This CVPR 2021 paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Mol2Image: Improved Conditional Flow Models for Molecule to Image Synthesis

Karren Yang[1] Samuel Goldman[1] Wengong Jin[1] Alex X. Lu[2]
Regina Barzilay[1] Tommi Jaakkola[1] Caroline Uhler[1]*
[1]Massachusetts Institute of Technology [2]University of Toronto

Abstract

In this paper, we aim to synthesize cell microscopy images under different molecular interventions, motivated by practical applications to drug development. Building on the recent success of graph neural networks for learning molecular embeddings and flow-based models for image generation, we propose Mol2Image: a flow-based generative model for molecule to cell image synthesis. To generate cell features at different resolutions and scale to high-resolution images, we develop a novel multi-scale flow architecture based on a Haar wavelet image pyramid. To maximize the mutual information between the generated images and the molecular interventions, we devise a training strategy based on contrastive learning. To evaluate our model, we propose a new set of metrics for biological image generation that are robust, interpretable, and relevant to practitioners. We show quantitatively that our method learns a meaningful embedding of the molecular intervention, which is translated into an image representation reflecting the biological effects of the intervention.

1. Introduction

High-content cell microscopy assays are gaining traction in recent years as the rich morphological data from the images proves to be more informative for drug discovery than conventional targeted screens [6, 12, 55]. Motivated by these developments, we aim to build, to our knowledge, the first generative model to synthesize cell microscopy images under different molecular interventions, translating molecular information into a high-content and interpretable image representation of the intervention. Such a system has numerous practical applications in drug development – for example, it could enable practitioners to virtually screen compounds based on their predicted morphological effects on cells, allowing more efficient exploration of the vast chemical space and reducing the resources required to perform extensive experiments [46, 53, 57]. Small molecules are known to enter cells and alter their biological functions and pathways, leading to changes in cell shape, structure, organization, *etc*., that are visible in microscopy images [14, 13]. In contrast to conventional models that predict specific chemical properties, a molecule-to-image synthesis model has the potential to produce a panoptic view of the morphological effects of a drug that captures a broad spectrum of properties such as mechanisms of action [33, 34, 45] and gene targets [4].

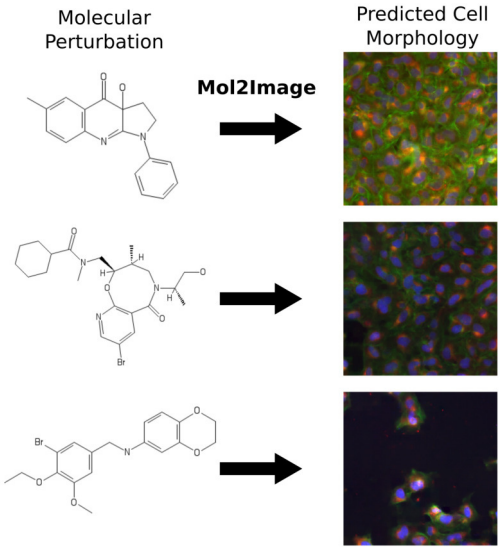


Figure 1: **Molecule to Image Synthesis.** High-content cell morphology images captured under different molecular interventions enable practitioners to assess a broad spectrum of drug effects. We improve on state-of-the-art flow-based generative models to build a molecule to image synthesis model with potential applications to virtual chemical screening.

To build our molecule-to-image synthesis model (Mol2Image), we integrate state-of-the-art graph neural networks for learning molecular representations with flow-based generative models. Flow-based models are a relatively recent class of generative models that learn the data distribution by directly inferring the latent distribution and maximizing the log-likelihood of the data [9, 10, 26]. Compared to other

*To whom correspondence should be addressed: KY, karren@mit.edu; CU, cuhler@mit.edu

classes of deep generative models such as variational autoencoders (VAEs) [27] and generative adversarial networks (GANs) [17], flow-based models do not rely on approximate posterior inference or adversarial training to learn the data distribution and are less prone to training instability and mode collapse, making them reliable and advantageous for biological practitioners [54].

However, molecule-to-image synthesis is a challenging task that highlights key, unsolved problems in flow-based image generation. First, state-of-the-art flow models such as NVP [10] and Glow [26] cannot be trained on large images such as full-resolution cell images (*e.g.*, 512×512) due to memory constraints, as flow models contain significantly more parameters than other generative models. Second, conditional variants of flow models are not nearly as well-developed as their counterparts in generative adversarial networks. While many variants of conditional GANs have been developed to synthesize images from complex information such as text [47], conditional flow models have so far synthesized images from binary vectors such as image classes or attributes [26, 32] or other images [1], where the correspondence between the image and the conditioning information is more straightforward. These open problems in flow-based models limit their application to real-world problems such as molecule-to-image synthesis.

Contributions. The contributions of this work are two-fold. (1) We improve on state-of-the-art flow-based generative models to develop a model that can generate high-resolution cell images conditioned on molecular interventions. Our methodological contributions to flow models include:

- A new multi-scale flow model based on the framework of a Haar wavelet image pyramid that is trained to generate images in a coarse-to-fine fashion and can scale to large, high-resolution cell images. The existing state-of-the-art model, Glow [26], cannot be trained on images larger than 256×256 due to memory constraints. Our principled choice of the Haar wavelet image pyramid enables us to scale training to large images, while preserving the original objective of maximizing the log-likelihood of the data.
- A training algorithm for conditional flow models that leverages contrastive learning to maximize the mutual information between the generated images and the conditioning molecules. Although we focus on molecule-to-image synthesis, this approach can potentially extend to other challenging applications of conditional flow models, e.g., text-to-image synthesis [47].

(2) We establish a new benchmark for molecule to image synthesis on the Cell Painting dataset [2], a high-content cell microscopy assay, motivated by practical applications to drug development and virtual chemical screening. To evaluate models on this task, we propose a new set of evaluation metrics specific to cell image generation that are robust, interpretable, and relevant to practitioners. We show that our approach outperforms the baselines on this task, indicating potential for virtual screening.

2. Related Work

Biological Image Generation. Osokin *et al*. [44] use GAN architectures to generate cellular images of budding yeast to infer missing fluorescence channels (stained proteins) in a dataset where only two channels can be observed at a time. Separately, Goldsborough *et al*. [16] qualitatively evaluate the use of different GAN variants in generating three-channel images of human breast cancer cell lines. While these works consider the task of generating single cell images, neither considers the generation of cells conditioned on complex inputs nor the generation of multi-cell images, which is useful in observing cell-to-cell interactions [42] and variability [39]. A separate, similar line of investigation in histopathology and medical imagery has used GAN models to refine and generate synthetic datasets for training downstream classifiers but does not address the difficulty of conditional image generation necessary to capture drug interventions [22, 38, 61]. While both high throughput image-based drug screens [5] and molecular structures [60] have been used to generate representations of small molecules, little work has focused on learning representations of these modalities jointly.

Graph Neural Networks for Molecules. A neural network formulation on graphs was first proposed by Gori *et al*. [18]; Scarselli *et al*. [51] and later extended to various graph neural network (GNN) architectures [31, 7, 41, 28, 19, 30, 56, 59]. In the context of molecule property prediction, Duvenaud *et al*. [11] and Kearns *et al*. [24] first applied GNNs to learn neural fingerprints for molecules. Gilmer *et al*. [15] further enhanced GNN performance by using set2set readout functions and adding virtual nodes into molecular graphs. Yang *et al*. [60] provided extensive benchmarking of various GNN architectures and demonstrated the advantage of GNNs over traditional Morgan fingerprints [49] as well as domain-specific features. While these works mainly focused on predicting numerical chemical properties, we here focus on using GNNs to learn rich molecular representations for molecule-to-image synthesis.

Flow-Based Generative Models. A flow-based generative model (e.g., Glow) is a sequence of invertible networks that transforms the input distribution to a simple latent distribution such as a spherical Gaussian [9, 10, 20, 26, 35, 48]. Conditional variants of Glow have recently been proposed for image segmentation [37, 58], modality transfer [29, 54], image super-resolution [58], and image colorization [1]. These applications are variants of image-to-image translation tasks and leverage the spatial correspondence between the conditioning information and the generated image. Other conditional models perform generation given an image class [26] or a binary attribute vector [32]. Since the condition is

categorical, these models apply auxiliary classifiers in the latent space to ensure that the model learns the correspondence between the condition and the image. Unlike these works, we generate images from molecular graphs; here spatial correspondence is not present and the conditioning information cannot be learned using a classifier. Therefore we must leverage other techniques to ensure correspondence between the generated images and the conditioning information.

In addition to conditioning on molecular structure, our flow model architecture is based on an image pyramid, which conditions the generation of fine features at a particular spatial resolution on a coarse image from another level of the pyramid. Flow-based generation of images conditioned on other images has been explored in various previous works [1, 29, 37, 54, 58], but different from these works, our flow-based model leverages conditioning to break generation into successive steps and refine features at different scales. Our approach is inspired by methods such as Laplacian Pyramid GANs [8] that break GAN generation into successive steps. A key design choice here is our use of a Haar wavelet image pyramid instead of a Laplacian pyramid, which is an important consideration that allows us to prove that our optimization procedure still maximizes the log-likelihood of the data. Ardizzone *et al*. [1] use the Haar wavelet transform to improve training stability, but they do not consider the framework of an image pyramid for separately generating features at different spatial resolutions.

3. Approach

In the following, we improve on state-of-the-art flow models to develop a flow-based generative model to synthesize cell images conditioned on molecular interventions. We first provide an overview of generative flows (Section 3.1). In Section 3.2, we describe our novel multi-scale flow model that generates images in a coarse-to-fine process based on the framework of a Haar wavelet image pyramid. Our flow model separates generation of image features at different spatial resolutions and scales to high-resolution cell images. In Section 3.3, we describe the full architecture of our conditional flow model that conditions image generation on the molecular embeddings of a graph neural network. We also propose an effective training strategy for conditional flow models that leverages contrastive learning to maximize the correspondence between generated images and molecular structure.

3.1. Preliminaries: Generative Flows

A generative flow consists of a sequence of invertible functions $f_1 \circ \cdots \circ f_L$ that transform an input variable $\mathbf{x}$ (*i.e.*, an image) to a latent variable $\mathbf{z}$. To generate an image, $\mathbf{z}$ is sampled from a Gaussian distribution and passed through the inverse of the flow functions:

$$\mathbf{z} \sim \mathcal{N}(\mu, \Sigma), \quad \mathbf{h}_L = \mathbf{z},$$
$$\mathbf{h}_{L-1} = f_L^{-1}(\mathbf{h}_L), \ \cdots, \ \mathbf{h}_0 = f_1^{-1}(\mathbf{h}_1), \quad \mathbf{x} = \mathbf{h}_0, \quad (1)$$

where $\{\mathbf{h}_i\}_{i \in 1 \cdots L}$ are the intermediate variables that arise from applying the inverse of individual flow functions $\{f_i\}_{i \in 1 \cdots L}$. During training, the log-likelihood of sampling target images $\mathbf{x}$ from the model is directly computed and optimized using the change-of-variables formula:

$$\log p(\mathbf{x}) = \log p_{\mathcal{N}}(\mathbf{z}; \mu, \Sigma) + \sum_{i=1}^{L} \log \left| \det \frac{d\mathbf{h}_i}{d\mathbf{h}_{i-1}} \right|, \quad (2)$$

where $p_{\mathcal{N}}$ is the Gaussian probability density function. Standard invertible functions for transforming the image include activation normalization, 1×1 convolution, and affine coupling layers [26]. The Jacobian matrices of these transformations are triangular, which makes the log-determinants in Equation 2 computationally tractable.

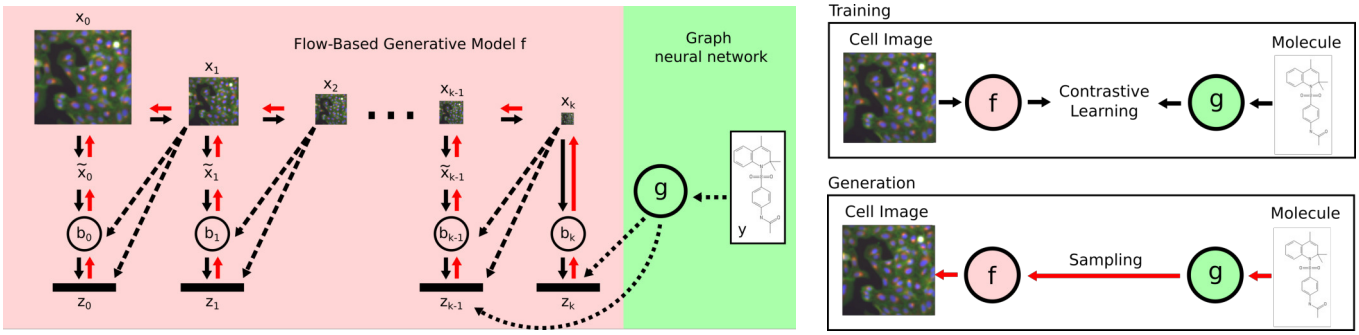
3.2. Haar Pyramid Flow Model

Transforming a Gaussian vector into an image using only invertible flow functions (Equation 1) requires us to compose many of these functions together. As a result, flow models contain significantly more parameters than other types of generative models, and existing multi-scale flow models for image generation that require end-to-end training [10, 26] cannot be applied to high-resolution (*i.e.*, 512×512) cell images due to memory constraints. Therefore, we propose a novel multi-scale flow model that successively generates images at multiple scales based on a Haar wavelet image pyramid, going from coarse-to-fine resolution. In contrast to existing work [10, 26], our model scales to high-resolution cell images, without changing the objective of maximizing the log-likelihood of the data.

Haar Wavelet Image Pyramid. Wavelets are functions that can be used to decompose an image into coarse and fine components. The Haar wavelet generates the coarse component in a way that is equivalent to nearest neighbor downsampling. The coarse component is obtained by convolving the image with an averaging matrix followed by sub-sampling by a factor of 2, and the fine components are obtained by convolving the image with three different matrices followed by sub-sampling by a factor of 2:

$$M_{\text{average}} = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, M_{\text{diff1}} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad (3)$$
$$M_{\text{diff2}} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, M_{\text{diff3}} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

To generate an image pyramid that captures features at different spatial resolutions, we recursively apply Haar



(a) Model Architecture

(b) Training Strategy

Figure 2: (a) (Red box) Our flow-based model architecture based on a Haar wavelet image pyramid. Information flow follows the black arrows during training/inference and the red arrows during generation. The dashed lines represent conditioning and are used in both training and generation. (Green box) Molecular information is processed and input to the network via a graph neural network g. (b) Our training strategy for effective molecule-to-image synthesis. See text for details.

wavelet transforms to the coarse image. Specifically, let $[\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k]$ be a pyramid of downsampled images, where $\mathbf{x}_i$ represents the image $\mathbf{x}_0$ after i applications of the coarse operation. We apply the fine operation to each downsampled image except the last, resulting in the image pyramid $[\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1, \cdots, \tilde{\mathbf{x}}_{k-1}, \mathbf{x}_k]$. The image at each spatial resolution can be reconstructed recursively,

$$\mathbf{x}_i = I([U(\mathbf{x}_{i+1}), \tilde{\mathbf{x}}_i]),$$

where U represents spatial upsampling, the brackets indicate concatenation, and I represents the inverse of the linear operation corresponding to the 2D Haar wavelet transform; see Equation (3).

Haar Pyramid Generative Flow. Our proposed multi-scale flow model f consists of multiple blocks $b_0, \cdots, b_k$, each responsible for generating the fine features for a different level of the Haar image pyramid conditioned on a coarse image from the next image in the pyramid; see Figure 2a, red box. Note that each block b_i consists of multiple invertible flow units, i.e., $b_i = f_1^{(i)} \circ \cdots \circ f_L^{(i)}$ and can be treated independently as a smaller generative flow from Section 3.1. The generative process is defined as follows. First we generate the final downsampled image of the pyramid,

$$\mathbf{z}_k \sim \mathcal{N}(\mu_k, \Sigma_k), \quad \mathbf{x}_k = b_k^{-1}(\mathbf{z}_k), \tag{4}$$

by sampling a latent vector that corresponds to the coarsest features and passing it through the first block. Then we recursively sample latent vectors corresponding to finer spatial features and generate the other images in the Haar image pyramid as follows:

$$\mathbf{z}_i \sim \mathcal{N}(\mu_i(\mathbf{x}_{i+1}), \Sigma_i(\mathbf{x}_{i+1})),$$
$$\tilde{\mathbf{x}}_i = b_i^{-1}(\mathbf{z}_i, \mathbf{x}_{i+1}), \quad \mathbf{x}_i = I([U(\mathbf{x}_{i+1}), \tilde{\mathbf{x}}_i]), \quad 0 \le i < k,$$

where $\mathbf{x} = \mathbf{x}_0$ is the final full-resolution image. To perform conditioning on the coarse image $\mathbf{x}_{i+1}$, we provide it as an additional input to both the prior distribution of $\mathbf{z}_i$ and to the individual flow units in b_i. We optimize the parameters of f by maximizing the conditional log-likelihood of the fine features $\tilde{\mathbf{x}}_i$ given the coarser image $\mathbf{x}_{i+1}$ for every level of the image pyramid (except the last layer, which uses standard log-likelihood):

$$\mathcal{L}(\mathbf{x}) = \sum_{i=0}^{k-1} \left(\log p_{\mathcal{N}}\big(\mathbf{z}_i; \mu_i(\mathbf{x}_{i+1}), \Sigma_i(\mathbf{x}_{i+1})\big) + \log \left| \det \frac{\partial \mathbf{z}_i}{\partial \tilde{\mathbf{x}}_i} \right| \right) + \left(\log p_{\mathcal{N}}\big(\mathbf{z}_k; \mu_k, \Sigma_k\big) + \log \left| \det \frac{d\mathbf{z}_k}{d\mathbf{x}_k} \right| \right)$$

The partial derivatives reflect that $\mathbf{z}_i$ depends on $\tilde{\mathbf{x}}_i$ as well as $\mathbf{x}_{i+1}$. Note that the optimization of each flow block b_i is uncoupled from the rest, which enables our model training to scale to high-resolution cell images. We now show that this procedure is more than a heuristic: optimizing these conditional log-likelihoods is equivalent to optimizing the log-likelihood of the data given in Equation (2).

Proposition 1 *Let f denote the multi-scale flow model based on a Haar image pyramid. Given an image $\mathbf{x} \in \mathbb{R}^{C \times W \times W}$ ($c \ge 1, W = 2^K, K \ge k$), the log-likelihood of sampling $\mathbf{x}$ from f can be computed exactly as,*

$$\log p(\mathbf{x}) = \mathcal{L}(\mathbf{x}) + CW^2 \log 2 \sum_{i=0}^{k-1} 2^{1-2(i+1)}.$$

In other words, the log-likelihood of sampling x is equal to the sum of (conditional) log-likelihoods of sampling each image in the Haar pyramid, up to a constant term given by the image dimension (CW^2) and depth k. The proof hinges on the observation that the Haar pyramid is a complete, invertible linear function of the original image and is provided in the Supplementary Material.

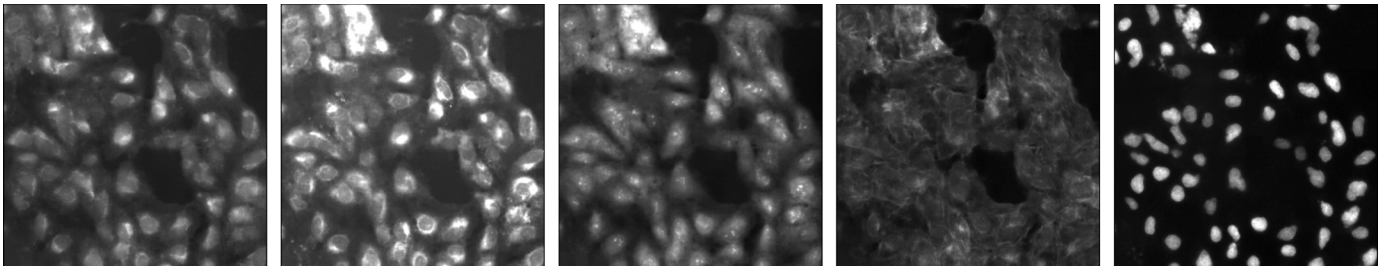


Figure 3: Example of a 5-channel 512 × 512 cell image generated by our multi-scale Haar image pyramid flow model. From left to right: mitochondria, endoplasmic reticulum, nucleoli/cytoplasmic RNA, actin (cytoskeleton), DNA (nucleus).

3.3. Conditioning on Molecular Graph

We now describe our full conditional flow model architecture and training strategy for conditioning the output of the multi-scale flow model on molecular structure.

Graph representation of molecules. A molecule $\mathbf{y}$ can be represented as a labeled graph $\mathcal{G}_\mathbf{y}$ whose nodes are the atoms in the molecule and edges are the bonds between the atoms. Each node v has a feature vector $\mathbf{f}_v$ including its atom type, valence, and other atomic properties. Each edge (u, v) is associated with a feature vector $\mathbf{f}_{uv}$ indicating its bond type.

Graph neural networks. A graph neural network (GNN) g learns to embed a graph $\mathcal{G}_\mathbf{y}$ into a continuous vector $g(\mathbf{y})$. We adopt the GNN architecture from [7, 60], which associates hidden states $\mathbf{h}_v$ with each node v and updates these states by passing messages $\mathbf{m}_{\vec{uv}}$ over edges (u, v). Each message $\mathbf{m}^{(0)}_{\vec{uv}}$ is initialized at zero. At time step t, the messages are updated as follows:

$$\mathbf{m}^{(t+1)}_{\vec{uv}} = \mathrm{MLP}\big(\mathbf{f}_u, \mathbf{f}_{uv}, \sum\nolimits_{w \in N(u), w \neq v} \mathbf{m}^{(t)}_{\vec{wu}}\big) \quad (5)$$

for all $(u, v) \in \mathcal{G}_\mathbf{y}$, where $N(u)$ is the set of neighbor nodes of u and MLP stands for a multilayer perceptron. After T message passing steps, we compute the hidden states $\mathbf{h}_v$ as,

$$\mathbf{h}_u = \mathrm{MLP}\big(\mathbf{f}_u, \sum\nolimits_{v \in N(u)} \mathbf{m}^{(t)}_{\vec{uv}}\big), \quad (6)$$

and we compute the final representation $g(\mathbf{y})$ as

$$g(\mathbf{y}) = \mathrm{MLP}\big(\sum\nolimits_{u \in \mathcal{G}_\mathbf{y}} \mathbf{h}_u\big). \quad (7)$$

Conditional Flow Model Architecture. To condition cell image generation by the flow model f on a molecular intervention $\mathbf{y}$, we provide the output of a graph neural network $g(\mathbf{y})$ as an additional input to μ_i, Σ_i, which govern the distribution of the latent variables $\mathbf{z}_i$ within each of our flow blocks b_i. Figure 2a illustrates the full architecture of our conditional flow model, with the GNN shown in green.

Training with Auxiliary Contrastive Loss. Existing formulations of conditional flow models propose to maximize the conditional log-likelihood of the data given the conditioning information $g(\mathbf{y})$. For our flow model, this means maximizing $\mathcal{L}$ with the modification that μ_i, Σ_i, and the flow blocks b_i now also take $g(\mathbf{y})$ as input; we denote this modified loss function as $\mathcal{L}_{\text{cond}}$. While maximizing the conditional log-likelihood has proven effective for generating images conditioned on binary vectors or conditioned on other images, we found that it does not sufficiently leverage the shared information between the input image and the molecule. Intuitively, the flow model is sufficiently powerful to achieve a high log-likelihood by converting the image distribution to a Gaussian distribution without using the conditioning information, especially when the effect of the molecular treatment on the cells is subtle in the image space.

To ensure that the conditional flow model extracts useful information from the molecular graph for generation, we propose a training strategy based on *contrastive learning*. Specifically, we add an auxiliary loss to our objective to maximize the mutual information between the image latent variables $[\mathbf{z}_0, \cdots, \mathbf{z}_k] = f(\mathbf{x})$ extracted by the flow model and the molecular embeddings $g(\mathbf{y})$ learned by the GNN:

$$\mathcal{L}^i_{\text{contrastive}} = -\mathbb{E}_{(\mathbf{x}^1, \mathbf{y}^1) \sim p_{xy}, \mathbf{y}^2 \cdots \mathbf{y}^N \sim p_y} \left[\log \frac{h_i(\mathbf{x}^1, \mathbf{y}^1)}{\sum_{j=1}^N h_i(\mathbf{x}^1, \mathbf{y}^j)} \right],$$

where $h_i(\mathbf{x}, \mathbf{y})$ is the cosine similarity of $\mathbf{z}_i$ and $g(\mathbf{y})$, $p_{\mathbf{xy}}$ is the joint distribution of the data, and $p_\mathbf{y}$ is the marginal distribution of $\mathbf{x}$ and $\mathbf{y}$. Concretely, the contrastive loss encourages $\mathbf{z}_i$ and $g(\mathbf{y})$ to be more aligned when $\mathbf{x}, \mathbf{y}$ are drawn from the same sample compared to when they are mismatched. Minimizing the contrastive loss in Equation (3.3) is equivalent to maximizing a lower bound on the mutual information between $\mathbf{z}_i$ and $g(\mathbf{y})$ and has been used in previous work for representation learning [43]. Our key insight is in leveraging contrastive learning in a conditional flow model. As shown in Figure 2b, during training, we use an additional contrastive learning loss to maximize the mutual information between the latent variables from the flow model f and the molecular embedding from the graph neural network g. During generation, information flow is reversed through the flow model to generate an image that is

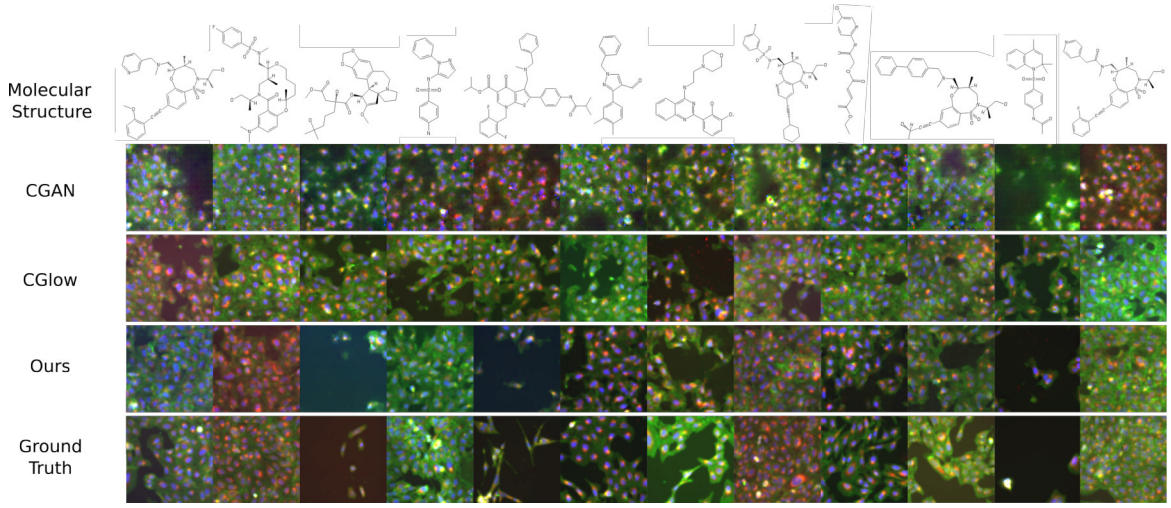


Figure 4: Examples of cell images generated by our method vs. the baselines.

tightly coupled to the conditioning molecular information. For a hyperparameter $\lambda > 0$ and some $0 \leq m \leq k$, the full objective of our conditional flow model is,

$$-\mathbb{E}_{\mathbf{x},\mathbf{y}\sim p_{\mathbf{xy}}} \mathcal{L}_{\text{cond}}(\mathbf{x},\mathbf{y}) + \lambda \sum_{i=m}^{k} \mathcal{L}^{i}_{\text{contrastive}}. \qquad (8)$$

4. Experiments

Dataset. We perform our experiments on the Cell Painting dataset introduced by Bray *et al*. [2, 3] and preprocessed by Hofmarcher *et al*. [21]. The dataset consists of 284K cell images collected from 10.5K molecular interventions. Each image contains five color channels that capture the structure of five cellular compartments: nucleus (DNA), mitochondria, endoplasmic reticulum (ER), nucleolus/cytoplasmic RNA, and actin (cytoskeleton). We divide the dataset into a training set of 219K images corresponding to 8.5K molecular interventions, and hold out the remaining of the data for evaluation. The held-out data consists of images corresponding to each of the 8.5K molecules in the training set as well as 2K molecules that are not in the training set.

Implementation Details. Our model for the molecule-to-image generation task consists of six flow modules that use the same flow units as Glow [26], which construct different levels of the Haar wavelet image pyramid, generating images from resolution of 16×16 to 512×512. The lowest resolution module consists of 64 flow units, and each of the other modules consists of 32 flow units. Each of the modules is trained to maximize the log-likelihood of the data (Equation 2). Additionally, the three flow modules that process low-resolution images (up to 64×64 resolution) are also trained to maximize the mutual information between the latent variables and the molecular features using contrastive learning with a weight of 0.1 and $\tau = 0.07$. We train each flow module for approximately 50K iterations using Adam [25] with initial learning rate of 10^{-4}, during which the highest resolution block sees over 1M images and the lowest resolution block sees over 10M images.

CellProfiler Biological Evaluation Metrics. For a molecule-to-image synthesis model to be useful to practitioners, it needs to generate image features that are meaningful from a biological standpoint. It has been shown that machine learning methods can discriminate between microscopy images using features that are irrelevant to the target content [52, 36]. Therefore, in addition to more conventional vision metrics, we propose a new set of evaluation metrics based on CellProfiler cell morphology features [40] that are more robust, interpretable, and relevant to practitioners [50]. We specifically consider the following morphological features: (1) **Coverage**, the total area of the regions covered by segmented cells; (2) **Cell/Nuclei Count**, the total number of nuclei/cells found in the image; (3) **Cell Size**, the average size of the segmented cells found in the image; (4) **Zernike Shape**, a set of 30 features that describe the shape of cells using a basis of Zernike polynomials (order 0 to order 9); (5) **Expression Level**, a set of five features that measure the level of signal from the different cellular compartments in the image. We extract these features from a subset of images and compute the Spearman correlation between the features of real and generated images corresponding to the same molecule, focusing on a subset of molecules that cause notable changes in cell morphology in real images, since these are the most useful molecules to capture in a virtual screen (see Supplementary Material for details). Due to space constraints, we show the mean correlation for the 30 Zernike shape features and the five expression level features.

	CellProfiler Metrics					Correspondence Accuracy						SWD
	Coverage	Count	Size	Zernike	Exp. Level	Mito	ER	RNA	Cyto	DNA	Overall	
Ground Truth (Upper Bound)	-	-	-	-	-	61.4	62.3	61.4	60.0	63.5	65.0	-
CGAN	7.0	4.8	-2.9	-3.9	7.4	53.4	52.3	55.9	51.3	55.7	56.6	5.65
CGlow	-1.3	3.8	5.8	2.2	6.6	50.8	51.1	52.0	52.6	53.9	55.5	5.01
CGlow+Contrast	28.5	36.1	17.5	8.7	26.7	55.9	53.9	55.4	55.8	58.6	60.0	4.96
Pyramid Flow	7.7	13.4	12.0	6.8	5.3	51.7	52.4	53.7	53.1	52.9	58.8	**3.68**
Pyramid Flow+Contrast (Mol2Image)	**44.6**	**54.4**	**27.5**	**15.8**	**37.3**	**56.7**	**56.1**	**56.3**	**56.8**	**59.1**	**63.2**	4.63

Table 1: Evaluation of Mol2Image (our model) vs. the baselines on images generated from molecules from the training set. "CellProfiler Metrics" are Spearman correlation coefficients ($\times 10^2$) between biological features from real and generated images; higher is better. "Correspondence accuracy" represents the accuracy of a pretrained correspondence classifier model evaluated on generated images; higher is better and ground truth (upper bound) achieves between 60.0 and 65.0. "SWD" is the sliced Wasserstein distance metric ($\times 10^{-2}$) from [23]; lower is better. See text for details.

Correspondence Classification Accuracy Metrics. In addition to CellProfiler metrics, to assess the correspondence between the generated images and the molecular information, we also compute the accuracy of pretrained correspondence classifiers on the generated images. These classifiers consist of a visual network and GNN that are trained to perform the following binary classification task: detect whether the input cell image matches the input molecular intervention (positive example) or whether they are taken from two different samples (negative example). We evaluate the correspondence classification accuracy for individual image channels to assess the generation quality of different cellular compartments (*e.g.*, DNA, mitochondria, ER, nucleolus/RNA, and actin), as well as 5-channel images that take all cellular compartments into account. We take the correspondence classification accuracy on real images (*i.e.*, ground truth data) to be the upper bound performance of the generated images.

Other Evaluation Metrics. In addition to these specialized metrics for biological images, we use the sliced Wasserstein Distance (SWD) [23] to measure unconditional image generation quality. We also compute the log-likelihood of the images to compare our proposed flow model based on the Haar image pyramid (the unconditional version) with the state-of-the-art Glow model [26].

Baselines. Since we present improvements to flow-based generative models, we primarily compare our approach to the state-of-the-art flow model, Glow, and its conditional variant CGlow [26], and perform ablations of our model without the proposed improvements. Furthermore, inspired by text-to-image synthesis models that combine generative adversarial networks with text feature extraction models [47], we develop and compare against a GAN-based approach that combines a generative adversarial network with a graph neural network for molecule-to-image synthesis (CGAN).

	CellProfiler Metrics					Corr
	Coverage	Count	Size	Zernike	Exp. Level	Overall
CGAN	6.4	1.9	-1.5	-1.0	9.2	56.1
CGlow	3.1	-3.7	-3.0	-3.1	3.7	54.5
CGlow+Contrast	9.2	1.7	**12.9**	**6.1**	8.6	59.1
Pyramid Flow	5.0	9.1	6.1	2.9	9.2	55.7
Pyramid Flow +Contrast (Mol2Image)	**15.8**	**19.7**	11.0	4.9	**13.4**	**62.6**

Table 2: Same as Table 1, but evaluated on images generated from held-out molecules. Ground truth (upper bound) achieves 64.2 on the correspondence accuracy (Corr) metric. See Supplementary Materials for full table.

5. Results

Tables 1 and 2 show the results of our model in comparison to the baselines. Note that since the baseline flow models are not capable of generating images at full 512 x 512 resolution, we compare all of the model results at 64 x 64 spatial resolution. Mol2Image, which uses the proposed multi-scale flow model based on the Haar image pyramid and is trained using contrastive learning, outperforms the baselines in generating cell images that reflect the effects of the molecular interventions for both molecules seen during training (Table 1) and held-out molecules (Table 2). Proposed improvements from Section 3 are analyzed below.

Haar Pyramid Flow vs. Glow [26]. One of our methodological contributions is a multi-scale flow model based on the framework of a Haar image pyramid (Section 3.2). The primary motivation for developing this model is to scale flow-based generation to high-resolution cell images. Existing state-of-the-art flow models such as Glow [26] cannot be trained on images larger than 256 x 256 due to memory limits (*i.e.*, in [26], the batch size per GPU was a single image of this size). In contrast, our multi-scale flow model successively generates images at multiple scales based on an image pyramid, going from coarse-to-fine resolution, and can be trained to generate full-resolution 512 x 512 cell

images as shown in Figure 4. We provided theoretical justification in Proposition 1, which states that our approach preserves the original log-likelihood objective of Glow, even though the training is decoupled at different image scales. Supplementary Table 2 shows empirical evidence that the log-likelihoods computed by Glow and our pyramid flow model are equivalent.

Although the primary motivation for our multi-scale flow model was scalability, we find that the image pyramid framework also improves the conditional generation of 64 x 64 images compared to the baseline model that directly generates images of this size (*i.e.*, in Table 1, compare "CGlow" to "Pyramid Flow", and compare "CGlow+Contrast" to "Pyramid Flow+Contrast (Mol2Image)"). We hypothesize that it is more efficient and easier to learn the relation between images and conditions when starting with the low-resolution images at the bottom of the image pyramid, which can be trained with larger batch size. Consistent with our observations, previous works have reported that training GANs starting from lower-resolution images [23, 8] is more effective than training directly on full-resolution images.

Training using Contrastive Learning. Our proposed training strategy for conditional flow models uses contrastive learning to maximize the mutual information between the image latent variables and the molecular embedding (Section 3.3). The results in Table 1 show that this is essential for effective generation of images conditioned on the molecular intervention. In particular, there is much lower correspondence between the images and the molecular intervention when contrastive learning is omitted. This result holds both in the case that we use the image pyramid framework (*i.e.*, compare "Pyramid Flow" to "Pyramid Flow+Contrast") and in the case that we directly generate 64 x 64 images using the baseline Glow model (*i.e.*, compare "CGlow" to "CGlow+Contrast"). On the other hand, contrastive learning does not appear to improve the unconditional quality of generated images (based on SWD).

Figure 4 shows a qualitative comparison between the baselines (CGAN, CGlow) and our method on generating images conditioned on molecular structure. The generated images from our method (Figure 4, row 3) more closely reflect the real effect of the intervention (Figure 4, row 4) compared to other methods, both in terms of cell morphology and in terms of channel intensities (representing expression of different cellular components). More qualitative examples (including full-resolution 512×512 images) are provided in the Supplementary Material.

We hypothesize that contrastive learning provides a strong signal for the conditional flow model to learn the relation between the treated cell image and the molecular structure, which leads to better conditional generation results. To this end, we also assess whether the molecular embeddings learned by the GNN of the conditional flow model are more reflective of the morphology they induce in treated cells. Specifically, we train linear classifiers on the molecular embeddings to predict a subset of 14 features curated from the morphological analysis of Bray *et al*. [2] (see the Supplementary Material). For comparison, we consider embeddings from a randomly initialized GNN, Morgan/circular fingerprints [49], and an ablation model trained without contrastive learning. The results suggest that contrastive learning learns molecular embeddings that highly reflect the morphological properties observed in the treated cells (Table 3, Column 1) and that these embeddings generalize to unseen molecules (Table 3, Column 2), which explains why this strategy is effective for improving conditional generation results.

	Embeddings (Validation)	Embeddings (Held out)
Random	0.569	0.578
Morgan Fingerprint	0.645	0.665
GNN	0.675	0.675
GNN+Contrast	**0.810**	**0.683**

Table 3: Evaluation of molecular embeddings on predicting morphological labels. Higher AUC is better. "Random" refers to embeddings from a randomly initialized GNN. "Held-out" refers to held-out molecules from the training set. For reference, a fully-supervised model (in which the parameters of the graph neural network are trained) achieves an AUC of 0.702 on held-out molecules.

6. Discussion

We have developed a new multi-scale flow-based architecture and training strategy for molecule-to-image synthesis and demonstrated the benefits of our approach on new evaluation metrics tailored to biological cell image generation. Our work represents a first step towards image-based virtual screening of chemicals and lays the groundwork for studying the shared information in molecular structures and perturbed cell morphology. A promising avenue for future work is integrating side information (e.g., known chemical properties or drug dosage) to impose constraints on the molecular embedding space and improve generalization to previously unseen molecules.

Acknowledgements. Karren Yang was supported by an NSF Graduate Research Fellowship and ONR (N00014-18-1- 2765). Alex X. Lu was funded by a pre-doctoral award from the National Science and Engineering Research Council. Regina Barzilay and Tommi Jaakkola were partially supported by the MLPDS Consortium and the DARPA AMD program. Caroline Uhler was partially supported by NSF (DMS-1651995), ONR (N00014-17-1-2147 and N00014-18-1-2765), and a Simons Investigator Award.

References

[1] Lynton Ardizzone, Carsten Lüth, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Guided image generation with conditional invertible neural networks. *arXiv preprint arXiv:1907.02392*, 2019. 2, 3

[2] Mark-Anthony Bray, Sigrun M Gustafsdottir, Mohammad H Rohban, Shantanu Singh, Vebjorn Ljosa, Katherine L Sokolnicki, Joshua A Bittker, Nicole E Bodycombe, Vlado Dančík, Thomas P Hasaka, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the cell painting assay. *Gigascience*, 6(12):giw014, 2017. 2, 6, 8

[3] Mark-Anthony Bray, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature protocols*, 11(9):1757, 2016. 6

[4] Marco Breinig, Felix A Klein, Wolfgang Huber, and Michael Boutros. A chemical–genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Molecular systems biology*, 11(12), 2015. 1

[5] Juan C Caicedo, Claire McQuin, Allen Goodman, Shantanu Singh, and Anne E Carpenter. Weakly supervised learning of single-cell feature embeddings. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9309–9318, 2018. 2

[6] Juan C Caicedo, Shantanu Singh, and Anne E Carpenter. Applications in image-based profiling of perturbations. *Current opinion in biotechnology*, 39:134–142, 2016. 1

[7] Hanjun Dai, Bo Dai, and Le Song. Discriminative embeddings of latent variable models for structured data. In *International Conference on Machine Learning*, pages 2702–2711, 2016. 2, 5

[8] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015. 3, 8

[9] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Nonlinear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 1, 2

[10] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 1, 2, 3

[11] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in neural information processing systems*, pages 2224–2232, 2015. 2

[12] Ulrike S Eggert. The why and how of phenotypic small-molecule screens. *Nature chemical biology*, 9(4):206, 2013. 1

[13] Yan Feng, Timothy J Mitchison, Andreas Bender, Daniel W Young, and John A Tallarico. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nature Reviews Drug Discovery*, 8(7):567–578, 2009. 1

[14] V Fetz, H Prochnow, Mark Brönstrup, and F Sasse. Target identification by image analysis. *Natural product reports*, 33(5):655–667, 2016. 1

[15] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017. 2

[16] Peter Goldsborough, Nick Pawlowski, Juan C Caicedo, Shantanu Singh, and Anne Carpenter. Cytogan: generative modeling of cell images. *bioRxiv*, page 227645, 2017. 2

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2

[18] Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 2, pages 729–734. IEEE, 2005. 2

[19] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216*, 2017. 2

[20] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019. 2

[21] Markus Hofmarcher, Elisabeth Rumetshofer, Djork-Arne Clevert, Sepp Hochreiter, and Gunter Klambauer. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *Journal of chemical information and modeling*, 59(3):1163–1171, 2019. 6

[22] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M Kurc, Rajarsi R Gupta, and Joel H Saltz. Unsupervised histopathology image synthesis. *arXiv preprint arXiv:1712.05021*, 2017. 2

[23] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *ICLR*, 2018. 7, 8

[24] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016. 2

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014. 6

[26] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018. 1, 2, 3, 6, 7

[27] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[28] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*, 2017. 2

[29] Ruho Kondo, Keisuke Kawano, Satoshi Koide, and Takuro Kutsuna. Flow-based image-to-image translation with fea-

ture disentanglement. In *Advances in Neural Information Processing Systems*, pages 4170–4180, 2019. 2, 3

[30] Tao Lei, Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Deriving neural architectures from sequence and graph kernels. *International Conference on Machine Learning*, 2017. 2

[31] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015. 2

[32] Rui Liu, Yu Liu, Xinyu Gong, Xiaogang Wang, and Hongsheng Li. Conditional adversarial generative flow for controllable image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7992–8001, 2019. 2

[33] 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. 1

[34] Lit-Hsin Loo, Hai-Jui Lin, Robert J Steininger III, Yanqin Wang, Lani F Wu, and Steven J Altschuler. An approach for extensibly profiling the molecular states of cellular subpopulations. *Nature methods*, 6(10):759, 2009. 1

[35] Christos Louizos and Max Welling. Multiplicative normalizing flows for variational bayesian neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2218–2227. JMLR. org, 2017. 2

[36] Alex Lu, Amy Lu, Wiebke Schormann, Marzyeh Ghassemi, David Andrews, and Alan Moses. The cells out of sample (coos) dataset and benchmarks for measuring out-of-sample generalization of image classifiers. In *Advances in Neural Information Processing Systems*, pages 1852–1860, 2019. 6

[37] You Lu and Bert Huang. Structured output learning with conditional generative flows. *arXiv preprint arXiv:1905.13288*, 2019. 2, 3

[38] Faisal Mahmood, Richard Chen, and Nicholas J Durr. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE transactions on medical imaging*, 37(12):2572–2581, 2018. 2

[39] Mojca Mattiazzi Usaj, Nil Sahin, Helena Friesen, Carles Pons, Matej Usaj, Myra Paz D Masinas, Ermira Shuteriqi, Aleksei Shkurin, Patrick Aloy, Quaid Morris, et al. Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. *Molecular systems biology*, 16(2):e9243, 2020. 2

[40] Claire McQuin, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A Cimini, Kyle W Karhohs, Minh Doan, Liya Ding, Susanne M Rafelski, Derek Thirstrup, et al. Cellprofiler 3.0: Next-generation image processing for biology. *PLoS biology*, 16(7), 2018. 6

[41] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *International Conference on Machine Learning*, pages 2014–2023, 2016. 2

[42] Laura H Okagaki, Anna K Strain, Judith N Nielsen, Caroline Charlier, Nicholas J Baltes, Fabrice Chrétien, Joseph Heitman, Françoise Dromer, and Kirsten Nielsen. Cryptococcal cell morphology affects host cell interactions and pathogenicity. *PLoS pathogens*, 6(6), 2010. 2

[43] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 5

[44] Anton Osokin, Anatole Chessel, Rafael E Carazo Salas, and Federico Vaggi. Gans for biological image synthesis. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 2233–2242, 2017. 2

[45] Zachary E Perlman, Michael D Slack, Yan Feng, Timothy J Mitchison, Lani F Wu, and Steven J Altschuler. Multidimensional drug profiling by automated microscopy. *Science*, 306(5699):1194–1198, 2004. 1

[46] A Srinivas Reddy, S Priyadarshini Pati, P Praveen Kumar, HN Pradeep, and G Narahari Sastry. Virtual screening in drug discovery-a computational perspective. *Current Protein and Peptide Science*, 8(4):329–351, 2007. 1

[47] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*, 2016. 2, 7

[48] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015. 2

[49] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010. 2, 8

[50] Mohammad Hossein Rohban, Shantanu Singh, Xiaoyun Wu, Julia B Berthet, Mark-Anthony Bray, Yashaswi Shrestha, Xaralabos Varelas, Jesse S Boehm, and Anne E Carpenter. Systematic morphological profiling of human gene and allele function via cell painting. *Elife*, 6:e24060, 2017. 6

[51] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. 2

[52] L Shamir. Assessing the efficacy of low-level image content descriptors for computer-based fluorescence microscopy image analysis. *Journal of microscopy*, 243(3):284–292, 2011. 6

[53] Brian K Shoichet. Virtual screening of chemical libraries. *Nature*, 432(7019):862–865, 2004. 1

[54] Haoliang Sun, Ronak Mehta, Hao H Zhou, Zhichun Huang, Sterling C Johnson, Vivek Prabhakaran, and Vikas Singh. Dual-glow: Conditional flow-based generative model for modality transfer. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10611–10620, 2019. 2, 3

[55] David C Swinney and Jason Anthony. How were new medicines discovered? *Nature reviews Drug discovery*, 10(7):507–519, 2011. 1

[56] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017. 2

[57] W Patrick Walters, Matthew T Stahl, and Mark A Murcko. Virtual screening—an overview. *Drug discovery today*, 3(4):160–178, 1998. 1

[58] Christina Winkler, Daniel Worrall, Emiel Hoogeboom, and Max Welling. Learning likelihoods with conditional normalizing flows. *arXiv preprint arXiv:1912.00042*, 2019. 2, 3

[59] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018. 2

[60] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, et al. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling*, 59(8):3370–3388, 2019. 2, 5

[61] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, page 101552, 2019. 2