Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Generative models and abstractions for large-scale neuroanatomy datasets

David Rolnick[1] and Eva L Dyer[2,3]

Neural datasets are increasing rapidly in both resolution and volume. In neuroanatomy, this trend has been accelerated by innovations in imaging technology. As full datasets are impractical and unnecessary for many applications, it is important to identify abstractions that distill useful features of neural structure, organization, and anatomy. In this review article, we discuss several such abstractions and highlight recent algorithmic advances in working with these models. In particular, we discuss the use of generative models in neuroanatomy; such models may be considered 'meta-abstractions' that capture distributions over other abstractions.

Addresses
[1] School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA
[2] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, GA, USA
[3] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Corresponding author: Dyer, Eva L (evadyer@gatech.edu)

Current Opinion in Neurobiology 2019, **55**:112–120

This review comes from a themed issue on **Machine learning, big data, and neuroscience**

Edited by **Jonathan Pillow** and **Maneesh Sahani**

For a complete overview see the Issue and the Editorial

Available online 15th March 2019

https://doi.org/10.1016/j.conb.2019.02.005

Introduction

As the scale of neuroanatomy data has grown, algorithms and abstractions have been developed to distill high-dimensional data into usable forms. Such approaches have allowed us to address questions such as: What is the density of synapses in a specific region of the brain? What is the connectivity between an area of interest and the rest of the brain? What is the best way to divide a brain area into subregions? As the number of data points grows yet further, however, it is possible to ask a different kind of question about variation across different samples or different individuals. These questions can be thought of as 'how' instead of 'what': How does neuroprotective treatment alter the density of synapses? How does learning affect the sparseness of connections in a network? How does the modularity of brain networks vary across subjects?

The goal of this article is to discuss how generative approaches in machine learning can be used to address such questions in large-scale neuroanatomy. A *generative model* captures the variability between samples in a dataset, or between entire datasets, by generating artificial examples with similar statistics to the real data. For example, a generative modeling approach can be used to sketch artificial neurons that are structurally similar to genuine ones, or to simulate a connectome for which the network properties match those observed from microscope data. Generally, the model itself incorporates randomness in order to simulate the true probability distribution over data. A perfect generative model would parameterize the underlying data distribution exactly, allowing the entire dataset to be recreated algorithmically.

We start by describing three main classes of abstractions widely used in neuroanatomy: *counts or densities* to model the spatial distribution of discrete objects like cells or synapses, *connectomes* to model the connectivity between either cells or brain areas, and *modular or hierarchical* models that describe how data are organized into groups. We then describe generative models that are matched to these various abstractions. For example, Poisson models can generate count data of objects such as cells or synapses [1], stochastic block models can be used to build graphs [2••], and hidden Markov models can be used to generate the dendritic trees of neurons [3]. In each case, we describe both the algorithmic approach and the conclusions that can be drawn from these abstractions.

After providing an overview of generative models that are built on top of these popular abstractions, we outline generative models that are not built upon any lower-level abstraction. Instead, models such as generative adversarial networks (GANs) [4••,5] and variational autoencoders (VAEs) [[4••,5]] can generate very high-dimensional data, including entire images. Such models can be used to analyze the sources of variability in observed images [7], to augment observed data, or to interpolate between different imaging modalities [8•].

Abstractions

In this section, we highlight key classes of abstractions used in neuroanatomy and describe approaches to

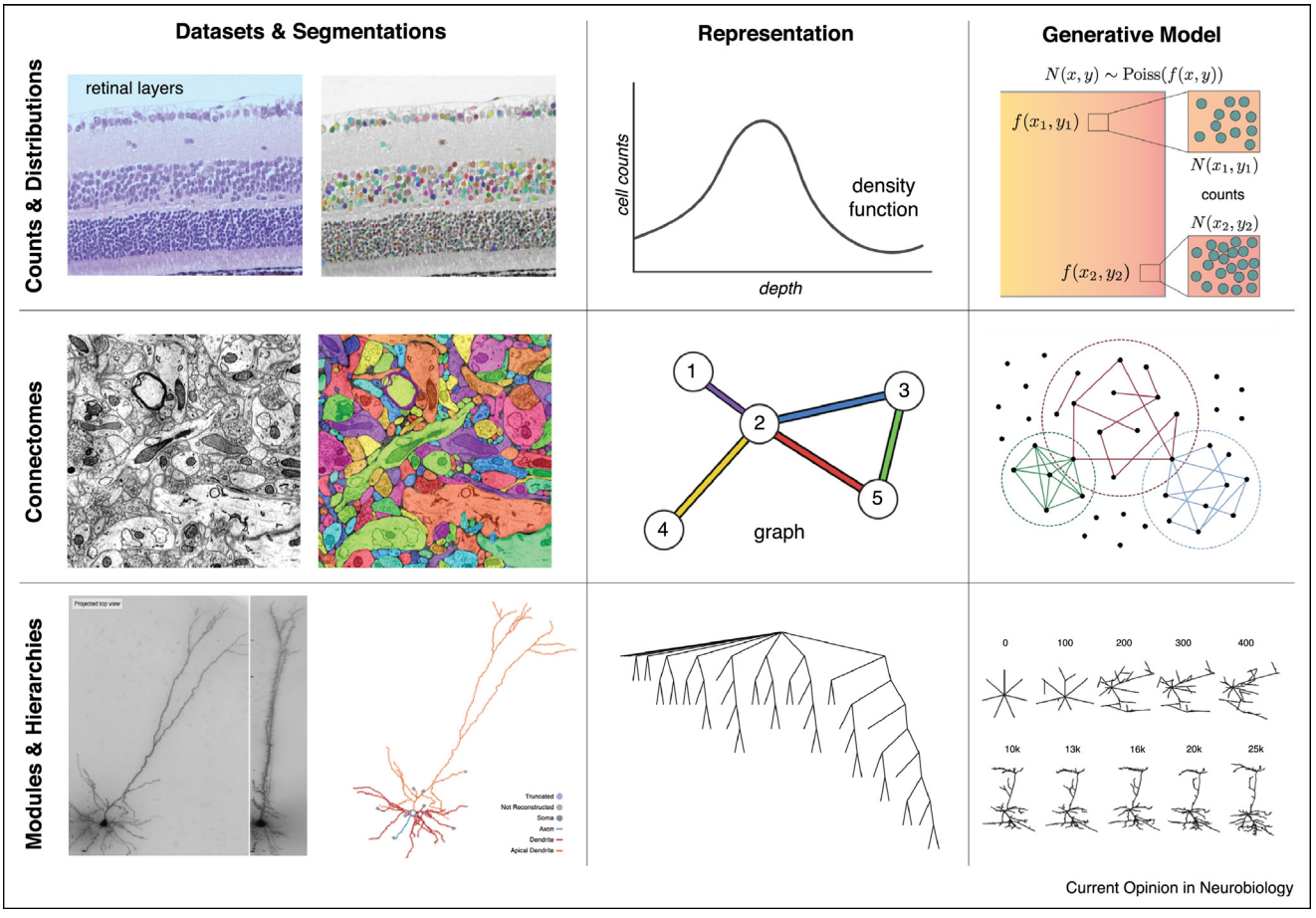
estimate these models from high-dimensional and complex brain data (see Figure 1). Each of these abstractions can be considered in terms of first, what data sources it is commonly derived from, second, what questions it can be used to answer, third, what information it retains and what it discards from the full-dimensional data, and fourth, algorithms used to derive the abstraction.

Counts and densities

In neuroanatomy, quantification of brain structure often starts by counting cells, synapses, spines, or other objects in the brain. Counts, or the number of discrete objects in a interval/bin of fixed size, provide the data necessary to compute density estimates from many samples. A large body of work in neuroanatomy involves modeling changes in densities across multiple samples or conditions.

- *Example data sources:* Cellular densities can be resolved in Nissl-stained or DAPI-stained brain images [12] and retinal datasets (Figure 1a) [13], as well as X-ray microCT [14]. Synapses can be resolved in electron microscopy (EM) [15,16] and array tomography [17] datasets. Individual mRNAs can now be resolved in brain tissue with spatial transcriptomics (mFISH) [18], multiplexed error-correcting FISH (merFISH) [19], and expansion microscopy-based FISH [20].
- *Type of conclusion drawn:* The spatial distribution and patterns of discrete objects like cells or synapses. Micro and macroscale architecture can also be detected by analyzing spatial patterns in the data. Counts can also be used to track changes to the nervous system in development [21], disease [22], or aging [23].
- *Information included:* The spatial position of objects is included but the connectivity between these objects is

Figure 1



Abstractions and generative models for neuroanatomy. (1) Counts & Distributions: From left to right, we show a (left) retinal dataset before and after cell detection [9], (middle) a depiction of how count data can be represented as a density function, and (right) a Poisson model for generating new count data. (2) Connectomes: From left to right, we show (left) an electron microscope image of a thin slice of cortical brain tissue before and after dense segmentation to build a connectome, (middle) a depiction of a connectome as a graph, and (right) an example of the random overlapping communities model for sparse graphs with three communities displayed as different colors [10]. (3) Modular and Hierarchical Representations: From left to right, we show (left) a light microscope image with a biocytin filled neuron in two views before and after tracing (from the Allen Institute for Brain Science's Cell Types Atlas [11]), (middle) a hierarchical representation of a dendrite, and (right) example neuronal morphologies generated after an iterative sampling procedure [3], where the iteration number is displayed over each generated morphology.

not modeled. In some cases, each count can also be associated with additional metadata or 'marks' like the object's size.
- *Algorithms used to create abstraction:* Segmentation and density-based methods have been developed to quantify the spatial organization and distributions of cells [14,24] (Figure 1c), synapses [25–28], neuronal arbors [29], organelles [30], and spines [31].

Connectomes

Graphs are some of the most widely used abstractions for neuroanatomical data. They are typically used to convey observed physical connectivity between individual neurons or neuronal assemblages. Such graphs are commonly referred to as 'structural' connectomes (in contradistinction to 'functional' connectomes, which capture correlations between observed *activity* of neurons). At the micro-scale, cellular connectomes have nodes for neurons and (weighted) edges for synapses. In meso-scale or macro-scale connectomes, nodes represent local or global brain areas, while edges represent projections between the areas. Such graphs are also referred to as 'projectomes'.

- *Example data sources:* At the microscale, connectomes can be extracted from EM [16] (Figure 1) and expansion microscopy (ExM) [32] datasets. Projectome mapping methods have made use of viral tracing methods and whole-brain serial two-photon microscopy (STP) and MOST [33,34••,35,36] to reveal long-range connections. Projectome data has also been obtained from humans using magnetic resonance imaging (MRI) [37–40], mainly through the use of diffusion tensor imaging.
- *Type of conclusion drawn:* Connectomes and projectomes can be used to understand learning and plasticity, as well as constrain models of neural information processing.
- *Information included:* The connectivity between neurons or brain areas is included in these models. In some cases, the strength of connections can also be estimated and included to produce a weighted graph. The spatial position of each node is often excluded in a graphical representation of the data.
- *Algorithms used to create abstraction:* There has been extensive recent work on automatic labeling of EM and ExM images to segment neurons [41–44,45•,46•,47–49] and synapses [26,50]. On the computational side, Majka *et al.* [51] demonstrate tools for coregistering projectomes to create a common map of primate (marmoset) cortex, while other algorithms have been developed to infer higher resolution completions of partial connectivity data [52,53].

Modular and hierarchical models

Finally, we consider modular and hierarchical abstractions which divide data into groups based upon which examples/segments have similar characteristics. One example is representing a large brain volume as a collection of brain regions, modules, or spatially defined regions of interest [54,55••]. This principle can be iterated by expressing data examples in terms of a hierarchical model, where discrete groups are divided into subgroups at many scales. For example, the morphology of a neuron can be described with a hierarchical format, with a coarse division into soma, axon, and dendrite which is further broken down into individual branches.

- *Example data sources:* Serial two-photon and fMOST for whole-brain imaging have been used to obtain parcellations of the brain [34••]. Morphological reconstructions for modeling the components of neurons can be extracted from light microscopy datasets [56].
- *Type of conclusion drawn:* The high-level organization of the data and which parts of the signal are similar and thus belong to the same group. A hierarchical format for data can be advantageous in representing similarities in data across multiple spatial or evolutionary divisions/scales.
- *Information included:* Modular representations group the structure of many nearby segments of a neuron (parts) or nearby parts of a brain region into one bulk class. The membership of data to a class is preserved and perhaps the average (centroid) of the class is also maintained. Hierarchical models further provide information about the distance between different groups as relative to their multi-scale dependencies.
- *Algorithms used to create abstraction:* To obtain an informative parcellation and simplification of the data, clustering algorithms [57•] such as k-means and spectral clustering methods [58] can be used to group spatial loci that have similar statistics in terms of their measured anatomical signal. Semi-automated approaches have recently been shown to provide new insights into structurally and functionally distinct areas in whole human brains with multi-modal measurements [59••].

Generative models for abstractions

In this section, we describe different generative models that are built on top of the previously discussed abstractions. Each generative model represents a design choice about what features of the true data are most important to capture, based upon the questions under consideration.

Generative models for count-valued data

A generative model for count-valued data (i.e. how many objects are in a region of interest) creates a synthetic dataset where objects are placed across space according to the underlying statistics of real data. Which statistics are important represent a design choice. For example, a model might be designed so that the density functions of real and synthetic data match or so as to preserve nearest-neighbor properties of the counts (e.g. the Ripley k-function [60]).

The simplest generative model used for count-valued data is a Poisson process. Here, we assume that the number of objects observed in a bin/interval is a Poisson-distributed random variable with mean given by an intensity (density) function, and where the numbers of objects in different bins are conditionally independent. Thus, given the potentially spatially varying intensity of the process, samples can be generated to create a simulated dataset. To extend the independence assumption of Poisson models to ensure that objects are separated by a minimum distance, random sequential adsorption (RSA) processes have been used to model synapses throughout all cortical layers [27]. Point process models can also be constrained to generate counts along a graph structure, for instance in the modeling of spines along a neurite [61]. See [1] for a review of spatial point process models and their applications in neuroanatomy.

To model more complex spatial properties of the data, the underlying intensity function can be approximated by a sparse combination of simpler functions. LaGrow *et al.* [62,9] show that by using a basis that can capture change points in the density, this enables the efficient estimation of mesoscopic properties of the density, like the layering structure in the cortex.

Generative models for connectomes

To create a realistic generative model for graphs, we need to first specify the property of the graph we wish to capture. One such property is the average degree of all nodes. Random graph theory provides a wealth of resources for generating graphs that have certain edge and high-level properties [63•]. A more complex generative model could ask that graph metrics like clustering and modularity match between real and synthesized data. Such generative models can be introduced by building on random graph models like the widely used Erdős-Renyi random graph model [64], in which each pair of nodes is assigned an edge with some fixed probability p. The random overlapping communities (ROC) model is a good example of a generative model that can generate overlapping communities as observed in neural circuits, and has provable convergence in terms of its desired properties [10]. In this model, many subsets of the overall graph are chosen at random, and dense Erdős-Renyi random graphs are constructed on these (possibly overlapping) subsets. Additionally, stochastic block models (SBMs) are a class of generative models for synthesizing graphs [65], which have been used to model hierarchical modules within a connectome. In an SBM, the nodes of the graph are divided into several *blocks*, and the probability of connection between two nodes depends only on the blocks in which they lie. Jonas and Kording [66••] introduce a variant of SBMs to model connectivity between neurons, where the blocks of the model correspond to cell types, and where distances also affect the probability of connection. They use Markov Chain Monte Carlo (MCMC) methods to fit the parameters of the model, thereby automatically inferring cell types from connectomics data.

Hidden Markov model (HMMs) have also been applied successfully to the graph structures representing the branching of individual neurons. HMMs model the growth of a graph or other data structure over time using a Markov chain that depends on *hidden variables* that can be statistically inferred but are not observed directly. For example, the *hidden state* of a neuron as it grows might include biochemical factors that are not directly observable, even though they lead to observable data such as the morphology of the neuron. In Farhoodi *et al.* [67], the branching patterns of different types of neurons are learned and incorporated into a generative model by analyzing single-neuron morphological data compiled by neuromorpho.org [68••]. The HMM inferred by Farhoodi *et al.* suggests that the probability of branching within a neuron depends on the distance to the soma, whether the branching occurs in a main branch or a side branch, and what the type of neuron is. The model thus yields both insights into the underlying factors that may be at play in neural branching and also a procedure for generative artificial neuronal morphologies. See also Farhoodi and Kording [3] for a generative approach to neuron morphologies based on Markov chain Monte Carlo (MCMC) sampling.

Modular and hierarchical generative models

Generative models built on top of hierarchical abstractions, typically will generate a sequence of items wherein the probabilistic model depends upon what was generated at previous generate samples. To ensure that our generative model matches the distribution of data, the sequence of steps must generate an output that matches the same sequence generation of real data.

SBMs (defined in the preceding section) are well-suited to dissecting graphical data into hierarchically organized modules. Lyzinski *et al.* [2••] combine SBMs with clustering algorithms to decompose a partial *Drosophila* connectome into blocks, which are then clustered into similar subnetworks (*motifs*). The process is then repeated to generate a *hierarchy* of motifs. Priebe *et al.* [69] apply another generalization of SBMs to the *Drosophila* connectome to explain variation in cells that is fit poorly by simple clusters.

Generative models for image data

Whereas the generative models highlighted in the previous section require (often intensive) pre-processing steps to first build an abstraction from image data, modern machine learning methods make it possible to learn a generative model from images directly. Learning models from images directly could potentially allow us to by-pass the initial steps of building an abstraction

(e.g. segmenting or finding objects in images). Rather than specifying what to look for in an image, the generative model would be able to automatically pull out features from the image data that are important for building realistic representations of neural structure and capturing variability across examples. In this section, we highlight matrix factorization and deep learning approaches for learning generative models from collections of images, and discuss their applications in neuroanatomy.

Latent variable models

Learning to model the distribution of high-dimensional image data is extremely challenging. A first step is often to form a low-dimensional representation that is easier to model. A simple and widely used linear approach for learning latent factors from data is principal component analysis (PCA); PCA fits a k-dimensional linear approximation to a dataset with many examples, such as a collection of many brain images. Other dimensionality reduction techniques such as non-negative matrix factorization [70], probabilistic PCA [71], and sparse PCA [72] can all be used to form a low-dimensional representation of collection of data (see [73] for a comprehensive review of dimensionality reduction techniques and their applications in analyzing measurements of neural activity).

After distilling data into a low-dimensional space, image data can be reconstructed by inverting the low-dimensional model learned in the analysis step (Figure 2, left). This synthesis operation is visualized in Figure 2 for a linear system learned in PCA. In this case, a new image is created by either: reconstructing an input (pass in a noisy signal and the output is a clean version) or generating a new sample in the low-dimensional space and then using the decoder to synthesize a new image as output. This interpretation of linear matrix factorization (PCA) as a generative model provides a simple strategy for creating high-dimensional images when the data lie near a linear subspace.

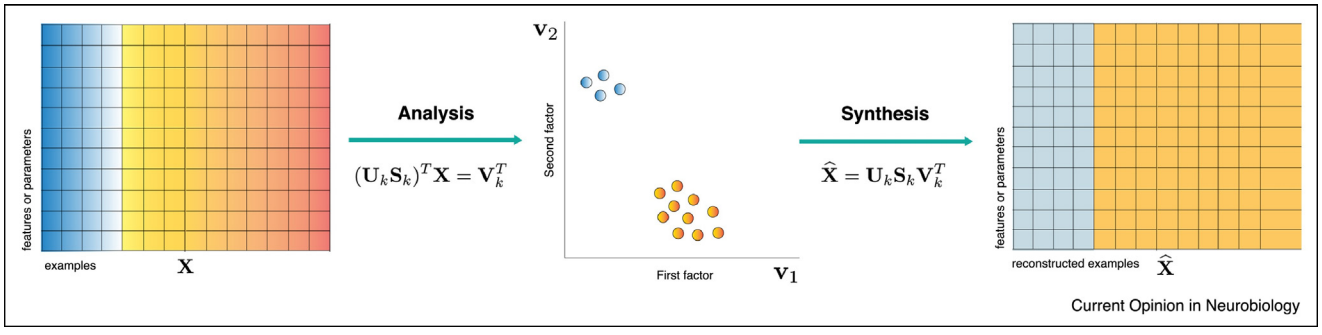
Autoencoders

Generative models that rely on PCA and other matrix factorization approaches use linear transformations of data. It is also possible to find nonlinear low-dimensional representations of data. *Autoencoders* are now routinely used for this task [74]. Autoencoders can be constructed through different neural network architectures, encompassing models such as stacked convolutional autoencoders [75] and variational autoencoders [6]. Essentially, an autoencoder functions by passing high-dimensional input through a sequence of layers, including a low-dimensional 'bottleneck' layer, then reconstructing the full-dimensional input again in the output layer (see Figure 3, left). The bottleneck layer thus learns a low-dimensional latent representation of the data. The first part (up to the bottleneck) is the *encoder* and the remainder (reconstructing the input) is the *decoder*. Thus, the encoder compresses the data to the latent representation, and the decoder is a generative model that recreates data from this latent representation. Thus autoencoders provide an analogous architecture for generative modeling for the nonlinear case as that depicted for the linear case in Figure 2.

Generative adversarial networks

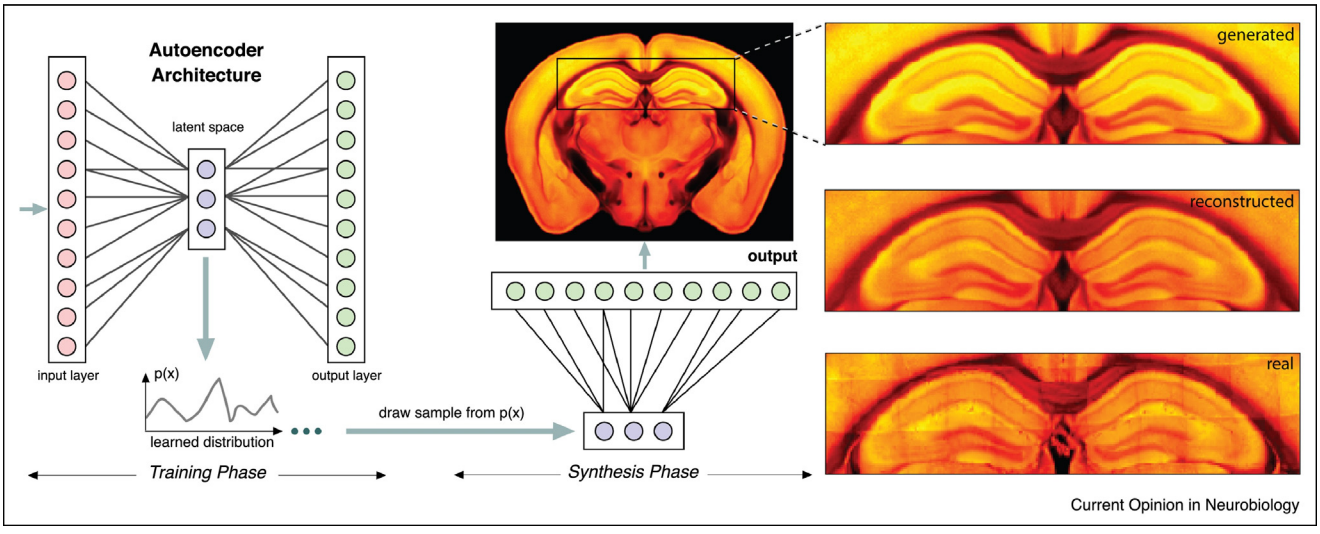
Within deep learning, generative adversarial networks (GANs) have recently been developed to learn from an unlabeled training dataset to generate artificial data resembling examples from the dataset. Like autoencoders, GANs learn a nonlinear generative process for data via an artificial neural network. However, unlike autoencoders, which learn both an 'encoding' step and a 'decoding' (generative) step, GANs learn by pitting two network algorithms against each other, with one

Figure 2



Linear matrix factorization methods like PCA and their interpretation as a generative modeling procedure. From left to right, we show the decomposition of a data matrix consisting of examples along its columns, into a low-dimensional format. In the middle, we depict the low-d representation of the data in two dimensions. On the right, we show a reconstructed or synthesized data matrix that uses the inverse mapping to expand the low-d data back into a high-dimensional space again.

Figure 3



Generative models for synthesizing structural brain images. On the left, we depict an autoencoder consisting of an input layer, a low-dimensional hidden layer (latent space), and output layer. In the training phase, a low-dimensional model for data is learned and in the synthesis phase, a sample from this model is generated and used to generate a new image. This architecture is applied to auto-fluorescence images of 1,700 different brains (25 micron resolution) to synthesize new images: on the right, a synthetically generated image (top), example of a real image used to train the network (bottom), and a denoised (reconstructed) version of the image displayed on the bottom.

(the *generator*) attempting to generate plausible examples from a dataset, while the other (the *discriminator*) tries to tell the difference between real and fake examples, thus forcing the generator to improve. While extensive applications to neuroanatomy have yet to be developed, GANs have already been used to simulate neuron morphologies [76] and spike trains [77]. A similar approach (using deep learning methods distinct from GANs) uses the output of one imaging modality to simulate the result of another imaging modality [78••].

It is tempting to consider using the output of a GAN to augment real data in fitting additional algorithms. However, there is so far no magical algorithm that replaces the power of large real datasets. For example, while a GAN might be used to learn from a thousand images and then create a million more similar-seeming images, the artificial images would likely either fail in subtle ways to be truly realistic or would fail to capture the full diversity of real-world data. We therefore believe the function of generative algorithms in neuroanatomy should be, for the moment, more in modeling than in augmenting data for training.

Conclusions

As neuroanatomy datasets become more numerous and higher-dimensional, there is increasing need for generative models that capture variability across data samples and subjects. Where traditional abstractions such as connectomes compress data, generative 'meta-abstractions' compress distributions over data or over abstractions. We believe that an understanding of the breadth of available abstractions and concomitant generative models, each suited to different questions and data modalities, is essential to present-day neuroanatomy.

Many of the generative models we have described make strong assumptions about the structure of the data — for example, that it is well-approximated by a density function or succinctly described by a Markov chain. By contrast, generative algorithms from deep learning typically have no such prior assumptions, and the models they learn are often 'black boxes' that are hard to interpret. Interpretation becomes increasingly difficult as we move to full images and raw data because it is not always clear what properties of the data are being modeled and how. New approaches for 'disentangling representations' [79•] aim to mitigate these issues and build architectures that reveal more interpretable factors. An important line of research is to build deep learning architectures that are interpretable and can be used to draw inferences about disease, inter-subject variability, and other changes to neural structure.

Traditionally, neuroscience has provided views of the structure of the nervous system that resolve or model one aspect of the anatomy at a time. Neuroscience methods are, however, increasingly moving towards resolving multiple types of structures simultaneously to provide multi-modal and multi-scale structural information for large volumes, in some cases up to whole brains [12,34••]. With increasing access to multi-modal

information, it is critical to develop abstractions and generative models that distill the data into a usable simplification that leverages the multi-modal data provided. Because traditional methods for modeling neuroanatomy have focused on modeling a single attribute of the data (a graph, or a density), in some cases it is not clear how best to integrate data formats and models across different modalities of information. It is exceedingly likely that different aspects of anatomy (change in density of synapses or cells, or strengthening of connections in a specific region of the brain) co-vary in complex and nonlinear ways and multi-modal datasets will be necessary to reveal these relationships.

Generative models are now being used to learn increasingly complex attributes of a wide range of datasets. We believe that they will be a useful tool moving forward for modeling variability in large-scale neuroanatomy datasets.

Conflict of interest statement

Nothing declared.

Acknowledgements

We would like to thank Konrad Körding and Chethan Pandarinath for helpful feedback on the manuscript. We would also like to acknowledge Samantha Petti and Santosh Vempala for providing us an example of a ROC (generative graph model), Yaron Meirovitch for segmented EM images for connectomics, and Roozbeh Farhoodi for providing us with the dendrogram and generative model figures for hierarchical models, all of which were used to create Figure 1. ED was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R24MH114799. DR was supported by the National Science Foundation under Award Number 1803547.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Budd JM, Cuntz H, Eglen SJ, Krieger: **Quantitative analysis of neuroanatomy**. *Front Neuroanat* 2015, **9**:143.

2. •• Lyzinski V, Tang M, Athreya A, Park Y, Priebe CE: **Community detection and classification in hierarchical stochastic blockmodels**. *IEEE Trans Netw Sci Eng* 2017, **4**:13-26.
This paper uses stochastic block models to identify a hierarchy of structural motifs within the Drosophila connectome.

3. Farhoodi R, Körding K: **Sampling neuron morphologies**. *BioRxiv* 2018:248385.

4. •• Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y: **Generative adversarial nets**. *Advances in Neural Information Processing Systems*. 2014:2672-2680.
This paper introduces a method to learn generative models using an adversarial process. The network architecture consists of a generator that produces new samples to match real data and a discriminator that tries to tell the differences between real and fake examples.

5. Radford A, Metz L, Chintala S: *Unsupervised representation learning with deep convolutional generative adversarial networks*. 2015. ArXiv preprint arXiv:1511.06434.

6. Kingma D, Welling M: *Auto-encoding variational Bayes*. 2013. ArXiv preprint arXiv:1312.6114.

7. Subakan C, Kojeyo O, Smaragdis: *Learning the base distribution in implicit generative models*. 2018. ArXiv preprint arXiv:1803.04357.

8. • Iglesias JE, Konukoglu E, Zikic D, Glocker B, Van Leemput K, Fischl B: **Is synthesizing MRI contrast useful for inter-modality analysis?** *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2013:631-638.
This paper shows that synthesized MRI imagery can improve coregistration of brain regions between different subjects.

9. LaGrow TJ, Moore M, Prasad JA, Davenport MA, Dyer EL: **Approximating cellular densities with sparse recovery techniques**. *BioArxiv* 2018.

10. Petti S, Vempala SS: *Approximating sparse graphs: the random overlapping communities model*. 2018. ArXiv preprint arXiv:1802.03652.

11. Allen Institute Cell Types Atlas, 2018 [Online]. Available at: http://celltypes.brainmaorg/experiment/morphology/607124114.

12. Amunts K, Lepage C, Borgeat L, Mohlberg H, Dickscheid T, Rousseau M-É, Bludau S, Bazin P-L, Lewis LB, Oros-Peusquens A-M *et al.*: **BigBrain: an ultrahigh-resolution 3D human brain model**. *Science* 2013, **340**:1472-1475.

13. Chang B, Hawes N, Pardue M, German A, Hurd R, Davisson M, Nusinowitz S, Rengarajan K, Boyd A, Sidney S, Phillips M: **Two mouse retinal degenerations caused by missense mutations in the beta-subunit of rod CGMP phosphodiesterase gene**. *Vision Res* 2007, **47**:624-633.

14. Dyer EL, Roncal WG, Prasad JA, Fernandes HL, Gürsoy D, De Andrade V, Fezzaa K, Xiao X, Vogelstein JT, Jacobsen C *et al.*: **Quantifying mesoscale neuroanatomy using x-ray microtomography**. *ENeuro, ENEURO-0195* 2017, **4**.

15. Merchán-Pérez A, Rodríguez J-R, González S, Robles V, DeFelipe J, Larrañaga C, Bielza: **Three-dimensional spatial distribution of synapses in the neocortex: a dual-beam electron microscopy study**. *Cereb Cortex* 2013, **24**:1579-1588.

16. Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, Lee D, Vázquez-Reina A, Kaynig V, Jones TR *et al.*: **Saturated reconstruction of a volume of neocortex**. *Cell* 2015, **162**:648-661.

17. Micheva KD, Smith SJ: **Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits**. *Neuron* 2007, **55**:25-36.

18. Shah S, Lubeck E, Zhou W, Cai L: **In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus**. *Neuron* 2016, **92**:342-357.

19. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X: **Spatially resolved, highly multiplexed RNA profiling in single cells**. *Science* 2015, **348**:aaa6090.

20. Chen F, Wassie AT, Cote AJ, Sinha A, Alon S, Asano S, Daugharthy ER, Chang J-B, Marblestone A, Church GM *et al.*: **Nanoscale imaging of RNA with expansion microscopy**. *Nat Methods* 2016, **13**:679.

21. Bandeira F, Lent R, Herculano-Houzel S: **Changing numbers of neuronal and nonneuronal cells underlie postnatal brain growth in the rat**. *Proc Natl Acad Sci* 2009, **106**:14108-14113.

22. Kreczmanski H, Heinsen V, Mantua F, Woltersdorf T, Masson N, Ulfig R, Schmidt-Kastner H, Korr HW, Steinbusch R, Hof *et al.*: **Volume, neuron density and total neuron number in five subcortical regions in schizophrenia**. *Brain* 2007, **130**:678-692.

23. Andrade-Moraes CH, Oliveira-Pinto AV, Castro-Fonseca E, da Silva CG, Guimaraes DM, Szczupak D, Parente-Bruno DR, Carvalho LR, Polichiso L, Gomes BV *et al.*: **Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles**. *Brain* 2013, **136**:3738-3752.

24. Zhang C, Yan C, Ren M, Li A, Quan T, Gong H, Yuan J: **A platform for stereological quantitative analysis of the brain-wide distribution of type-specific neurons**. *Sci Rep* 2017, **7**:14334.

25. Li L, Tasic B, Micheva KD, Ivanov VM, Spletter ML, Smith SJ, Luo L: **Visualizing the distribution of synapses from individual neurons in the mouse brain**. *PLoS ONE* 2010, **5**:e11503.

26. Roncal WG, Pekala M, Kaynig-Fittkau V, Kleissas DM, Vogelstein JT, Pfister H, Burns R, Vogelstein RJ, Chevillet MA, Hager GD: *Vesicle: volumetric evaluation of synaptic interfaces using computer vision at large scale*. 2014. ArXiv preprint arXiv:1403.3724.

27. Anton-Sanchez L, Bielza C, Merchán-Pérez A, Rodríguez J-R, DeFelipe J, Larrañaga: **Three-dimensional distribution of cortical synapses: a replicated point pattern-based analysis**. *Front Neuroanat* 2014, **8**:85.

28. Domínguez-Álvaro M, Montero-Crespo M, Blazquez-Llorca L, Insausti R, DeFelipe J, Alonso-Nanclares L: **Three-dimensional analysis of synapses in the transentorhinal cortex of Alzheimer's disease patients**. *Acta Neuropathol Commun* 2018, **6**:20.

29. Sümbül U, Zlateski A, Vishwanathan A, Masland RH, Seung HS: **Automated computation of arbor densities: a step toward identifying neuronal cell types**. *Front Neuroanat* 2014, **8**:139.

30. Perez AJ, Seyedhosseini M, Deerinck TJ, Bushong EA, Panda S, Tasdizen T, Ellisman MH: **A workflow for the automatic segmentation of organelles in electron microscopy image stacks**. *Front Neuroanat* 2014, **8**:126.

31. Anton-Sanchez L, Larrañaga R, Benavides-Piccione I, Fernaud-Espinosa J, DeFelipe C, Bielza: **Three-dimensional spatial modeling of spines along dendritic networks in human cortical pyramidal neurons**. *PLOS ONE* 2017, **12**:e0180400.

32. Chen F, Tillberg W, Boyden ES: **Expansion microscopy**. *Science* 2015, **347**.

33. Economo MN, Clack NG, Lavis LD, Gerfen CR, Svoboda K, Myers EW, Chandrashekar J: **A platform for brain-wide imaging and reconstruction of individual neurons**. *Elife* 2016, **5**:e10566.

34. •• Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S, Wang Q, Lau C, Kuan L, Henry AM *et al.*: **A mesoscale connectome of the mouse brain**. *Nature* 2014, **508**:207.

This paper introduces a pipeline for collecting and analyzing whole-brain datasets to infer the mesoscale connectome of the mouse brain.

35. Ragan T, Kadiri LR, Venkataraju KU, Bahlmann K, Sutin J, Taranda J, Arganda-Carreras I, Kim Y, Seung HS, Osten: **Serial two-photon tomography for automated ex vivo mouse brain imaging**. *Nat Methods* 2012, **9**:255.

36. Li A, Gong H, Zhang B, Wang Q, Yan C, Wu J, Liu Q, Zeng S, Luo Q: **Microoptical sectioning tomography to obtain a high-resolution atlas of the mouse brain**. *Science* 2010, **330**:1404-1408.

37. Gray WR, Bogovic JA, Vogelstein JT, Landman BA, Prince JL, Vogelstein RJ: **Magnetic resonance connectome automated pipeline: an overview**. *IEEE Pulse* 2012, **3**:42-48.

38. Craddock RC, Jbabdi S, Yan C-G, Vogelstein JT, Castellanos FX, Di Martino A, Kelly C, Heberlein K, Colcombe S, Milham M: **Imaging human connectomes at the macroscale**. *Nat Methods* 2013, **10**:524.

39. Hagmann L, Cammoun X, Gigandet R, Meuli CJ, Honey VJ, Wedeen O, Sporns: **Mapping the structural core of human cerebral cortex**. *PLoS Biol* 2008, **6**:e159.

40. He Y, Chen ZJ, Evans AC: **Small-world anatomical networks in the human brain revealed by cortical thickness from MRI**. *Cereb Cortex* 2007, **17**:2407-2419.

41. Lee K, Zlateski A, Ashwin V, Seung HS: **Recursive training of 2D-3D convolutional networks for neuronal boundary prediction**. *Advances in Neural Information Processing Systems*. 2015:3573-3581.

42. Meirovitch Y, Matveev A, Saribekyan H, Budden D, Rolnick D, Odor G, Jones SK-BTR, Pfister H, Lichtman JW, Shavit N: *A multi-pass approach to large-scale connectomics*. 2016. ArXiv preprint arXiv:1612.02120.

43. Maitin-Shepard JB, Jain V, Januszewski M, Li, Abbeel: **Combinatorial energy learning for image segmentation**. *Advances in Neural Information Processing Systems*. 2016:1966-1974.

44. Januszewski M, Maitin-Shepard J, Li J, Kornfeld W, Denk V, Jain: *Flood-filling networks*. 2016. ArXiv preprint arXiv:1611.00421.

45. • Januszewski M, Kornfeld J, Li H, Pope A, Blakely T, Lindsey L, Maitin-Shepard JB, Tyka M, Denk W, Jain V: **High-precision automated reconstruction of neurons with flood-filling networks**. *BioRxiv* 2017, **200**:675.

This paper uses recurrent neural networks to build 3D reconstructions of individual neurons from electron microscope imagery with very low error rate.

46. • Matveev A, Meirovitch Y, Saribekyan H, Jakubiuk W, Kaler T, Odor G, Budden D, Zlateski A, Shavit N: **A multicore path to connectomics-on-demand**. *Proceedings of the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, ACM* 2017:267-281.

This paper presents highly optimized and parallelized algorithms for obtaining connectomes swiftly from electron microscope imagery.

47. Rolnick D, Meirovitch Y, Parag T, Pfister H, Jain V, Lichtman JW, Boyden ES, Shavit N: *Morphological error detection in 3D segmentations*. 2017. ArXiv preprint arXiv:1705.10882.

48. Yoon Y-G, Dai J, Wohlwend J-B, Chang AH, Marblestone ES, Boyden: **Feasibility of 3D reconstruction of neural morphology using expansion microscopy and barcode-guided agglomeration**. *Front Comput Neurosci* 2017, **11**.

49. Zung J, Tartavull I, Lee K, Seung HS: **An error detection and correction framework for connectomics**. *Advances in Neural Information Processing Systems*. 2017:6821-6832.

50. Santurkar S, Budden D, Matveev A, Berlin H, Saribekyan H, Meirovitch Y, Shavit N: *Toward streaming synapse detection with compositional ConvNets*. 2017. ArXiv preprint arXiv:1702.07386.

51. Majka TA, Chaplin H-H, Yu A, Tolpygo, Mitra DK, Wójcik MG, Rosa: **Towards a comprehensive atlas of cortical connections in a primate brain: mapping tracer injection studies of the common marmoset into a reference digital template**. *J Comp Neurol* 2016, **524**:2161-2181.

52. Harris KD, Mihalas S, Shea-Brown E: **High resolution neural connectivity from incomplete tracing data using nonnegative spline regression**. *Advances in Neural Information Processing Systems*. 2016:3099-3107.

53. Knox JE, Harris KD, Graddis N, Whitesell JD, Zeng H, Harris JA, Shea-Brown E, Mihalas S: **High resolution data-driven model of the mouse connectome**. *BioRxiv* 2018, **293**:019.

54. Van Essen DC: **Windows on the brain: the emerging role of atlases and databases in neuroscience**. *Curr Opin Neurobiol* 2002, **12**:574-579.

55. •• Amunts K, Zilles K: **Architectonic mapping of the human brain beyond Brodmann**. *Neuron* 2015, **88**:1086-1107.

This paper reviews efforts in mapping the cytoarchitecture and organization of the human brain.

56. Armañanzas R, Ascoli GA: **Towards the automatic classification of neurons**. *Trends Neurosci* 2015, **38**:307-318.

57. • Jain AK: **Data clustering: 50 years beyond k-means**. *Pattern Recogn Lett* 2010, **31**:651-666.

This paper provides a comprehensive review of clustering methods and their applications in a wide range of scientific fields.

58. Von Luxburg U: **A tutorial on spectral clustering**. *Stat Comput* 2007, **17**:395-416.

59. •• Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M *et al.*: **A multi-modal parcellation of human cerebral cortex**. *Nature* 2016, **536**:171-178.

This paper introduces a semi-automated approach for integrating multi-modal anatomical and functional datasets to parcellate the human brain.

60. Ripley BD: **Modelling spatial patterns**. *J R Stat Soc Ser B: Methodol* 1977:172-212.

61. Baddeley A, Jammalamadaka A, Nair G: **Multitype point process analysis of spines on the dendrite network of a neuron**. *J R Stat Soc Ser B: Appl Stat* 2014, **63**:673-694.

62. LaGrow TJ, Moore M, Prasad JA, Davenport MA, Dyer EL: **Approximating cellular densities from high-resolution**

neuroanatomical imaging data. *Proceedings of the IEEE Engineering in Medicine and Biology Society Conference* 2018.

63. Sporns O, Tononi G, Edelman GM: **Theoretical neuroanatomy:**
• **relating anatomical and functional connectivity in graphs and cortical connection matrices**. *Cereb Cortex* 2000, **10**:127-141.
This paper introduces an approach for relating anatomical and functional connectivity by implementing dynamical systems on graphs. Their analysis reveals many similarities between the anatomical connectivity of macaque and cat cortex.

64. Erdós A, Rényi: **On the evolution of random graphs**. *Publ Math Inst Hung Acad Sci* 1960, **5**:17-60.

65. Holland PW, Laskey KB, Leinhardt S: **Stochastic blockmodels: first steps**. *Soc Netw* 1983, **5**:109-137.

66. Jonas E, Körding K: **Automatic discovery of cell types and**
•• **microcircuitry from neural connectomics**. *Elife* 2015, **4**.
This paper uses a variant of stochastic block models to infer neuron types with high accuracy based on connectivity. The algorithm models the probability of connection between cells as a function of their distance apart and their underlying types.

67. Farhoodi R, Rolnick D, Körding K: **Neuron dendrograms uncover asymmetrical motifs**. *Computational Systems Neuroscience (COSYNE) Annual Meeting* 2018.

68. Parekh R, Ascoli GA: **Neuronal morphology goes digital: a**
•• **research hub for cellular and system neuroscience**. *Neuron* 2013, **77**:1017-1038.
This paper provides a new resource and open dataset which catalogs neuronal morphologies. In addition to providing neuron models for single cells, they also provide tools for extracting morphological features from neurons for organizing and categorizing cells into types.

69. Priebe CE, Park Y, Tang M, Athreya A, Lyzinski V, Vogelstein JT, Qin Y, Cocanougher B, Eichler K, Zlatic M *et al.*: *Semiparametric spectral modeling of the Drosophila connectome*. 2017. ArXiv preprint arXiv:1705.03297.

70. Churchland MM, Cunningham J, Kaufman MT, Foster JD, Nuyujukian SI, Ryu KV, Shenoy: **Neural population dynamics during reaching**. *Nature* 2012, **487**:51.

71. Kao JC, Nuyujukian SI, Ryu MM, Churchland J, Cunningham KV, Shenoy: **Single-trial dynamics of motor cortex and their applications to brain-machine interfaces**. *Nat Commun* 2015, **6**:7759.

72. Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR *et al.*: **Inferring single-trial neural population dynamics using sequential auto-encoders**. *Nat Methods* 2018.

73. Dyer EL, Azar MG, Perich MG, Fernandes HL, Naufel S, Miller LE, Körding K: **A cryptography-based approach for movement decoding**. *Nat Biomed Eng* 2017, **1**:967.

74. Vincent H, Larochelle Y, Bengio P-A, Manzagol: **Extracting and composing robust features with denoising autoencoders**. *Proceedings of the 25th International Conference on Machine Learning, ACM* 2008:1096-1103.

75. Masci J, Meier U, Cireşan D, Schmidhuber J: **Stacked convolutional auto-encoders for hierarchical feature extraction**. *International Conference on Artificial Neural Networks, Springer* 2011:52-59.

76. Farhoodi R, Ramkumar K, Körding: **Deep learning approach towards generating neuronal morphology**. *Comput Syst Neurosci* 2018.

77. Molano-Mazon M, Onken A, Piasini E, Panzeri S: **Synthesizing realistic neural population activity patterns using generative adversarial networks**. *Proceedings of the International Conference on Learning Representations* 2018.

78. Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G,
•• Lipnick S, Mount E, O'Neil A, Shah K, Lee AK *et al.*: **In silico labeling: predicting fluorescent labels in unlabeled images**. *Cell* 2018, **173**:792-803.
This paper uses deep learning algorithms to transform real transmitted-light images into simulated fluorescence microscopy images. This approach holds across different fluorescent labels, cell types, and datasets.

79. Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel:
• **InfoGAN: interpretable representation learning by information maximizing generative adversarial nets**. *Adv Neural Inf Process Syst* 2016:2172-2180.
This paper introduces a generative adversarial network architecture for interpretable machine learning.