This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits noncommercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.
10.1017/S2633903X24000084

Reconstructing Interpretable Features in Computational Super-Resolution microscopy via Regularized Latent Search

Marzieh Gheisari and Auguste Genovesio

Institut de Biologie de l'ENS, Ecole Normale Supérieure, 46 rue d'Ulm, PSL Research University, 75005 Paris

Abstract. Supervised deep learning approaches can artificially increase the resolution of microscopy images by learning a mapping between two image resolutions or modalities. However, such methods often require a large set of hard-to-get low-res/high-res image pairs and produce synthetic images with a moderate increase in resolution. Conversely, recent methods based on GAN latent search offered a drastic increase in resolution without the need of paired images. However, they offer limited reconstruction of the high-resolution image interpretable features. Here, we propose a robust super-resolution method based on regularized latent search (RLS) that offers an actionable balance between fidelity to the ground-truth and realism of the recovered image given a distribution prior. The latter allows to split the analysis of a low-resolution image into a computational super-resolution task performed by deep learning followed by a quantification task performed by a handcrafted algorithm and based on interpretable biological features. This two-step process holds potential for various applications such as diagnostics on mobile devices, where the main aim is not to recover the high-resolution details of a specific sample but rather to obtain high-resolution images that preserve explainable and quantifiable differences between conditions.

Keywords: Microscopy· Super-resolution· Generative-prior· Diagnostic.

1 Introduction

Various deep learning models have shown excellent performance in the single-image super-resolution (SISR) task which aims to restore a high-resolution (HR) image from its low-resolution (LR) counterpart. Deep-learning SISR models have been applied few years ago to enhance the resolution of microscopy images [1–3]. More recently, studies have explored image-to-image translation models, that are trained to learn a parameterized function between two different image resolutions or modalities. These supervised approaches require a large number of paired images and rely on generative models that output artificial images [4]. These artificial images were accepted by the community as real mainly because they were measured as approaching real image instances. However, for the latter to

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

hold, these approaches only offered a moderate increase in resolution of up to 4x.

Importantly, learning a parameterized mapping from a low- to a high-resolution image is an ill-posed problem: a single low-resolution image corresponds to infinitely many highly resolved ones. Therefore, the super-resolution task cannot only consist of maximizing the fidelity of the recovered high-resolution image; it must be further constrained to be well-posed. In this work, we propose to constrain this task by imposing that the reconstructed super-resolution image be realistic, *i.e.* , belongs to a given image distribution, a principle that is also explored by utilizing gradient distribution prior in the context of biomedical images [5]. In short, on the one hand, fidelity refers to recovering a super-resolution image that, once downgraded, is close to the original image. On the other hand, realism refers to keeping the image within a given image domain.

In this paper, we aim to better define the super-resolution task by enforcing the solution to be a trade-off between fidelity to the original low-resolution image and realism, that includes the preservation of biologically relevant content, which we describe as "interpretable features". Assessing these features is essential for accurate phenotypic interpretation and discrimination. By guiding the super-resolution process towards this end, we generate images that are both faithful to the sample and biologically interpretable and quantifiable. In this way, we anticipate super-resolution in biology could exploit more than recovering only the details of a given sample, but benefit from features measured over a set of recovered images.

Our approach, Regularized Latent Search (RLS), consists of a regularized search in the latent space of a pre-trained generative model for high-resolution images. By doing so, we both suppress the need for paired images and make the problem well-posed as we search for the closest image a generator can produce that, when down-scaled, matches the low-resolution image input. Moreover, as we keep the super-resolved image in the original domain, instead of producing a moderate increase in resolution, we propose to push further the synthesis: we anticipate that creating very highly resolved (up to 32x) but controlled artificial images could be of great interest for applications such as diagnostic. This is because what is at stake, in this case, is the preservation of measurable and interpretable features of microscopy images, not the absolute matching with real image samples.

Similar to image-to-image translation methods like CycleGAN[6], which operate without the need for paired images, our method differentiates itself in its approach to super-resolution. It is specifically designed to address the ill-posed nature of super-resolution by utilizing a GAN's latent space to identify plausible HR images for any given LR input. Once the model is trained, it enables the super-resolution of any LR image with just an adjustment in the degradation function during inference. Unlike CycleGAN, which is deterministic and tailored for specific domain translations requiring individual models for each task, our method offers versatility and is suitable for a broad spectrum of super-resolution tasks. It eliminates the need to learn domain-specific features and is intended for

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

widespread application in super-resolution, capable of being trained on a single dataset to enhance any LR image.

2 Related Work and Background

2.1 Super-resolution of microscopy images

While optical super-resolution techniques such as STED [7] and PALM [8] can break the diffraction limit, they are limited by the need for specialized equipment and complex sample preparation. Computational super-resolution methods on the other hand, while they rely on existing training data, can represent a cheap way to enhance the resolution of images acquired with conventional microscopes. Several approaches were developed to address the problem of computational super-resolution of microscopy images. Deep-STORM [9] uses an encoder-decoder network to localize emitters in super-resolved images. Content-aware image restoration [1] uses a U-Net architecture and is trained with low signal-to-noise ratio (input) and high signal-to-noise ratio (target) image pairs. These approaches produced interesting results but recovered images often lacking high-frequency details due to the MSE loss. To balance this issue, other approaches such as ANNA-PALM [3] are based on generative adversarial network (GAN). The authors trained a U-Net, and in contrast with Deep-STORM, a combination of pixel-wise reconstruction loss and adversarial losses is used to obtain image reconstructions of better quality. Using a similar network architecture and training loss, Wang et al. [2] achieved super-resolution in fluorescence microscopy across different modalities. Overall, using an adversarial loss results in output images that are sharper and of better perceptual quality and these methods demonstrated the potential of deep learning to improve the spatial resolution of fluorescence microscopy images. While the deep learning architectures and the applications of these methods differ, they all require hard-to-get paired image data for training and offer a moderate increase in resolution of up to 4x.

It is essential to note that computational super-resolution techniques effectively enhance the resolution of images by optimizing the use of available data in the original images. However, these methods are constrained by the existing information and cannot add details beyond what the original optical systems could capture.

2.2 Style-based generative models

StyleGAN models [10, 11] are well-known for their ability to generate highly realistic images. The StyleGAN architecture consists of two sub-networks: a mapping network denoted by $\mathsf{G}_m : \mathbb{R}^d \to \mathbb{R}^d$, and a synthesis network consisting of L layers denoted by $\mathsf{G}_s : \mathbb{R}^{L \times d} \to \mathbb{R}^n$. Here, d represents the dimensionality of the latent space. The mapping network takes a sample $\mathbf{z} \in \mathbb{R}^d$ from a standard normal distribution and maps it to a vector $\mathbf{w} \in \mathcal{W}$, where $\mathcal{W}$ denotes the intermediate latent space. StyleGAN2 introduced path length penalty, which

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

encourages a fixed-size step in $\mathbf{w}$ to result in a fixed-magnitude change in the image. The regularization term is computed using the Jacobian determinant, and penalizing changes in it promotes the generation of smoother and more realistic images. L copies of the d-dimensional vector $\mathbf{w}$ are fed to the L-layer synthesis network G_s, with each copy representing the input to the corresponding layer of G_s. The G_s network controls the level of detail in the generated image at each layer. Individual modification of these L layers, by adjusting the latent vector copy of $\mathbf{w}$ for each layer, extends the latent space into$\mathcal{W}^+$. This extended latent space enhances the model's capability for accurate image reconstruction, which is vital for super-resolution tasks, offering more nuanced control over the image generation process [12, 13].

2.3 GAN-based high-resolution image reconstruction

The problem of obtaining a super-resolution (SR) image of dimension n from a low resolved (LR) image of dimension m is ill-posed as, for a non-invertible forward operator D with $m < n$, there are infinitely many high resolution (HR) images that match a given LR image. Thus the reconstruction procedure must be further constrained by prior information to better define the objective and lead to a stable solution. One such prior consisted of considering the reconstructed HR image to be part of a given domain. First, the distribution of HR images is learned in an unsupervised fashion, thanks to a GAN, then the latent space of this trained GAN is searched to find a latent vector producing an HR image that, once down-scaled, is the closest to the LR image input. GAN-prior-based images reconstruction was first introduced by Bora *et al.* [14] and further improved using StyleGAN [10, 11] by Menon *et al.* in PULSE [15] by constraining the search to remain on the image manifold. To this end, PULSE and two other studies [16, 13] use an invertible transformation of the intermediate latent space $\mathcal{W}^+$ which includes a leaky rectified linear unit (ReLU) [17] followed by an affine whitening transformation, so that transformed latent vectors approximately followed the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Sampled vectors are then constrained to lie around a hypersphere with radius $\sqrt{d}$ hypothesizing that most of the mass of a high-dimensional Gaussian distribution is located at or near $\sqrt{d}\mathcal{S}^{d-1}$, where $\mathcal{S}^{d-1}$ is the d-dimensional unit hypersphere. Constraining samples to lie in a dense area of the StyleGAN style distribution resulted in increased realism of the generated images.

Although the above approach showed major improvements over previous work, it also presents important caveats that in practice led to image artifacts. Here we show that transforming the intermediate latent space in this way does not lead to an accurate standard Gaussian distribution, and so prevents proper regularization based on this hypothesis. Moreover, as the search is strictly limited to the spherical surface $\mathcal{S}^{d-1}$, this limitation may prevent the search from reaching the closest reconstruction of the high-resolution image.

In this work, we regularize the search in the latent space for a latent code located in "healthy" regions of the latent space. In this way, the system is constrained to produce images that belong to the original image domain StyleGAN

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

was trained on. To do so, we take advantage of normalizing flow to gaussianize the latent style sample distribution which leads to a much closer standard Gaussian distribution. We then use this revertible transformation to regularize the search in $\mathcal{W}^+$ such that it remains in a high-density area of the style vector distribution. We then show experimentally that the latter produces reconstructed images that are not only realistic but also more faithful to the original HR image.

3 Method

3.1 Super-resolution by Regularized Latent Search

Super-resolution aims to reconstruct an unknown high resolution-image $\mathbf{x} \in \mathbb{R}^{n\times n}$ from a low-resolution image $\mathbf{y} \in \mathbb{R}^{m\times m}$, which is related to the HR image by a down-scaling process described by $\mathbf{y} = \mathsf{D}(\mathbf{x}) + \delta$, where $\mathsf{D} : \mathbb{R}^{n\times n} \rightarrow \mathbb{R}^{m\times m}$ is a non-invertible down-scaling forward operator and δ is an independent noise with distribution p_δ. We can formulate this task in terms of Maximum A Posteriori (MAP) estimation [18]. Given an LR image $\mathbf{y}$, our goal is to recover the HR image $\mathbf{x}$ as the MAP estimate of the conditional distribution $p_\mathsf{G}(\mathbf{x}|\mathbf{y})$:

$$\arg\max_{\mathbf{x}} \log p_\mathsf{G}(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} [\log p_\delta(\mathbf{y} - \mathsf{D}(\mathbf{x})) + \log p_\mathsf{G}(\mathbf{x})]. \tag{1}$$

The first term is the likelihood term describing the image degradation process D and the second is the image prior, describing the manifold of real HR images.

Image prior Let G_s be the synthesis network of a StyleGAN [11] pretrained on the considered image domain. G_s takes as input $\mathbf{w}$, produced by the mapping network, and outputs an image. The image prior $\log p_\mathsf{G}(\mathbf{x})$ can be expressed with respect to the latent variables $\mathbf{w}$:

$$\log p_\mathsf{G}(\mathsf{G}_s(\mathbf{w})) = \log p_\mathbf{w}(\mathbf{w}) + \log |\det J_{\mathsf{G}_s^{-1}}(\mathbf{w})|. \tag{2}$$

The second term can be dropped as the path length penalty in StyleGAN2 implies that the Jacobian determinant is constant for all $\mathbf{w}$. The first term $p_\mathbf{w}(\mathbf{w})$ is the image prior we define on $\mathbf{w} \in \mathcal{W}^+$ by:

$$\log p_\mathbf{w}(\mathbf{w}) = \lambda_w \mathcal{P}_w + \lambda_c \mathcal{P}_{cross} \tag{3}$$

where:

- $\mathcal{P}_w$ is a prior that keeps $\mathbf{w}$ in the area of high density in $\mathcal{W}^+$: $\mathcal{P}_w = \frac{1}{L}\sum_{i=1}^{L} \log p_\mathsf{F}(\mathbf{w}_i)$, where $p_\mathsf{F}(\mathbf{w})$ is estimated by a normalizing flow model F explained in further detail below.
- $\mathcal{P}_{cross}$ is a pairwise euclidean distance prior on $\mathbf{w} = [\mathbf{w}_1, \ldots, \mathbf{w}_L] \in \mathcal{W}^+$ that ensures $\mathbf{w} \in \mathcal{W}^+$ remains close to the trained manifold in $\mathcal{W}$: $\mathcal{P}_{cross} = -\sum_{i=1}^{L-1}\sum_{j=i+1}^{L} ||\mathbf{w}_i - \mathbf{w}_j||_2^2$
- λ_w and λ_c are hyperparameters that control the relative importance of the two priors.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

Normalizing Flow Using a sequence of invertible mappings, a Normalizing Flow $\mathsf{F} : \mathbb{R}^d \to \mathbb{R}^d$ is a transformation of an unknown complex distribution into a simple probability distribution that is easy to sample from and whose density is easy to evaluate such as standard Gaussian [19].

Let $\mathbf{z} = \mathsf{F}(\mathbf{w})$ with probability density function $p(\mathbf{z})$. Using the change-of-variable formula, we can express the log-density of $\mathbf{w}$ by [20]:

$$\log p_{\mathsf{F}}(\mathbf{w}) = \log p(\mathbf{z}) + \log |\det J_{\mathsf{F}}(\mathbf{w})|, \qquad \mathbf{z} = \mathsf{F}(\mathbf{w}) \tag{4}$$

where $J_{\mathsf{F}}(\mathbf{w})$ is the Jacobian of F evaluated at $\mathbf{w}$. In practice the Jacobian determinant in Eq. (4) should be easy to compute so that the density $p_{\mathsf{F}}(\mathbf{w})$ can be evaluated. Furthermore, as a generative model, the invertibility of F allows new samples $\mathbf{w} = \mathsf{F}^{-1}(z)$ to be drawn through sampling from the base distribution. In the literature, several flow models were proposed, such as Real Non-volume Preserving Flow [21] and Masked Auto-regressive Flow (MAF) [22].

Optimization process In the likelihood term presented in Eq. (1), we assume that the noise follows a Laplace distribution, that is $\delta \sim \mathit{Laplace}(0, \lambda_l I)$. This assumption simplifies the log-density of δ to: $\log p_\delta(\delta) = -||\delta||_1 - C$, where C is a constant. Thus the optimization problem in Eq. (1) is effectively transformed into an optimization over $\mathbf{w}$, resulting in the final objective function:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathsf{D}(\mathsf{G}_s(\mathbf{w}))||_1 - \log p_{\mathbf{w}}(\mathbf{w}) \tag{5}$$

The selection of the Laplace distribution for noise modeling is favored due to its simple form, which simplifies the log-density function and makes the loss function easier to optimize. Moreover, the L_1 norm, which arises in the log-density of the Laplace distribution, offers greater robustness to outliers than the L_2 norm. This characteristic is particularly beneficial in the context of super-resolution, where the ability to handle irregular noise or artifacts in low-resolution images is crucial, especially given the challenges commonly associated with microscopy data.

3.2 Evaluation of the Gaussianization process

To assess the quality of the gaussianization process, we sampled 5000 vectors $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. These vectors were then transformed into style vectors $\mathbf{w} = \mathsf{G}_m(\mathbf{z})$ using the mapping network G_m of StyleGAN. We then gaussianized the distribution of $\mathcal{W}$ using alternatively PULSE and our method. The quality of the gaussianization process was evaluated by computing the squared norm for all of these vectors within the transformed distribution (see Fig. 1). Ideally, the squared norm of the standard Gaussian distribution $\mathcal{Z}$ should follow a chi-squared distribution *i.e.* $||\mathbf{z}_n||_2^2 \sim \chi_d^2$ and thus forms a narrow distribution centered around $d = 512$, the dimensionality of $\mathcal{W}$. We can see that this is the case for $\mathcal{Z}$ which is Gaussian. However, the $\mathcal{W}$ does not follow this pattern, which is inconsistent with the prior assumption held by BRGM [18]. Furthermore, while the PULSE method results in a broader distribution, our method more accurately narrows the squared norm distribution to match the expected chi-squared distribution.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

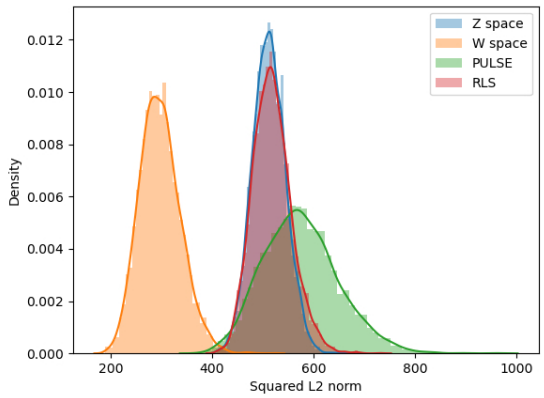


Fig. 1: Distribution analysis of squared L_2 norms demonstrating the gaussianization of latent style vectors. The graph compares the density of squared norms from the original $\mathcal{Z}$ space (blue), untransformed $\mathcal{W}$ space (orange), and the distributions resulting from the PULSE (green) and our method (red).

4 Experiments

4.1 Experimental settings

A critical aspect of our study was the necessity to work with high downscaling factors, such as 16x and 32x, which are central to our application scenario. These factors are not commonly used in the literature, to the best of our knowledge there are no experimental datasets available with such high downscaling factors. Most existing datasets typically focus on moderate downscaling factors of 2x or 4x, which is not aligned with the needs of our study, where we aim to address more extreme cases of resolution enhancement. Therefore, we had to generate our own low-resolution images.

As for implementation details, we began by training a StyleGAN2-ADA [23] model on a subset of the BBBC021 dataset [24, 25] which comprises cells treated with drugs acquired following the cell painting assay [26]. That is, cells were fluorescently stained with markers for F-actin, B-tubulin, and DNA, as described in [26]. The dataset comprises wide-field epi-fluorescence images, captured using the automated ImageXpress imaging platform. From this dataset, we extracted images of size 128×128 centered around each cell nucleus and used 400 images per compound treatment (approximately 28,000 in total) to train StyleGAN2-ADA [23]. We then proceeded to generate 100000 random samples $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and used their associated style vector $\mathbf{w} = \mathsf{G}_m(\mathbf{z})$ to train the normalizing flow model. We opted for MAF, as it tends to perform better than RealNVP for density estimation tasks. Our normalizing flow model comprised five flow blocks with all hidden dimensions set to 1024. For super resolving the images using our

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

regularized latent search algorithm, we employed an Adam optimizer over 200 iterations with a learning rate of 0.5 and initialized the search using the mean of 10,000 randomly generated latent vectors.

To fine-tune the regularization parameters λ_w and λ_c, we empirically evaluated a spectrum of values, eventually setting them to 5e-5 and 0.01, respectively. This process revealed that the algorithm's performance remained relatively stable across a broad range of these parameters, indicating a lack of sensitivity as long as the values fell within a specific boundary. Precisely, we observed optimal performance when λ_w was between 1e-6 and 5e-4, and λ_c ranged from 0.005 to 0.05. These findings suggest that while the exact values of λ_w and λ_c are flexible, maintaining them within these determined ranges ensures the algorithm functions effectively.

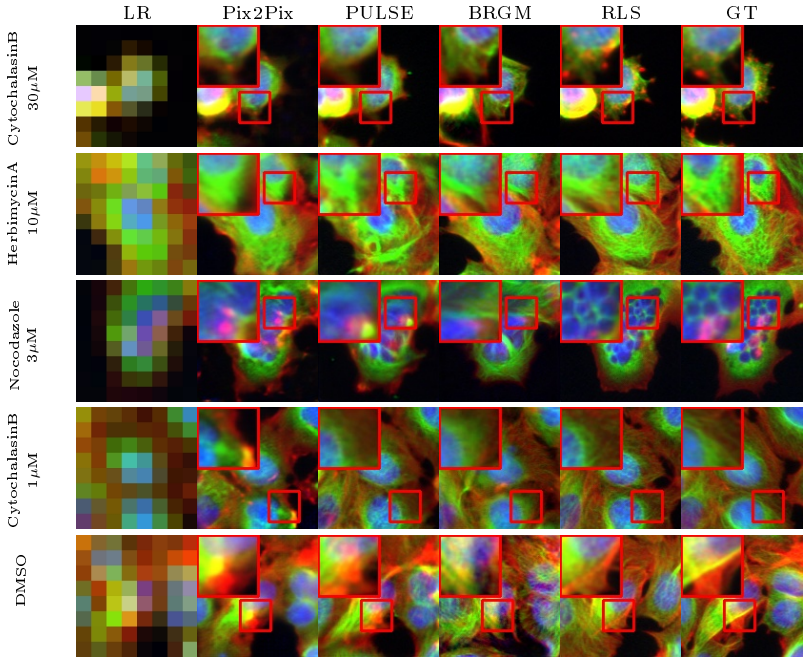


Fig. 2: Qualitative comparison of super-resolution reconstructions on the BBBC021 dataset: visualizing the performance of RLS against baseline methods in reconstructing cellular structures and phenotypes under negative control (DMSO) and various treatment conditions at a 16x upscaling factor.

For evaluation, we used 100 random samples from various compound treatments, ensuring samples were not included in the training of the generative network. We simulated degraded images from these high-resolution images using bicubic down-sampling, which is a common choice in the literature. We then

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

compared the results of our algorithm with those of Pix2Pix and two state-of-the-art unsupervised image reconstruction methods based on StyleGAN inversion: PULSE [15] and BRGM [18]. For the baseline methods, we used the same parameter settings reported by the original papers.

4.2 Results

RLS recovers high-quality cell images We reconstructed images using RLS and compared them to images generated with baseline methods. The results of this comparison is presented in Fig. 2. At these high upscaling factors, due to the lack of proper regularization, the competing methods fail to produce reasonable cellular details and tend to produce images that can be accurately downscaled to the LR image at the cost of generating distorted cellular structures. In contrast, our approach produces highly realistic images for most examples and can reproduce cell phenotypes induced by compound treatment even if the fine-grain details are not similar. For instance, when reconstructing the images of cells treated with Nocodazole, a known microtubule destabilizer, RLS captures the typical fragmented microtubule phenotype, resulting in a more accurate texture. Reconstructing images of cells at higher resolution obviously provides access to finer-grain measurable features.

To quantitatively assess this gain in performance, we evaluate the SR images using Frechet Inception Distance (FID) [27] and Kernel Inception Distance (KID) [28] to measure the discrepancy between the real HR images and the reconstructed one. As expected, the scores of FID and KID listed in Table 1 show that with both upscaling factors, our method significantly improves realism.

Upscaling factor	Method	FID↓	KID$^{(\times 10^3)}$↓	MSSIM↑	LPIPS↓	PSNR↑
32x	Pix2Pix [29]	67.141	66.515	0.507	0.420	15.846
	PULSE [15]	*13.629*	*9.056*	*0.526*	*0.378*	*16.469*
	BRGM [18]	15.862	13.293	0.400	0.475	14.207
	RLS	**7.571**	**3.209**	**0.528**	**0.343**	**16.536**
16x	Pix2Pix [29]	45.503	41.700	*0.708*	0.274	*19.983*
	PULSE [15]	*17.354*	*14.937*	**0.723**	**0.244**	**20.806**
	BRGM [18]	19.862	17.254	0.608	0.301	17.914
	RLS	**6.983**	**3.347**	0.660	*0.262*	19.177

Table 1: Quantitative evaluation of RLS and baseline methods for super-resolution on the BBBC021 dataset at 32x and 16x upscaling factors. The **best** and the *second-best* are highlighted in bold and italic respectively.

RLS achieves better perceptual scores We evaluated the reconstruction accuracy of RLS using LPIPS [30], PSNR, and MS−SSIM [31] metrics to compare the reconstructed SR images to the real HR images. As expected, RLS did not

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

achieve the best pixel-wise reconstruction losses but achieved either the lowest or the second-lowest LPIPS scores on both 32x and 16x upscaling factors. The latter suggested that while the recovered SR image is not exactly the same as the HR image, it is perceptually closer. It is worth noting that although Pix2Pix produced images with high PSNR and MSSIM scores, it struggled to accurately reconstruct cellular details. This observation highlights the fact that PSNR and SSIM metrics may not be fully appropriate to evaluate the performances of the super-resolution tasks.

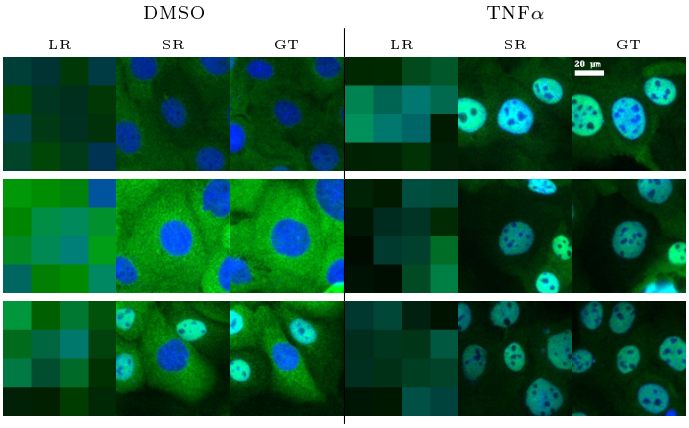


Fig. 3: Visual examples of super-resolving Translocation assay low-resolution images at a 32x upscaling factor under negative control (DMSO) and TNFα treatment conditions: left: low-resolution image, middle: super-resolution reconstruction, right: ground truth.

RLS preserves interpretable features RLS achieves a balance between realism and fidelity when super-resolving microscopy images. However, as any deep learning method for super-resolution, it cannot generate details it would not have seen during training. Therefore it cannot be used in a context where novel events can be expected. However, it could possibly be used to perform measurements to quantify expected phenotypic changes. This is the case for many assays used in basic research in biology. It is also the case in diagnostics such as parasitemia for instance where the tool must assess if a parasite is present or not. Instead of relying on a blind classification or regression of low-resolution images, these tasks on low-resolution images in such a context could be decomposed in two steps. A first step based on our deep learning approach would consist of reconstructing a high-resolution image, while a second step would use a dedicated handcrafted analysis to quantify a phenotypic feature, making the analysis explicitly interpretable.

To evaluate to what degree the information conveyed by the SR images can be used for such quantitative assays and maybe later for interpretable diagnos-

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

tics, we applied it to two assays that allowed straightforward quantification of "interpretable features" on HR images but not on LR images. Here, interpretable features refer to the measurable and explainable phenotypic changes crucial for biological assessment, such as the nucleo-cytoplasmic ratio in response to TNFα treatment and alterations in the Golgi apparatus morphology in reaction to nocodazole. Quantitative analysis was conducted on the SR images to assess the reconstruction of these features, and results were compared against those obtained from HR images as well as baseline methods, including BRGM, PULSE, and an unregularized latent search ('w/o regu.').

The first assay aimed to track the location of the NF-κB (nuclear factor kappa B) protein within the cell. Upon treatment with TNFα, a pro-inflammatory cytokine, the protein moves to the nucleus, causing a shift in fluorescence signal from the cytoplasm to the nuclear area, resulting in bright green nuclei. We observed that the nucleo-cytoplasmic fluorescence ratio measured on 1000 high-resolution (HR) treated and untreated images, could be replicated when computed from super-resolved (SR) images (Fig. 5a). We also provided some visual examples of super-resolving the low-resolution images *i.e.* the first step in Fig. 3 , which shows that our method can reconstruct images of wild-type and treated cell images obtained with this common assay.

The second assay we conducted aimed to monitor changes in the morphology of the Golgi apparatus. When cells are exposed to nocodazole, microtubules disassemble, causing the Golgi, originally located near the center of the cell, to break up into smaller stacks. Fig. 4 shows that a standard assay such as the nocodazole-induced Golgi scattering (green) were reproduced with SR images. Moreover, as illustrated in Fig. 5b, when computed only from SR images, a straightforward average spot size difference measured on 1000 HR treated and 1000 untreated images could be retrieved.

Furthermore, we utilized cells treated with DMSO as negative controls and cells treated with nocodazole (or TNF in the case of the translocation assay) as positive controls. We developed a classifier trained on high-resolution (HR) images to distinguish between the two phenotypes and evaluated its performance on both HR and super-resolved (SR) images, ensuring that the datasets for GAN training and classifier training were distinct to maintain the integrity of our results. Additionally, we trained a separate classifier on low-resolution (LR) images for phenotype differentiation, and its testing was confined to LR images. The results as shown in Table 2, indicate that the super-resolution operation does not degrade the discriminative signal contained in LR images, the classification accuracy is about the same on SR and LR images. Furthermore, we show evidence that this signal difference can be retrieved when quantifying the same interpretable feature variation present in HR images using handcrafted image analysis (see Fig. 5)

It is worth noting that in Fig. 5 the similarity in performance between the "w/o regu" model and RLS during the Translocation assay reveals an interesting insight. It suggests that, for this specific assay, the additional regularization does not significantly change the outcome. However, it's important to note that this

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

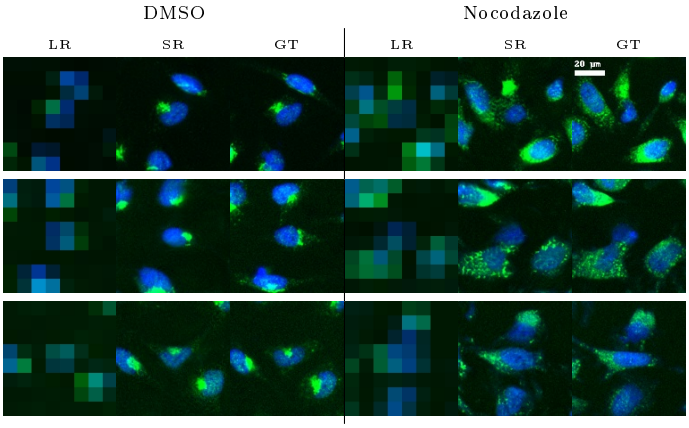


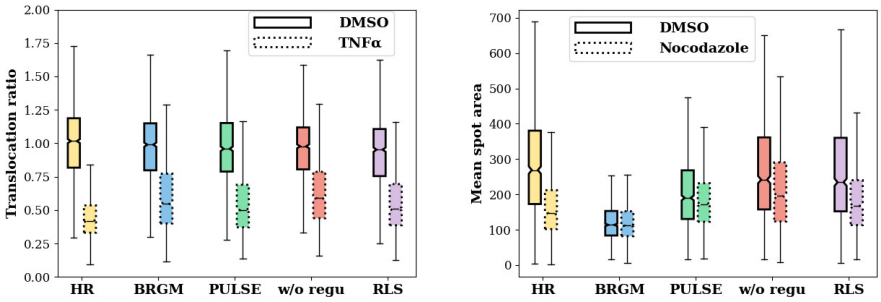
Fig. 4: Super-resolution of images from a Golgi assay at a 16x upscaling factor under negative control (DMSO) and Nocodazole treatment conditions. The left column shows the low-resolution images, the middle column shows the super-resolution reconstructions and the right column shows the ground truth images.

assay was designed to evaluate strong phenotypic changes where the expected changes are substantial and easy to detect. In this case, the non-regulated version can also accurately capture the phenotypic changes, since the differences between the images are large. This is evidenced by the comparative classification accuracies in Table 2 for the Translocation assay, where RLS achieves similar performance to the "w/o regu" version and the baseline methods. The advantages of regularization in RLS are more apparent in assays with more subtle phenotypic variations, where detecting nuanced differences is more challenging. For example, in the Golgi morphology assay, where detecting changes requires more detailed analysis, RLS's regularization reconstructs realistic images that more faithfully represent biological structures required for detailed quantitative evaluations. Furthermore, considering the overall performance, RLS demonstrates consistent and enhanced performance compared to both the baselines and the non-regularized version, as indicated by the superior classification results in Table 2. This suggests that RLS is a more robust method for super-resolving microscopy images, especially in assays where the phenotypic changes are subtle. Overall, the experiments confirm that the quality of super-resolved images is ad-

	LR	BRGM	PULSE	w/o Regu.	RLS	HR
DMSO vs TNFα (32x)	85.60	79.20	85.60	85.80	87.00	97.60
DMSO vs Nocodazole (16x)	72.80	53.00	62.60	67.80	74.20	96.80

Table 2: Comparison of classification accuracy for identifying phenotypic changes between negative control (DMSO) and positive control (TNFα and Nocodazole and Nocodazole conditions) in the translocation and Golgi assays respectively, using super-resolved (SR) images and high-resolution (HR) images as a benchmark.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press



(a) Translocation (32x) (b) Golgi (16x)

Fig. 5: Making interpretable measurements from low-resolution images: first increasing the resolution by a super-resolution method, then measuring a handcrafted interpretable feature. Each pair in the boxplots displays the distribution of handcrafted interpretable measurements, where the solid box represents the negative control (DMSO) and the dotted box signifies the positive controls (TNFα for translocation and Nocodazole for Golgi), across various super-resolution methods including RLS, BRGM, PULSE and "w/o regu" alongside with the high-resolution (HR) images for benchmarking. **(a)** Translocation ratio measurement: The y-axis quantifies the translocation ratio, an interpretable metric indicating TNF-induced NFκB translocation (green). The translocation ratio can be differentiated between two conditions from real high-resolution images (HR), but also from super-resolution images (SR). **(b)** Mean spot area measurement: The y-axis quantifies the mean spot area, an interpretable metric indicating nocodazole-induced Golgi spreading (green), distinguishable between two conditions from real high-resolution images (HR), but also from super-resolution images (SR) reconstructed by our method.

equate for further analysis. Our method enables a two-stage process: employing deep learning for the challenging task of super-resolution followed by a handcrafted, interpretable method for the subsequent quantitative measurements.

4.3 Robustness

In opposition to supervised methods that are sensitive to the input image domain, this approach is not restricted to a particular degradation operator that is used during training. To evaluate this aspect we applied additional degradation operators such as Gaussian noise, Salt and Pepper, and Gaussian blur to a bicubic downscaled image DS before reconstruction. As depicted in Fig. 6, the reconstruction closely matches the DS image. This result validates our choice of using the bicubic downscaling operator during training instead of more complicated specific degradation.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

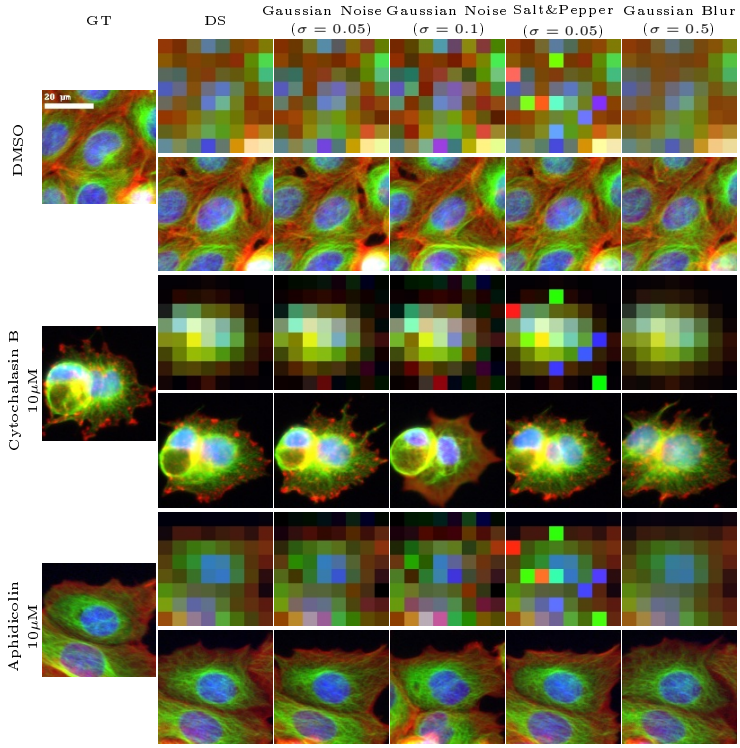


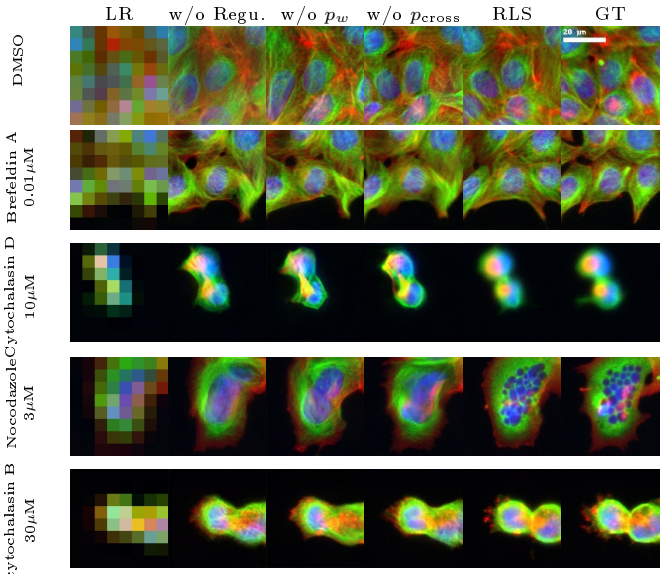
Fig. 6: Evaluation of RLS performance under various degradation conditions: images downscaled by bicubic method are further altered with Gaussian noise, Salt and Pepper noise, and Gaussian blur to assess the stability of the proposed method across a range of image perturbations (at a 16x upscaling factor).

4.4 Ablation study

Fig. 7 demonstrates ablation experiments highlighting the impact of the components of our image prior. First, "w/o Regu." searches the latent space without any regularization for the image that, once downscaled matches the LR image. The second variant is denoted "w/o $\mathcal{P}_w$", *i.e.* , the image prior does not include the prior term $\mathcal{P}_w$. Similarly, "w/o $\mathcal{P}_{\text{cross}}$" refer to the suppression of $\mathcal{P}_{cross}$.

To evaluate the three variants, we use the same set of parameters described in Section 4.1. One can see that searching the latent space without any regularization produces images that do not necessarily belong to the original image manifold and therefore do not appear realistic. It also tends to generate images that are accurately downscaled to the LR image but at the cost of generating distorted images when $\mathcal{P}_{\text{cross}}$ or $\mathcal{P}_w$ is discarded. This implies that both priors play an important role in generating realistic details.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press



Variant	FID↓	PSNR↑
w/o Regu.	14.69	20.56
w/o p_w	7.96	20.21
w/o p_{cross}	8.99	21.24
RLS	6.98	19.18

Fig. 7: Ablation study showcasing the impact of regularization components on RLS performance, with qualitative results in the left column and quantitative results in the right column. The variants include: "w/o Regu." (searching the latent space without any regularization), "w/o p_w" (the image prior does not include the prior term p_w), "w/o p_{cross}" (the image prior does not include the prior term p_{cross}), and "RLS" (the full RLS model) (at a 16x upscaling factor).

4.5 Uncertainty.

An important challenge of the super-resolution task is that it is an ill-posed problem. Although we can improve this aspect by using an image prior constraint, several closely related highly resolved images could still be consistent with a single LR image. To generate n realistic SR images, we sample n latent codes denoted as $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$. We assume that their distribution follows a Gaussian distribution $\mathcal{N}(\mu, \sigma)$, where the parameter σ follows an inverse gamma distribution. Using the Bayes rule, we estimate the distribution's parameters by:

$$\max_{\mu,\sigma} \left[\sum_{i=1}^{n} \log p(\mathbf{w}_i|y) - \frac{n}{2}\log\sigma^2 - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{w}_i - \mu)^2 + \log p(\sigma) \right]$$

Here, the first term is the log-likelihood of the posterior distribution of $p(\mathbf{w}|\mathbf{y})$, which is defined in Eq. (5), the second term is the regularization term, which penalizes large values of σ and $\log p(\sigma)$ is the log prior distribution of σ. Fig. 8 displays multiple solutions for a given LR image we can obtain with our approach. Sampling close variations of HR images from a single LR input can be used to enhance the robustness of the estimation of an image-based quantitative

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

feature by reducing the effects of noise or artifacts that may exist in the input LR image.

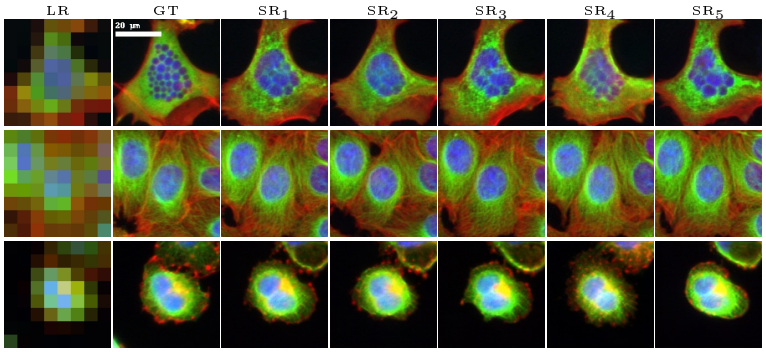


Fig. 8: Visualizing the ill-posed nature of the super-resolution task and the uncertainty associated with SR reconstruction. Five distinct SR images (SR_1 to SR_5) are generated for each LR image by sampling five different latent codes from the latent space, alongside the ground truth (GT) (at a 16x upscaling factor).

5 Conclusion

In this paper, we propose a robust super-resolution method based on regularized latent search within a pre-trained StyleGAN. It does not require coupled image pairs for training and constraints the super-resolution task to a given image prior to offer a trade-off between fidelity and realism of the SR reconstruction. Furthermore, we show that such a method could be used to split analyses, such as the classification of low-resolution images, into a reconstruction of super-resolution images performed by deep learning and a simple dedicated handcrafted analysis of an interpretable feature.

The latter could be used for instance for rapid diagnostic based on smartphone directly available on the field. In this case, a dataset of low-resolution images of slides acquired from fast and/or cheap solutions available on the field could be coupled with the acquisition of the same slides using high-end expensive high-resolution microscopes with limited access. In this application, high-resolution images could then be reconstructed directly on the field to perform an explainable diagnostic such as a parasite count.

6 Limitations and future work

One common challenge with deep learning-based methods, especially those using generative priors like in our study, is their limited generalizability to unfamiliar and unseen data. Our approach, which utilizes StyleGAN for unsupervised

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

learning within a specific domain, may not encompass the full diversity encountered in real-world scenarios. To enhance the model's generalizability, strategies such as incorporating data augmentation to introduce training data variability, or applying transfer learning for domain adaptation, could be beneficial.
Our current findings serve as an initial validation of our method's capabilities. However, we acknowledge the necessity of further evaluations using more varied datasets, encompassing a broader range of imaging techniques and sample preparation methods. The absence of such diverse experimental data in our current research is due to the specific requirement of our study to explore super-resolution at high upscaling factors, like 16x and 32x. These factors, which are crucial to our application's needs, are rarely addressed in the available literature, and to our knowledge, datasets with such extreme upscaling factors are not yet available. Existing datasets generally focus on more moderate upscaling factors, such as 2x or 4x, which do not meet the demands of our research that targets significantly higher levels of resolution enhancement. We are committed to extending our validation to include real-world imaging conditions as soon as datasets meeting our high upscaling factor requirements become available. This will enable a more comprehensive assessment of our method's applicability and performance in practical scenarios.

7 Acknowledgments

We thank GENCI for the access to the HPC resources of IDRIS under the allocation 2020-AD011011495.

8 Competing interests

The authors declare no competing interests.

9 Author contributions

MG proposed and implemented the RLS method. AG hypothesized that RLS reconstructs interpretable features and suggested experiments to validate it. MG run the experiments. MG and AG wrote the manuscript.

10 Funding statement

This work was supported by the ANR VISUALPSEUDOTIME, ANR–10–LABX–54 MEMOLIFE and ANR–10 IDEX 0001–02 PSL* Université Paris.

11 Data availability

We used the BBBC021 image set available from the Broad Bioimage Benchmark Collection (https://bbbc.broadinstitute.org/).

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

References

1. Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, 15(12):1090–1097, 2018.
2. Hongda Wang, Yair Rivenson, Yiyin Jin, Zhensong Wei, Ronald Gao, Harun Günaydın, Laurent A Bentolila, Comert Kural, and Aydogan Ozcan. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature methods*, 16(1):103–110, 2019.
3. Wei Ouyang, Andrey Aristov, Mickaël Lelek, Xian Hao, and Christophe Zimmer. Deep learning massively accelerates super-resolution localization microscopy. *Nature biotechnology*, 36(5):460–468, 2018.
4. David P Hoffman, Isaac Slavitt, and Casey A Fitzpatrick. The promise and peril of deep learning in microscopy. *Nature methods*, 18(2):131–132, 2021.
5. Yuanhao Gong and Ivo F Sbalzarini. A natural-scene gradient distribution prior and its application in light-microscopy image processing. *IEEE Journal of Selected Topics in Signal Processing*, 10(1):99–114, 2015.
6. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2223–2232, 2017.
7. Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782, 1994.
8. Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.
9. Elias Nehme, Lucien E Weiss, Tomer Michaeli, and Yoav Shechtman. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458–464, 2018.
10. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.
11. Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.
12. Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4432–4441, 2019.
13. Jonas Wulff and Antonio Torralba. Improving inversion and generation diversity in stylegan using a gaussianized latent space. *arXiv preprint arXiv:2009.06529*, 2020.
14. Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *Proceedings of the International Conference on Machine Learning*, pages 537–546. PMLR, 2017.
15. Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2445, 2020.

16. Peihao Zhu, Rameen Abdal, Yipeng Qin, John Femiani, and Peter Wonka. Improved stylegan embedding: Where are the good latents? *arXiv preprint arXiv:2012.09036*, 2020.
17. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
18. Razvan Marinescu, Daniel Moyer, and Polina Golland. Bayesian image reconstruction using deep generative models. In *NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.
19. Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. 43(11):3964–3979, 2020.
20. George Papamakarios, Eric T Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. 22(57):1–64, 2021.
21. Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.
22. George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. 2017.
23. Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. pages 12104–12114, 2020.
24. Juan C Caicedo, Claire McQuin, Allen Goodman, Shantanu Singh, and Anne E Carpenter. Weakly supervised learning of single-cell feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9309–9318, 2018.
25. Vebjorn Ljosa, Peter D Caie, Rob Ter Horst, Katherine L Sokolnicki, Emma L Jenkins, Sandeep Daya, Mark E Roberts, Thouis R Jones, Shantanu Singh, Auguste Genovesio, et al. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *Journal of biomolecular screening*, 18(10):1321–1329, 2013.
26. Peter D Caie, Rebecca E Walls, Alexandra Ingleston-Orme, Sandeep Daya, Tom Houslay, Rob Eagle, Mark E Roberts, and Neil O Carragher. High-content phenotypic profiling of drug response signatures across distinct cancer cells phenotypic profiling across cancer cell types. *Molecular cancer therapeutics*, 9(6):1913–1926, 2010.
27. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 30, 2017.
28. Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.
29. Hao-Chih Lee, Sarah T Cherng, Riccardo Miotto, and Joel T Dudley. Enhancing high-content imaging for studying microtubule networks at large-scale. In *Proceedings of Machine Learning for Healthcare Conference*, pages 592–613. PMLR, 2019.
30. Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press

31. Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Proceedings of Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1398–1402. IEEE, 2003.

https://doi.org/10.1017/S2633903X24000084 Published online by Cambridge University Press