Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)

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

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, Université Côte d'Azur, Sophia Antipolis, France
[2]Sorbonne Université, 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 Network (GAN) for the problem of image deconvolution. Differently from standard approaches combining a least-square data fitting term based on one (long-time exposure) image with sparsity-promoting regularisation terms, FluoGAN relies on a data fitting term defined as a distribution distance between the fluctuating observed timelapse (short-time exposure images) and the generative model. The distance between these two distributions is computed using adversarial training of two competing architectures: a physics-inspired *generator* simulating the fluctuation behaviour as a Poisson process of the observed images combined with blur and undersampling, and a standard convolutional *discriminator*. FluoGAN is a fully unsupervised approach requiring only a fluctuating sequence of blurred, undersampled and noisy images of the sample of interest as input and it can be complemented with prior knowledge on the desired solution such as sparsity, non-negativity etc. After having described in depth the main ideas behind FluoGAN, we formulate the corresponding optimisation problem and report several results on a simulated and real phantoms used by microscopy engineers to quantitatively assess spatial resolution. The comparison of FluoGAN with state-of-the-art methodologies shows unprecedented resolution and allows for high-precision reconstruction of very fine structures in challenging real *Ostreopsis cf Ovata* data.

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 this limit 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 *fluorescence microscopy*, which in a nutshell, uses special proteins, called *fluorescent dyes* or *fluorophores*. They bind to the molecule of interest in the sample and, when exposed to light excitation at specific wavelengths, emit photons which are then captured by CCD or sCMOS camera or photo-multiplicator sensor after passing through special lenses and optical devices. The resolution of the captured image is then limited due to diffraction of the light. The diffraction (Airy) pattern acts as a convolution operator (the Point Spread Function PSF) on the emitters giving a blurred 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 system and, possibly, some statistical prior information on the noise in the system. More sophisticated approaches aim at estimating partially or entirely the PSF and the noise moments, together with the desired solution and thus formulate the task as a *blind* inverse problem. In this work, we assume to know (up to a sufficient accuracy) the PSF of the microscope which is typically modelled as a Gaussian function or as an Airy pattern and the statistic of the noise. When formulated in a discrete setting, the deconvolution/super-resolution inverse problem further introduces a 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), where the difference in size is 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 undersampled vectorised image $y \in \mathbb{R}^m$ and given $H \in \mathbb{R}^{m\times n}$ to be the circulant-block-circulant matrix corresponding to the 2D PSF of the system and assuming that the noise is additive and white with Gaussian distribution, i.e. $e \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$ with $\mathrm{I} \in \mathbb{R}^{m\times m}$ being the identity matrix, the inverse problem reads:

$$\text{find} \quad (x, b) \qquad \text{s.t.} \qquad y = UHx + e + b, \tag{1}$$

where the operator $U \in \mathbb{R}^{m\times n}$ with $n = L^2 m$ is an under-sampling operator mapping images from the fine to the coarse grid and $b \in \mathbb{R}^m_{\geq 0}$ is an unknown and possibly space-variant background term. Model-based regularisation approaches for (1) formulate the task as a minimisation problem where a sparsity-promoting term (in the image space or w.r.t. some redundant basis) is combined with a least-square data fidelity. Standard compressed-sensing techniques 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 dense samples with high concentration of fluorophores, however, the use of sparsity appear however not suited.

In order to have super-resolution and not only extrapolation due to a regularising additional term, additive information from the sample must be provided. The interplay

between the applied mathematical approach and the microscopy imaging community for solving (1) in such challenging scenarios thus favoured the design of sophisticated acquisition and sample-preparation techniques simplifying the problem in some way. A popular class of such 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 the reconstructed sparse images. Mathematically, such problem can be cast in the form (1) (see [18] for more complicated forward models) and its regularised version can be solved at each acquisition time point. 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 [17] 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 measured frame, which is of course very time consuming in large-scale scenarios. A rather recent approach proposed in the field to mitigate such demanding requirements consists in exploiting, rather than that sparse sequential activation of specific fluorophores over time, the independent stochastic intensity fluctuations of common fluorescence molecules. By acquiring a video of stochastichlly 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 have been obtained. It is the case, for instance, of SOFI (Super-resolution Optical Fluctuation Imaging) [8] where temporal information are integrated by computing a sample covariance image where, by independence, the PSF appears to be squared so that a reduction of a factor $\sqrt{2}$ is obtained on the width of the PSF. Higher-order statistics can also be computed [10]. A similar approach reducing the computational costs of such covariance calculation is SCORE (Spatial COvariance REconstructive) [7] where only pixels with relevant contributions to such calculation are computed. A different class of methods, still taking sequences of fluctuating images as an input but based on a different type of analysis is SRRF (Super-Resolution Radial Fluctuations) [15] where degrees of symmetry are calculated at each frame. The mathematical modelling of these problems relies on model-based sparse regularisation approaches. In particular, in [26] the authors propose to formulate the problem (1) in the covariance domain using an ℓ_1 regularisation and call it sparsity-based super-resolution correlation microscopy (SPARCOM). In recent works [32, 30] the authors improve upon the artefacts created by SPARCOM on noisy simulated and real data, by considering a non-convex continuous exact approximation of the ℓ_0 pseudo-norm previously proposed 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 contexts is that, due to their mathematical structure, they tend to promote point-like structures also in the case of continuous samples. Moreover, they appear quite sensitive to the choice of model hyperparameters, which limits their actual applicability to some extent.

We propose in this work an hybrid strategy where we keep the protocol of acquiring a video of short-time exposure images, and we use this fluctuating fluorescence image video in an optimisation procedure where the data term is given by a distance between distributions, one given by the observed video considered as samples of the observed (real) distribution and one is given by the physical model of stochastic fluctuations. This distance and optimisation procedure is realised by a Generative Adversarial Network (GAN) architecture where the generator network is replaced by the generative physical model. GANs [12] 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, 13] and some others more focused on applications to learning the distribution of human faces [16] or medical data [20]. 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. Recently, a hybrid approach allowing for the use of a physically-inspired simulator in place of the generator has been proposed for the problem of volume reconstruction from noisy projections in Cryo Electron Microscopy (Cryo-EM) and called CryoGAN [14]. 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 [22, 29] have been proposed. Up to our knowledge, no generative approaches have been considered in this field.

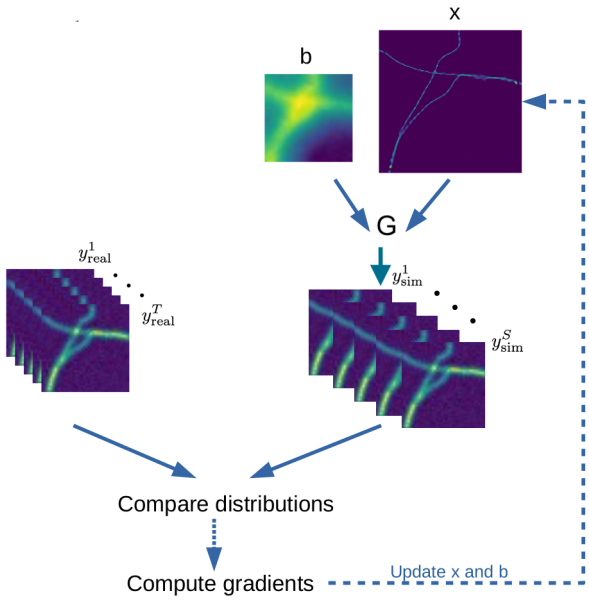


Figure 1: Schema of FluoGAN: Given a temporal sequence of blurred, undersampled and noisy images (left) and a simulator of the fluorescence fluctuation behaviour and optical system (right), represented here by G, FluoGAN compare the empirical distribution of the real images with the output distribution of the samples generated by G. Then, it 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.

Contribution and structure of the paper. We propose an hybrid approach combining physical modelling with data-driven learning and propose FluoGAN (Fluorescent image deconvolution/super-resolution microscopy via GAN learning), a novel approach where in the usual structure of a GAN, the generator network is replaced by a non-linear model of intensity fluctuation and light diffraction. From a given set of blurred, undersampled and noisy images where the noise models stochastic fluctuations of intensities, 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, see Figure 1. The paper is structured as followed: 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 simplified forward modelling used in the following as a reference simulator mimicking the physical processes creating blur and noise – including fluorophore fluctuations – in the observed image data. The fluorescence process is a stochastic process which we model by means of a Poisson distribution. It can be interpreted as the intensity (e.g., number of photons, whence the non-negativity) emitted by a biological structure of interest, sampled on a *fine* grid of size $n \in \mathbb{N}$. In [23, 11] a different model for stochastic fluctuations was considered. Differently from our model, the simulator proposed therein allows only for on-off switching of the fluorophores (so-called *blinking*) and stochastic intensity fluctuations are included in the model only in the form of noise. We believe that having a Poisson modelling is indeed more consistent with the actual physical modelling of fluorescence, since adding this we are able model the fluctuations of a sum of molecules that all lie in a pixel of the fine grid image $x \in \mathbb{R}^n_{\geq 0}$ we want to recover. We thus assume that the fluctuations in time of an image x can be modelled by Poisson(x), the notation Poisson(x) standing for the random vector drawn from a multidimensional Poisson distribution with vector parameter $x \geq 0$. These Poisson variables are independent of each other (in time and space).

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

From the sample to the sensor: blur, undersampling and background. Due to light diffraction, the light emitted from the fluorescent sample is subject to blur. We model it by means of 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 this kernel on the vectorised image Poisson(x) by means of the 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 lie and denote by $L > 1$ the super-resolution factor along both the horizontal and vertical direction so that $n = L^2 m$. Mathematically, as shown in Figure 2 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 the image 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, thus, however, preserving the total amount.

The photons captured by the sensor do not come all 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

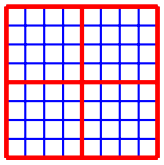


Figure 2: An example of the fine grid (blue) and the course grid (red) with $L = 4$.

fact relatively high. We include them in the model using a non-constant non-negative vector of low-resolution entries denoted by $b \in \mathbb{R}^m_{\geq 0}$.

Overall, we thus have that the blurred, undersampled and subject to background flux arriving at the receiving sensor is (proportional to) :

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

Recording noise. We have now 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 be described as a counting process with the following characteristics :

(i) It is possible to find an interval short enough so that the probability to absorb more than one photon is negligible

(ii) The number of photons absorbed during disjoint time intervals are independent;

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

Because of these properties, this counting process can be mathematically modelled as a Poisson counting process. In particular, 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 is a useful parameter depending on the specific setup considered 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 $\sigma^2\text{I}$, $e \sim \mathcal{N}(0, \sigma^2\text{I})$.

Collecting altogether, we can then define the following forward observation model

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

This model has been used in fact to generate simulated data in the experimental section 5.1. However, due to the presence of the double Poisson distributions, it is difficult to handle in computations. To alleviate such limitation, we will make use of the following approximation model:

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

which we claim to provide an acceptable approximation 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 model then adds further Poisson fluctuations. We thus use (4) in the estimation process.

Summarising, the setup is composed of *generated images* which are the realisations y^t_{sim}, $t = 1, \dots, T$ of the multi-variate random variable Y_{sim} following a distribution $\mathcal{D}_{\text{sim}}(x, b)$ given by (4) with quantities x and b that will play the role of learnable parameters to be improved within the simulator so as to generate samples closer and closer to the measured data. More precisely, we have:

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

From an experimental viewpoint, we remark that to avoid sample deterioration, the laser power is usually kept as low as possible. Furthermore, since the sample may move and fluorescent molecules progressively lose they capability to fluoresce when exposed to light stimulation for long time (*photobleaching*), the acquisition duration for each image is typically limited (20-100 images/s).

3. The inverse problem formulation

Interpreting the given temporal sequence of $T > 1$ noisy, blurred and undersampled images $\{y^t_{\text{real}}\}_{t=1}^T$ as the acquired T realisations of an unknown distribution $\mathcal{D}_{\text{real}}$, the data-driven inverse problem formulation of the problem thus reads:

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

where $\mathcal{D}_{\text{sim}}(x, b)$ is defined in (5). In other words, given the measurements $\{y^t_{\text{real}}\}_{t=1}^T$, we aim for estimating learnable parameters (x, b) such that the distribution $\mathcal{D}_{\text{sim}}(x, b)$ is as close as possible (in a sense which has to be specified) to the unknown underlying distribution $\mathcal{D}_{\text{real}}$ so that samples drawn by both distributions are 'close' in a distributional sense. Note that since the problem is formulated in a discrete setting, the choice of n (or, similarly, of the super-resolution factor L) is crucial as it determines the maximum resolution level we want to reach.

Since the sensor has limited resolution, information loss on the precise location of the sources is observed: problem (6) is thus *under-determined*. Similarly as in [8, 26, 32, 30] the analysis of stochastic fluctuations will be used to estimate a super-resolved image x where fluorescent molecules too close to each other are better resolved. Regarding the estimation of the background term $b \in \mathbb{R}^m_{\geq 0}$, we recall that this includes out-of-focus fluorescent molecules but also different unmodelled errors , and interferences which may make the reconstruction process even more unstable. As previously shown [32, 30], the estimation of this further ingredient significantly improves the quality of the reconstruction in comparison to competing approaches [26] providing only an estimate of the super-resolved image.

We can thus aim at solving (6) by formulating the problem from an optimisation perspective as the minimisation of a suitable distance between $\mathcal{D}_{\text{sim}}(x,b)$ and $\mathcal{D}_{\text{real}}$ over x and b. Due to ill-posedness, we may further incorporate additional and physically-consistent assumptions on the desired signal x and the background b, which stabilise the optimisation and eliminate potential solutions without physical meaning. We thus consider the following 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.

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 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 being both the expected value and the variance. For the simulated image samples such parameter is thus given by $\phi(x,b)$ and thus depend 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 nonetheless by the sample average $\bar{y}_{\text{real}} := \frac{1}{T}\sum_{t=1}^{T} y^t_{\text{real}}$. Denoting similarly by $\bar{y}_{\text{sim}}(x,b) = \frac{1}{T}\sum_{t=1}^{T} y^t_{\text{sim}}(x,b)$, we can choose

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

where Y_{real} is the multidimensional random variable associated to the observed samples. Among the many other distances that could be here considered (see the following sections) the advantage of this choice is that it can be computed explicitly, so its minimisation is straightforward. Of course it has the major limitation of comparing only two mean images instead of the full datasets. So it does not exploit all the information contained in the data. As we will see in the following, this method is in nonetheless able to recover a good deconvolved version of the image but seems to lack precision to achieve super-resolution.

It is possible to slightly improve the amount of information carried out by (8) and compare distributions by computing the mean ℓ_2 distance between images, i.e. by choosing:

$$d(\mathcal{D}_{\text{sim}}(x,b),\mathcal{D}_{\text{real}}) = \frac{1}{T}\sum_{t=1}^{T} \|y^t_{\text{sim}} - y^t_{\text{real}}\|_2^2. \tag{9}$$

3.1.2. From KL divergence to Wasserstein distance. The use of the Kullback-Leibler divergence to compare two probability distributions is rather classical. For two

absolutely continuous probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ with respect to a measure on Ω and with densities p_1 and p_2, such quantity is defined by:

$$\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2) = \int_\Omega \log\left(\frac{p_1(w)}{p_2(w)}\right) p_2(w) d\mu(w).$$

where μ is a measure defined on Ω. In their vanilla form, Generative Adversarial Networks (GANs) [12] rely also on the asymptotic minimisation of such quantity to minimise the distance between $\mathcal{D}_{\mathrm{sim}}$ and $\mathcal{D}_{\mathrm{real}}$. However, it is often not possible to compute this divergence explicitly in cases when the two distributions have disjoint support and/or when either of the two densities is unknwon, as it happens in our case for $\mathcal{D}_{\mathrm{real}}$. When it is possible, moreover, such quantity turns out to be often infinity [2].

Another candidate which allows to overcome such limitations takes its functional form from the field of optimal transport with the so-called 1-Wasserstein functional also called as *earth-mover distance*, see, e.g., [1] 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. Intuitively, the W_1 distance can be understood as the cost corresponding to an optimal transport plan a worker should do to transform one earth heap into another with different position or shape. Contrarily to the KL divergence, 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 particularly suited for comparing distributions, as it was showed in [2]. Note, however, that is in general not possible to compute this distance explicitly although it is sometimes possible to define relaxed optimisation problems corresponding to the minimisation of W_1 [13]. This point is discussed in section 4 below.

3.2. Choosing physically-consistent regularisation terms

Recalling (7), it is now beneficial to provide some rationale on how to choose the regularisation terms R_1 and R_2 favouring *a-priori* regulartity on the desired super-resolved image x and on the background image b, respectively.

Since in our applications we typically seek a super-resolved image made of fine structures (typically, filaments of proteins, point-like sources), we can saftely assume that x 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 $\|x\|_0 := |\{i \in \{1, \ldots, n\} : x_i \neq 0\}|$. Such approach is known to be NP-hard [21], hence 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, among which the inferior limit is the so-called CEL0 (continuous-exact relaxation penalty) [27]. For what follows, we will set

$R_1(x) = \lambda_1 \|x\|_1$ and deal with the non-smoothness associated to such choice by means of proximal-gradient algorithms, see, e.g. [3].

As far as the background b is concerned, such quantity 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. A standard approach favouring its smoothness and its consistency with the surrounding sample structures consists in choosing a gradient-smoothing term, i.e. $R_2(b) = \frac{\lambda_2}{2}\|\nabla b\|_2^2$. This is not atypical in applications. In previous work (e.g., [30]) it has been shown that assuming a smooth background is indeed a reasonable assumption.

4. FluoGAN: a GAN-inspired approach for super-resolution fluorescence microscopy

We choose a generative adversarial learning approach to combine distribution fit between $\mathcal{D}_{\text{sim}}(x, b)$ and $\mathcal{D}_{\text{real}}$, with regularisation. Generative Adversarial Networks (GANs) were introduced in firstly in [12] by Goodfellow et al. and have been successfully applied to many different problems in imaging (see for example [14]). 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 simulating sampling from a random variable of a given law by which the inverse of its cumulative distribution function is applied to a random variable uniformly distributed in $[0, 1]$. In the case of GANs, the inverse distribution function is modelled by the *generator* network G and it is trained until the distribution of its outputs cannot be distinguished by the one of the real data.

To complete such training a loss function is needed for measuring how far from real data the ones generated by the network are. 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 the *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 the distribution of the training data with high probability while $D(y) \approx 0$ meaning that y is likely to be a sample created by the generator. 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 [12] is done by enforcing a comparison between distributions in terms of the Jensen-Shannon divergence, although, as recalled in Section 3.1.2 other choices are possible such as the Kullback-Leibler divergence and the Wasserstein distance [2].

Recalling (7), we observe that this problem is closely related to a GAN-type problem

since it involves the minimisation of a distance between distributions. However, two significant differences have to be highlighted

(i) since we have formulated in (5) the physical model of fluorescence fluctuations, our knowledge of the unknown random variable $Y(x,b)$ depending on the two quantities of interest x and b does not start from scratch. Instead of generator network and inspired by previous works [14, 35], it thus appears natural to consider (5) as a simulator of the direct problem. The input learnable parameters $(x,b) \in \mathbb{R}^n_{\geq 0} \times \mathbb{R}^m_{\geq 0}$ of this "network" are 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 empirical distribution of the generated data matches the empirical one of given data, the corresponding input parameters of the process are the super-resolved image x and the background image b which are improved during the training process.

The proposed method, which we call FluoGAN given the particular context considered here, is thus inspired by GANs but, similarly as for the case of CryoGAN [14] has the goal of solving an ill-posed inverse problems given a large training set of measured images and by exploiting the physical knowledge of the inverse problem considered. In Figure 3 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 generator and the discriminator networks of parameters (x,b) and φ respectively.

4.1. Formulation as an optimisation problem

Inspired by the GAN modelling, we formulate the adversarial training process as a competing procedure defined in terms of two functions to be optimised alternatively and corresponding to the generator and the discriminator, respectively. To maintain the same notation as for GANs, we will denote in the following by $G_{x,b}$ the physically-inspired generator (i.e. the simulator) providing for a given input (x,b) an output y_{sim} being the realisation of the random variable $Y(x,b)$ in (5). 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:

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

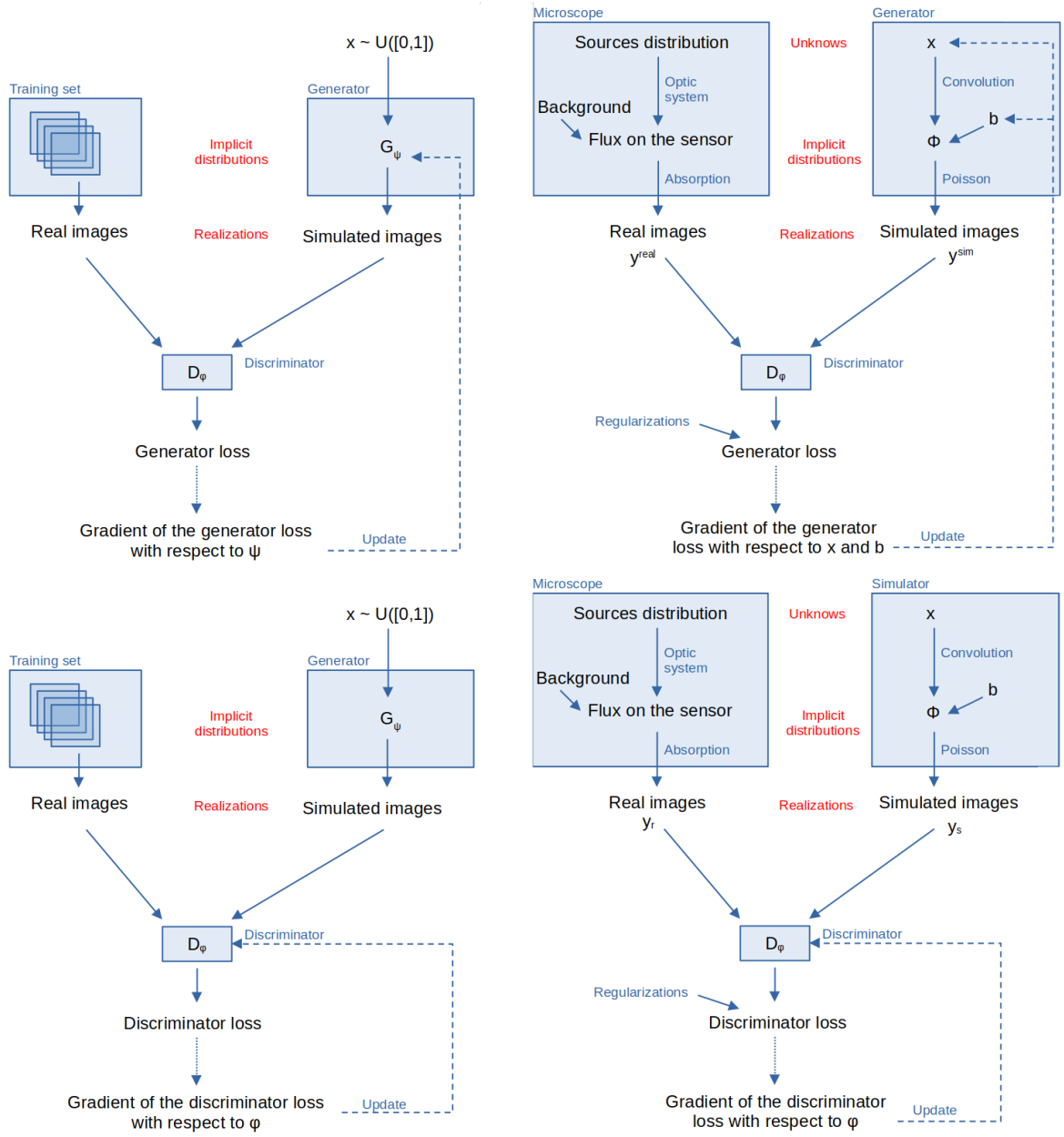


Figure 3: Comparison between a regular GAN architecture (left) and FluoGAN (right). In both cases the generator G_ψ (or the simulator) is trained in an adversarial manner with the discriminator D_φ. The proposed method can be considered as a specialised GAN architecture where the additional knowledge available on the direct problem enables us to replace G_ψ by a more interpretable physics-based function.

where the term depending on the classifier D_φ should enforce here $D_\varphi(y^t_{\text{sim}}) \approx 1$ for samples $t = 1, \dots, T$ generated by $G_{x,b}$.

- L_{D_φ} is a loss term which should enforce 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:

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

To avoid the well-known convergence instabilities of such adversarial training, it is rather classical to further introduce in such loss another term, called *gradient penalty*, to promote gradient updates with norm close to 1, see [13]. This is in fact a penalised formulation of the constraint of the discriminator to be 1-Lipschitz which was observed to correspond to the minimisation of the 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 computed on images y_{mix} uniformly chosen at random between simulated images and real images as: $y_{\text{mix}} = \eta y_{\text{real}} + (1 - \eta) y_{\text{sim}}$ 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 the empirical risk formulation of (11)-(12) which reads:

$$\min_{x \in \mathbb{R}^n_{\geq 0},\ b \in \mathbb{R}^m_{\geq 0}} \frac{\gamma}{2} \sum_{t=1}^{T} \|y^t_{\text{sim}}(x, b) - y^t_{\text{real}}\|_2^2 - \delta \sum_{t=1}^{T} D_\varphi(y^t_{\text{sim}}(x, b)) + \lambda_1 ||x||_1 + \frac{\lambda_2}{2} ||\nabla b||^2 \tag{14}$$

$$\min_{\varphi \in \Phi} \sum_{t=1}^{T} D_\varphi(y^t_{\text{sim}}(x, b)) - \sum_{t=1}^{T} D_\varphi(y^t_{\text{real}}) + \lambda \left(\|\nabla_y D_\varphi(y^t_{\text{mix}}(x, b))\| - 1 \right)^2, \tag{15}$$

which, recalling (7), corresponds to choose as fitting term between distributions:

$$d(\mathcal{D}_{\text{sim}}(x, b), \mathcal{D}_{\text{real}}) = \frac{\gamma}{2} \sum_{t=1}^{T} \|y^t_{\text{sim}}(x, b) - y^t_{\text{real}}\|_2^2 - \delta \sum_{t=1}^{T} D_\varphi(y^t_{\text{sim}}(x, b)),$$

where the former term runs average comparisons between $\mathcal{D}_{\text{sim}}(x, b)$ and $\mathcal{D}_{\text{real}}$, while the latter is linked to the particular use of a GAN discriminator whose parameters have to be learned by solving (15) which includes also a Wasserstein-type stabilisation. Practically, and similarly as for GANs, problem (11) is minimised for few (say, $\bar{k}$) epochs, then, similarly, (12) is minimised for $\bar{k}$ epochs using the values $x^{\bar{k}}, b^{\bar{k}}$ 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 of the quantities $y^t_{\text{sim}}(x, b)$ through the Poisson model (6).

4.2. Computing gradients for a Poisson random variable

Recalling (13), we notice that the computation of gradients w.r.t. x and b in (11) requires the quantities $\nabla_x\, y^t_{\text{sim}}(x,b)$ and, similarly, $\nabla_b\, y^t_{\text{sim}}(x,b)$ which both rely on the non-linear Poisson model given by (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 $f : \mathbb{R}^m_{>0} \to \mathbb{R}$ be a continuous and bounded function and 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:
$$v_0'(t) = -e^{-t} = -v_0(t). \tag{17}$$

We would like to apply the dominated convergence theorem to the sequence $(v_k)_k$. For that, 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.

As a consequence of the dominated convergence theorem it is possible to switch the derivation and integral 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:

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

Then a variable change is 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 can now write:

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

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: incremental 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 [19], 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, thus finally getting:

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

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 it is a sufficient condition to bound the second-order error of this approximation. However, in general we do not know whether f is twice differentiable. In our context, this condition can thus be relaxed by assuming that we have a sufficiently fine quantisation compared to the variations of $\partial f/\partial q_i$.

4.3. Discriminator architecture

Inspired by [35] the discriminator is a regular convolutional neural network $\mathbb{R}^n \to [0,1]$ with 3 main layers and 2 fully connected layers, as shown in Figure 4. 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

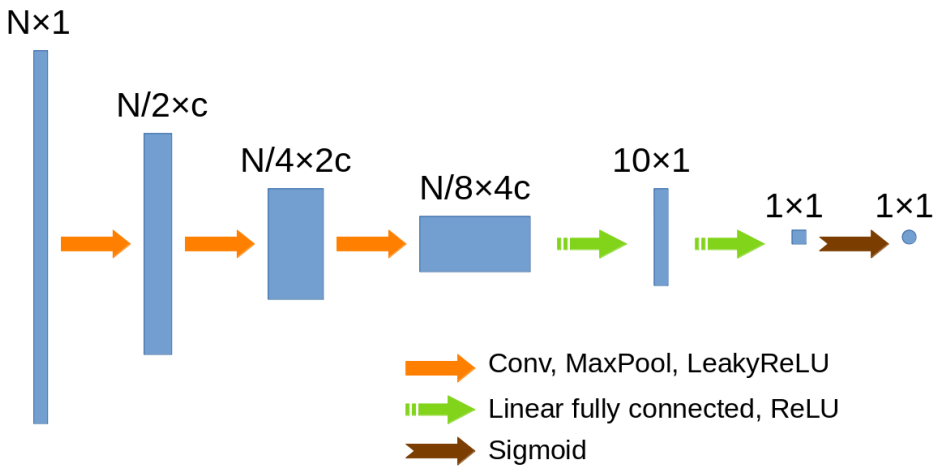


Figure 4: Discriminator architecture

the range 0 to 1. This setup is a common choice for this kind of network 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, we use the FISTA accelerated proximal algorithm studied in [3] which can deal with the non-smoothness of the ℓ_1 norm. 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, 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

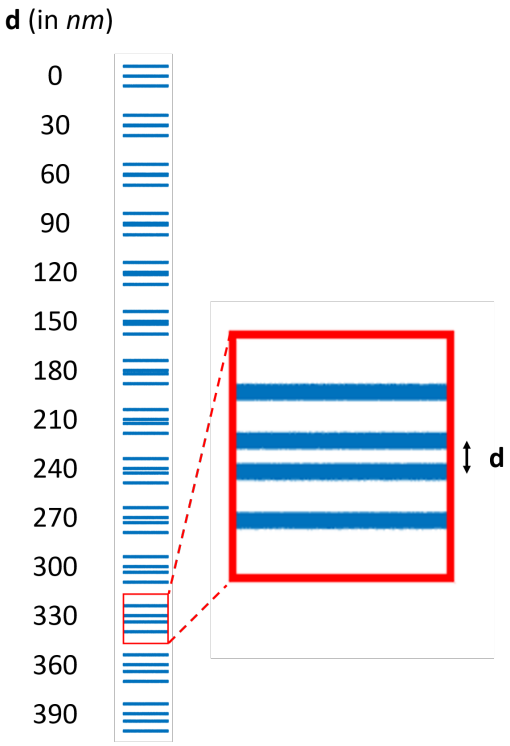


Figure 5: Spatial structure of simulated/real calibrated sample (ARGO-CR slide, Argolight)

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.

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 [15]. 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).

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 generated by us after considering parameters

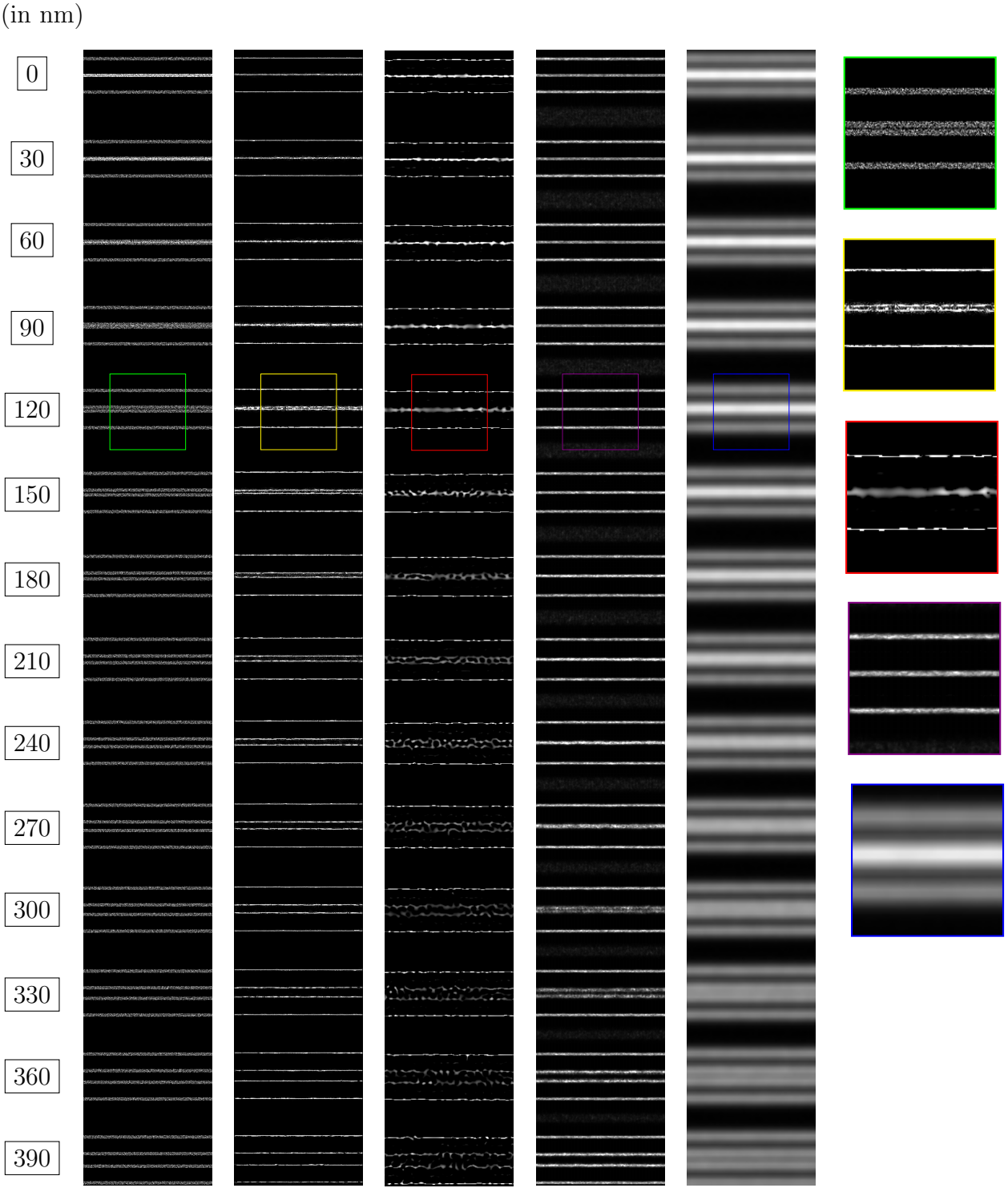
of the microscope configuration as well as properties of the sample (e.g. fluctuation behaviour, out-of-focus fluorescence, etc.). The spatial pattern used is shown in Figure 5 and is similar to the pattern of a calibrated sample ARGO-CR (Argolight, Pessac, France). To be more precise, the pattern used 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 big advantage we have with this kind of sample is that we can easily see 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 at frame rate of 100 frames per second (fps). 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 [15] methods in Figure 6, using a super-resolution factor of 6 ($L = 6$) for all three approaches. The method COL0RME allows us to select between different regularisation penalties in the support estimation step of the method (see [32, 30] for more details) and for the purpose of this experiment the ℓ_1-norm penalty was chosen due to the better and most time-efficient reconstruction. Regarding the method SRRF, we make use the freely available Fiji plugin ([15]). Especially, for the temporal analysis of the radiality image sequence we are using the auto-correlation analysis similar to that used in the SOFI [8] and many other methods that according to the authors works successfully for dense data as the ones we are using.

From the reconstruction shown in Figure 6, 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 ($80-110$nm), 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 are acquiring images of a ARGO-CR calibrated sample using an epifluorescence microscope. Images were acquired on an AxioObserver Z1 (Zeiss, Germany) with 63x/1.15W Korr LD C-Apochromat objective, ORCA Flash 4.0 camera (Hamamatsu, Japan), 540-570nm LED excitation and 581-619nm 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



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

Figure 6: 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.

from 0 to 390nm (see Figure 5 for a graphical representation). Here we can estimate the resolution reached by the methods 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 270nm. 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 of 6 was used for both methods. For our approach, FluoGAN, the 500 acquired images constituted the training set while 5000 iterations were performed in order to obtain the reconstruction presented in Figure 7. Compared to SRRF, FluoGAN achieves significantly better resolution levels. However the structures reconstructed by FluoGAN are a bit thinner than the real ones and this is the reason that it is able to separate filaments that are slightly overlapping, i.e. when the centre-to-centre distance is equal to 60nm.

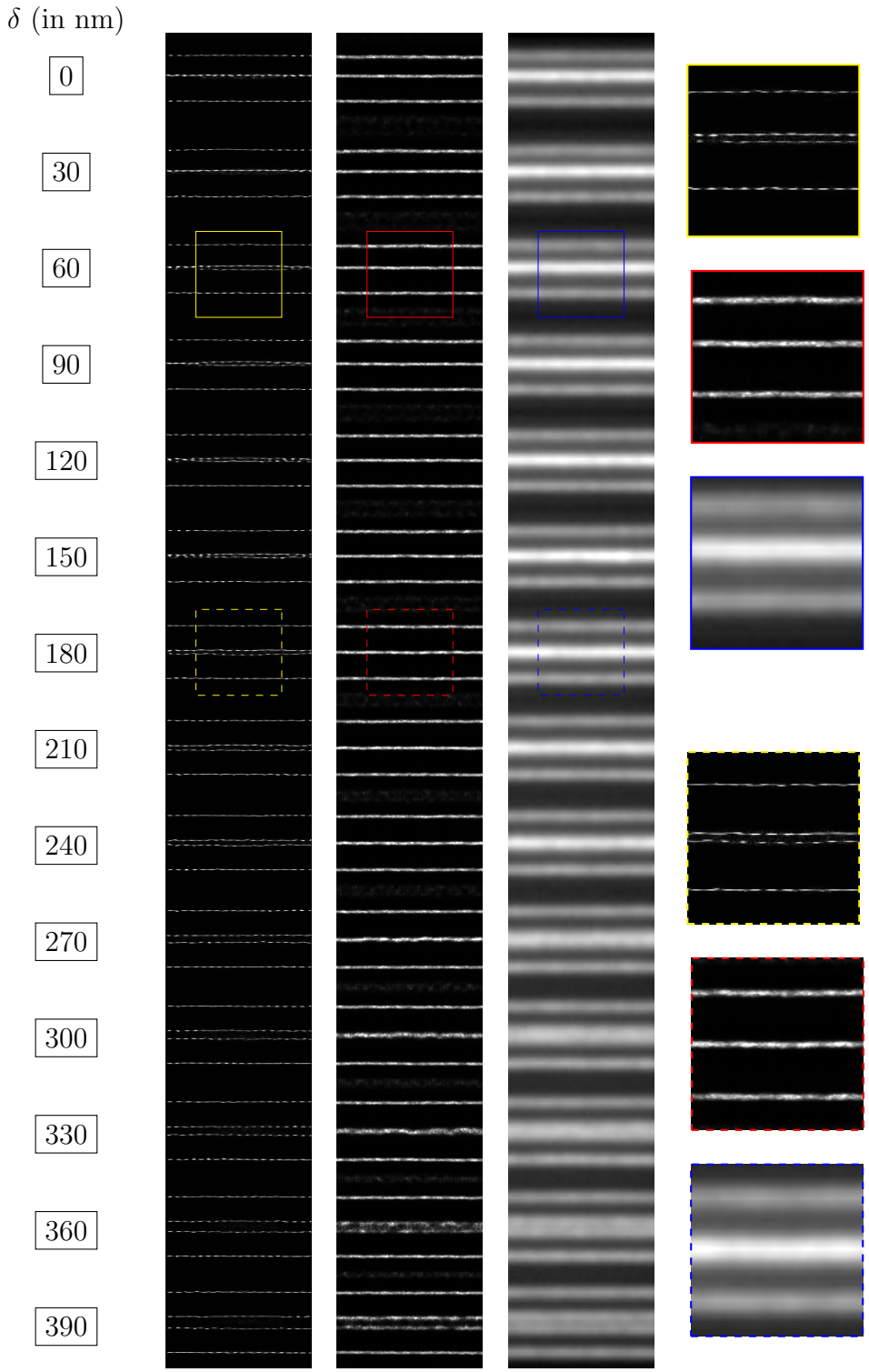
A nice observation in this experiment, is the different result we obtain with ($\delta \neq 0$) and without ($\delta = 0$) the presence of the discriminator in (14). The two results are available in Figure 8. Even if the ℓ_2 distance is enough to recover the main pattern, the explicit presence of the discriminator allows to achieve a greater precision. The result with no discriminator can nonetheless be used as an educated guess for the initialisation of FluoGAN.

Finally, unlike other optimisation problems, it is difficult to visualise convergence from the loss functions graphs (see Figure 9). Since simulator and discriminator have adverse goals the difficulty is to maintain both in competition preventing one from overtaking the other.

5.3. Ostreopsis images

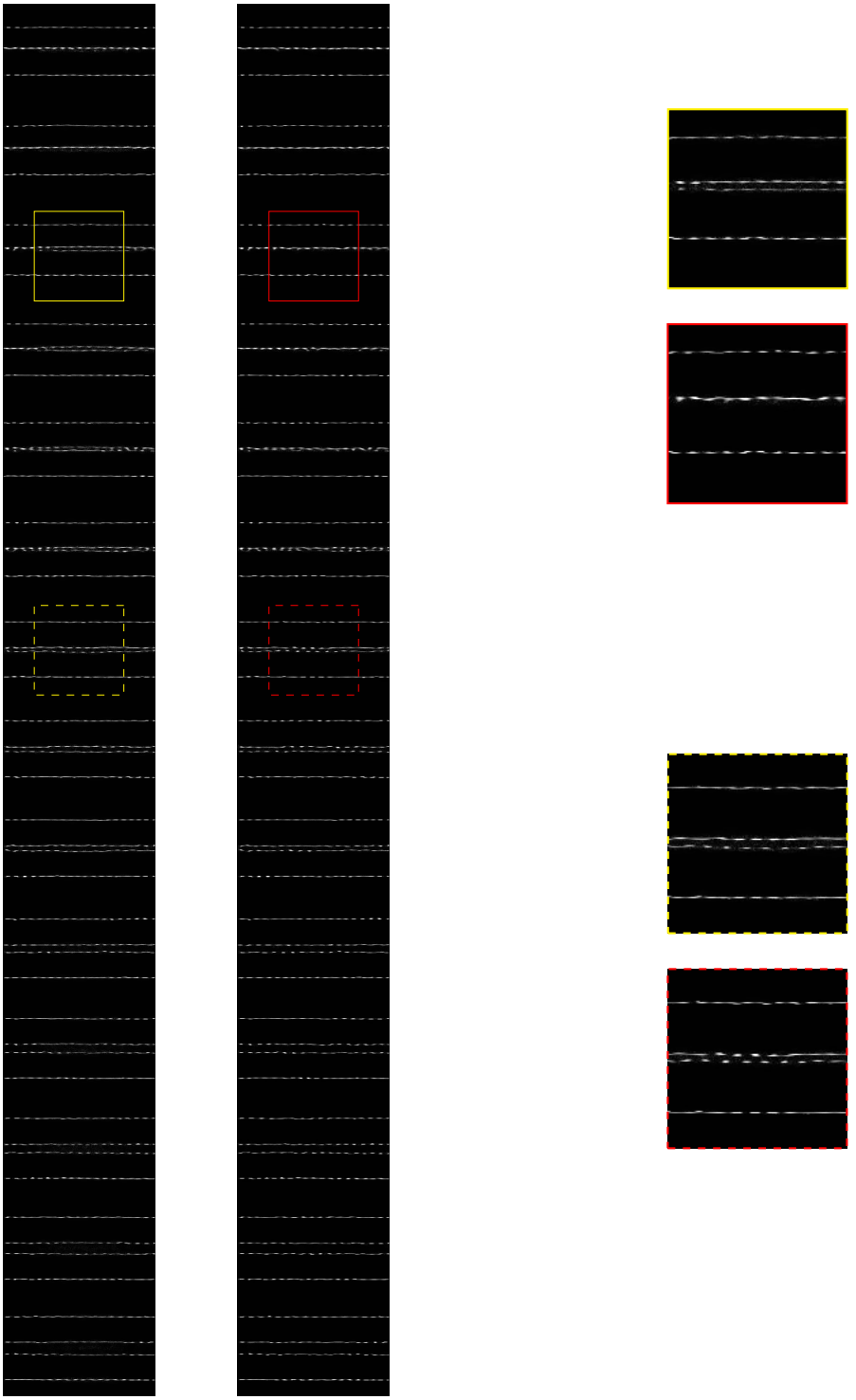
To test FluoGAN on challenging real biological samples, we considered a dataset of the unicellular alga *Ostreopsis cf. Ovata* (see Figure 10a). Such dataset combines several difficulties. First, it shows a 20µm thick sample which can exhibit a strong out-of-focus signal which is clearly the case here 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. Nonetheless, the flexibility and the data-adaptivity of FluoGAN allows to obtain accurate results on such data, whereas standard 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). The reconstruction performed by FluoGAN is shown in Figure 10c, while the parameters



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

Figure 7: 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 $\delta \neq 0$ (b) FluoGAN $\delta = 0$

Figure 8: Comparison of the results with (left) and without (right) discriminator.

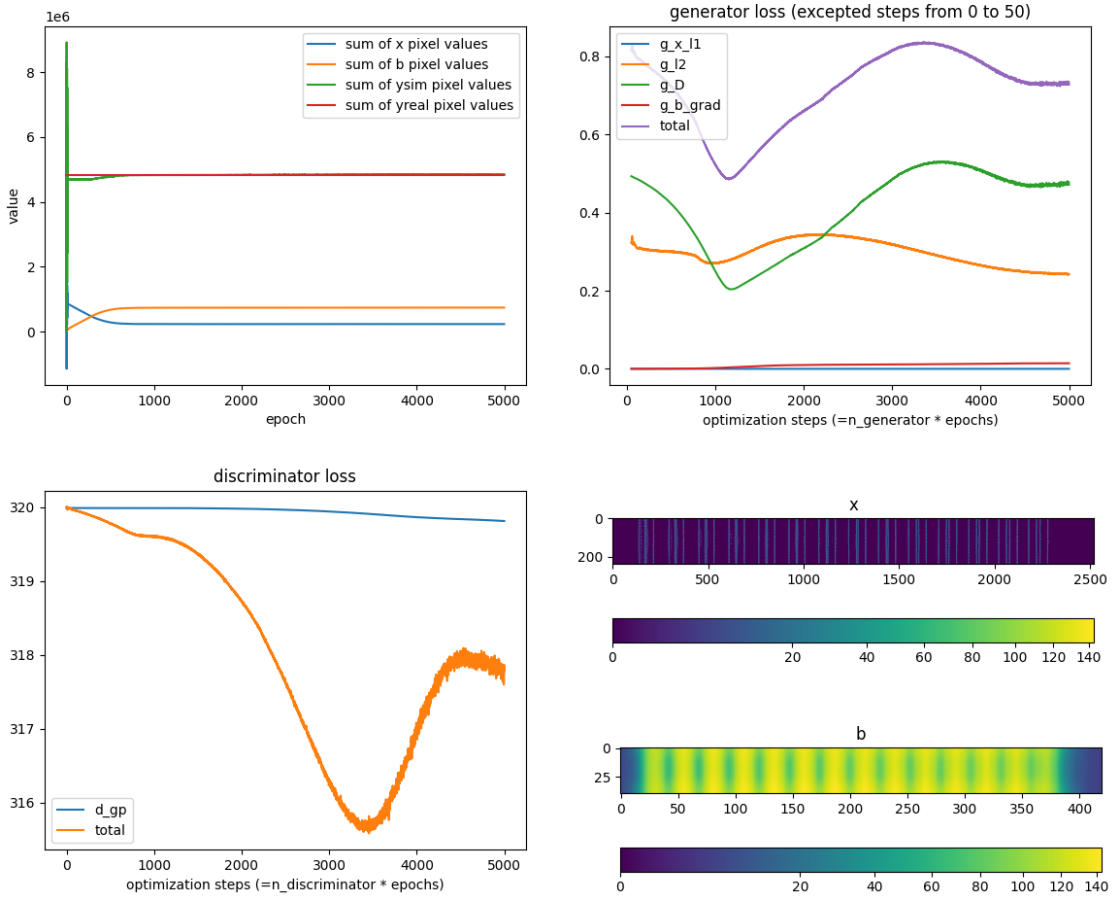


Figure 9: Top left : sum of pixel values (i.e., total photon count) VS. iteration number. After about 1000 iterations the sum of pixels for simulated images (y^{sim}) equals the sum of pixels for real input images (y^{real}). Top right and bottom left : balance between between simulator and discriminator training. Bottom right : super resolution image x and reconstructed background b.

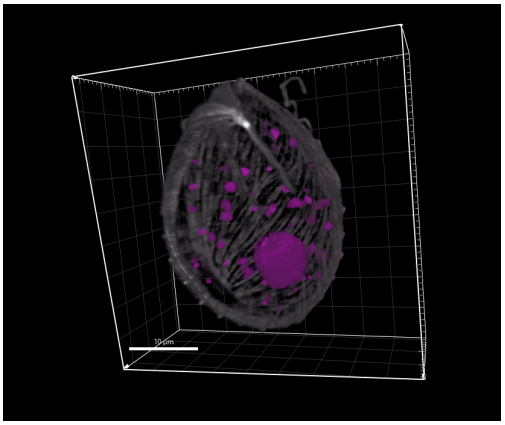
used are reported in Table 1. We compare the results obtained by FluoGAN with the ones obtained by SRRF which is the only one capable of dealing with confocal data. Regarding SRRF and differently from the two previous experiments, gradient weighting is performed, as suggested by the authors, to deal with the low signal-to-noise ratio (SNR) regime. The reconstruction obtained by SRRF is reported in 10d. Overall, the reconstruction obtained by FluoGAN compared to the one obtained by SRRF is more accurate, as it preserves true signal intensity and it does not suffer from any background artefacts.

6. Conclusions

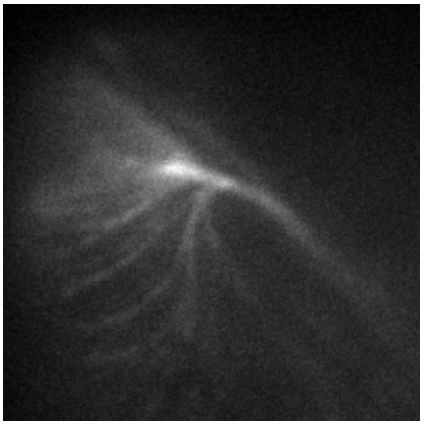
We presented FluoGAN, a novel framework for fluctuation-based super-resolution fluorescence microscopy combining the physical modelling of optical system with

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

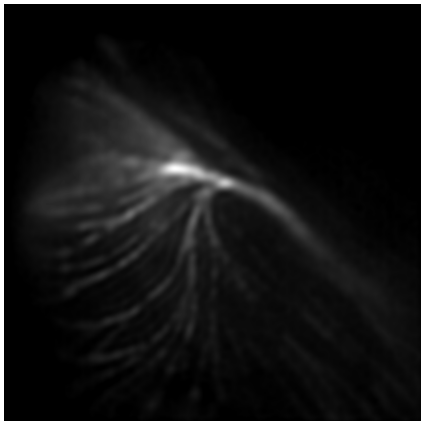
Table 1: Parameters used for ostreopsis image reconstruction



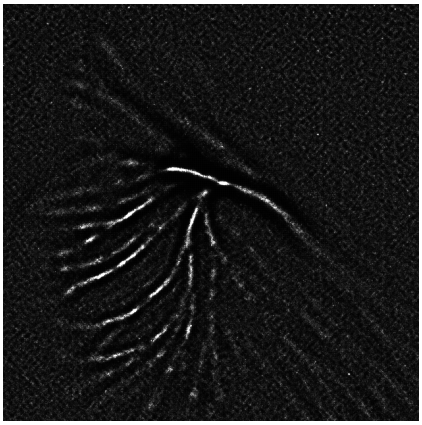
(a) Confocal 3D image of *Ostreopsis* with microtubules (white) and DNA (magenta).



(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.

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 along with a background image containing out-of-focus molecules, ambient fluorophores and other errors. 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 [14], 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 convergence 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.

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 d'Azur 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 academy on complex systems of UCA JEDI, the one 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*. Birkhäuser 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 of *Proceedings of Machine Learning Research*, 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-Féraud. 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] Geissbuehler et al. Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI). *Opt.l Nanoscopy*, 1(1):4, 2012.

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

[12] 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.

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

[14] 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.

[15] 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.

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

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

[18] 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.

[19] 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.

[20] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb. Learned reconstruction methods with convergence guarantees, 2022. to appear in IEEE Signal Processing Magazine.

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

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

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

[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 l0 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-Féraud. 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-Féraud. 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.