors obtained ground truths and several noise levels.

Images are dominated by Poissonian noise because of the weak number of photons captured by a microscopic detector (PMT or CCD). The microscopy imaging system is modeled with a Poisson-Gaussian noise model [4]. The model is composed of a Poisson noise component that accounts for the signal-dependent uncertainty, i.e., shot noise, and an additive Gaussian noise component which represents the signal independent uncertainty such as thermal noise.

4.3. Metrics

All the papers cited use the PSNR metric to rate the denoising. We decided to complete the study by using the SSIM metric [11] which aim at giving a better indication of the visual result. This index is based on three key components : the contrast, the luminance and the structure of the image. The index ranges between 0 (very different) and 1 (perfect similarity). As it is bounded, it allows an easier interpretation.

PSNR : The term peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation.

$$PSNR = 20log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right)$$

with :

$$MSE = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1}\|f(i,j) - g(i,j)\|^2$$

SSIM : The Structural Similarity Index (SSIM) is a perceptual metric that quantifies image quality degradation caused by processing such as data compression or by losses in data transmission. It is a full reference metric that requires two images from the same image capture. Unlike PSNR (Peak Signal-to-Noise Ratio), SSIM is based on visible structures in the image.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\mu_x y + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

With $c1$ and $c2$ two variables to stabilize the division with weak denominator.

5. Evaluation

Firstly, results of the SSDN paper [8] are reproduced to ensure that the official implementation [7] that we will use for this report is working correctly. Precisely, the evaluation on Set14, BSDS300 and Kodak datasets with synthetic white Gaussian noise with $\sigma = 25$ is performed. Results are shown in table 1. Three configurations have been tested : with σ known, σ evaluated as a parameter during training and μ only which corresponds to ssdn without the Bayesian inference.

method	PSNR
ssdn (σ known)	31.52 dB
ssdn (σ eval)	31.50 dB
ssdn (μ only)	29.54 dB

Table 1. Averaged PNSR on Set14, Kodak and BSD300 for a Gaussian noise ($\sigma = 25$)

PSNR scores on Set14, Kodak and BSDS300 are exactly the sames as in the article.

5.1. Artificial noise

The first point of this study is to test the three algorithms on medical datasets with artificial noises with 1k images. Even if the model has been trained on the right noise model, particularities of MRI and X-rays imaging (lot of dark areas) may complicate the task for the algorithm. Noise in X-rays radiography will be a Poisson distribution. On the contrary, noise in MRI will be approximated as a Gaussian distribution. Those datasets are used to compare the SSDN algorithm to another well known self-supervised technic : noise-2-noise and a supervised one (noise-2-clean). The sigma for the artificial Poisson noise is equal to 25 and the sigma for artificial the Gaussian noise is equal to 30. On the

figure 1 and on the figure 2 can be observed the qualitative results of the algorithms :

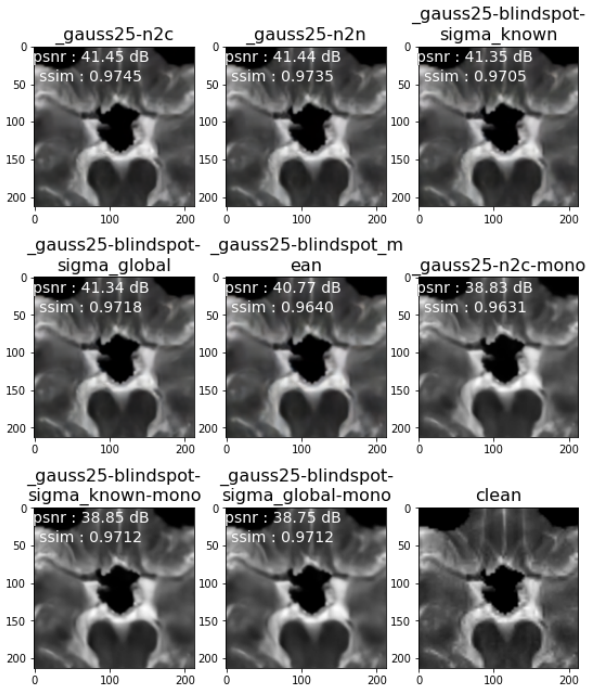

Figure 1. Zoom in the same image of the RSNA Brain tumor dataset denoised with different algorithms

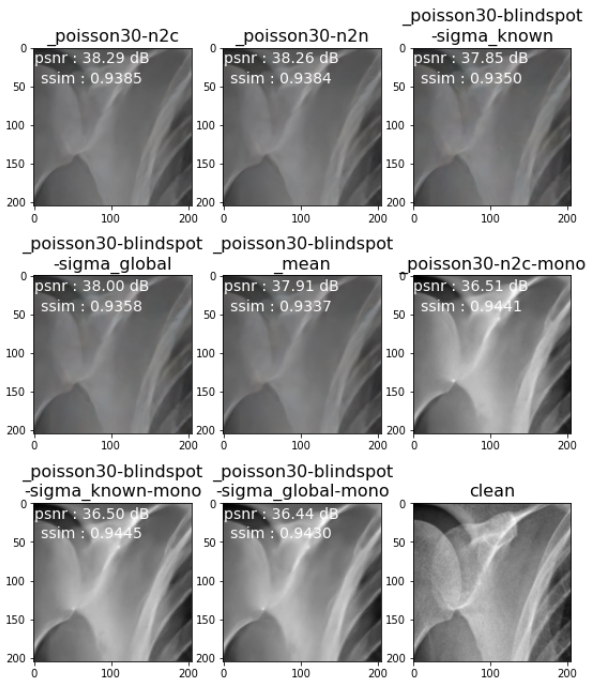

Figure 2. Zoom in the same image of the NIH Chest X-ray dataset denoised with different algorithms

There is a distinct difference between algorithm in mono and not, and in particular on the brightness of the image. The mono architecture gives an image that is brighter and more accurate compared to the clean image. Nevertheless, mono images seem to have smooth too much certain parts and borders of the image that makes the border between two bones impossible to distinguish.

And as quantitative results, the table 2 and 3 below resumes respectively PSNR and SSIM results for the different architectures for the different datasets :

Method	NIH (PSNR)	RSNA (PSNR)
n2c	39.2381	41.8208
n2n	39.2183	41.7935
ssdn - sigma known	38.6657	41.606
ssdn - sigma global	38.8456	41.6313
ssdn - sigma mean	38.9058	40.3533
n2c - mono	37.6943	39.2137
ssdn - sigma known mono	37.6689	39.1999
ssdn - sigma global mono	37.6171	39.1118

Table 2. Average PSNR on 1k samples of NIH Chest X-rays and RSNA Brain tumor datasets

The PSNR between both datasets cannot be compared, but it can be used to compare methods on the same dataset. For example, ssdn algorithms have worse results than n2c or n2n, but not that far. Qualitative results seem very close when comparing those.

Method	NIH (SSIM)	RSNA (SSIM)
n2c	0.9423	0.9745
n2n	0.9424	0.9721
ssdn - sigma known	0.9384	0.9672
ssdn - sigma mean	0.9387	0.9597

Table 3. Average SSIM on 1k samples of NIH Chest X-rays and RSNA Brain tumor datasets

The SSIM is a better metric to compare between datasets. As it can be observedn the SSIM seems better for the RSNA dataset and for the Gaussian model.

Although ssdn does not have equal or better performance than n2c or n2n, ssdn has the undeniable advantage to use only a single noisy image to process it. That's why ssdn can be considered as a strong algorithm, considering its close performance with other more complex ones.

5.2. Real noise

The second step of this study is to investigate the generalization of self supervised algorithms trained on standard photographs with artificial noises to Microscopy images with real Poisonian-Gaussian noise. The purpose is to show the limitations of models trained on artificial noises. In total, 11 variants of the SSDN algorithms have been tested on the four different levels of the Fluorescence Microscopy

Dataset (FMD), they are also compared to the ssdn baseline with noise-2-clean and noise-2-noise training. All the dataset have not been denoised, 48 representative samples are used for the tests. The model are referenced with the noise model used for training : either Poisson ($\lambda = 30$) or Gauss ($\sigma = 25$); the ablation used : None, *diag* meaning that the Covariance matrix of the distribution $p(x|\Omega_y)$ is enforced to be diagonal (this setup corresponds to treating each color component of the prior as a univariate, independent distribution) *mean* which ablates the Bayesian inference (This variant is similar to the noise-2-void algorithm) ; and the noise parameter estimation fashion : *known* means that the parameter σ or λ is fixed in advance, *global* means that the noise parameter is estimated during training as a constant over all the dataset by the network. The last parameter is the number of channel of the image : *mono* means that the network is working on grayscale images. All the images of the dataset are grayscale images, but networks trained on RGB images have also been tested. They sometimes have remarkable results. In addition, the noise levels are referred as *fmdS* where *S* is the number of images used for the averaging.

5.2.1 Global results

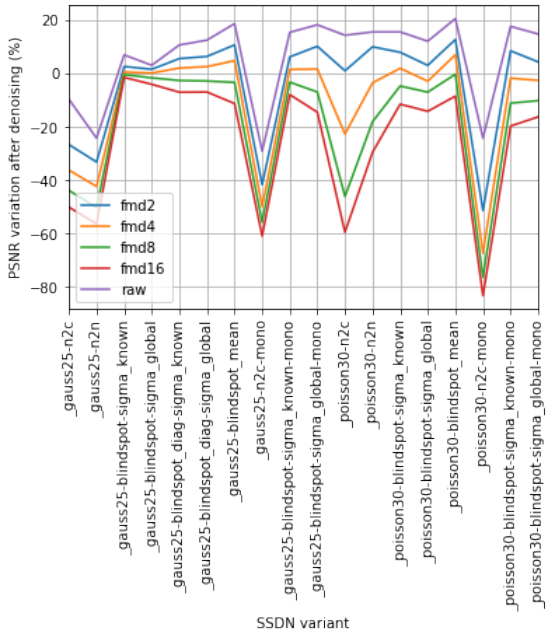

Figure 3. PSNR of ssdn variants measured after denoising.

Figures 3 & 4 show the PSNR and SSIM variation after denoising of each SSDN variant. Instead of plotting the PSNR and SSIM we decided to plot the evolution of these indexes after the denoising, the formula used is : $PSNR_{var} = \frac{PNSR_{denoised} - PNSR_{noisy}}{PNSR_{noisy}}$, where $PNSR_{noisy}$ is the PSNR of the noisy image w.r.t the clean one and $PNSR_{denoised}$ is the PSNR of the output of the network w.r.t to the clean one, same for the computation of $SSIM_{var}$. Since, we are comparing several noise levels, this metric make more sens because if we only diplay the PSNR and SSIM, we will just demonstrate that the scores are better for lower noise levels. However, we want to assess the improvement after denoising.

Surprisingly, the baselines trained with the noise-2-clean (n2c) scheme systematically obtained the lowest PSNR on real noises while they obtained the best results on artificial ones; except for the networks trained on RGB images with a Poisson noise. As the Poissonian component is dominant on those images, this result simply asserts that supervised networks trained with a noise model near to the real one are performing better. However, as supervised technics strongly rely on data, they easily specialize to the artificial noise used during training and are also more disposed to learn the underlying bias of this noise and then generalize poorly. In addition, both figures 3 & 4 demonstrates that when more than 4 images are used for the averaging the networks are not able to improve the image, the opposite in fact. Notes that as Poisson-Gaussian random processes are independent, averaging the images results in reducing the variance of the noise by the factor *S* which is the number of images used for the averaging. Therefore, certain noise levels should be closer than other to the artificial noise used to train networks. This may partially explain why models are having better PSNR and SSIM variations on RAW data. Given the results, self-supervised technics are generalizing better on average. Networks trained on grayscale images have among the best PSNR variations, but interestingly the best scores are obtained when the Bayesian inference is ablated. This might be due to the incomplete model used to describe the noise distribution : the Bayesian inference is therefore bringing wrong information that can mislead the denoising of the image. The exact same remarks are valid for figure 4 which presents the SSIM variation after denoising.

Also, networks that have to estimate the noise parameter (σ or λ) have lower scores than for networks where the noise parameter is fixed. Authors of the SSDN paper [8] also provide pre-trained networks trained on Gaussian and Poissonian noise but with noise parameter varying in a certain range, at test time the model uses an auxiliary network to estimate the noise parameter per image. We tested these models since they should give better results because they are more adaptable. Unfortunately, as the noise model is not matching the real one, the auxiliary network were unable to retrieve a good parameter and the algorithms completely failed to denoise the image.

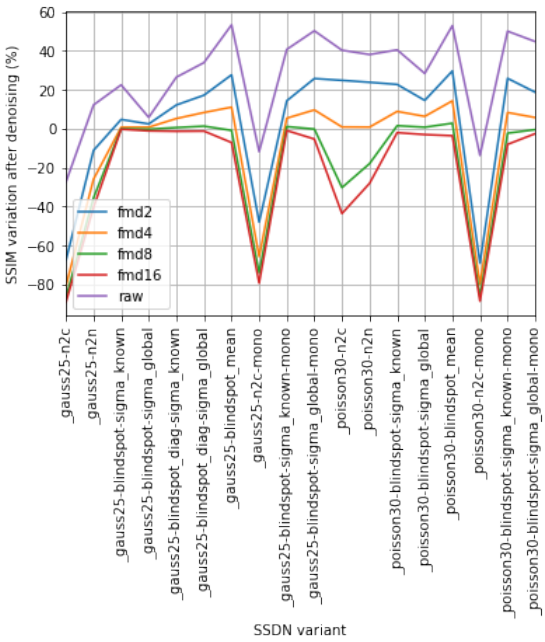

Figure 4. SSIM of ssdn variants measured after denoising.

5.2.2 Focusing on some images

Figure 5 illustrates the results of each SSDN variants trained on a Gaussian noise. The best results are obtained for the configuration where σ is fixed (*known*). The model trained on RGB images (image 3) seems to better preserve the sharpness of the image, but at a lower intensity. As contrary, the model trained on grayscale images (image 9) retains the right intensity but partially denoise the image. Chromatically correlated artifacts appears for some models. This is due to the application of a technic that is specialized to a certain type of noise to another incompatible noise, for self-supervised algorithms it might be due to the Bayesian inference which bring incorrect information and for supervised ones it is surely due to the RGB correlation bias learned during training.

On figure 6, the same artifacts can be observed. One can also note the presence of strong artifacts for the n2c baseline trained with grayscale images. This time, the networks trained on grayscale images with a λ evaluated and the RGB model with the Bayesian inference ablated are giving the best scores. One striking result is that the area denoised for the grayscale models with Poisson and Gaussian noise are not the same. This induces that the noise is spatially correlated. In fact, the Poissonian component is signal dependent, but the Gaussian part is not. Besides, the noise model can be described by a Heteroscedastic Gaussian, see section 5.3.

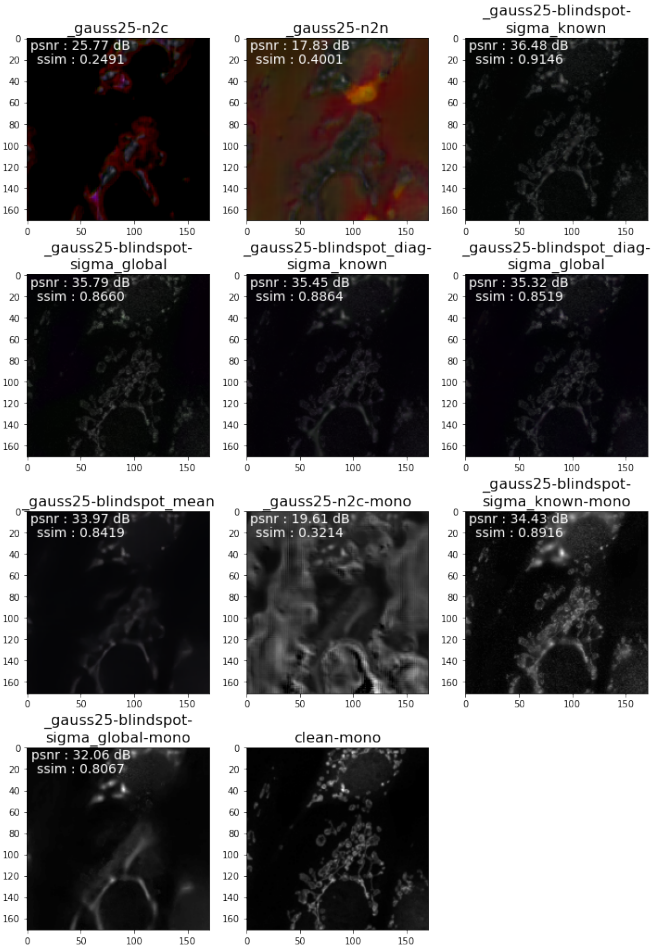

Figure 5. SSDN variants trained with Gauss ($\sigma = 25$) noise model. Zoom x2 of Confocal BPAE image. Fmd2 noise level.

Figure 7 shows the denoising results of a network trained on Poissonian noise and with λ estimated at test time by the network on two different noise level. The second raw of the figure shows the corresponding noisy images. This figure illustrates that the network does not enhance the image for the fmd16 level, the opposite in fact. However, the network performs well when only two images are used for the averaging.

Figure 8 demonstrates the same phenomenon. However, white and gray lines are disappearing in some areas. This type of error when denoising could easily mislead the diagnosis linked to the analysis of the image.

Finally, figure 9 presents the comparison on RAW data of two networks where the noise parameter is estimated and trained either with a Gaussian or a Poissonian noise. In both case, significant improvement are visible after denoising, but noise remain in the image. The network trained on Gaussian noise is doing better, but a lot of information is lost. This study on real noise shows that networks trained on artificial noise can generalize to other noise types but are

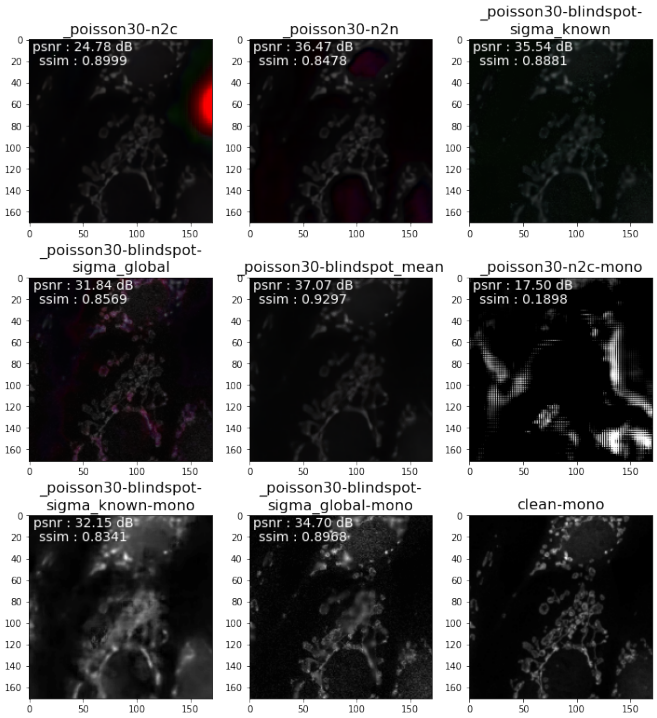

Figure 6. SSDN variants trained with Poisson ($\lambda = 30$) noise model. Zoom x2 of Confocal BPAE image. Fmd2 noise level.

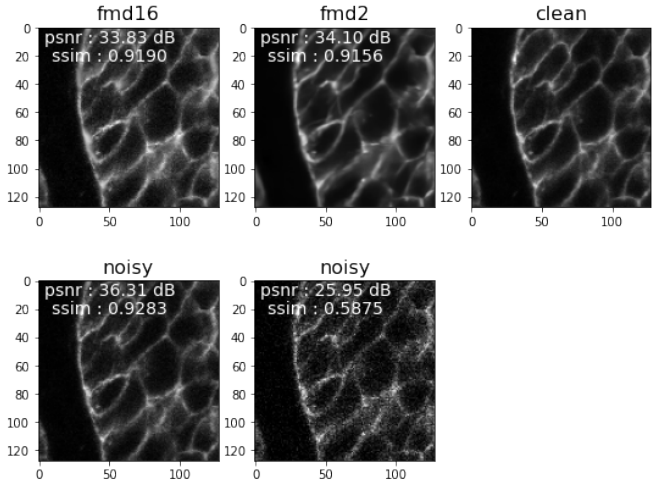

Figure 7. SSDN with Poisson (λ evaluated) noise model. Zoom x4 of Confocal fish image

limited, especially for medical imaging, where information is crucial to analyze and interpret the images.

Another interesting experience would have been to finetune a pretrained-model on real noisy data to firstly see if the ssdn algorithm converge even if the noise model does not match the one in the images. Likewise, the implementation of the Heteroscedastic Gaussian as a noise model in SSDN should have been intriguing, one paper have already

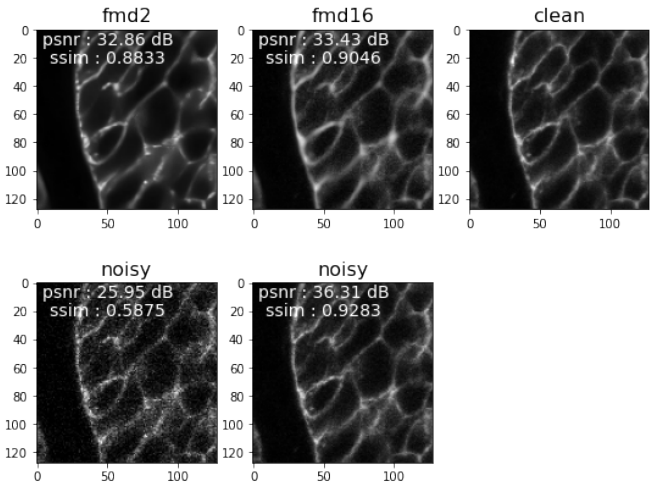

Figure 8. SSDN with Gauss (σ evaluated) noise model. Zoom x4 of Confocal fish image

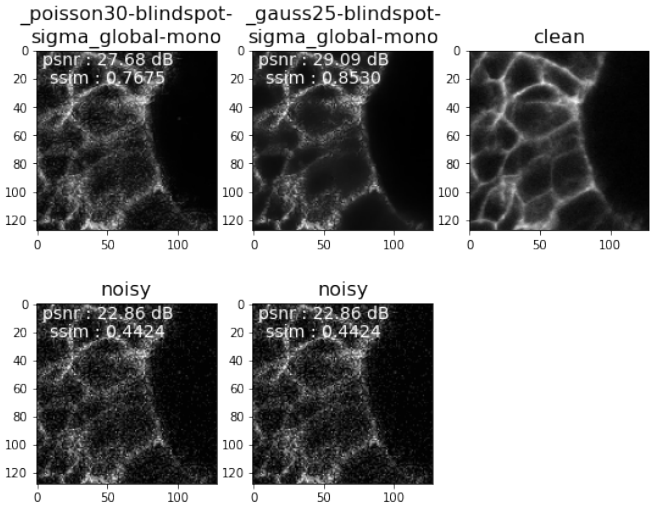

Figure 9. Poisson and Gauss noise model comparison on raw data. Zoom x4 of Confocal fish image.

performed this implementation and obtained excellent results on the FMD dataset [5]. Unfortunately, we did not manage to modify the code in time to train it with real noisy data.

5.3. A further noise modelisation

The noise model which can be implemented for the FMD dataset is the following : the noisy observation y_i is sampled by first applying Poisson corruption to x_i and then adding Gaussian noise which is independent of x_i :

$$y_i = aP(x_i/a) + N(0, b)$$

where $a > 0$ is a scaling factor (related to the gain of the camera) and b is the variance of the Gaussian noise compo-

nent, which models other sources of noise such as electric and thermal noise.

An approximation of the Poisson distribution as Gaussian with equal mean and variance can be applied :

$$\begin{aligned} y_i &\approx aN(x_i/a, x_i/a) + N(0, b) \\ &= N(x_i, ax_i + b) \end{aligned}$$

The noise model is then simply a Gaussian noise model whose variance is an affine transformation of the clean value. Note that in practice, we allow b to be negative; this models the effect of an offset or "pedestal" value in the imaging system. This general formulation encompasses both pure Gaussian $(a = 0)$ and Poisson noise $(b = 0)$.

6. Conclusion

Self-supervised denoising algorithms with explicit noise model may be suited for medical images : the noise model of imagery machines are often well known, and clean data are very scarce or time-consuming to obtain. However, these technics have to be enhanced with more complete noise models and a better parameter estimation fashion. To deal with spatially variant noise, ideas like noise level map proposed in the advanced supervised algorithm FFDnet [12] could be investigated and extended to self-supervised ones. Finally, given the relatively good generalization results of self supervised technics and also the scarcity of real noisy dataset in the medical field, the training of these algorithms on more complex artificial noises on nonmedical images could be considered.

References

[1] nih-chest-xrays. https://www.kaggle.com/nih-chest-xrays/data, 2018. 2

[2] rsna-miccai-brain-tumor-radiogenomic-classification. ://www.kaggle.com/c/rsna-miccai-brain-tumor-radiogenomic-classification, 2021. 3

[3] Ujjwal Baid et al. The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *CoRR*, abs/2107.02314, 2021. 3

[4] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 17:1737–54, 11 2008. 3

[5] Wesley Khademi, Sonia Rao, Clare Minnerath, Guy Hagen, and Jonathan Ventura. Self-supervised poisson-gaussian denoising, 2020. 7

[6] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images, 2019. 2

[7] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. Self-supervised denoising. https://github.com/NVlabs/selfsupervised-denoising, 2019. 3

[8] Samuli Laine, Jaakko Lehtinen, and Timo Aila. Self-supervised deep image denoising. *CoRR*, abs/1901.10277, 2019. 1, 2, 3, 5

[9] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *CoRR*, abs/1803.04189, 2018. 2

[10] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017. 2

[11] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 3

[12] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, Sep 2018. 8

[13] Yide Zhang, Yinhao Zhu, Evan L. Nichols, Qingfei Wang, Siyuan Zhang, Cody J. Smith, and Scott S. Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. *CoRR*, abs/1812.10366, 2018. 3