

JOINT CRYO-ET ALIGNMENT AND RECONSTRUCTION WITH NEURAL DEFORMATION FIELDS

Valentin Debarnot[†], Sidharth Gupta*, Konik Kothari*, and Ivan Dokmanic^{†*}

[†]University of Basel, *University of Illinois at Urbana-Champaign

ABSTRACT

We propose a framework to jointly determine the deformation parameters and reconstruct the unknown volume in electron cryotomography (CryoET). CryoET aims to reconstruct three-dimensional biological samples from two-dimensional projections. A major challenge is that we can only acquire projections for a limited range of tilts, and that each projection undergoes an unknown deformation during acquisition. Not accounting for these deformations results in poor reconstruction. The existing CryoET software packages attempt to align the projections, often in a workflow which uses manual feedback. Our proposed method sidesteps this inconvenience by automatically computing a set of undeformed projections while simultaneously reconstructing the unknown volume. We achieve this by learning a continuous representation of the undeformed measurements and deformation parameters. We show that our approach enables the recovery of high-frequency details that are destroyed without accounting for deformations.

Index Terms—CryoET imaging, unknown deformations, registration, implicit neural networks, neural fields.

1. INTRODUCTION

Tomographic imaging plays a central role in science, medicine, and engineering. An emerging representative in biological imaging is electron cryotomography (CryoET). Unlike single-particle cryoelectron microscopy, CryoET can image entire cells under cryogenic conditions. The three-dimensional volume is tilted around an axis relative to a probing electron beam. A sensor array then collects a series of two-dimensional projections—a *tilt-series*—at a discrete set of tilt angles.

More formally, we measure M projection images of resolution $N \times N$ of an unknown volume $\rho \in \mathbb{R}^{N \times N \times N}$,

$$\mathbf{y}_m = \mathbf{D}(\phi_m^*) \mathbf{P} \mathbf{R}(\theta_m) \rho + \boldsymbol{\eta}_m, \quad m = 1, \dots, M. \quad (1)$$

This research was supported by the European Research Council (ERC) Starting Grant 852821-SWING.

The authors wish to thank Ben Engel, Ricardo Diogo Righetto and Lorenz Lamm for preliminary discussion about the cryo-ET problem.

In (1) $\mathbf{R}(\theta_m)$ denotes the rotation (tilt) by an angle θ_m and \mathbf{P} denotes a projection from $\mathbb{R}^{N \times N \times N}$ to $\mathbb{R}^{N \times N}$ which is a simple summation over the last coordinate. Due to the mechanical stage drift and beam-induced sample motion, the CryoET projections are affected by deformations such as shifts, shears, and rotations [1]. We model the deformations by the operator $\mathbf{D}(\phi_m^*)$, with ϕ_m^* the deformation parameters for the m th projection. The noise $\boldsymbol{\eta}_m$ is iid Gaussian. We aim at recovering ρ from $\{\mathbf{y}_m\}_{m=1}^M$. However, we face two challenges: 1) the tilt θ_m can only vary between -70 and $+70$ degrees resulting in a *missing wedge* of measurements, and 2) the deformation parameters, ϕ_m^* , are unknown.

If we ignore the deformation operator, the reconstructed volume (*tomogram*) can be obtained by filtered-back projection (FBP) [2]. In order to account for the unknown deformations, many popular CryoET reconstruction packages first perform a tilt series alignment in order to invert the degradation caused by $\mathbf{D}(\phi_m^*)$ [3, 4, 5]. These packages show that deformation estimation is vital for accurate reconstruction, but they are often based on geometric heuristics which do not guarantee an optimal reconstruction. This motivates our work: we build a framework to jointly recover the deformation parameters and the unknown volume which minimize a data consistency loss. We adapt the recent framework of Gupta et al. [6] to leverage neural fields (or implicit neural representations) [7] with their key property that they enable automatic differentiation with respect to input coordinates. The coordinate-based framework allows us to effectively parameterize various classes of deformations.

1.1. Related work

Existing CryoET software such as IMOD [3, 5], TomoAlign [4], and Warp [5] handles deformation using fiducial markers in the specimen. Some recent software such as AreTomo estimates the deformation parameters without fiducial markers [8], by leveraging the geometry of structured misalignments and by tracking patches from tilt to tilt. While this leads to an automated reconstruction pipeline, it is still based on heuristics. What is more, we could not find a precise mathematical description of the method and there exists no open source

code ¹ A recent work by Liu et al. [9] that appeared during preparation of this manuscript similarly proposes to jointly estimate the unknown object and the deformation parameters in optical tomography. In this paper, we represent the deformation parameters and the projection images by a continuous neural field which allows us to optimize over them using standard optimizers (without alternating between the deformation and the volume), while being able to plug-and-play almost arbitrary deformations.

Neural fields (or implicit neural networks) represent continuous signals as maps from coordinates to function values [7]. Their use cases include fitting 3D radiance maps to 2D images [10] and solving partial differential equations [11]. Sun et al. applied implicit networks to interpolate and upsample measurements in 2D computed tomograph [12]. Rather than to obtain denser measurements, we use automatic differentiation *with respect to coordinates* to optimize over parameters in the measurement space and thus fit the *unobserved* measurements.

2. LEARNING DEFORMATIONS

We represent the measurements $\{\mathbf{y}_m\}_{m=1}^M$ in (1) by an implicit neural network [10]. These networks parameterize a continuous representation of the observed measurements. Automatic differentiation, available in all major deep learning libraries, then allows us to compute the gradients of this continuous representation with respect to the measurement coordinates—a three-dimensional coordinate comprising a tilt angle and a location on the two-dimensional sensor array. In this work, we train an implicit neural network, $f_\gamma : [-\pi, \pi) \times \mathbb{R}^2 \rightarrow \mathbb{R}$, parameterized by $\gamma \in \Gamma$, where Γ is the space of feasible parameters. For each tilt angle θ_m we denote by \mathbf{X} a uniform sampling of the sensor array. We train the network to reproduce the undeformed measurements, that is, for each point $\mathbf{x} \in \mathbf{X}$ of the sensor array grid we want

$$f_\gamma(\theta_m, \mathbf{x}) \approx (\mathbf{PR}(\theta_m)\boldsymbol{\rho})(\mathbf{x}). \quad (2)$$

Importantly, since the non-deformed projections $\mathbf{PR}(\theta_m)\boldsymbol{\rho}$ cannot be observed, we also learn the unknown deformation parameters (ϕ_m^*) jointly with the implicit neural network. This is the crux of the proposed method: since the deformation parameters are functions of the measurements coordinates (tilt and position within a projection), we can directly use automatic differentiation to optimize over them, simultaneously with optimizing the weights γ of the neural field network. The objective we minimize has two components. The first one is simply the usual interpolation (fitting) loss subject to *unknown* deformations,

$$\mathcal{L}_{\text{data}}(\phi, \gamma) \stackrel{\text{def.}}{=} \sum_{m=1}^M \|\mathbf{D}(\phi_m)f_\gamma(\theta_m, \cdot) - \mathbf{y}_m\|_{\ell^2(\mathbf{X})}^2. \quad (3)$$

¹At the time of writing, this seems to hold more generally: there are no open source CryoET packages that handle deformations.

Here ϕ denotes all deformation parameters, ϕ_1, \dots, ϕ_M , and $\|\cdot\|_{\ell^2(\mathbf{X})}^2$ is the usual 2-norm computed with f_γ sampled on \mathbf{X} .

Minimizing $\mathcal{L}_{\text{data}}(\phi, \gamma)$ alone is clearly insufficient to learn correct deformations. We thus restrict the class of implicit network with parameters γ so that the deformation parameters can be estimated within a physically meaningful range,

$$\mathcal{L}_{\text{op}}(\gamma) \stackrel{\text{def.}}{=} \|f_\gamma - A(\boldsymbol{\theta})A^\dagger(\boldsymbol{\theta})f_\gamma\|_{\ell^2(\mathbf{X})}^2, \quad (4)$$

where $A(\boldsymbol{\theta}) \stackrel{\text{def.}}{=} [a(\theta_1), \dots, a(\theta_M)]$ with $a(\theta_m) \stackrel{\text{def.}}{=} \mathbf{PR}(\theta_m)$ and $A^\dagger(\boldsymbol{\theta})$ denotes the filtered backprojection.

This method is inspired by recent work on implicit representations for correcting operator error [6] and extends the framework to handle random deformations seen in practice in CryoET. Since CryoET operates at very low SNRs we use an additional total variation norm regularizer,

$$\mathcal{L}_{\text{reg}}(\gamma) \stackrel{\text{def.}}{=} \lambda_\theta \|\nabla_\theta f_\gamma\|_{\ell_1(\mathbf{X})} + \lambda_x \|\nabla_x f_\gamma\|_{\ell_1(\mathbf{X})}, \quad (5)$$

where $\nabla_\theta f_\gamma$ and $\nabla_x f_\gamma$ refer respectively to the gradient of $\theta \mapsto f_\gamma(\theta, \cdot)$ and $\mathbf{x} \mapsto f_\gamma(\cdot, \mathbf{x})$. We empirically verify that this term helps obtain accurate reconstructions. The regularization of the implicit network along the coordinate modeling the tilt angle is important to correctly estimate the parameters ϕ . It ensures that two estimates of consecutive undeformed observations do not differ drastically. Our experiments suggest that a small value of λ_x and λ_θ suffices to stabilize the joint measurement representation and deformation parameter learning (cf. Section 3).

Summarizing, we compute $(\hat{\gamma}, \hat{\phi})$ that solve

$$\min_{\gamma \in \Gamma, \phi \in \Phi} \lambda_1 \mathcal{L}_{\text{data}}(\phi, \gamma) + \lambda_2 \mathcal{L}_{\text{op}}(\gamma) + \mathcal{L}_{\text{reg}}(\gamma), \quad (6)$$

where Φ is the space of admissible deformations and $\lambda_1, \lambda_2 \geq 0$ and empirically chosen. We use the Adam algorithm to minimize (6). The final tomogram is given by

$$\hat{\rho} \stackrel{\text{def.}}{=} A^\dagger(\hat{\boldsymbol{\theta}})f_{\hat{\gamma}}. \quad (7)$$

One advantage of our approach is that we do not need to explicitly invert the deformation operator: a potentially problematic task for nonlinear or non-invertible deformations. Furthermore, our framework allows us flexibility in budgeting computation. For example, if repeatedly applying the forward operator to a high-resolution volume is computationally expensive, we can initially set $\lambda_2 = 0$ to obtain a coarse measurement representation and then refine it with $\lambda_2 > 0$.

3. EXPERIMENTS

3.1. Influence of noise on individual cells

We use the volume density of a native *M. pneumoniae* cell treated with chloramphenicol [13] (dataset DOI on EMPIAR 10.6019/EMPIAR-10499); see Fig. 1a.

3.1.1. Experimental parameters

We simulate CryoET acquisition using 60 projections at angles between -70 and $+70$ degrees and using the deformation and noise model (1) at various signal-to-noise ratios (SNRs),

$$\text{SNR}(\mathbf{y}_0, \boldsymbol{\eta}) = 10 \log_{10} \left(\frac{\text{Var}(\mathbf{y}_0)}{\text{Var}(\boldsymbol{\eta})} \right).$$

We experiment with (post-deformation) SNRs of -10 dB, 0 dB, and 10 dB. The observed projections with the volume side length $N = 64$ are displayed in Fig. 1c. The deformations comprise shifts between ± 10 pixels, shears between $\pm 10\%$ of the sensor array, and rotation between ± 10 degrees. We run 1500 iterations of Adam to solve (6) with $\lambda_1 = 10$, $\lambda_2 = 1$, $\lambda_\theta = 10^{-5}$ and $\lambda_x = 10^{-5}$.

3.1.2. Results

In Fig. 1b, we report the Fourier Shell Correlation (FSC), a common metric to assess the quality of CryoET reconstruction [2]. The FSC measures the correlation between the frequencies of the estimated volume and the original volume. We see that our method provides a significant gain compared to directly applying the FBP reconstruction algorithm on the raw observation. Even with large measurement noise we successfully recover projections close to the non-deformed ones.

In Table 1 we report the average error over the M projections between the true and the estimated deformation parameters given by solving (6). The initialization for the deformation parameters is $\phi = 0$, that is to say, no deformation. This quantitative inspection confirms the ability of the proposed approach to identify the deformation parameters. We display the reconstructed volume in Fig. 1a. While the overall structure is well-retrieved at reasonable SNRs, we observe a severe loss of fine details at SNR -10 dB. However, at SNR 0 dB and 10 dB, as indicated by the FSC scores, the smallest details are correctly reconstructed; see insets in Fig. 1a.

3.2. Volume density from a real CryoET acquisition

Finally, we experiment on a volume obtained using in a real CryoET acquisition. We use the cryo-electron tomogram of mouse hippocampal neurons [14] (dataset DOI on EMPIAR 10.6019/EMPIAR-10923) display in Fig. 2a.

Table 1: Average error on deformation of the volume in Fig. 1a.

	shift [px]	shear [%]	rotation [deg]
<i>Init</i>	3.36	5.1	5.3
-10 dB	0.86	5.8	4.0
0 dB	0.56	3.7	2.2
10 dB	0.36	2.9	1.4

3.2.1. Experimental parameters

We simulate CryoET acquisition (1) by collecting 50 projections from angles between -70 to 70 degrees. The measurement SNR, after adding noise to the deformed projections, is 10 dB. Note that this SNR is comparatively more severe for this experiment than in the previous experiments because the volume is much less sparse. The volume density is of size $128 \times 128 \times 90$ and the projections are of size 128×128 . The deformations comprise shifts between ± 5 pixels, shear between $\pm 5\%$ and rotation between ± 5 degrees. We run 2000 iterations of Adam to solve the optimization problem (6) with $\lambda_1 = 100$, $\lambda_2 = 10^{-2}$, $\lambda_\theta = 10^{-6}$ and $\lambda_x = 10^{-5}$.

3.2.2. Results

We display several slices of the true and the estimated volume in Fig 2a. The FSC in Fig. 2b shows that both the coarse structure and fine details are reconstructed well. We obtain accurate estimates of the deformation as shown in Table 2.

4. CONCLUSIONS

We demonstrated how coordinate-based neural representations can be made into an effective tool for joint projection alignment and calibration in CryoET. The fact that we can seamlessly build diverse deformation models as coordinate transformations in neural fields and then automatically differentiate with respect to their parameters gives us a simple and powerful framework. Ongoing work includes extensions of deformation classes to model the full complexity of those encountered in CryoET (although the three we use are known to be the most important ones). In order to keep complexity under control it may be helpful to borrow the parametric classes used to model spatially-varying blur in light microscopy [15, 16]. Here we again benefit from the flexibility of the introduced framework.

Table 2: Average error on deformation of the volume in Fig. 2a.

	shift [px]	shear [%]	rotation [deg]
<i>Initialization</i>	1.7	2.5	2.7
<i>Estimate</i>	0.2	1.9	1.1

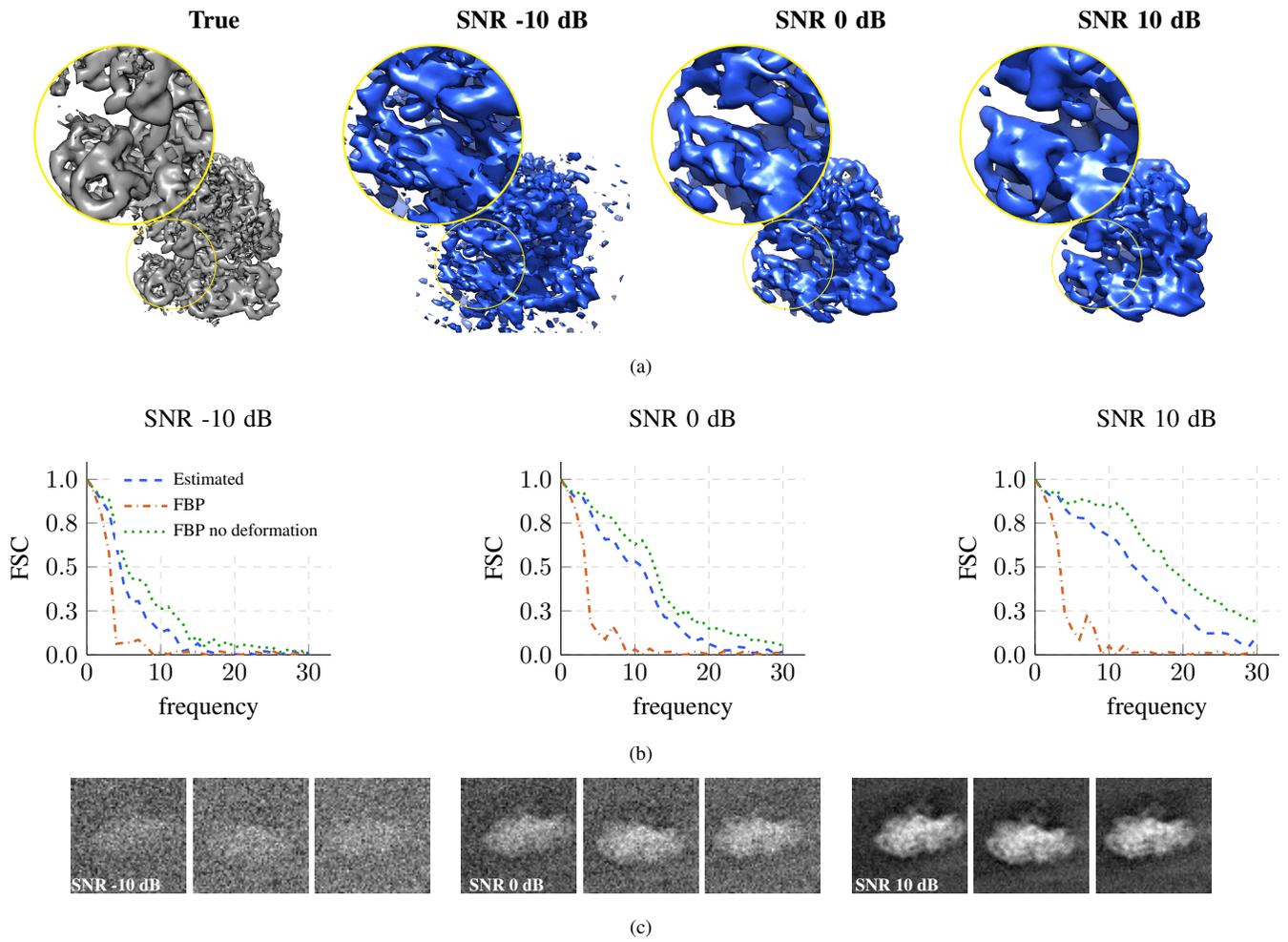


Fig. 1: a) 3D density estimation at different SNR. b) FSC of the proposed approach compared to FBP reconstruction when measurements are perturbed or not by deformations. c) Three projections corresponding to consecutive viewing directions.

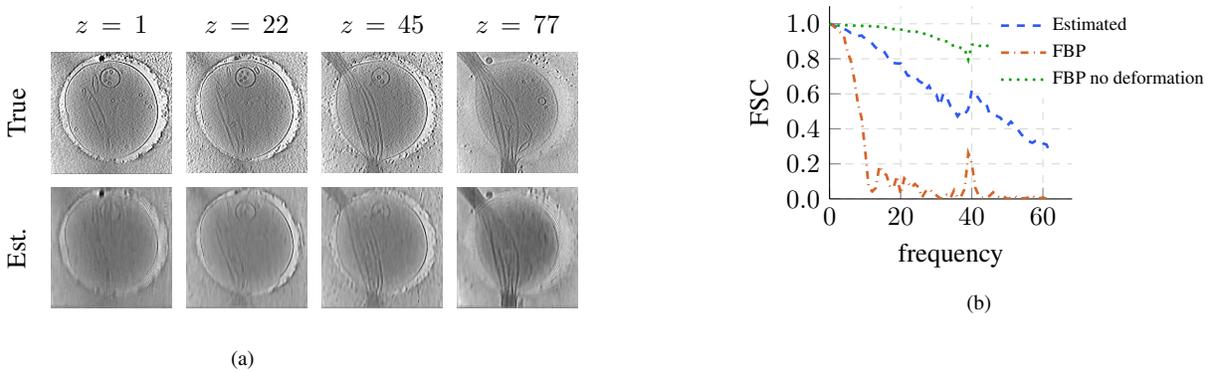


Fig. 2: a) Reconstruction of CryoET volume [14] (second row) at different depths (z), compared with the original volume (first row). b) FSC for the neuron volume of Fig. 2a for the proposed approach compared to FBP reconstruction when measurements are perturbed or not by deformations.

5. REFERENCES

- [1] David N Mastronarde, “Fiducial marker and hybrid alignment methods for single-and double-axis tomography,” in *Electron tomography*, pp. 163–185. Springer, 2007.
- [2] George Harauz and Marin van Heel, “Exact filters for general geometry three dimensional reconstruction.,” *Optik.*, vol. 73, no. 4, pp. 146–156, 1986.
- [3] David N Mastronarde and Susannah R Held, “Automated tilt series alignment and tomographic reconstruction in imod,” *Journal of structural biology*, vol. 197, no. 2, pp. 102–113, 2017.
- [4] Jose-Jesus Fernandez and Sam Li, “Tomoalign: A novel approach to correcting sample motion and 3d ctf in cryoet,” *Journal of Structural Biology*, vol. 213, no. 4, pp. 107778, 2021.
- [5] Dimitry Tegunov and Patrick Cramer, “Real-time cryo-electron microscopy data preprocessing with warp,” *Nature methods*, vol. 16, no. 11, pp. 1146–1152, 2019.
- [6] Sidharth Gupta, Konik Kothari, Valentin Debarnot, and Ivan Dokmanić, “Differentiable uncalibrated imaging,” *arXiv preprint arXiv:2211.10525*, 2022.
- [7] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar, “Neural fields in visual computing and beyond,” in *Computer Graphics Forum*. Wiley Online Library, 2022, vol. 41, pp. 641–676.
- [8] Shawn Zheng, Georg Wolff, Garrett Greenan, Zhen Chen, Frank GA Faas, Montserrat Bárcena, Abraham J Koster, Yifan Cheng, and David A Agard, “Aretomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction,” *Journal of Structural Biology: X*, vol. 6, pp. 100068, 2022.
- [9] Yan Liu, Jonathan Dong, Thanh-an Pham, François Marelli, and Michael Unser, “Mechanical artifacts in optical projection tomography: Classification and automatic calibration,” *arXiv preprint arXiv:2210.03513*, 2022.
- [10] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *European conference on computer vision*. Springer, 2020, pp. 405–421.
- [11] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, “Implicit neural representations with periodic activation functions,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.
- [12] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov, “Coil: Coordinate-based internal learning for tomographic imaging,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.
- [13] Dimitry Tegunov, Liang Xue, Christian Dienemann, Patrick Cramer, and Julia Mahamid, “Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.5 Å in cells,” *Nature Methods*, vol. 18, no. 2, pp. 186–193, 2021.
- [14] Hana Nedozralova, Nirakar Basnet, Iosune Ibiricu, Satish Bodakuntla, Christian Biertümpfel, and Naoko Mizuno, “In situ cryo-electron tomography reveals local cellular machineries for axon branch development,” *Journal of Cell Biology*, vol. 221, no. 4, 2022.
- [15] Valentin Debarnot, Paul Escande, and Pierre Weiss, “A scalable estimator of sets of integral operators,” *Inverse Problems*, vol. 35, no. 10, pp. 105011, 2019.
- [16] Valentin Debarnot and Pierre Weiss, “Deep-Blur : Blind Identification and Deblurring with Convolutional Neural Networks,” preprint, 2022.