Learning Binary Sampling Patterns for Single-Pixel Imaging using Bilevel Optimisation

Anonymous Author(s)
Affiliation
Address
email

Abstract

Single-Pixel Imaging enables reconstructing objects using a single detector through sequential illuminations with structured light patterns. We propose a bilevel optimisation method for learning task-specific, binary illumination patterns, optimised for applications like single-pixel fluorescence microscopy. We address the non-differentiable nature of binary pattern optimisation using the Straight-Through Estimator and leveraging a Total Deep Variation regulariser in the bilevel formulation. We demonstrate our method on the CytoImageNet microscopy dataset and show that learned patterns achieve superior reconstruction performance compared to baseline methods, especially in highly undersampled regimes.

1 Introduction

Single-Pixel Imaging (SPI) is a technique that allows imaging using a single non-spatial detector that measures the total intensity of transmitted light [7, 9]. The lack of spatial resolution is resolved by illuminating the object with a sequence of structured light patterns. The object is then reconstructed using the sequence of corresponding measurements along with the associated illumination patterns. The measurement process can be described by a forward model $\mathbf{y} = P(\mathbf{A}\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^N$ represents the (vectorised) object, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the sensing matrix, with rows representing illumination patterns, $\mathbf{y} \in \mathbb{R}^M$ are the measurements, and P is some noising process, for which in this work we use additive Gaussian noise.

SPI scan and reconstruction times are directly related to the number of illumination patterns M. Thus, a primary goal is to reduce the number of patterns, ideally achieving an undersampling regime $M \ll N$. Reconstruction of $\mathbf{x}$ from (noisy) measurements $\mathbf{y}$ in this regime is an ill-posed inverse problem, and it is commonly formulated using variational regularisation as

$$\hat{\mathbf{x}} = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\mathcal{J}(\mathbf{x}), \tag{1}$$

where $\mathcal{J}$ is a regularisation term which incorporates prior knowledge about the object [20]. The choice of both the sensing matrix $\mathbf{A}$ and the regulariser $\mathcal{J}$ is crucial for accurate reconstruction. While patterns like Hadamard or random matrices are common [5], they may not be optimal for a specific imaging task, and reconstructions can be improved by using data-adaptive patterns.

This work focuses on designing optimal illumination patterns for image modalities, like single-pixel fluorescence microscopy, that impose physical constraints on admissible patterns. Specifically, we consider pattens $\mathbf{A}$ whose elements are restricted to $\{-1, 1\}$. These patterns are standard in SPI and are acquired by measuring a pair of complementary $\{0, 1\}$ patterns and subtracting the results. In SPI systems, values 0 and 1 in the sensing matrix correspond to blocking and transmitting light, respectively. We tackle this problem by framing pattern design as a bilevel optimisation problem,

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

where we jointly learn the optimal $\{-1, 1\}$ patterns and specific hyperparameters of the reconstruction process. This approach has been successfully employed in domains like magnetic resonance imaging (MRI), sparse recovery or model selection [23, 19, 4]. Our main contributions are:

- Use the Straight-Through Estimator (STE) [3] to handle the discrete nature of pattern optimisation;
- Integrate a pre-trained Total Deep Variation (TDV) regulariser [15] in the lower-level problem to enhance reconstruction quality.

Related Work Higham et al. [11] introduced the first data-driven framework for learning SPI illumination patterns, showing that learned patterns can improve both reconstruction quality and compression efficiency. Their method formulates pattern design as training an autoencoder, where the linear encoder defines the illumination patterns and the nonlinear decoder performs image reconstruction. To improve the decoder, Wu et al. [26] propose an unrolling approach, while Wang et al. propose a physics-informed architecture based on differential ghost imaging [24]. Optimising sampling patterns also arises in different applications, such as minimising matrix coherence for compressive sensing [1] and k-space sampling in MRI [21, 23]. In MRI, sampling pattern design requires binary masks, for which the (Gumbel) STE [14] is commonly employed [19, 27].

2 Learning Sampling Patterns via Bilevel Learning

We aim to find an optimal sensing matrix $\mathbf{A}$ by minimising the reconstruction error over a representative dataset of images $\{\mathbf{x}^{(i)}\}_{i=1}^n$. We consider a bilevel problem given by

$$\min_{\substack{\mathbf{A}\in\{-1,1\}^{M\times N}\\ \alpha>0}} \left\{ L(\theta) := \sum_{i=1}^{n} \mathcal{L}\Big(\mathbf{x}^{(i)}, \hat{\mathbf{x}}\big(\theta; P(\mathbf{A}\mathbf{x}^{(i)})\big)\Big) \right\}, \text{ where } \theta = (\mathbf{A}, \alpha), \tag{2}$$

$$\text{such that } \hat{\mathbf{x}}(\theta; \mathbf{y}) \in \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\mathcal{J}(\mathbf{x}). \tag{3}$$

The upper-level problem (2) aims to find a sensing matrix $\mathbf{A}$ and the regularisation parameter $\alpha > 0$ that minimise the discrepancy between ground truth data and reconstructions from noisy measurements. The reconstructions are obtained by solving the lower-level problem (3), which is defined by a reconstruction method $\hat{\mathbf{x}}(\theta; \cdot)$ for a given $\mathbf{A}$ and α. The success of bilevel optimisation relies on the quality of lower-level solutions: if the regulariser $\mathcal{J}$ is a poor match for the data, the learned patterns will be suboptimal. Prior work has often relied on classical regularisers like Total Variation (TV) [23] or the ℓ_1-norm [25]. Instead, we leverage TDV [15], a powerful, data-driven regulariser that has shown superior performance over TV in many linear inverse problems.

The key challenge in the bilevel formulation is that the upper-level problem (2) is over a discrete set, making standard gradient-based methods inapplicable. We explore two methods to address this.

Relax and Penalise (RnP) is an approach that replaces the binary constraint $\mathbf{A} \in \{-1, 1\}^{M\times N}$ with $\mathbf{A} \in [-1, 1]^{M\times N}$ and drives the matrix entries towards $\{-1, 1\}$ by adding a penalty term

$$r_\epsilon(\mathbf{A}) = \frac{1}{\epsilon}\sum_{i,j} 1 - a_{i,j}^2. \tag{4}$$

It follows from [17, Theorem 1] that with an appropriate schedule of the penalty strength $\epsilon > 0$, the relaxed problem has the same minimiser as the binary one. However, in practice, this requires careful parameter tuning. The constraint $\mathbf{A} \in [-1, 1]^{M\times N}$ can be enforced by projecting the matrix entries onto the constraint set after each gradient step or by a reparameterisation $\mathbf{A} = \tanh(\mathbf{Z})$, applied entry-wise, using a real-valued latent matrix $\mathbf{Z} \in \mathbb{R}^{M\times N}$. While these methods allow computing the gradient exactly, a notable drawback is that $\mathbf{A}$ is not strictly binary during optimisation.

STE enables gradient-based optimisation for binary variables by using a surrogate gradient during backpropagation [3]. We represent the binary matrix $\mathbf{A} \in \{-1, 1\}^{M\times N}$ using a real-valued matrix $\mathbf{Z} \in \mathbb{R}^{M\times N}$ and the sign function as $\mathbf{A} = \mathtt{sgn}(\mathbf{Z})$, applied entry-wise with $\mathtt{sgn}(0) = 1$. The $\mathtt{sgn}$ function has zero gradient almost everywhere, so it cannot be used for gradient-based optimisation. To avoid this, STE replaces the activation function with a differentiable surrogate during the backwards pass. A common choice is the derivative of the hyperbolic tangent, $\tanh'(\mathbf{Z})$, allowing the latent

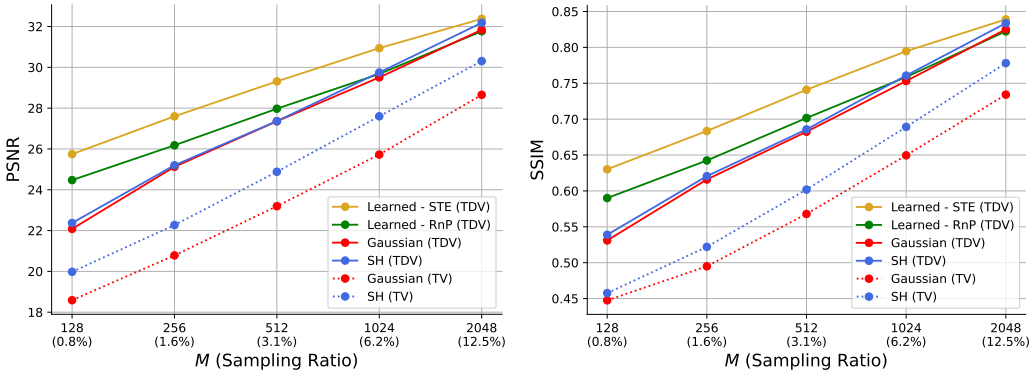
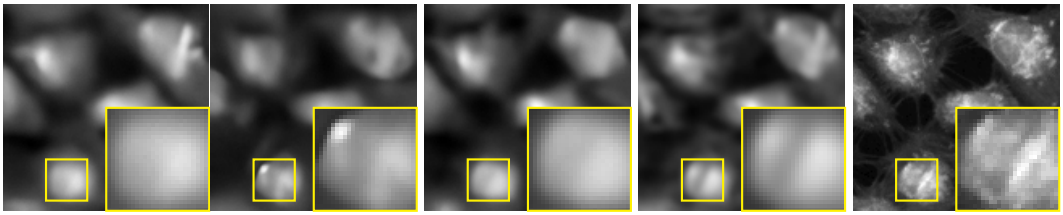


Figure 1: PSNR (left) and SSIM (right) for reconstructions using different illumination patterns with respect to an increasing number of patterns M.



Gaussian (25.08, 0.68) | SH (25.26, 0.68) | Learned - RnP (25.72, 0.70) | Learned - STE (27.05, 0.73) | Ground truth

Figure 2: Reconstructions using TDV regulariser with different illumination patterns for the same number of measurements $M = 512$ (3.125% subsampling ratio) with (PSNR, SSIM).

matrix $\mathbf{Z}$ to be updated as if the objective were differentiable. As an extension, we can introduce a scaling parameter controlling the asymptotics, though this was not tested. The key trade-off is that while STE preserves the binary nature of $\mathbf{A}$ during training, it relies on inexact gradient updates.

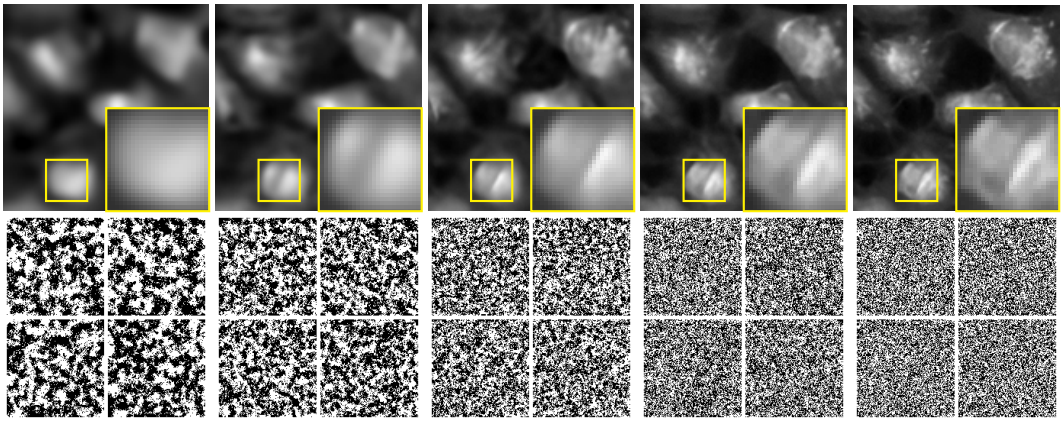
3 Numerical Experiments

Dataset We validate our approach using a subset of the CytoImageNet dataset [12], which contains cell microscopy images sourced from a range of publicly available datasets. The full dataset contains 890K images. To mimic the limited data setting often encountered in applications, we use only 1000 images to estimate the optimal patterns. We use a second, independent subset of 100 images from CytoImageNet to evaluate the learned patterns. We consider 128×128 px^2 images, which is at the physical limit in microscopy SPI [8], with a BRISQUE score [18] lower than 25.0. When simulating the measurements $\mathbf{y}$ for a given matrix $\mathbf{A}$ we add 5% relative Gaussian white noise.

Bilevel Optimisation Minimising the objective (2) requires computing the gradient of the upper-level problem with respect to parameters $\theta = (\mathbf{A}, \alpha)$. The main challenge is computing the Jacobian $\frac{\partial \hat{\mathbf{x}}(\theta)}{\partial \theta}$ of the solution $\hat{\mathbf{x}}(\theta)$ to the lower-level problem (3). We obtain $\hat{\mathbf{x}}(\theta)$ via fixed-point iterations $\mathbf{x}^{(k+1)} = T_\theta(\mathbf{x}^{(k)})$ with a suitable operator T_θ that is defined by the used iterative scheme. Differentiating the fixed-point equation $\hat{\mathbf{x}}(\theta) = T_\theta(\hat{\mathbf{x}}(\theta))$ gives the Jacobian

$$\frac{\partial \hat{\mathbf{x}}(\theta)}{\partial \theta} = \left(\mathbf{I} - \frac{\partial T_\theta(\hat{\mathbf{x}}(\theta))}{\partial \hat{\mathbf{x}}(\theta)} \right)^{-1} \frac{\partial T_\theta(\hat{\mathbf{x}}(\theta))}{\partial \theta},$$

see, e.g. [2]. To reduce the computational cost, we use Jacobian-Free Backpropagation [10], which relies on the zero-order Neumann series approximation of the inverse term. This yields an approximate gradient that is efficient to compute and works well in practice [22, 28].



$M = 256$ (≈1.6%) $M = 512$ (≈3.1%) $M = 1024$ (≈6.3%) $M = 2048$ (12.5%) $M = 4096$ (25.0%)

Figure 3: Reconstructions using Learned-STE with TDV (top row), and the first four learned patterns (bottom row) over five values of M.

Results: TDV We first compare the proposed TDV regulariser with the standard TV regulariser. For TV, the lower level variational problem (1) is solved using the Primal-Dual Hybrid Gradient algorithm [6], while TDV is optimised with the nonmonotonic Accelerated Proximal Gradient method [16], which thus defines T_θ. Unlike TV, TDV allows for a fully gradient-based optimisation since the regulariser is smooth. Results are presented in Figure 1, where we can see a clear performance increase, with respect to both PSNR and SSIM, for all tested undersampling ratios. Additional qualitative results are provided in Figure 6 in the Appendix, which shows the staircasing artefacts, characteristic of TV, and the improved reconstruction quality achieved by TDV.

Results: Learned Patterns We compare our approach at different numbers of measurements M against random Gaussian and scrambled Hadamard (SH) sampling patterns [13], which are common choices in compressive sensing and SPI. Images are reconstructed using the TDV regulariser and nmAPG to minimise the variational objective (1). Regularisation parameter α is selected by maximising SSIM on a batch of images from the training set. Results in Figure 1 show that learning the sampling pattern provides a significant increase in reconstruction quality, especially in the highly undersampled regime, which is of particular significance for fluorescence microscopy. Moreover, Learned - STE is noticeably better at capturing finer image structures (cf. Figure 2 zoom). In Figure 3 we show Learned - STE reconstructions at different sampling ratios (first row), and the learned patterns (second row). Learned patterns exhibit more structure at lower sampling ratios, and less structure at higher sampling ratios (also observed by [11]). We show the best, median and worst reconstructions with respect to PSNR in Figure 4 in the Appendix. The code will be publicly released at a later stage.

4 Conclusions and Further Work

We introduce a bilevel framework for learning binary illumination patterns for SPI, with a particular focus on fluorescence microscopy. In addition, we incorporate a data-driven TDV regulariser into the variational reconstruction method. Experiments on the CytoImageNet dataset demonstrate clear quantitative and qualitative performance boosts, especially in the highly undersampled regime.

In the current manuscript, our validation is limited to simulated measurements. Applying the learned patterns to experimental single-pixel fluorescence microscopy data, with a realistic noise model, will be a critical next step. Moreover, our comparisons are restricted to variational regularisation–based methods; future work will include benchmarking against fully learned approaches, such as [11], with particular emphasis on stability and generalisability to unseen data. In addition, this work uses a specific data-driven regulariser with weights trained on natural images. Investigating alternative learned regularisers and retraining them on a relevant dataset (e.g. CytoImageNet) represents another promising research direction. Finally, while this work focuses on binary illumination patterns, extending the framework to ternary $\{-1, 0, 1\}$ patterns is another important line of future work.

References

[1] Vahid Abolghasemi, Saideh Ferdowsi, and Saeid Sanei. "A Gradient-Based Alternating Minimization Approach for Optimization of the Measurement Matrix in Compressive Sensing". In: *Signal Processing* 92.4 (Apr. 2012), pp. 999–1009.

[2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. *Deep Equilibrium Models*. 2019.

[3] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. *Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation*. 2013.

[4] K.P. Bennett et al. "Model Selection via Bilevel Optimization". In: *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. Vancouver, BC, Canada: IEEE, 2006, pp. 1922–1929.

[5] Emmanuel J. Candes and Terence Tao. "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" In: *IEEE Transactions on Information Theory* 52.12 (Dec. 2006), pp. 5406–5425.

[6] Antonin Chambolle and Thomas Pock. "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging". In: *Journal of Mathematical Imaging and Vision* 40.1 (May 2011), pp. 120–145.

[7] Marco F. Duarte et al. "Single-Pixel Imaging via Compressive Sampling". In: *IEEE Signal Processing Magazine* 25.2 (Mar. 2008), pp. 83–91.

[8] Alexander Ebner et al. "Diffraction-Limited Hyperspectral Mid-Infrared Single-Pixel Microscopy". In: *Scientific Reports* 13.1 (Jan. 2023), p. 281.

[9] Matthew P. Edgar, Graham M. Gibson, and Miles J. Padgett. "Principles and Prospects for Single-Pixel Imaging". In: *Nature Photonics* 13.1 (Jan. 2019), pp. 13–20.

[10] Samy Wu Fung et al. "JFB: Jacobian-Free Backpropagation for Implicit Networks". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 36.6 (June 2022), pp. 6648–6656.

[11] Catherine F. Higham et al. "Deep Learning for Real-Time Single-Pixel Video". In: *Scientific Reports* 8.1 (Feb. 2018), p. 2369.

[12] Stanley Bryan Z. Hua, Alex X. Lu, and Alan M. Moses. *CytoImageNet: A Large-Scale Pretraining Dataset for Bioimage Transfer Learning*. 2021.

[13] Nam Huynh et al. "Single-Pixel Camera Photoacoustic Tomography". In: *Journal of Biomedical Optics* 24.12 (Sept. 2019), p. 1.

[14] Eric Jang, Shixiang Gu, and Ben Poole. *Categorical Reparameterization with Gumbel-Softmax*. 2016.

[15] Erich Kobler et al. "Total Deep Variation for Linear Inverse Problems". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 7549–7558.

[16] Huan Li and Zhouchen Lin. "Accelerated Proximal Gradient Methods for Nonconvex Programming". In: *Advances in neural information processing systems* 28 (2015).

[17] S. Lucidi and F. Rinaldi. "Exact Penalty Functions for Nonlinear Integer Programming Problems". In: *Journal of Optimization Theory and Applications* 145.3 (June 2010), pp. 479–488.

[18] A. Mittal, A. K. Moorthy, and A. C. Bovik. "No-Reference Image Quality Assessment in the Spatial Domain". In: *IEEE Transactions on Image Processing* 21.12 (Dec. 2012), pp. 4695–4708.

[19] Sriram Ravula et al. *Optimizing Sampling Patterns for Compressed Sensing MRI with Diffusion Generative Models*. 2023.

[20] Otmar Scherzer et al. *Variational Methods in Imaging*. Vol. 167. Springer, 2009.

[21] Matthias Seeger et al. "Optimization of k -space Trajectories for Compressed Sensing by Bayesian Experimental Design". In: *Magnetic Resonance in Medicine* 63.1 (Jan. 2010), pp. 116–126.

[22] Amirreza Shaban et al. "Truncated Back-Propagation for Bilevel Optimization". In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Vol. 89. Proceedings of Machine Learning Research. PMLR, Apr. 2019, pp. 1723–1732.

[23] Ferdia Sherry et al. "Learning the Sampling Pattern for MRI". In: *IEEE Transactions on Medical Imaging* 39.12 (Dec. 2020), pp. 4310–4321.

[24] Fei Wang et al. “Single-Pixel Imaging Using Physics Enhanced Deep Learning”. In: *Photonics Research* 10.1 (Jan. 2022), p. 104.

[25] Yair Weiss, Hyun Sung Chang, and William T Freeman. “Learning Compressed Sensing”. In: *Snowbird Learning Workshop, Allerton, CA*. Citeseer, 2007.

[26] Shanshan Wu et al. “Learning a Compressed Sensing Measurement Matrix via Gradient Unrolling”. In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 6828–6839.

[27] Penghang Yin et al. *Understanding Straight-Through Estimator in Training Activation Quantized Neural Nets*. 2019.

[28] Zihao Zou et al. “Deep Equilibrium Learning of Explicit Regularization Functionals for Imaging Inverse Problems”. In: *IEEE Open Journal of Signal Processing* 4 (2023), pp. 390–398.

A Additional Results

Table 1: PSNR and SSIM for sampling patterns with respect to M, using the TV regulariser.

	$M = 128$		$M = 256$		$M = 512$		$M = 1024$		$M = 2048$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Gaussian	18.59	0.448	20.78	0.495	23.19	0.568	25.72	0.650	28.65	0.734
SH	19.98	0.458	22.27	0.522	24.88	0.602	27.60	0.689	30.31	0.778
Learned - RnP	23.48	0.544	24.56	0.579	25.98	0.630	27.64	0.689	29.89	0.765
Learned - STE	**24.62**	**0.580**	**26.59**	**0.641**	**28.14**	**0.697**	**29.61**	**0.752**	**30.95**	**0.797**

Table 2: PSNR and SSIM for sampling patterns with respect to M, using the TDV regulariser.

	$M = 128$		$M = 256$		$M = 512$		$M = 1024$		$M = 2048$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Gaussian	22.08	0.531	25.13	0.616	27.36	0.682	29.51	0.753	31.84	0.825
SH	22.37	0.529	25.20	0.621	27.36	0.686	29.75	0.761	32.18	0.834
Learned - RnP	24.47	0.590	26.18	0.642	27.97	0.702	29.68	0.759	31.77	0.822
Learned - STE	**25.75**	**0.630**	**27.60**	**0.684**	**29.31**	**0.741**	**30.94**	**0.795**	**32.38**	**0.839**

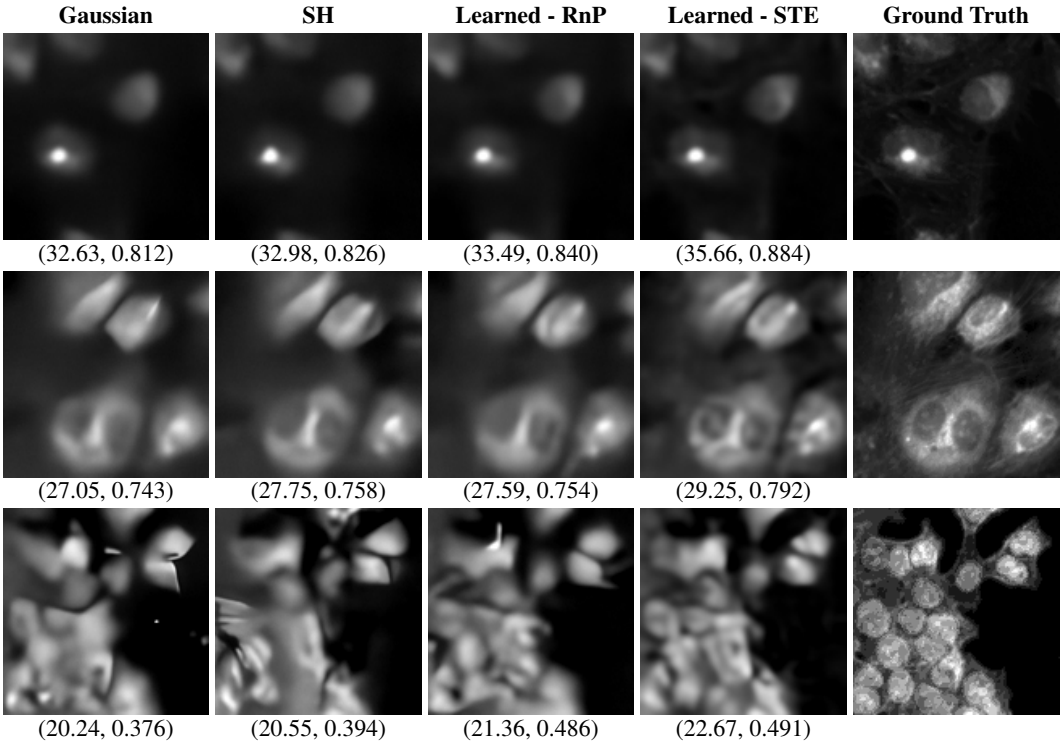


Figure 4: Comparison of the different sampling patterns for $M = 512$ and the TDV regulariser. We order the rows by best, median, and worst reconstruction PSNR of Learned - STE. We report the reconstruction quality metrics in brackets (PSNR, SSIM).

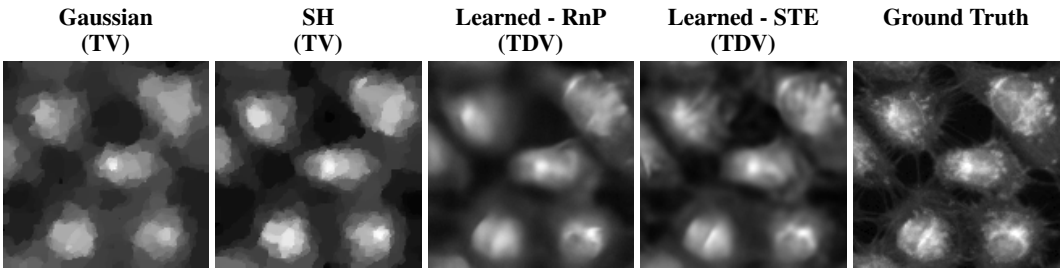


Figure 5: Comparison between classical approaches (Gaussian and SH using TV regularisers), and learned approaches (RnP and STE with TDV regulariser) for $M = 1024$. We report reconstruction quality metrics in brackets (PSNR, SSIM).

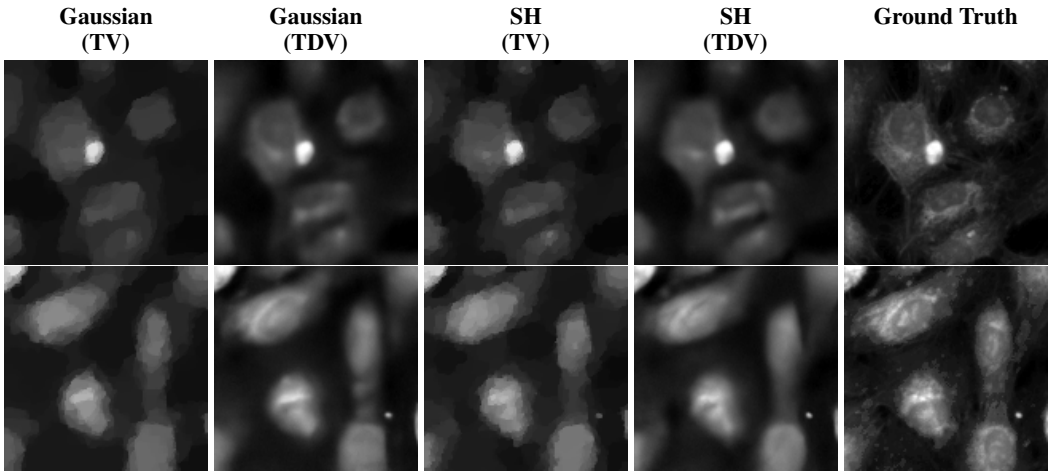


Figure 6: Comparison of the TDV and TV regulariser during reconstruction for both the Gaussian and SH patterns for $M = 1024$.