

Physics-informed loss function for a deep neural network for Poissonian noise reduction

Emmanouil Xypakis (✉ emmxypakis@gmail.com)

Center for Life Nano- and Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy

Valeria De Turis

Center for Life Nano- and Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy

Fabrizio Gala

Crestoptics S.p.A. Italy

Giancarlo Ruocco

Center for Life Nano- and Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy

Marco Leonetti

Soft and Living Matter Laboratory, Institute of Nanotechnology, Consiglio Nazionale delle Ricerche, 00185, Rome, Italy

Research Article

Keywords: machine learning, microscopy, denoising, Poisson distribution

Posted Date: September 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2049014/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Physics-informed loss function for a deep neural network for Poissonian noise reduction

Emmanouil Xypakis [1,2], Valeria deTuris[1], Fabrizio Gala[3], Giancarlo Ruocco[1], Marco Leonetti[1,2,4]

We developed a physics-informed deep neural network architecture able to achieve signal to noise ratio improvements starting from low exposure noisy data. Our model is based on the nature of the photon detection process characterized by a Poisson probability distribution which we included in the training loss function. Our approach surpasses previous algorithms performance for microscopy data, moreover, the generality of the physical concepts employed here, makes it readily exportable to any imaging context.

In any optical and non-optical imaging technologies, measurement comes with a noise addition producing a signal that follows a Poisson probability distribution (PPD). Signal enhancement algorithms increase the amount of information by increasing the signal to noise ratio (SNR), making them useful for modeling and visualization of biological data including: microscopy images representing light; medical imaging; computer tomography; positron emission tomography and other in-vivo imaging technologies.

Deep neural networks (DNNs)¹⁻⁷ based algorithms achieve the best results in signal enhancement. The performance and the ability to train DNNs, however, depends both on the chosen loss function—a quantity comparing predictions and target that DNN minimizes to learn its internal parameter—but also on the normalization of the network inputs and targets⁸. Two popular used loss functions for denoising and other image enhancement tasks are the L1-norm and the L2-norm (MSE)^{2,4,9,10}, where the data are arbitrary normalized. On the other hand, when the desired output comes from a known probability distribution like in semantic segmentation (U-net¹¹), or other classification tasks, an entropic loss function, and a probabilistic normalization of data, is of great significance. Although accounting the physics of the camera detection process is known to significantly improve imaging efficiency¹², little research has been done in applying this physical properties in DNNs.

Physics-informed machine learning is a new trend in artificial intelligent¹³ and encoding the physics and statistics of light in a DNN is a missing puzzle. Here we report about physically informed DNN which builds on the PPD of signal detections. Our approach aims to provide a general and exportable approach to deal with Poisson distributed signals: I) we use a non-arbitrary and physics-based normalization process, II) we employ a physically informed loss function, and III) we design DNN architecture which takes advantage of the previous features. First, we remove any arbitrariness on the normalization just working with images in which each pixel count represents the photon number. Then we design a loss function which considers distance between probability distributions instead of the distance between count number (we employ the Kullback-Leibler divergence (KL)¹⁴, see also online-methods). This enables the algorithm to work with the same efficiency in all dynamic range windows. We employ a

custom DNN architecture capable of classifying each pixel on a predicted photon number, thus preserving the photon number encoding and meaning for the output images. Our custom DNN employs structures from RCAN employed previously for denoising, and U-net which is employed for semantic segmentation.

We first illustrate how the PPD information can be encoded in the DNN architecture in Fig.1. In panel A) we review the detection process of a PPD signal. The sample emits photons from volume V in all direction and a fraction of them, N_{ph} , arrive at the detector. It is of great importance to characterize the signal as the photon numbers (see Methods and ¹²), by calibrating the detector with an efficiency g plus an offset. The detection of photons follows a PPD of average μ which depends on the detector exposure τ as $\mu = \tau V \rho$, while the variance is $\sigma = \sqrt{\mu}$. Thus, the SNR for each voxel scales as $1/\sqrt{\mu}$ and as an effect has a different accuracy in the detection (see panel B) and the reconstruction of the density (see panel C). In Fig 1. Panel D we plot two different DNN Loss functions, that quantify the DNN output W with the GT mean value μ as a function of the PPD relative error $(\mu - W)/\sigma$: the KL divergence (ours, see Methods), and the MSE (other state-of-the-art DNN). The difference between the two losses is that MSE loss strongly depends on the GT mean value while KL is only affected at large relative errors; the curves are diverging for different μ . In this way, the KL loss gives the same relative loss at the full dynamic range and does not penalize the absolute different between the GT and the output of the DNN; a small absolute difference between the GT and DNN gives a large Loss value for small μ of the GT and a big loss for large μ . In the literature KL loss has been used for classification. Here, we identified the classes as the number of photons. In panel E) we show different architectures of the DNNs. A state-of-the-art architecture that is used for classification is the U-net, that uses a fine-tuned contracted and expansive path between convolutional blocks, while RCAN employs the residual convolutional blocks to perform denoising. We employ the ResUnet, which differs from the U-net only in the fact that instead of having convolutional blocks has residual blocks and that uses the KL loss modified for photon classification. We train both the RCAN, and ResUnet in the same dataset (DS1) and evaluate in a different dataset (DS2). DS1 and DS2 are both generated by experimental data for confocal fluorescence microscopy but differ in: a) the camera; b) the fluorophores; c) the objective; d) the biological samples (See Methods). In panel F) we show that ResUnet produce artifact free reconstructions when exported to a different imaging condition.

In Figure 2 we further quantify the performance of ResUnet and compare it with RCAN and the NLM classical denoising algorithm. In panel A) we show the ResUnet, RCAN (trained in DS1 and evaluated in DS2) and NLM denoised images for

1. Center for Life Nano- and Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy

2. D-TAILS srl, 00161, Rome, Italy

3. Crestoptics S.p.A., Italy

4. Soft and Living Matter Laboratory, Institute of Nanotechnology, Consiglio Nazionale delle Ricerche, 00185, Rome, Italy

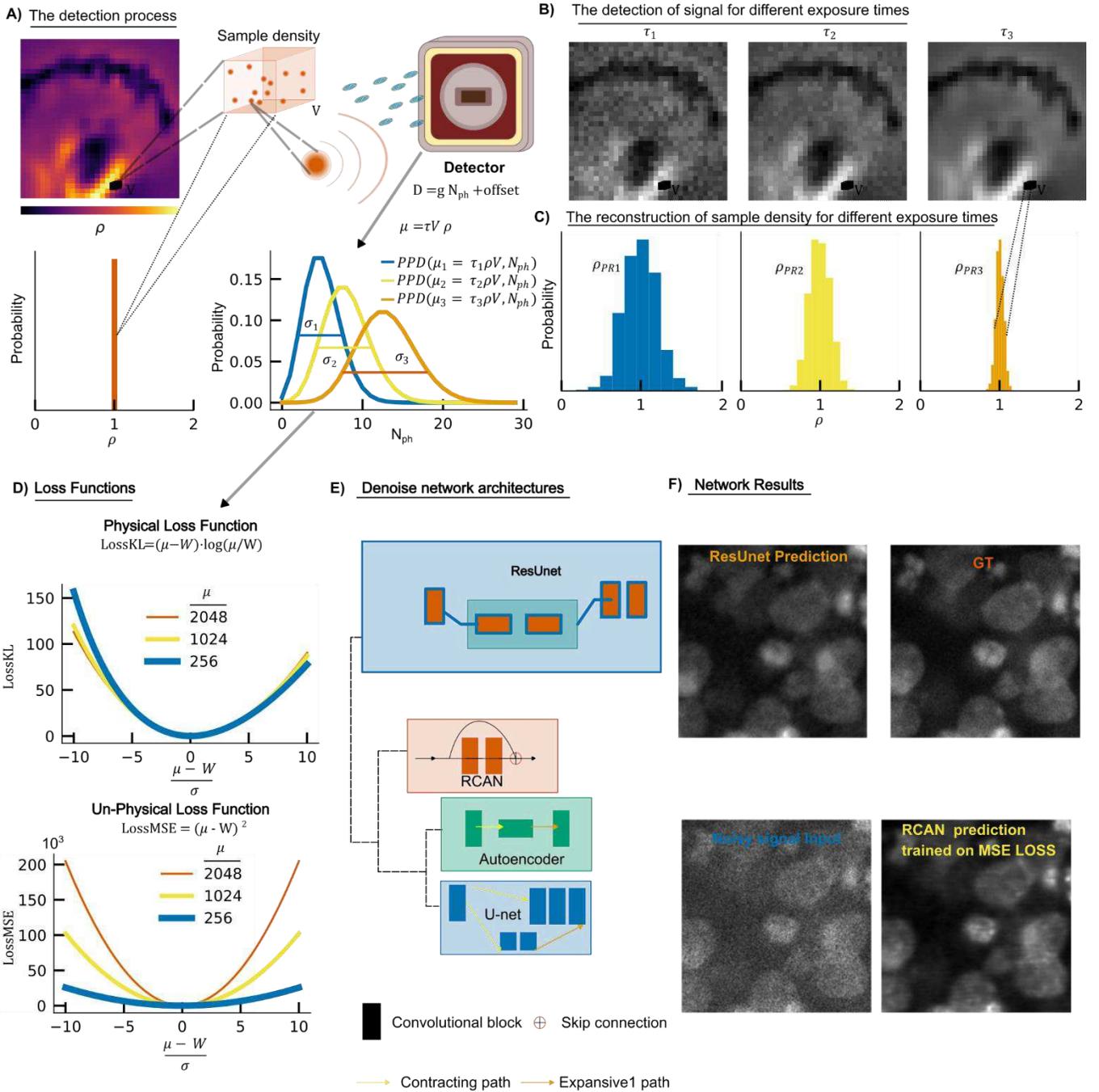


Figure 1: A) From a small volume V of the density ρ of the sample N_{ph} photons are absorbed in the detector and transformed into electrical signal with efficiency g plus an offset. The longer the exposure τ the higher the average signal μ originating from the same density value ρ . On the left lower corner the real probability of the density value inside the volume V . On the right lower corner, the detector absorbs N_{ph} for different exposure times $\tau = \tau_1 < \tau_2 < \tau_3$ producing a PPD with average value μ and variance $\sigma = \sqrt{\mu}$ B) Three different measurements of the sample density for the exposure times of panel A C) The reconstruction of the density inside V for the exposure times of panel A. D) The physical loss function Loss_{KL} and the unphysical Loss_{MSE} for a single voxel of the sample for three different GT average values μ as a function of the relative error $\frac{\mu - W}{\sigma}$ between the predicted DNN output W for three different average values μ . E) A schematic representation of different DNN architectures: ResNet, RCAN, autoencoder, U-net. ResNet is a combination of the other three and uses the physical Loss function F) Different predictions between the RCAN and the ResNet (both trained in dataset 1, and evaluated in dataset 2) for a Noisy signal Input and a comparison with the GT.

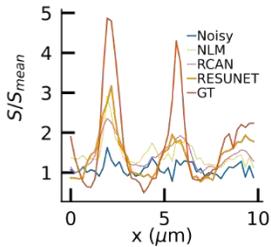
the Noisy input (10 msec exposure time image slice taken from dataset 2) and the GT image (500 msec exposure time, same laser power). Both ResNet and RCAN has been trained on the same dataset (Dataset 1) for the same number of epochs. ResNet, produce a high contrast, artifact-free smooth image. RCAN produces artifacts that originate from dataset 1 (horizontal lines in the lower part of the image) and a grainy image. NLM produce a high contrast image but tends to highlight more the high intensity values and loses some information on the structure of the image (legs on the right lower corner). In panel B we show the line profiles of the line in panel A. ResNet, in comparison with the other algorithms produce a higher contrast between the deeps and the peaks

along the lines. In panel C) we show the cumulative probability (Cum. Prob.) of the intensity histogram for the images shown in panel A. The vertical line represents the Kolmogorov distance. Both RCAN and ResNet do not surpass the GT Cum. Prob. of the GT, with ResNet approaching the GT at around 20 % of the total counts and RCAN at around 50 %, while NLM surpass the Cum. Prob. at around 10 % of the total counts while approach the GT Cum. Prob. at 50 %. In other words, around 40 % of the NLM denoised values are over-estimated, 50 % of the RCAN are under-estimated, while only around 20 % of ResNet are underestimated. In panel D) we compare different score of the three algorithms for the same

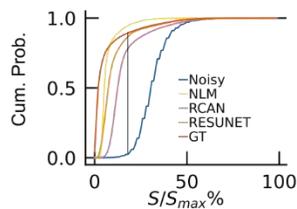
A) Comparison of DNN trained in Dataset 1 and evaluated on Dataset 2 with the classical NLM denoise algorithm



B) Line Profiles from panel A



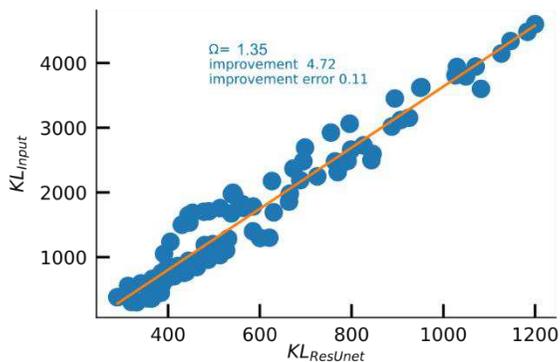
C) Cumulative probability of the intensity histograms in panel A



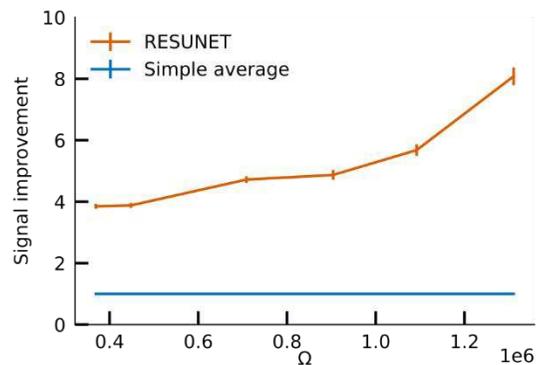
D) Scores of panel A

	ResUnet	NLM	RCAN
MSE Improvement	3.13	2.27	2.18
SSIM Improvement	1.63	1.54	1.54
Cum. Prob. Improvement	1.55	1.45	1.17

E) Signal Improvement of ResUnet trained in DS1 and evaluated in other DS



F) Signal Improvement for different datasets



G) Comparison of ResUnet trained in Dataset 1 and Dataset 3 and evaluated in Dataset 3 with the classical NLM

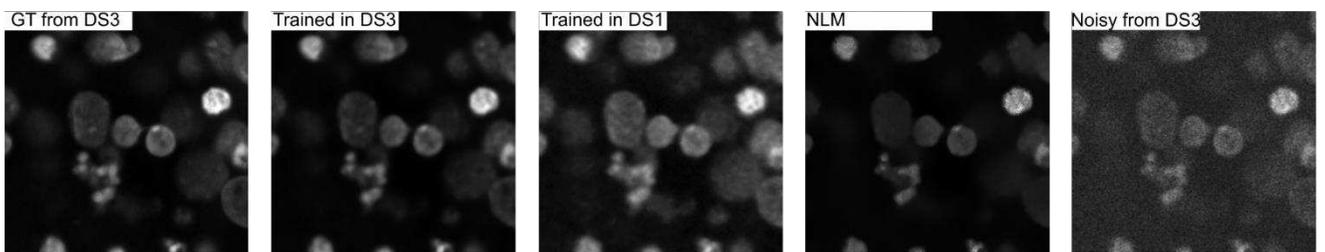


Figure2: A) The comparison of the ResUnet with RCAN and NLM algorithm reconstruction of a Noisy in input 10msec exposure (Noisy) for a Noci cell image slice from Dataset 2 with the corresponding GT (500msec exposure). Both ResUnet and RCAN have been trained in a different dataset 1. Characterized by different optical systems, fluorophores, and cameras. B) the line profile of the panel A for the line shown above C) The Cumulative probably of the intensity which is characterized by the Kolmogorov distance (vertical line) for panel A. D) The MSE, SSIM, Cumulative probability improvements for Panel A. E) We define the signal Improvement as the slope between the KL_{Input} and the KL_{ResUnet} scatter plot linear fit, which we find for each different dataset here shown just for one characterized by the parameter F) The Signal improvements for different datasets with different $\Omega = \frac{V_{input}}{V_{psf}}$. G) Comparison of ResUnet for the same DS trained and different and comparison with the NLM. The signal improvement is 54.2 (trained in DS3) and 4.4 (Trained in DS2)

image, showing that ResUnet achieves better scores. In panel E) we quantify the signal improvement of the ResUnet (trained in DS1) for a different dataset by performing a linear fit on the scatter plot of the KL divergence between the GT and Noisy input and the KL between the GT and the output of ResUnet. The slope of the linear fit is the signal improvement. The parameter $\Omega = V_{input}/V_{psf}$ identifies the information contained in the image volume. In panel F) we plot the Signal improvement for different dataset with different Ω showing that signal improvement increases as Ω increases, while a simple average of the Input image brings no advantages in the signal improvement. Finally, in panel G) we show that when trained on the same dataset, same imaging conditions, the performance of ResUnet increases even further.

In summary, we showed that a physics-informed Loss function employed in a novel DNN architecture, outperforms the state-of-the-art denoising algorithms for fluorescence

microscopy. We overcome the arbitrariness of the data normalization in denoising DNNs, by using a photon model that do not produce instabilities, thus showing portability when use in different imaging conditions. Our Loss function is based on the statistics of Poisson distribution and could be used in DNNs performing tasks in other than microscopy optical or non-optical imaging technologies that have Poissonian signal.

Acknowledgment: This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 713694. GR acknowledges the support by European Research Council Synergy grant ASTRA (No. 855923). ML thanks Project LOCALSCENT, Grant PROT. A0375-2020-36549, Call POR-FESR “Gruppi di Ricerca 2020”.

Competing interests: The authors declare no competing interests.

Bibliography

1. Belthangady, C. & Royer, L. A. learning for fluorescence image reconstruction. *Nat. Methods* (2009) doi:10.1038/s41592-019-0458-z.
2. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
3. Laine, R. F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nat. Methods* **18**, 1136–1144 (2021).
4. Chen, J. *et al.* Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat. Methods* **18**, (2021).
5. Gurrola-ramos, J., Dalmau, O. & Alarcón, T. E. A Residual Dense U-Net Neural Network for Image Denoising. **9**, (2021).
6. Byun, J., Cha, S. & Moon, T. FBI-Denoiser: Fast Blind Image Denoiser for Poisson-Gaussian Noise. (2021).
7. Mayorov, A. S. *et al.* Interaction-Driven Spectrum Reconstruction in Bilayer Graphene. *Science (80-.)*. **333**, 860–863 (2011).
8. Shanker, M., Hu, M. Y. & Hung, M. S. Effect of data standardization on neural network training. *Omega* **24**, 385–397 (1996).
9. Qiao, C. *et al.* Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* (2021) doi:10.1038/s41592-020-01048-5.
10. Xypakis, E. *et al.* Deep learning for blind structured illumination microscopy. *Sci. Rep.* **12**, 8623 (2022).
11. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
12. Mandracchia, B. *et al.* Fast and accurate sCMOS noise correction for fluorescence microscopy. *Nat. Commun.* 1–12 doi:s41467-019-13841-8.
13. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
14. Kullback, S. & Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
15. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
16. Demidenko, E. Kolmogorov-smirnov test for image comparison. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **3046 LNCS**, 933–939 (2004).

Methods

Normalization and camera calibration.

One known issue in the training process is the normalization of the data with improper normalization to lead to a failure in the training⁸. To overcome this issue denoising DNN for microscopy an (arbitrary) percentile normalization of the images has been proposed, however on the evaluation of the algorithms the efficiency scores that are reported^{2,4} not only depend on the normalization choice but an extra normalization to the ground truth image (GT) has to be performed in order to

produce the best scores. This ambiguity on the results becomes important especially when a GT is not provided or unknown. Overall, a robust image normalization scheme is missing from the literature. To overcome these issues, we first calibrate the camera as described in¹² and use the photon numbers as the input and target of the ResUnet so that each pixel/voxel represents the number of photons. Both the GT and the Output of the ResUnet represent the average value of the PPD.

Physical Loss function.

The Kullback-Leibler divergence is a measure of comparing two probability distributions functions $p(n), q(n)$, for the same sample space n . It is defined as

$$KL(p, q) = \sum_n p(n) * \log \left(\frac{p(n)}{q(n)} \right)$$

For two PPD with mean values μ, W , $p(n, \mu) = \mu^n e^{-\mu}/n!$ $q(n, W) = W^n e^{-W}/n!$, it can be shown that the $KL(p, q)$ depends only on the mean values μ, W

$$KL(\mu, W) = \mu \log \left(\frac{W}{\mu} \right) - (W - \mu)$$

In the paper we used the symmetrized version of $KL(\mu, W)$ as the Loss function.

$$Loss_{KL} = \frac{1}{2} (KL(\mu, W) + KL(W, \mu)) = (\mu - W) \log \left(\frac{\mu}{W} \right)$$

Since the signal is coming from a probabilistic distribution, we treat the photon number for each pixel of the input and the target of ResUnet as the mean values of the PPD.

Res-U-net

Regarding the architectures, a promising approach is to use of skip connections between network layers to bypass residual contents (RCAN⁴) and an encoder-decoder like architecture solves efficiently semantic segmentation /classification tasks (U-net¹¹). In our case the classes are the number of photons, so we use the efficiency of U-net to classify photons and the power of residues to perform denoising. To do that we substitute the Convolutional layers of U-net with residual blocks.

Evaluation datasets and reproducibility.

To ensure reproducibility³, we evaluated and compared our architecture with the state of the art DNN architecture and a non DNN classical denoising algorithm. The dataset 1 is the dataset that RCAN is originally trained for Tubulins and can be found in⁴. The evaluation was performed on confocal spinning disk microscopy experimental data acquired in our lab where we systematically change the exposure time (1,2,5,10,20,50,100 and 500 msecs). and thus the SNR. We differentiate each dataset by the parameter $\Omega = \frac{V_{input}}{V_{psf}}$ as the

sampling rate between the volume of the input and the psf, by using three different fluorophores (DAPI, Alexa488, Alexa647) emission wavelength (420nm, 525nm, 665nm). and three different objectives with different depth of field (UPLSAPO 30XS, UPLSAPO 20X, LUCPLFLN40XPH 40X).

Evaluation metrics.

To provide the best comparison in addition to the widely used structural similarity index (SSIM¹⁵) and mean square error (MSE) we use also the Kolmogorov-Smirnov index¹⁶ and the Kullback-Leibler divergence. In this way we provide the maximum information for image comparison.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Xmodel.png](#)