Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)

Mayeul Cachia, Vasiliki Stergiopoulou, Luca Calatroni, Sébastien Schaub, Laure Blanc-Féraud

To cite this version:

Mayeul Cachia, Vasiliki Stergiopoulou, Luca Calatroni, Sébastien Schaub, Laure Blanc-Féraud. Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN). Inverse Problems, 2023, 39 (5), pp.054006. 10.1088/1361-6420/acc889. hal-03790156v2

HAL Id: hal-03790156
https://hal.science/hal-03790156v2

Submitted on 29 Mar 2023

Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)

Mayeul Cachia[1], **Vasiliki Stergiopoulou**[1], **Luca Calatroni**[1], **Sebastien Schaub**[2], **Laure Blanc-Féraud**[1]

[1]CNRS, INRIA, I3S, Universit Cte dAzur, Sophia Antipolis, France
[2]Sorbonne Universit, CNRS, LBDV, Villefranche-sur-Mer, France

E-mail: `mayeulcachia@gmail.com`, `vasiliki.stergiopoulou@i3s.unice.fr`, `calatroni@i3s.unice.fr`, `sebastien.schaub@imev-mer.fr`, `blancf@i3s.unice.fr`

Abstract. We propose FluoGAN, an unsupervised hybrid approach combining the physical modelling of fluorescence microscopy timelapse acquisitions with a generative adversarial learning procedure for the problem of image deconvolution. Differently from standard approaches combining a least-square data term based on one (long-time exposure) image with sparsity-promoting regularisation terms, FluoGAN relies on a data term being the distributional distance between the fluctuating observed timelapse (short-time exposure images) and the generative model. Such distance is computed by adversarial training of two competing architectures: a physics-inspired *generator* simulating the fluctuating behaviour as a Poisson process of the observed images combined with blur and undersampling, and a standard convolutional *discriminator* network. FluoGAN is a fully unsupervised approach requiring only a fluctuating sequence of blurred, undersampled and noisy images of the sample of interest as input. It can be complemented with prior knowledge on the desired solution such as sparsity, non-negativity etc. After having described the main ideas behind FluoGAN, we formulate the corresponding optimisation problem and report several results on simulated and real phantoms used by microscopy engineers to quantitatively assess spatial resolution. The comparison of FluoGAN with state-of-the-art methodologies shows improved resolution, allowing for high-precision reconstructions of fine structures in challenging real *Ostreopsis cf Ovata* data. The FluoGAN code is available at: https://github.com/cmayeul/FluoGAN.

1. Introduction

The physical limit imposed by light diffraction in the context of light microscopy still poses a major challenge for the accurate reconstruction and analysis of small samples: structures closer to such barrier (around 250 nm in the lateral x-y plane) cannot be distinguished. Since several biological quantities of interests (such as viruses, proteins and molecules) have size significantly smaller than this limit, it is thus crucial to overcome it by means of advanced image reconstruction approaches.

The reconstruction of a sample of interest from its blurred, noisy and under-sampled data is an ill-posed inverse problem for which numerous approaches lying at the interface between applied mathematics and microscopy imaging have been proposed. In this paper, we are interested in image reconstruction problems arising in *fluorescence microscopy*. In a nutshell, in this field special chemical compounds, called *fluorescent dyes* or *fluorophores* are used to bind to the molecules in the sample of interest. When exposed to light excitation at specific wavelengths, these molecules emit photons which are captured by CCD or sCMOS camera or photo-multiplicator sensors after passing through special lenses and optical devices. As a consequence, the resolution of the captured image is thus limited due to diffraction of the emitted light. The diffraction (Airy) pattern acts as a convolution operator (the Point Spread Function, PSF) on the emitters, resulting in a blurred and noisy observed image. The standard deconvolution/super-resolution inverse problem thus consists in retrieving the spatial structure of the sample under observation by knowing the Point Spread Function (PSF) of the optical system and, possibly, some statistical prior information on the noise. In this work, we assume to know (up to a sufficient accuracy) the PSF of the microscope under consideration. Typically, such function is modelled as a Gaussian function or as an Airy pattern. As far as the noise is concerned, we consider a mixed scenario where both a signal-dependent Poisson component and an additive white Gaussian perturbation is add to model photon counting and electronic noise processes, respectively.

In a discrete setting, the mismatch between the 2D spatial grid where measured data lie (often called the coarse grid) and the one where we look for solutions (the fine grid) may be explicitly quantified by a super-resolution factor $L \in \mathbb{N}, L > 1$ for both the horizontal and vertical direction. In mathematical terms, for a given blurred, noisy and under-sampled vectorised image $y \in \mathbb{R}^m$, a convolution matrix $H \in \mathbb{R}^{n \times n}$ corresponding to the 2D PSF of the system and an additive Gaussian noise component $e \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$ with $\mathrm{I} \in \mathbb{R}^{m \times m}$ being the identity matrix, the task is to retrieve a super-resolved image $x \in \mathbb{R}^n$ and a background image $b \in \mathbb{R}^m_{\geq 0}$ such that:

$$y = \text{Poisson}(UHx + b) + e, \tag{1}$$

where the operator $U \in \mathbb{R}^{m \times n}$ with $n = L^2 m$ denotes an under-sampling operator mapping images from a fine to the L-coarser grid and where, for $z \in \mathbb{R}^m_{\geq 0}$, $\text{Poisson}(z)$ denotes a Poisson random vector of parameter z. The space-variant background term $b \in \mathbb{R}^m_{\geq 0}$ models here out-of-focus and/or ambient molecules. Model-based regularisation approaches for solving (1) formulate the task as a minimisation problem where a sparsity-promoting term is combined with a least-square data fidelity. Standard compressed-sensing techniques, for instance, rely on the use of the ℓ_1 norm as regulariser possibly combined with redundant representation techniques [34]. In later years, however, non-convex regularisation approaches relying on the minimisation of continuous relaxations of the ℓ_0 pseudo-norm have also been considered [27, 28]. In the case of samples with wide support (not points nor curves), the sparsity assumption on the signal is, however, not suited. In these cases, a tailored choice is indeed far from being trivial.

In order to have super-resolution and not only extrapolation enforced by an additional regularisation term, additional information on the sample preparation and on its physical properties must be provided. A popular class of super-resolution techniques achieving a resolution of about 20 nm is the family of Single Molecule Localisation Microscopy (SMLM) [25], for which the idea is to sequentially activate only a small number of fluorophores at a time so that their detection can be made with high precision. A super-resolved image of the sample of interest can thus be found by simply combining together all the reconstructed sparse images. Mathematically, such problem can be cast in the form (1) (see [17] for more complicated forward models) and its regularised solution can be computed at each time stamp. Popular SMLM approaches are, for instance PALM (Photo-Activated Localisation Microscopy) [4] and STORM (STochastic Optical Reconstruction Microscopy) [24]. From a practical viewpoint, the main disadvantage in using SMLM techniques is the fact that they require fluorophores with specific chemical and physical characteristics [16] as well as a significant number (typically of the order of thousands) of sequential activations and sparse acquisitions which may significantly harm the sample under observation. Furthermore, for a computational viewpoint, a typically non-smooth and non-convex regularised problem has to be solved for each acquired frame, which is of course very time consuming in large-scale scenarios. A rather recent approach mitigating such demanding requirements consists in exploiting, rather than that sparse sequential activation of specific fluorophores over time, the independent stochastic intensity fluctuations of *common* fluorescencent molecules. By acquiring a video of stochastichally fluctuating images with an acquisition rate in the range 20-100 images/s with conventional microscopes and exploiting in different ways the independence of fluctuations between emitters, significant improvements in resolution can indeed be obtained without specific fluorophores. It is the case, for instance, of SOFI (Super-resolution Optical Fluctuation Imaging) [8] where temporal information is integrated by computing a sample covariance image in which, by independence, the PSF appears squared so that a reduction of a factor $\sqrt{2}$ on the width of the PSF is obtained. A similar approach reducing the computational costs of covariance calculation is SCORE (Spatial COvariance REconstructive) [7] where only pixels with relevant contributions to such computation are used. A different method taking sequences of fluctuating images as an input, but based on a different type of analysis is SRRF (Super-Resolution Radial Fluctuations) [14] where degrees of symmetry are calculated at each frame. A mathematical modelling of these problems based on sparse regularisation approaches is proposed in [26], where the authors propose to formulate the problem (1) in the covariance domain using an ℓ_1 regularisation for sparsity-based super-resolution correlation microscopy (SPARCOM) problems. In recent works [32, 30], the authors improved upon the artefacts created by SPARCOM on simulated and real noisy data by considering a non-convex continuous exact approximation of the ℓ_0 pseudo-norm previously employed in [27, 9] and incorporating a further background and noise optimisation step within a two-step procedure allowing also for intensity estimation. Such approach, called COL0RME (COvariance-based ℓ_0 super-Resolution Microscopy with intensity Estimation), has been also applied to 3D image data in [31] showing high-quality reconstructions and high applicability in common scenarios. A major limitation of SPARCOM and COL0RME when applied to challenging real-word contexts is that, due to their mathematical structure, they tend to promote point-like structures also in the case of samples characterised by the presence of different structures (e.g., thin filaments).

Motivated by the recent advances in the field of deep learning models for imaging, we review in the following some particular reconstruction models based on the use of generative approaches. Generative Adversarial Networks (GANs) [11] are nowadays very powerful tools in the field of imaging. They aim at estimating the unknown distribution of a class of images given as input by the adversarial training of two different networks: the former (the *generator*) aiming at creating images with similar properties to the given ones and the latter (the *discriminator*) which, during training, distinguishes whether a generated image is a fake one or not, so that convergence is achieved when such distinction is not possible anymore. There is a vast literature on the use of GANs in imaging: we recall here some nowadays pretty established strategies for improving training stability [2, 12] and some others more focused on applications to learning the distribution of human faces [15] or medical data [19]. In their vanilla formulation, GANs do not incorporate in their structure any physical modelling, their performance being characterised only by the (often not interpretable) parameters of the generator and the discriminator. A hybrid approach allowing for the use of a physically-inspired simulator in place of a GAN generator has recently been proposed for the problem of volume reconstruction from noisy projections in Cryo Electron Microscopy (Cryo-EM) under the name CryoGAN [13]. In CryoGAN the volume of interest becomes the learnable parameter to estimate itself upon an adversarial training where, given noisy projections $\{y_i\}_i$, the current estimate of the desired volume is passed through the simulator to obtain simulated noisy projections $\left\{y_k^{\text{sim}}\right\}_k$ being 'close enough' to the given data in a distributional sense. The approach is proved to be very effective, allowing for a very efficient volume estimation using a simple linear physical model which allows to derive reconstruction guarantees. A similar approach is used in [35] for multi-segment reconstruction. Note that in the context of super-resolution microscopy, data-driven methods relying on algorithmic unrolling [5] and convolutional neural networks [21, 29] have been proposed. Up to our knowledge, no generative approaches have been considered in this field.

Contribution. Given a video of blurred, under-sampled and noisy short-time exposure and stochastically fluctuating images, we consider in this work a reconstruction procedure based on the minimisation of the distance between two distributions: the former associated to the observed video, the latter describing image examples simulated by the known physical model of stochastic fluctuations. The optimisation procedure is realised by means of a generative adversarial learning strategy, in which the generator

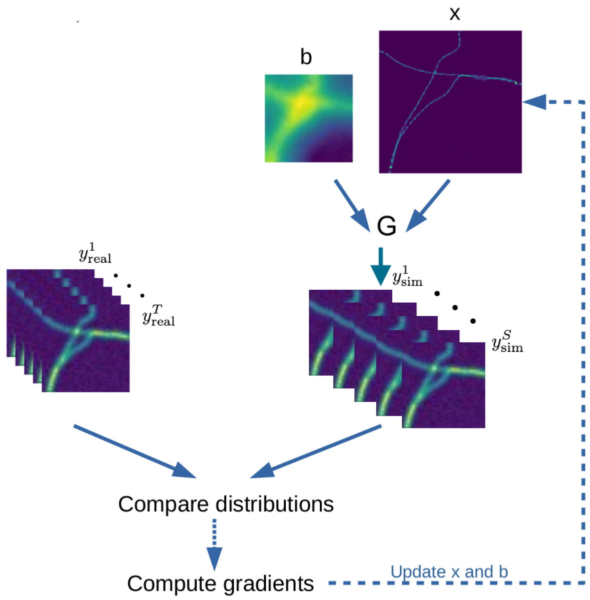


Figure 1: Schema of FluoGAN: Given a temporal sequence of blurred, under-sampled and noisy images (left) and a physical-inspired simulator of fluorescence fluctuations and of the optical system (right), denoted by G, FluoGAN compares the distribution of the acquired images with the distribution of the samples generated by G. It then computes gradients of a suitable loss functional minimising such distance and updates accordingly the simulator inputs. At convergence, the approach outputs the super-resolved image x and the background image b.

network is replaced by a generative physical model of stochastic fluorecence fluctuations. The proposed approach is named FluoGAN (Fluorescent image deconvolution/super-resolution microscopy via GAN learning). It combines physical modelling with data-driven learning. Differently from the the usual structure of a GAN, the generator network is here replaced by a non-linear model of intensity fluctuations and light diffraction. From the given video, the method estimates one single high-resolution image of the sample of interest by the adversarial update of the simulator and the discriminator network parameters, as schematically described in Figure 1.

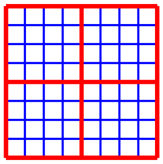
Structure of the paper. In Section 2 we detail the physical model of video image acquisition used as simulator, then, in Section 3 we revise the GAN formulation from an optimisation perspective while in Section 4 we describe in detail the modelling and optimisation aspects of the proposed FluoGAN approach. Numerical results are reported in Section 5: the proposed method is validated first on realistic simulated spatial patterns for assessing the resolution achieved, also in comparison to state-of-the art approaches. Next, reconstructions computed from real data of a commercial phantom acquired on standard microscopes and of the *Ostreopsis cf Ovata* alga are reported. The proposed approach achieves notable improvement in resolution and appears much more robust to noise perturbations and modelling errors, being effective also on challenging scenarios.

2. A non-linear model for stochastic fluctuations

We present in this section the model used in the following for simulating the physical processes creating blur and noise – including fluorophore fluctuations – in the observed image data. Fluorescence fluctuations are modelled here as a stochastic process which, for simplicity, we consider here as a Poisson model. It can be interpreted as the number of photons emitted by a biological structure of interest sampled on a *fine* grid of size $n \in \mathbb{N}$. In [22, 10] a different model for stochastic fluctuations was considered. Differently from our model, in those works the proposed simulator models the on-off switching (a.k.a. *blinking*) of individual fluorophores and stochastic intensity fluctuations are included in the model in the form of noise. In our modelling, we describe the fluctuations of a sum of molecules lying in a pixel of the fine grid image $x \in \mathbb{R}^n_{\geq 0}$ as a stochastic Poisson process, denoted by Poisson(x). These Poisson variables are independent of each other (in time and space). We remark that we will work with short acquisition times, which allows us to neglect the modelling of photobleaching, i.e. the loss of fluorescence capability.

We now introduce the modelling of the optical device (i.e. microscope) considered.

From the sample to the sensor: blur, under-sampling and background. We model blur due to light diffraction as a convolution with a 2D Gaussian PSF denoted by $h \in \mathbb{R}^{k\times k}$ where k^2 denotes the size of the support of the kernel, and we endow it with periodic boundary conditions. By linearity, we then represent the action of h on the vectorised image Poisson(x) by means of a block circulant with circulant block matrix $H \in \mathbb{R}^{n\times n}$. The incident flux to the receiving sensor is then under-sampled to fit the size of the sensor pixels which form a low-resolution (coarse) grid. We denote by $m < n$ the size of the coarse grid where the measured data are acquired and denote by $L > 1$ the super-resolution factor along both the horizontal and vertical direction so that $n = L^2 m$. We model such undersampling by means of a rectangular matrix $U \in \mathbb{R}^{m\times n}$ whose entries are simply 1's and 0's and acts by summing the intensity values defined on $L \times L$ patches (blue grid) of x so as to produce one single coarse pixel (red grid) in the measurement space. This process has the physical meaning of summing up the number of photons emitted by the unknown sample x within the measurement space, while preserving their total amount. The photons captured by the sensor do not all come from the fluorescent molecules we want to observe. Some of them are produced by out-of-focus emitters or even from outside the sample. Depending on the acquisition conditions, their proportion can be in fact relatively high. We include these molecules introducing in the model a non-constant and non-negative background term denoted by $b \in \mathbb{R}^m_{\geq 0}$.



$L = 4$

The blurred and under-sampled flux arriving at the receiving sensor is (proportional to) therefore:

$$\phi(\text{Poisson}(x), b) := UH\text{Poisson}(x) + b \in \mathbb{R}^m. \tag{2}$$

Recording noise. We now want to simulate the stochastic behaviour of the recording process. Typically, each pixel records only a few dozens of photons per image. The photon absorption can thus be described as a counting process with the following characteristics:

(i) There is a time interval short enough so that the probability to absorb more than one photon is negligible;

(ii) The number of photons absorbed during one time interval is independent to another;

(iii) During a short time interval the probability to absorb a photon is proportional both to the interval duration and to the incident flux.

This counting process can thus be mathematically modelled as a further Poisson counting process. For a given image, the number of photons captured by one pixel during the acquisition time can be simulated by random sampling a Poisson variable whose expected value is proportional to the flux at the sensor, i.e. at (2). The variance of this variable depends nevertheless on the gain $\alpha > 0$ of the sensor, i.e. the amplification factor applied to the image by the image sensor, which depends on the specific setup considered and which can be easily incorporated in the model. Last but not least, a further term describing electronic noise should be added as in (1). To do so, we thus consider a realisation a multidimensional random vector with Gaussian distribution of zero mean and covariance matrix $\sigma^2\mathrm{I}$, $e \sim \mathcal{N}(0, \sigma^2\mathrm{I})$.

Collecting altogether, we can finally define the complete forward observation model

$$y = \alpha \ \text{Poisson}\left(\phi(\text{Poisson}(x), b)\right) + e, \tag{3}$$

which is used in the following to generate simulated data in the experimental section 5.1. However, due to the presence of the double Poisson distributions, model (3) is difficult to handle in computations during the reconstruction process. This model will be used to generate simulated data but a simpler one will be used for the reconstruction process. In the following, we will thus make use of the following image formation model:

$$y = \alpha \ \text{Poisson}\left(\phi(x, b)\right) + e, \tag{4}$$

which we claim to provide an acceptable approximation of (3) in the estimation process. Fluctuations of fluorophores are in fact passed through the PSF and undersampling operations, which both are processes smoothing out the fluctuations themselves. The Poisson noise modelling then adds further Poisson fluctuations.

Summarising, the setup is composed of *generated images* which are the realisations y^t_{sim}, $t = 1, \ldots, T$ of the multi-variate random variable Y_{sim} following a distribution $\mathcal{D}_{\text{sim}}(x, b)$ so that:

$$Y_{\text{sim}}(x, b) \sim \mathcal{D}_{\text{sim}}(x, b) := \alpha \ \text{Poisson}\left(\phi(x, b)\right) + e. \tag{5}$$

Note that the quantities x and b play here the role of the unknown learnable parameters to be optimised, while we assume that $\alpha > 0$, the physical process $\phi(\cdot,\cdot)$ and the variance σ^2 characterising the additive noise component e are known.

3. The inverse problem formulation

Interpreting the given temporal sequence of $T > 1$ noisy, blurred and under-sampled images $\{y^t_{\text{real}}\}_{t=1}^T$ as the acquired T realisations of a random variable Y_{real} with an unknown distribution $\mathcal{D}_{\text{real}}$, the distributional inverse problem formulation of the problem reconstruction procedure thus reads:

$$\text{given } \left\{y^t_{\text{real}}\right\}_{t=1}^T \quad \text{find } (x,b) \in \mathbb{R}^n \times \mathbb{R}^m \qquad \text{s.t.} \qquad \mathcal{D}_{\text{sim}}(x,b) \sim \mathcal{D}_{\text{real}} \tag{6}$$

where $\mathcal{D}_{\text{sim}}(x,b)$ is as in (5). From an optimisation perspective, we thus formulate problem (6) as the minimisation of a suitable distance between $\mathcal{D}_{\text{sim}}(x,b)$ and $\mathcal{D}_{\text{real}}$ over x and b. Further additional and physically-consistent regularisation terms on both x and b can be can be incorporated to stabilise the optimisation procedure and eliminate potential solutions without any physical meaning. We thus consider the general optimisation problem:

$$\min_{x\in\mathbb{R}^n_{\geq 0},\ b\in\mathbb{R}^m_{\geq 0}} \ d(\mathcal{D}_{\text{sim}}(x,b), \mathcal{D}_{\text{real}}) + R_1(x) + R_2(b), \tag{7}$$

where the relevant distance d comparing $\mathcal{D}_{\text{sim}}(x,b)$ and $\mathcal{D}_{\text{real}}$ and the assumptions on x and b with the corresponding choice of R_1 and R_2 have to be specified, along with a tailored optimisation algorithm for their computation.

3.1. Comparing distributions

3.1.1. ℓ_2 distance. The easiest way to compare $\mathcal{D}_{\text{sim}}(x,b)$ to $\mathcal{D}_{\text{real}}$ is to compute the difference between their empirical mean. In our modelling, we suppose that both distributions are, up to some zero-mean Gaussian noise modelling, Poisson-like distributions and, as such, characterised by their multidimensional parameter corresponding both to their expected value and variance. For the simulated image samples such parameter depends on the desired parameters x and b, but for the real ones this is of course unknown. Given a batch of images, an unbiased estimate of such value is given by the sample average $\bar{y}_{\text{real}} := \frac{1}{T}\sum_{t=1}^T y^t_{\text{real}}$. Denoting similarly $\bar{y}_{\text{sim}}(x,b) = \frac{1}{T}\sum_{t=1}^T y^t_{\text{sim}}(x,b)$, we can thus choose:

$$d(\mathcal{D}_{\text{sim}}(x,b), \mathcal{D}_{\text{real}}) = \|\mathbb{E}[Y_{\text{sim}}(x,b)] - \mathbb{E}[Y_{\text{real}}]\|_2^2 \simeq \|\bar{y}_{\text{sim}}(x,b) - \bar{y}_{\text{real}}\|_2^2. \tag{8}$$

The advantage of this loss is that it can be computed explicitly. However, comparing only two mean images is not very representative of the diversity in the data. A better choice improving the amount of information carried out by (8) consists in comparing distributions using the mean ℓ_2 distance between individual samples drawn uniformly from a batch $B \subset \{1, \ldots, T\}$ and the mean of the observed data, i.e.:

$$d(\mathcal{D}_{\text{sim}}(x,b), \mathcal{D}_{\text{real}}) = \mathbb{E}\left[\|Y_{\text{sim}}(x,b) - \mathbb{E}[Y_{\text{real}}]\|^2\right] \simeq \frac{1}{|B|}\sum_{t\in B}\|y^t_{\text{sim}}(x,b) - \bar{y}_{\text{real}}\|_2^2 \tag{9}$$

where $\{y^t_{\text{sim}}(x,b)\}_{t\in B}$ being a set of realisations of the random variable $Y_{\text{sim}}(x,b)$.

3.1.2. From KL divergence to Wasserstein distance. Another way to compare two probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ absolutely continuous with respect to a measure on Ω consists in computing their Kullback-Leibler (KL) divergence. In their vanilla form, Generative Adversarial Networks (GANs) [11] rely in fact on the asymptotic minimisation of a symmetrised version of such distance between $\mathcal{D}_{\text{sim}}$ and $\mathcal{D}_{\text{real}}$ (the Jensen-Shannon divergence). The direct minimisation of KL divergence is not possible when the underlying densities are unknown (as it happens for $\mathcal{D}_{\text{real}}$). Moreover, even in the case of known density functions, such minimisation is generally challenging due, for instance, to the fact that the KL divergence is equal to infinity when one of the two densities vanishes and/or when they have disjoint supports [2]. A more effective choice motivated by the field of optimal transport relies on the minimisation of a different loss, the 1-Wasserstein distance also called as *earth-mover distance*, see, e.g., [1] for a survey. Its functional form reads:

$$W_1(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\gamma\in\Pi(\mathbb{P}_1,\mathbb{P}_2)} \mathbb{E}_{(u,v)\sim\gamma}\left[\|u - v\|\right], \tag{10}$$

where $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ stands for the set of all joint distributions $\gamma(x, y)$ with marginals equal to $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. Such distance is well defined also for distributions with disjoint support as it depends on the Euclidean distance between supports. The W_1 distance is therefore better suited for comparing general distributions. In [2], it was to improve their performance of GANs and celebrated under the name Wasserstein-GANs. Note, that while it may not be possible to compute W_1 explicitly, efficient strategies based on suitable gradient penalties enforcing its minimisation can be defined [12]. In Section 4 we will consider such relaxations and incorporate them in the general optimisation problem considered.

3.2. Choosing the regularisation terms

Recalling (7), we now provide some details on how to choose physically-consistent regularisation terms R_1 and R_2 favouring *a-priori* regularity on the desired super-resolved image x and on the background image b, respectively.

In fluorescence microscopy applications, the task often consists in seeking a super-resolved image made of fine structures (typically, thin filaments of proteins, point-like sources...). Hence, we can safely assume that $x \in \mathbb{R}^n_{\geq 0}$ has only a few non-zero pixels corresponding to these structures. Enforcing sparsity corresponds, in a rather natural way, to minimise the non-continuous, non-convex ℓ_0 pseudo-norm defined by $\|x\|_0 := |\{i \in \{1,\dots,n\} : x_i \neq 0\}|$. The minimisation of the ℓ_0 pseudo-norm is a NP-hard problem [20] for which several relaxations have been proposed in the literature relying either on convex ℓ_1 minimisation [6] or on several non-convex but continuous penalties, see [28] for a unified approach. For what follows, we will set $R_1(x) = \lambda_1\|x\|_1, \lambda_1 > 0$ and deal with the non-smoothness associated to such choice by means of an accelerated proximal-gradient algorithm [3]. The background term $b \in \mathbb{R}^m_{\geq 0}$ is supposed to provide information on the photons emitted by out-of-focus sources as well as other sources of noise including ambient light. It is therefore not likely to contain high-frequency details, hence a smooth regularisation term, i.e. $R_2(b) = \frac{\lambda_2}{2}\|\nabla b\|_2^2, \lambda_2 > 0$ can be considered, see [30] for analogous choices.

4. FluoGAN: formulation and optimisation details

We now describe how to combine generative adversarial learning approaches for evaluating the distributional fit between $\mathcal{D}_{\text{sim}}(x, b)$ and $\mathcal{D}_{\text{real}}$. Generative Adversarial Networks (GANs) were introduced firstly in [11] by Goodfellow et al. and have been successfully applied to several inverse problems in imaging (see for example [13] for Cryo-EM). Given a training set of images interpreted as the many realisations of a multidimensional random variable with unknown probability density, GANs aim at train two competing networks so that, at convergence, new samples from the unknown distribution can be drawn. The idea generalises the standard approach for sampling from a random variable of a given law by applying the inverse of its cumulative distribution function to a random variable uniformly distributed in $[0, 1]$. In the case of GANs, the inverse distribution function is modelled by a *generator* network G and it is trained until the distribution of its outputs matches the one of the real data. To define such training procedure, a loss function measuring how far data generated by the network are from real ones is needed. Ideally, one would also like such loss to be nicely differentiable, in order to allow network parameters update by means of gradient-based algorithms. Since it may be hardly possible to find such a convenient loss function, the particularity of GANs is to model it by means of a *discriminator* network D which has to decide whether an input data is real (i.e. belongs to the training set) or simulated (it is an output of the generator). Generator and discriminator are thus trained together but with opposed goals. Since D acts essentially as classifier, it is standard to assume that its range is within the interval $[0, 1]$ with $D(y) \approx 1$ denoting that y is drawn from $\mathcal{D}_{\text{real}}$ with high probability while $D(y) \approx 0$ meaning that y is likely to be a sample from $\mathcal{D}_{\text{sim}}$. The training of the two networks is thus adversarial and can be modelled mathematically as a min-max optimisation problem. The standard formulation of GANs in [11] is done by enforcing a comparison between distributions in terms of the Jensen-Shannon divergence, although, as seen in Section 3.1.2 other choices are possible.

Problem (7) is closely related to a GAN-type problem since it involves the minimisation of a distance between distributions. However, two significant differences can be highlighted:

(i) given the physical model of fluorescence fluctuations (5), our knowledge of the unknown random variable $Y_{\text{sim}}(x,b)$ depending on the two quantities of interest x and b does not start from scratch. Inspired by previous works [13, 35], it thus appears natural to consider (5) as a physically-inspired generator, i.e. a simulator of the direct problem. The learnable parameters $(x,b) \in \mathbb{R}^n_{\geq 0} \times \mathbb{R}^m_{\geq 0}$ of this "network" are therefore not any longer parameters of a neural network, but the quantities of interest we want to retrieve.

(ii) a GAN is usually trained to get, at convergence, a satisfying generator which is used to create data distributed like the training ones. In our approach the goal is different since we look into a way of producing fake blurred, noisy and undersampled data by means of the given simulator, so that, when the $\mathcal{D}_{\text{sim}}(x,b)$ matches $\mathcal{D}_{\text{real}}$, the corresponding x and b are the desired solutions of the inverse problem (6).

We call FluoGAN (Fluorescence image deconvolution microscopy via generative adversarial learning) the proposed method, given the particular applicative context considered here. It is therefore inspired by GANs but it has the goal of solving an ill-posed inverse problems given a training set of measured images and by exploiting the underlying physical knowledge. In Figure 2 we draw a comparison between how a standard GAN works in comparison to FluoGAN. For what follows, we will denote by $G_{x,b}$ and D_φ the physically-inspired generator and the discriminator network of parameters (x,b) and φ, respectively.

4.1. Formulation as an optimisation problem

Inspired by GANs, we formulate an adversarial training process defined in terms of two functions to be optimised alternatively and corresponding to the generator and the discriminator, respectively. We thus consider the problem:

$$\min_{x,b} \ \mathbb{E}\big[L_{G_{x,b}}(x,b,\varphi)\big], \tag{11}$$

$$\min_{\varphi} \ \mathbb{E}\big[L_{D_\varphi}(x,b,\varphi)\big]. \tag{12}$$

In such formulation:

- $L_{G_{x,b}}$ is a loss term enforcing equality in distribution between $\mathcal{D}_{\text{real}}$ and $\mathcal{D}_{\text{sim}}(x,b)$ in some sense. Recalling (9), we can choose for instance:

$$L_{G_{x,b}}(x,b,\varphi) := \|Y_{\text{sim}}(x,b) - \mathbb{E}[Y_{\text{real}}]\|^2 - D_\varphi(Y_{\text{sim}}(x,b)). \tag{13}$$

- L_{D_φ} is a loss term enforcing the discriminator to be maximised (i.e. $D_\varphi \approx 1$) on real images, while it should be minimised ($D_\varphi \approx 0$), in an adversarial manner w.r.t. to what mentioned above for the generator loss, on simulated images. A natural choice would thus be:

$$L_{D_\varphi}(x,b,\varphi) := D_\varphi(Y_{\text{sim}}(x,b)) - D_\varphi(Y_{\text{real}}).$$

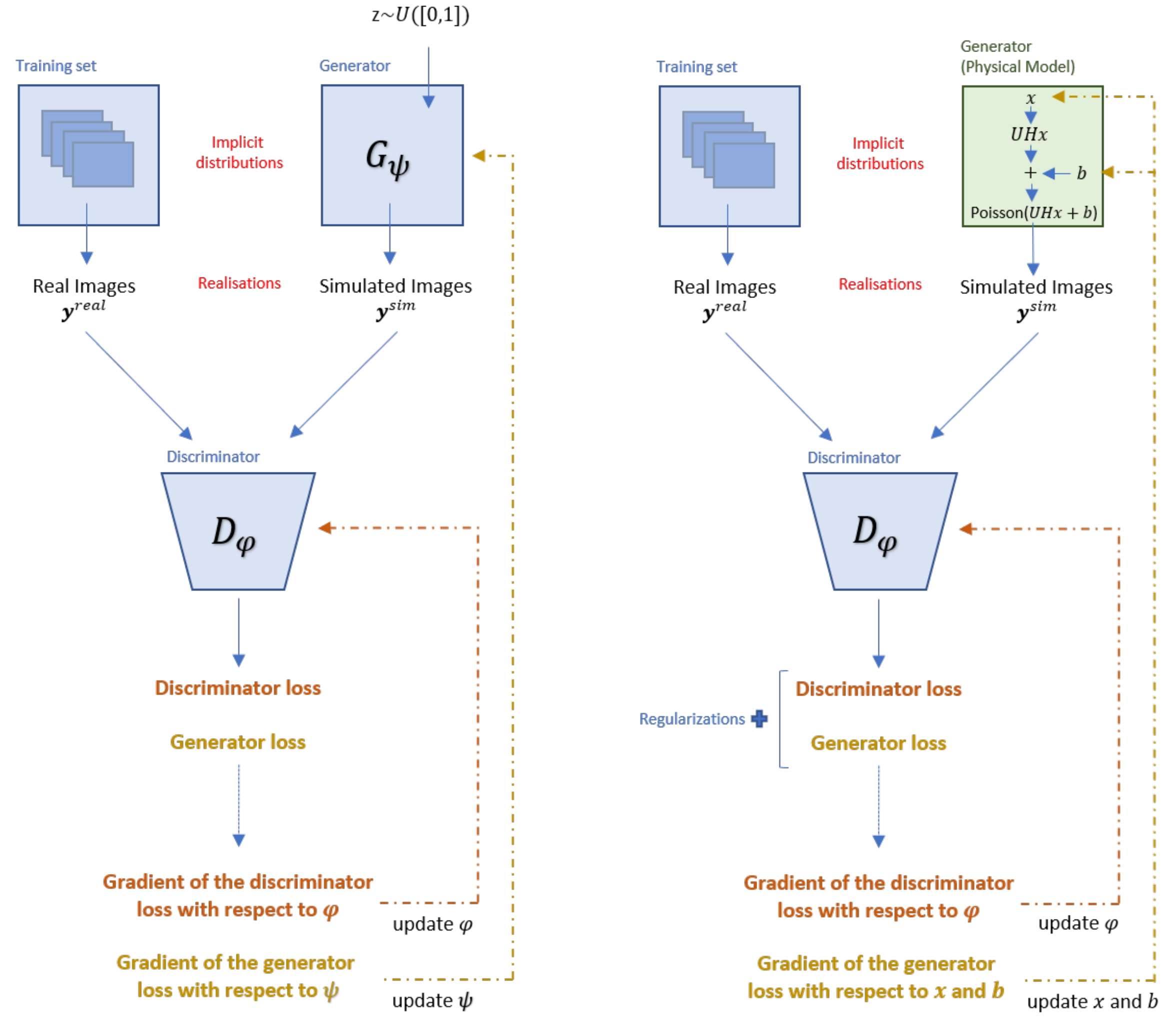


Figure 2: Comparison between a standard GAN architecture (left) and FluoGAN (right). In both cases the generator G_ψ (or physically-inspired generator) is trained in an adversarial manner against the discriminator D_φ. FluoGAN can be considered as a specialised GAN architecture where the additional knowledge available on the direct problem enables to replace G_ψ by a more interpretable physics-based function.

To avoid the well-known convergence instabilities of such adversarial training, it is rather classical to further introduce a regularisation term called *gradient penalty*, to promote gradient updates with norm close to 1, see [12]. This is in fact a penalised formulation of the constraint on the discriminator to be 1-Lipschitz which was observed to correspond to the minimisation of the 1-Wasserstein distance (10) by duality arguments [2]. Note, that the gradient penalty term should be applied on the domain where the discriminator is applied. To enforce that, it is therefore applied on images Y_{mix} uniformly chosen at random between simulated images and real images as: $Y_{\text{mix}}(x,b) := \eta Y_{\text{real}} + (1-\eta) Y_{\text{sim}}(x,b)$ with $\eta \sim \mathcal{U}([0,1])$, where $\mathcal{U}([0,1])$ stands for the uniform distribution on $[0,1]$.

Introducing now parameters $(\gamma, \lambda, \lambda_1, \lambda_2) \in \mathbb{R}^4_{\geq 0}$ and $\delta \in \{0,1\}$ (we make this choice to assess the effect of the presence of the discriminator term in the loss) and incorporating in (11) the physical-inspired regularisation terms discussed in (3.2) we thus consider regularised version of (11)-(12) which reads:

$$\min_{x\in\mathbb{R}^n_{\geq 0},\ b\in\mathbb{R}^m_{\geq 0}} \ \mathbb{E}\left[\frac{\gamma}{2}\ \|Y_{\text{sim}}(x,b) - \mathbb{E}[Y_{\text{real}}]\|^2 - \delta D_\varphi(Y_{\text{sim}}(x,b))\right] + \lambda_1 ||x||_1 + \frac{\lambda_2}{2}||\nabla b||^2 \tag{14}$$

$$\min_{\varphi\in\Phi} \ \mathbb{E}\left[D_\varphi(Y_{\text{sim}}(x,b)) - D_\varphi(Y_{\text{real}}) + \lambda\left(\|\nabla_y D_\varphi(Y_{\text{mix}}(x,b))\| - 1\right)^2\right], \tag{15}$$

where during optimisation an empirical estimation on batches $B \subset \{1,\dots,T\}$ drawn uniformly at random can be performed for approximating the expected values for both Y_{sim} and Y_{real}. Recalling (13), we observe that in (14) we are enforcing distributional equality using (9) and a GAN discriminator with parameters to be trained via (15), but, furthermore, we are enforcing physically-motivated regularisation. Practically, and similarly as for GANs, problem (14) is minimised at each epoch for few (say, k_{G}) iterations, then, similarly, (15) is minimised for k_{D} iterations by freezing the quantities $x^{k_{\text{G}}}, b^{k_{\text{G}}}$ previously computed. This alternate procedure continues till convergence. In order to resort to gradient-based solvers for the problem (14), some details should be given concerning the computation of gradients w.r.t. the variables x and b due to the non-linear dependence between (x,b) and $Y_{\text{sim}}(x,b)$ through the Poisson model (6).

4.2. Computing gradients for a Poisson random variable

For optimising the loss function in (14) we need to compute quantities in the form $\nabla_x\mathbb{E}\left[f(Y_{\text{sim}}(x,b))\right]$ and $\nabla_b\,\mathbb{E}\left[f(Y_{\text{sim}}(x,b))\right]$ where $f(Y_{\text{sim}}(x,b)) = \frac{\gamma}{2}\ \|Y_{\text{sim}}(x,b) - \bar{y}_{\text{real}}\|^2 - \delta D_\varphi(Y_{\text{sim}}(x,b))$ and the dependence of $Y_{\text{sim}}(x,b)$ on both x and b is non-linear due to the Poisson model (5). The gradient of the expected value of a Poisson random variable with respect to its parameter(s) can be computed directly. Let $z \in \mathbb{R}^m_{>0}$ and $Q \sim \text{Poisson}(z)$. For every component $i = 1,\dots,m$, q_i is then a discrete random variable with univariate Poisson density given by $p(q_i = k|z_i) : k \mapsto \frac{1}{k!}e^{-z_i}z_i^k$, for $k \in \mathbb{N}$. Let now v_k be the function defined by $v_k : t \mapsto p(q_i = k|t) = \frac{1}{k!}e^{-t}\,t^k$. Clearly, v_k is differentiable on $\mathbb{R}_{>0}$ for all $k \in \mathbb{N}$. We distinguish two cases:

- For $k \geq 1$ there holds:

$$v_k'(t) = \frac{-e^{-t}t^k}{k!} + \frac{e^{-t}kt^{k-1}}{k!} = v_{k-1}(t) - v_k(t) \tag{16}$$

- For $k = 0$ we have:

$$v_0'(t) = -e^{-t} = -v_0(t). \tag{17}$$

For every $i = 1,\dots,m$, let now be $A_i \in \mathbb{R}$ such that $A_i > \max(z_i, 1)$. For $x \in [0, A_i]$ we notice that the following properties hold:

- Since the sequence $(v_k(x))_k$ comes from a Poisson density, it has finite sum and for all $k \in \mathbb{N}$, v_k is differentiable on $[0, A_i]$.

- For $k \geq 1$ there holds $v_k'(x) = \frac{1}{k!}e^{-x}x^{k-1}(k-x) < \frac{A_i^{k-1}}{(k-1)!}$. The sequence $(v_k'(x))_k$ is thus dominated by a summable sequence.

Let now $f : \mathbb{R}^m_{>0} \to \mathbb{R}$ be a continuous and bounded function. By above, the dominated convergence theorem applies hence it is possible to switch the derivation with the integral (sum) signs to obtain:

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(q)] = \frac{\partial}{\partial z_i}\sum_{k=0}^{\infty} \mathbb{E}[f(q)|q_i = k]p(q_i = k|z_i) = \sum_{k=0}^{\infty}\mathbb{E}[f(q)|q_i = k]\frac{\partial}{\partial z_i}p(q_i = k|z_i)$$

We can now use the recursion formulas (16)-(17) for v_k' as follows:

$$\begin{aligned}\frac{\partial}{\partial z_i}\mathbb{E}[f(q)] &= \sum_{k=0}^{\infty}\mathbb{E}[f(q)|q_i = k]v_k'(z_i)\\ &= \mathbb{E}[f(q)|q_i = 0]v_0'(z_i) + \sum_{k=1}^{\infty}\mathbb{E}[f(q)|q_i = k](v_{k-1}(z_i) - v_k(z_i))\\ &= -\mathbb{E}[f(q)|q_i = 0]v_0(z_i) - \sum_{k=1}^{\infty}\mathbb{E}[f(q)|q_i = k]v_k(z_i)\\ &+ \sum_{k=0}^{\infty}\mathbb{E}[f(q)|q_i = k+1]v_k(z_i)\end{aligned}$$

A variable change is then applied. Denoting by $1_i = \{\delta_{i,j}\}_{j=1}^n \in \mathbb{R}^n$ (with $\delta_{i,j}$ being the Dirac delta function) the vector of only zeros except a one in the i-th position, we have:

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(q)] = -\sum_{k=0}^{\infty}\mathbb{E}[f(q)|q_i = k]p(k|z_i) + \sum_{k=0}^{\infty}\mathbb{E}[f(q + 1_i)|q_i = k]p(q_i = k|z_i)$$

Finally, the i-th component of the gradient of expected value of $f(q)$ can be simply written as a finite difference :

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(q)] = \mathbb{E}[f(q + 1_i) - f(q)] \tag{18}$$

This expression is easily interpreted: increments of z_i correspond to increments of 1 of the component q_i in expectation. This is in fact a consequence of the definition of Poisson law, whose parameter equals its expected value.

Unlike other methods such as the one of score functions in [18], formula (18) can deal with the case $z_i = 0$. However, its application is computationally demanding: computing the expectancy in (18) by empirical means of B realisations, requires in fact $B(m + 1)$ evaluations of f. To reduce computations, and by assuming that f is differentiable, we can make however the following approximation:

$$f(q + 1_i) - f(q) \simeq \frac{\partial f}{\partial q_i}(q) \tag{19}$$

which can be plugged in (18) and estimated empirically by means of B samples:

$$\nabla_z \mathbb{E}[f(q)] \simeq \mathbb{E}[\nabla_q f(q)] \simeq \sum_{t=1}^{B} \nabla_q f(q^t),$$

which provides a handy way of approximating the desired quantity.

Remark. The approximation (19) is valid only when $\frac{\partial f}{\partial q_i}$ does not change too much between $f(q)$ and $f(q + 1_i)$. If all the second derivatives of f are bounded, then the second-order error of this approximation can be indeed bounded. However, in general f may not be twice differentiable. In our context, this condition can thus be relaxed by assuming a sufficiently fine quantisation compared to the variations of $\partial f / \partial q_i$.

4.3. Discriminator architecture

Inspired by [35], we use as a discriminator a standard convolutional neural network $D_\varphi : \mathbb{R}^n \to [0, 1]$ with 3 main layers and 2 fully connected layers, as shown in Figure 3. Each convolutional layer is followed by a max pooling layer and a non-linear ReLU (rectified linear unit) activation function. The number of channels increases and is doubled after each convolutional layer. The last layer is a sigmoid function that returns an output value in the range $[0, 1]$. This setup is a common choice for this type of networks since it gives more capacity to the network while reducing the resolution of the intermediate layers. However the number of convolutional layers is limited by the input size and the size of the kernel. Indeed it is not possible to apply a convolutional kernel on an image smaller than the kernel. For this reason there are only 3 convolutional layers here. In practice and for the size of images considered, we have seen that the chosen architecture gives enough capacity to the network. The discriminator is trained to yield 0 given a simulated image as input and 1 in case the input is a real image. However, exactly like the original GAN discriminator, it is used passively to improve the quality of simulated images.

4.4. Optimisation algorithms

The concurrent minimisation problems (14) and (15) have different regularisation terms. To train the generator, a FISTA type proximal gradient algorithm [3] with a stochastic computation of the gradient (in the spirit of [23]) is considered to deal effectively with the non-smoothness of the ℓ_1 norm and the non-negativity constraints. The choice of the learning rates has critical importance to achieve convergence. If such parameter is too large than the algorithm may completely diverge, while if it is too small the convergence is too slow. Such parameter depends also on the proportion of noise and signal in the data. More precisely there is a trade-off between the sparsity constraint parameter and the learning rates. For our experiments, we choose the learning rates manually, making sure to obtain a converging algorithm. We tested also backtracking strategies to automatise this choice. In practice, however,

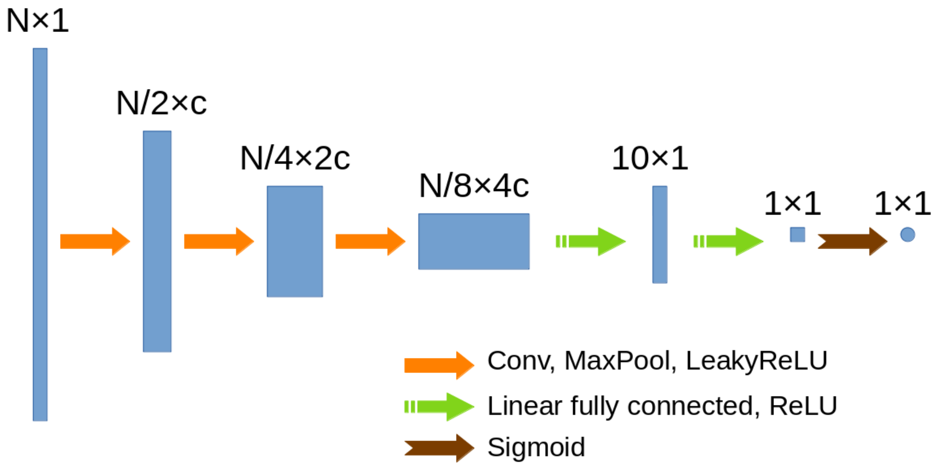


Figure 3: Discriminator architecture

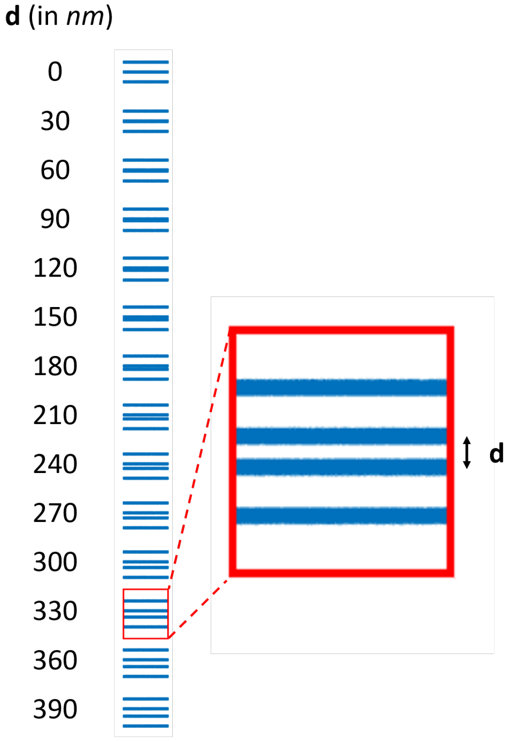
the constant values chosen here after empirical tuning seem to be general enough to process all the different kinds of data we present in our results without any change. For the discriminator loss in (12), we do not have the same differentiability problem and the choice of the corresponding learning rate is not data-constrained. To perform such optimisation, we thus used stochastic ADAM algorithm, which is a common choice to train a neural network of the form described in Section 4.3. In Table 1 we report the algorithmic parameters (batch size B for ADAM, learning rate τ_φ for the optimisation of (15) and initial learning rates τ_x and τ_b for the optimisation in (14)) used in the experiments below. Note that due to the use of backtracking, a possible overestimation of both τ_x and τ_b can be corrected along the iterations.

B	32
τ_φ	10^{-6}
τ_x	1
τ_b	1

Table 1: Optimisation parameters.

5. Numerical results

The proposed method is applied to three different types of data in order to evaluate the performance of FluoGAN and to compare it with state-of-the-art approaches, namely COL0RME [32, 30] and SRRF [14]. We first use synthetic data (see section 5.1) by simulating standard microscopic data acquired using conventional fluctuating/blinking fluorophores. This type of data is useful because it allows us to access the true signal and evaluate the reconstruction quantitatively. Next, we apply FluoGAN on images of a phantom sample with known structure acquired by a real fluorescence microscope. The results are presented in section 5.2. Finally, for a more difficult reconstruction, we applied FluoGAN on images of a real biological sample and more specifically of the unicellular algae *Ostreopsis cf. Ovata*, acquired by an epifluorescent microscope (see Section 5.3).



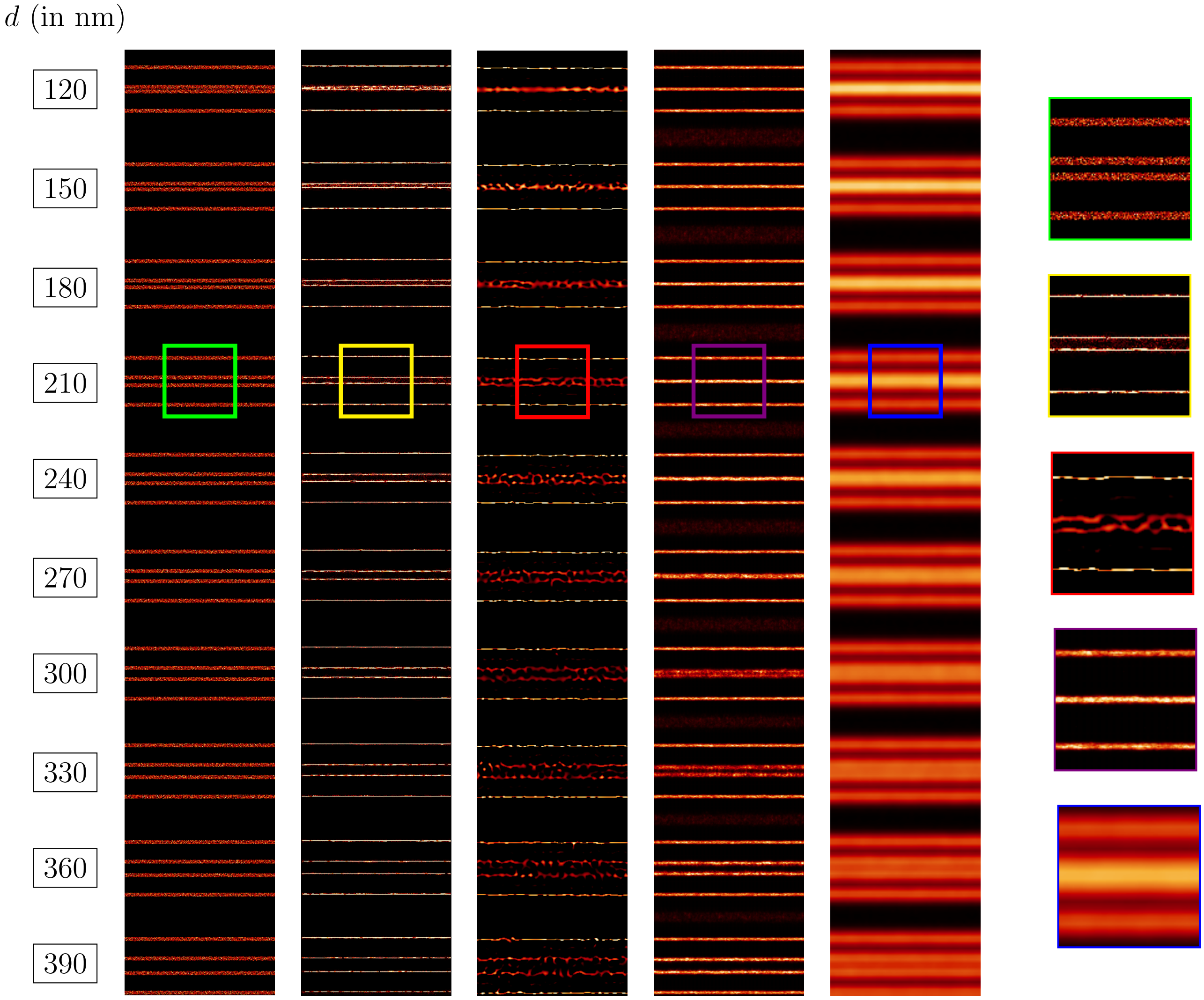
schema.png

Figure 4: Spatial structure of simulated/real calibrated sample (ARGO-CR slide, Argolight)

5.1. A simulated ARGO-CR-*type validation dataset*

We started by applying FluoGAN to synthetic images for simple reconstruction quality evaluation. The synthetic images have been manually generated after considering parameters of the microscope configuration as well as properties of the sample (e.g. fluctuation behaviour, out-of-focus fluorescence, etc.). The spatial pattern used is shown in Figure 4 and is similar to the pattern of a calibrated sample ARGO-CR (Argolight, Pessac, France). To be more precise, the pattern used consists of 14 sets of parallel lines of 100nm width. The separation distance d (center-to-center distance) between the two middle lines of each set is gradually increasing with a rate of 30nm. The knowledge of these details allows us to exactly quantity the resolution level reached by each approach.

For this first experiment, we simulate the temporal fluctuations using the Poisson model in (3) and make a video of 500 frames, assuming a frame rate of 100 frames per second (fps) to avoid photobleaching. The pixel size is equal to 100nm while the full width at half maximum (FWHM) of the PSF is 324nm. A spatially varying background is added to simulate the presence of the out-of-focus fluorescent molecules and an average bleaching time (i.e. time which the fluorophore stays emissive) of 20s has been used. Reconstructions of the simulated diffraction-limited data are available for the FluoGAN, COL0RME ([32, 30]) and SRRF [14] methods in Figure 5, using a super-resolution factor $L = 6$ for all three approaches. In the method COLORME [32, 30] we use for this



(a) GT (b) FluoGAN (c) COL0RME (d) SRRF (e) $\bar{y}$

Figure 5: Numerical results on simulated dataset with zooms, with enhanced contrast for better visualization. The separation distance d, used to measure the resolution, is given for each set of lines. The FWHM of the PSF in this experiment is equal to 324nm.

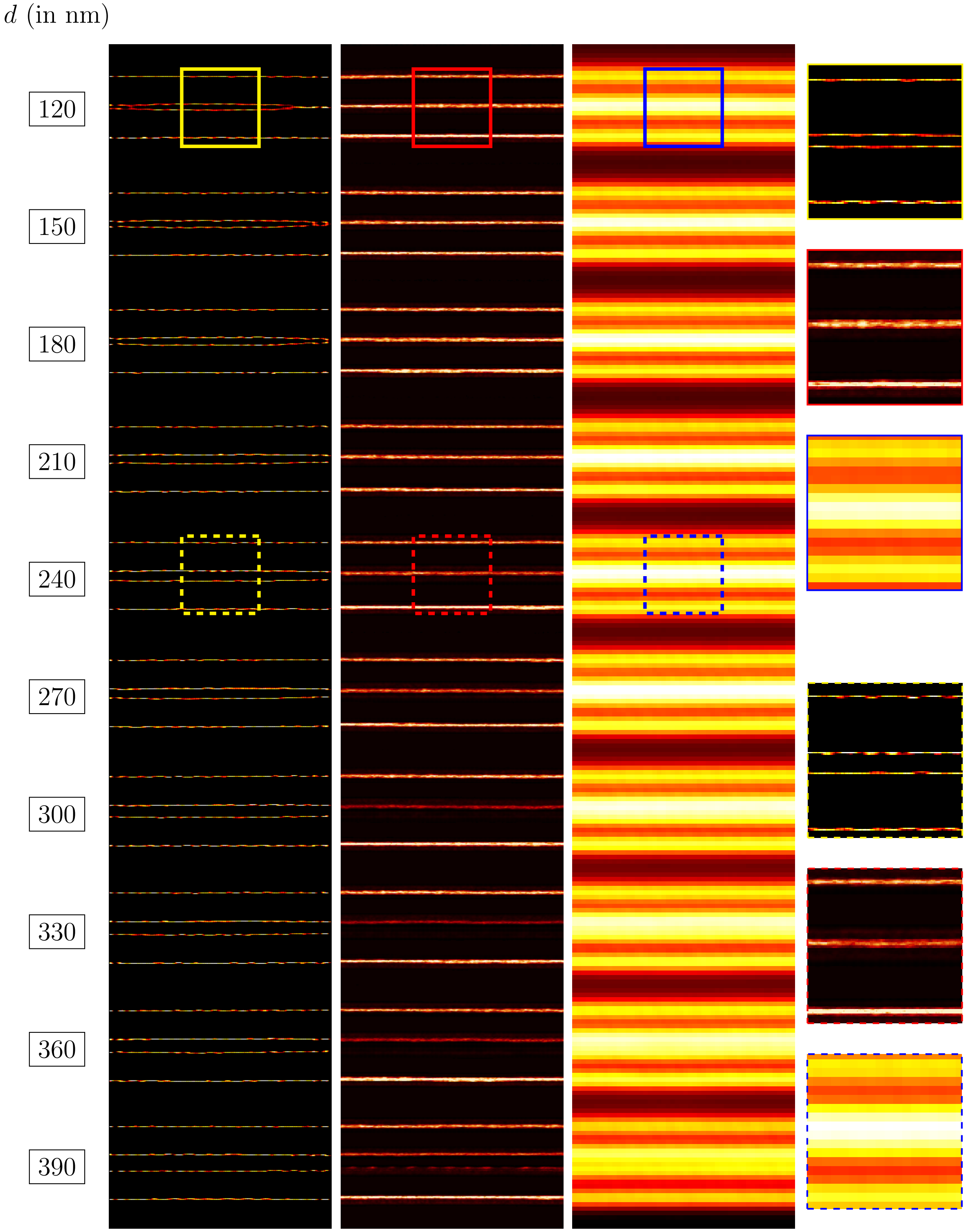
experiment the ℓ_1-norm penalty. Regarding the method SRRF, we make use the freely available Fiji plugin ([14]). Especially, for the temporal analysis of the radiality image sequence we are using the auto-correlation analysis similar to the one used in [8]. From the reconstruction shown in Figure 5, it is clear that the proposed method achieves better results than both COL0RME and SRRF. FluoGAN is able to reconstruct two parallel lines with separation distance of 120nm between their centers or only 20nm between their closest edges (given that the with of the lines is 100nm). COL0RME is more resolutive than SRRF, achieving a resolution of $80-110$nm computed as the ability to reconstruct two lines with closest edge-to-edge distance of such value, while SRRF reaches a resolution of around 200nm measured in the same way. Finally, FluoGAN and COL0RME more than SRRF estimate real intensity values, while SRRF preserves better the "continuous" structure of the sample.

5.2. Real ARGO-CR *dataset*

In the second experiment, we consider images of a ARGO-CR calibrated sample acquired by an epifluorescence microscope. Images were acquired on an AxioObserver Z1 (Zeiss, Germany) with 63x/1.15W Korr LD CApochromat objective, ORCA Flash 4.0 camera (Hamamatsu, Japan), 540570nm LED excitation and 581619nm emission filter. The spatial pattern is similar to the one used in our simulated data, with the distance between the two middle filaments of each set increasing by 30nm and varying from 0 to 390nm (see Figure 4 for a graphical representation). In this second example the resolution reached by the methods can be estimated under real conditions. Only FluoGAN and SRRF are applied to these data, as due to a slight sample shift during acquisition, these data are not adaptable for COL0RME, which is based on the estimation of temporal covariance matrices. The pixel size of the sCMOS camera used is equal to 103nm, while the FWHM of the PSF is estimated to be 318nm. Only 500 images were acquired with a frame rate of 10fps, i.e. an acquisition time of less than 1 minute. For the reconstruction, a super-resolution factor $L = 6$ was used for both methods. For FluoGAN, the 500 acquired images constituted the training set while 5000 iterations were performed in order to obtain the reconstruction presented in Figure 6. Compared to SRRF, FluoGAN achieves significantly better resolution levels. Structures reconstructed by FluoGAN are a bit thinner than real ones; this is the reason why it is able to separate filaments that are slightly overlapping, i.e. when the centre-to-centre distance is equal to 60nm.

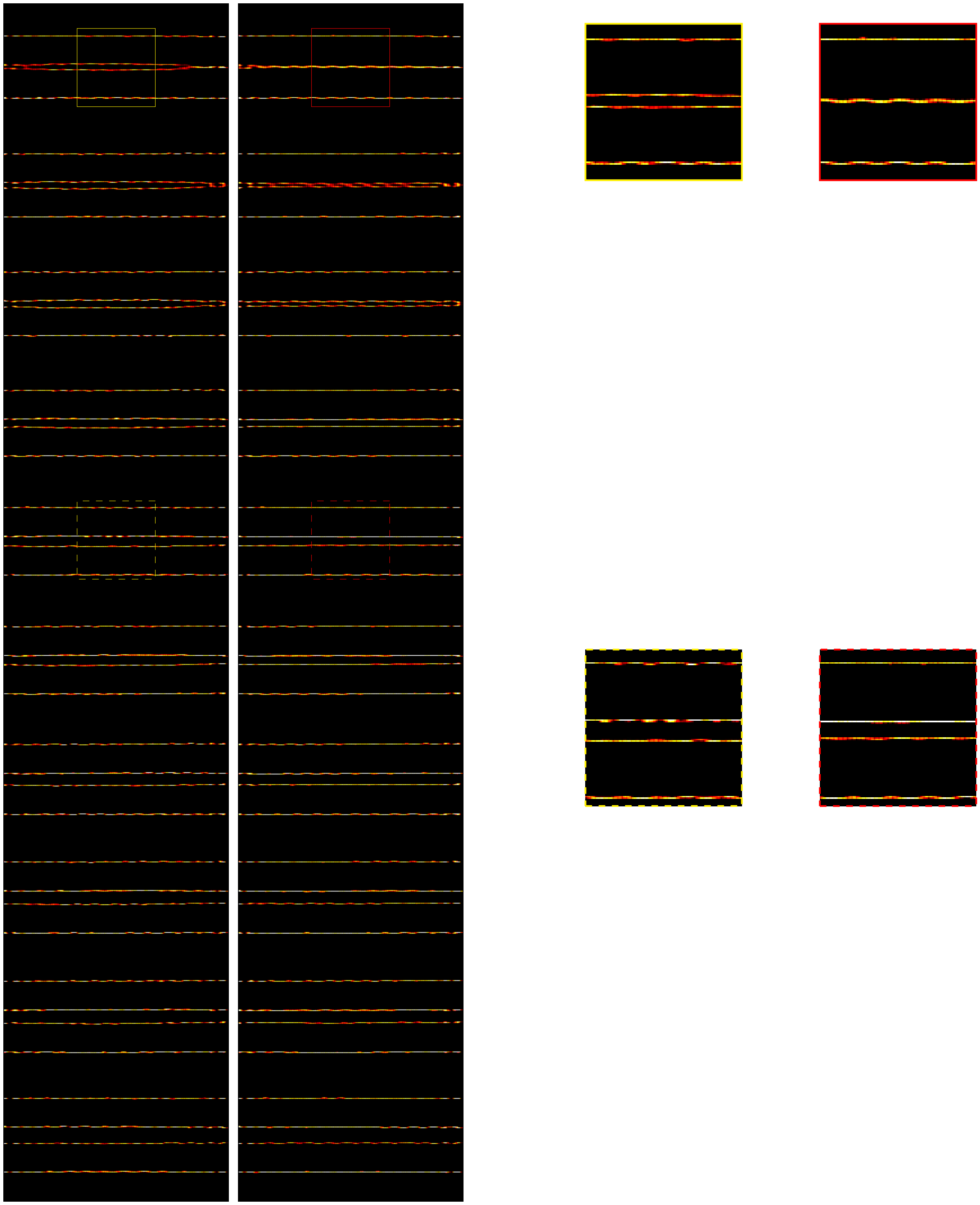
We then tested the performance of FluoGAN with ($\delta = 1$) and without ($\delta = 0$) the presence of the discriminator in problem (14) to assess how important is role is in the overall estimation process. The two results are available in Figure 7. We observed that even if the ℓ_2 distance (9) allows to recover the main pattern, the explicit presence of the discriminator significantly improves the reconstruction. The result with no discriminator can nonetheless be used as an educated guess for the initialisation of FluoGAN. In Figure 8 we provide a schema assessing the reconstrution precision and the resolution achieved in both cases. We note that without discriminator the distance between filaments reconstructed by FluoGAN is systematically under-estimated. The discriminator plays here the double role of both allowing the separation of filaments with 120 nm resolution and of improving the reconstruction precision for all filaments. In Table 2 the computational costs required to run FluoGAN on image stacks of different sizes with/without the discriminator are reported.

In Figure 9 we report the convergence graphs along the iterations. Unlike other optimisation problems, it is difficult to visualise convergence since simulator and discriminator have adverse goals and the difficulty is to maintain both in competition preventing one from overtaking the other.



(a) FluoGAN (b) SRRF (c) $\bar{y}$

Figure 6: Numerical results for *ARGO-CR* data with zooms. The distance d is given for each set of lines. The FWHM of the PSF is estimated to be equal to 270nm.

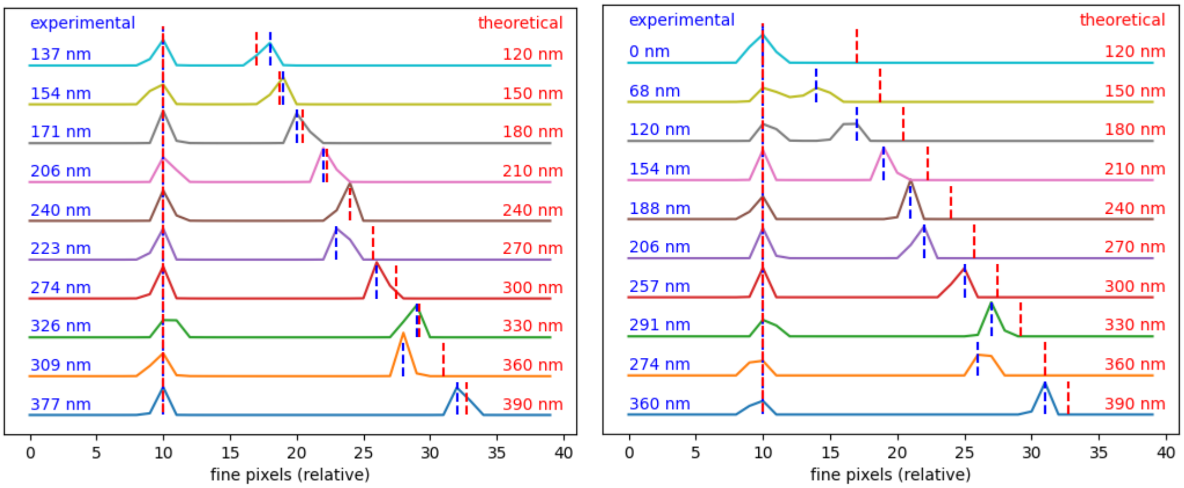


(a) FluoGAN $\delta = 1$ (b) FluoGAN $\delta = 0$

Figure 7: Comparison of FluoGAN results with (left, $\delta = 1$) and without (right $\delta = 0$) discriminator. Since the discriminator relies on covolutionnal layers, it improves first the centre of the image. Due to the boundary effects related to the width of convolution kernels, boundary regions are not compared.

input images size	FluoGAN ($\delta = 1$)	FluoGAN ($\delta = 0$)
50×50	10 min	3 min
50×266	60 min	20 min

Table 2: Computation time for 5000 iterations of FluoGAN on a small GPU Nvidia GeForce GTX 950M from 2015 with 2GB of dedicated memory.



(a) FluoGAN, $\delta = 1$ (b) FluoGAN, $\delta = 0$

Figure 8: Distances between filaments in Figure 6 reconstructed by FluoGAN, with (left) and without (right) discriminator. Each coloured 1D profile corresponds to the y-averaged intensity of the reconstructed image plotted along the x-axis (excluding boundaries). Taking the left-most peak as a reference, the red dotted lines denote the theoretical location of the filament to reconstruct, while the blue dotted lines is placed in correspondence of the maximum value of the filament reconstructed by FluoGAN.

5.3. Ostreopsis images

To test FluoGAN on challenging real biological samples, we considered a dataset of the unicellular alga *Ostreopsis cf. Ovata* (see Figure 10a). Such dataset combines several difficulties. First, it shows a 20m thick sample which can exhibit a strong out-of-focus signal which is clearly the case here for the microtubules staining as they form a cortical structure all around the alga. Another difficulty is the reduced transparency and none-negligible turbidity, which induces PSF distortion. Due to those complications, confocal microscopes cannot acquire through the whole thickness of the alga. To avoid this, we considered epifluorescence acquisitions. The flexibility and the data-adaptivity of FluoGAN allows to deal with such data, whereby other approaches fail.

For the experimental design, microtubules were stained with TRITC dye, but also with Hoechst for DNA (for details see [33]). We focused on the tip of a microtubule bundle at the ventral pole (see Figure 10b). Epifluorescence images have been acquired with the same protocol as for ARGO-SR sample (see details in section 5.2).

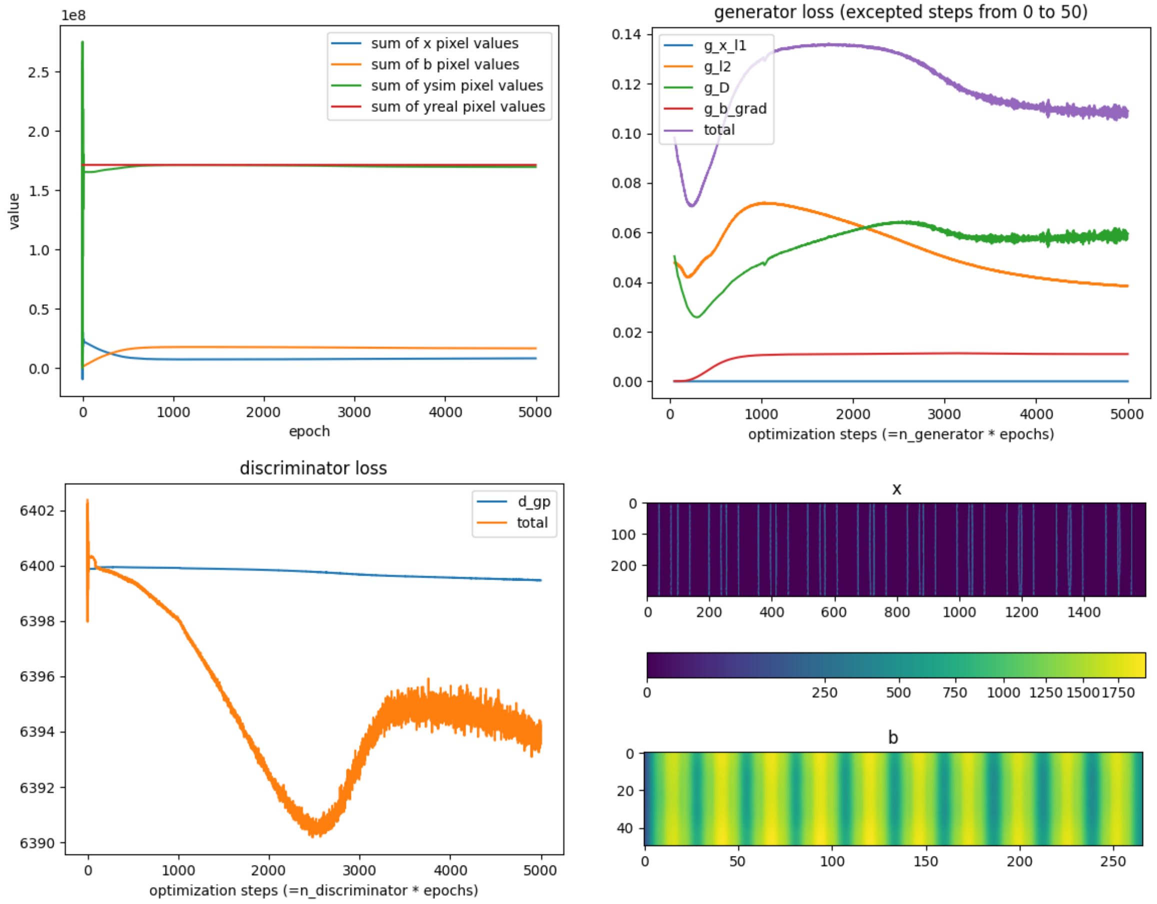
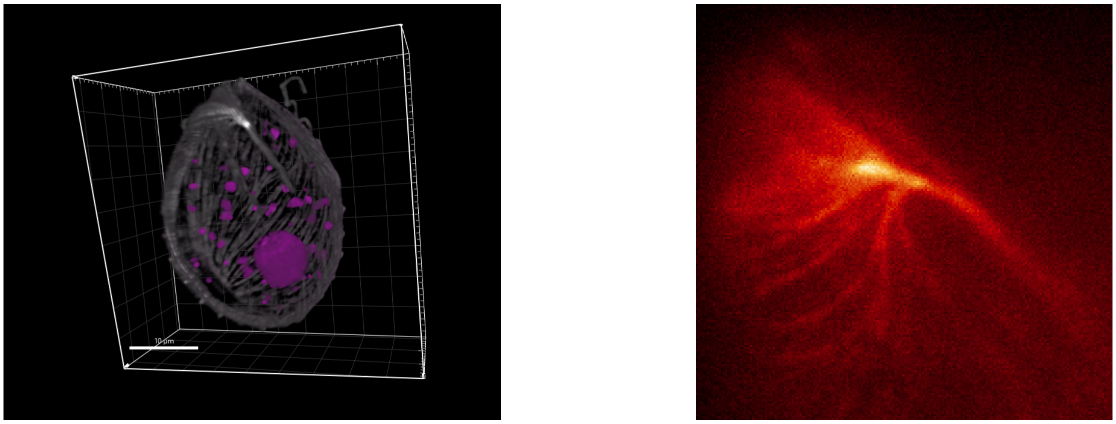


Figure 9: Top left : sum of pixel values (i.e., total photon count) VS. iteration number. After about 1000 iterations the sum of pixels for simulated images (y^{sim}) equals the sum of pixels for real input images (y^{real}). Top right and bottom left : balance between simulator and discriminator training. Top right : contributions to generator loss function : discriminator distance, ℓ^2 distance, regularizations on b and x. Bottom left : discriminator loss function (orange) including the gradient penalty regularization (blue). Bottom right : super resolution image x and reconstructed background b.

Sensor's pixel size	103nm
FWHM of the PSF (estimated)	325nm
Undersampling rate	6
Iteration number	1000
Number of images in training set	500

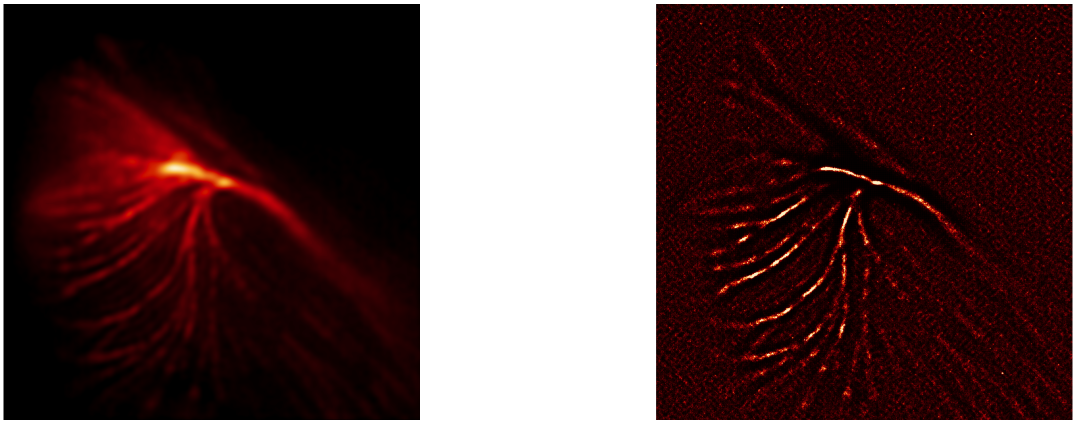
Table 3: Parameters used for *Ostreopsis* Fig. 10.

The reconstruction performed by FluoGAN is shown in Figure 10c, while the parameters used are reported in Table 3. We compare the results obtained by FluoGAN with the ones obtained by SRRF. Regarding SRRF and differently from the two previous experiments, gradient weighting is performed, as suggested by the authors, to deal with the low signal-to-noise ratio (SNR) regime. The reconstruction obtained by SRRF is reported in 10d. Compared to the solution obtained by SRRF, the FluoGAN result shows less background artefacts and promotes better continuity of filaments and intensity reconstruction. Furthermore, the reconstruction obtained by FluoGAN reduces the appearance of ghost filaments.



(a) 3D image of *Ostreopsis* with microtubules (white) and DNA (magenta).

(b) Low-resolution epifluorescent image (sample average) of the microtubules tip.



(c) FluoGAN reconstruction of Fig. 10b.

(d) SRRF reconstruction of Fig. 10b.

Figure 10: Real *Ostreopsis cf Ovata* data.

6. Conclusions

We presented FluoGAN, a novel framework for fluctuation-based super-resolution fluorescence microscopy combining the physical modelling of optical system with data-driven adversarial learning. Hand-crafted (often sparsity-based) regularisation approaches formulate the super-resolution inverse problems by means of tailored regularisation terms, thus requiring time-consuming model and parameter tuning. On the other hand, FluoGAN computes the desired super-resolved image with high precision along with a background image containing out-of-focus molecule and ambient fluorophores by comparing, in a suitable sense, the empirical distribution of observed data with the one of samples generated by a physically-grounded simulator. Due to its alternate minimisation formulation, the proposed approach is flexible as it allows to the possible introduction of further regularisation terms such as sparsity, smoothness and non-negativity constraints. In comparison to standard GANs and inspired by recent work on Cryo-EM [13], FluoGAN replaces the model-blind generator network with a simulator model encoding biophysical expertise in its structure and having as learnable parameters the desired quantities of interest. On simulated data, the proposed approach allows to achieve better resolutions than standard model-based and state-of-the-art approaches. On real data, we first proved the efficiency of our algorithm on a 2D phantom, then we validated FluoGAN on real challenging *Ostreopsis* data. Future work should be addressed towards the proof of rigorous reconstruction guarantees certifying the quality of the solution at convergence, which, at the moment, remains empirical. Another prospective part is to implement FluoGAN to improve resolution for full 3D super-resolution problems.The FluoGAN code is available on GitHub at the address https://github.com/cmayeul/FluoGAN.

Acknowledgments

The authors would like to thank D. Velasquez and S. Castagnetti from the Developmental Biology Laboratory of Villefranche sur Mer (LBDV) who kindly prepared and provided the experimental samples. Image acquisitions were conducted in the microscopy platform PIM (member of MICA microscopy platform labelled by IBiSA), supported by EMBRC-France, whose French state funds are managed by the ANR within the Investments of the Future program under reference ANR-10-INBS-0. The work of VS and LBF has been supported by the French government, through the 3IA Côte dAzur Investments in the Future project managed by the National Research Agency, grant ANR-19-P3IA-0002. VS, LBF and LC acknowledge the support by the ANR MICROBLIND grant ANR-21-CE48-0008. LC acknowledges the support received by the EU H2020 RISE program NoMADS, GA 777826, the one received by the GdR ISIS grant SPLIN and by the IEA CNRS grant VaMOS.

References

[1] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhuser Basel, 2005.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223. PMLR, 2017.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 2(1), 2009.

[4] E. Betzig et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793), 2006.

[5] G. Dardikman-Yoffe and Y. C. Eldar. Learned SPARCOM: unfolded deep super-resolution microscopy. *Opt. Express*, 28(19), 2020.

[6] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure and Appl. Math.*, 57(11), 2004.

[7] Y. Deng, M. Sun, P. Lin, J. Ma, and J. Shaevitz. Spatial covariance reconstructive (SCORE) super-resolution fluorescence microscopy. *PloS one*, 9, 04 2014.

[8] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *PNAS, 106*, 106, 2009.

[9] S. Gazagnes, E. Soubies, and L. Blanc-Fraud. High-density molecule localization for super-resolution microscopy using CEL0 based sparse approximation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 28–31, 2017.

[10] A. Girsault et al. SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging. *PLOS ONE*, 11(9), 2016.

[11] I. Goodfellow et al. Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of Wasserstein GANs. *CoRR*, abs/1704.00028, 2017.

[13] H. Gupta, M. T. McCann, L. Donati, and M. Unser. CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning. *IEEE Transactions on Computational Imaging*, 7, 2021.

[14] N. Gustafsson, S. Culley, G. Ashdown, D. M. Owen, P. M. Pereira, and R. Henriques. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature communications, 7(1)*, pages 12471–12471, 2016.

[15] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation, 2017.

[16] H. Li and J. Vaughan. Switchable fluorophores for single-molecule localization microscopy. *Chem. Rev.*, 118, 2018.

[17] M. Lopez-Martinez, G. Mercier, K. Sadiq, O. Scherzer, M. Schneider, J. C. Schotland, G. J. Schütz, and R. Telschow. *Inverse Problems of Single Molecule Localization Microscopy.* Springer International Publishing, Cham, 2021.

[18] S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih. Monte Carlo Gradient Estimation in Machine Learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

[19] S. Mukherjee, A. Hauptmann, O. ktem, M. Pereyra, and C.-B. Schnlieb. Learned reconstruction methods with convergence guarantees: A survey of concepts and applications. *IEEE Signal Processing Magazine*, 40(1):164–182, 2023.

[20] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2), 1995.

[21] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4), 2018.

[22] M. Ovesn, P. Kek, J. Borkovec, Z. vindrych, and G. M. Hagen. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging . *Bioinformatics*, 30(16), 2014.

[23] L. Rosasco, S. Villa, and B. C. Vũ. Convergence of stochastic proximal gradient algorithm. *Applied Mathematics & Optimization*, 82(3):891–917, 2020.

[24] M. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. methods*, 3, 11 2006.

[25] D. Sage et al. Super-resolution fight club: Assessment of 2D & 3D single-molecule localization microscopy software. *Nat. Methods*, 16, 2019.

[26] O. Solomon, Y. C. Eldar, and M. Mutzafi, M.and Segev. SPARCOM: Sparsity Based Super-resolution Correlation Microscopy. *SIAM J. Imaging Sci.*, 12(1), 2019.

[27] E. Soubies, L. Blanc-Féraud, and G. Aubert. A Continuous Exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM J. on Imaging Sci.*, 8(3), 2015.

[28] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for ℓ_2-ℓ_0 minimization. *SIAM J. on Optim.*, 27(3), 2017.

[29] A. Speiser et al. Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat. Meth.*, 18(9), 2021.

[30] V. Stergiopoulou, L. Calatroni, J. H. de Morais Goulart, S. Schaub, and L. Blanc-Fraud. COL0RME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation. *Biological Imaging*, 2, 2022.

[31] V. Stergiopoulou, L. Calatroni, S. Schaub, and L. Blanc-Féraud. 3D Image Super-Resolution by Fluorophore Fluctuations and MA-TIRF Microscopy Reconstruction (3D-COL0RME). In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022.

[32] V. Stergiopoulou, J. H. de Morais Goulart, S. Schaub, L. Calatroni, and L. Blanc-Fraud. COL0RME: Covariance-Based ℓ_0 Super-Resolution Microscopy with Intensity Estimation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021.

[33] D. Velasquez. *Cellular and molecular characterization of Ostreopsis cf. ovata cell cycle.* Theses, Sorbonne Université, May 2021.

[34] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. on Image Process.*, 19(11), 2010.

[35] M. Zehni and Z. Zhao. MSR-GaN: Multi-segment reconstruction via adversarial learning. *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021.