Applications and Challenges of Machine Learning to Enable Realistic Cellular Simulations

R. Vasan [1], **M. P. Rowan** [2], **C. T. Lee** [1], **G. R. Johnson** [3], **P. Rangamani** [1], **and M. Holst** [4,5,*]

[1] *Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92093, USA*

[2] *Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA*

[3] *Allen Institute of Cell Science, Seattle, WA 98109, USA*

[4] *Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA*

[5] *Department of Physics, University of California San Diego, La Jolla, CA 92093, USA*

Correspondence*:
Michael Holst
mholst@ucsd.edu

2 ABSTRACT

In this perspective, we examine three key aspects of an end-to-end pipeline for realistic cellular simulations: reconstruction and segmentation of cellular structures; generation of cellular structures; and mesh generation, simulation, and data analysis. We highlight some of the relevant prior work in these distinct but overlapping areas, with a particular emphasis on current use of machine learning technologies, as well as on future opportunities.

Keywords: **machine learning, cellular structures, segmentation, reconstruction, meshing, simulation**

1 INTRODUCTION

Machine learning (ML) approaches, including both traditional and deep learning methods, are revolutionizing biology. Owing to major advances in experimental and computational methodologies, the amount of data available for training is rapidly increasing. The timely convergence of data availability, computational capability, and new algorithms is a boon for biophysical modeling of subcellular and cellular scale processes such as biochemical signal transduction and mechanics [1]. To date, many simulations are performed using idealized geometries that allow for the use of commonly used techniques and software [2, 3, 4, 5, 6]. This is historically due to the lack of high-resolution structural data as well as the theoretical and computational challenges for simulations in realistic cellular shapes, due to the complexity of generating high-quality, high-resolution meshes for simulation, and the need to develop specialized fast numerical solvers that can be used with very large unstructured mesh representations of the physical domain.

As biophysical methods have improved, the complexity of our mathematical and computational models is steadily increasing [7, 8, 9]. A major frontier for physics-based study of cellular processes will be

21 to simulate biological processes in realistic cell shapes derived from various structural determination
22 modalities [10, 11]. For biological problems ranging from cognition to cancer, it has long been understood
23 that cell shapes are often correlated with mechanism [12, 13, 14, 15, 16]. Despite such clear correlations,
24 there remain gaps in our understanding of how cellular ultrastructure contributes to cellular processes
25 and the feedback between cellular structure and function. Challenges such as the diffraction limit of light
26 and difficulties in manipulation of intracellular ultrastructure constrain the potential scope of what can be
27 achieved experiments. Much like the partnership between biophysics and molecular dynamics simulations
28 have enabled the modeling of invisible protein motions to shed insights on experimental observations,
29 simulations of cellular processes can also aid in the validation and generation of hypothesis currently
30 inaccessible by experimental methods. Recently, we and others have shown that, for example, cell shape
31 and localization of proteins can impact cell signaling [4, 12, 17, 5, 18].

32 The major bottleneck for the widespread use of cell scale simulations with realistic geometries is not
33 the availability of structural data. Indeed, there exist many three-dimensional imaging modalities such as
34 confocal microscopy, multiphoton microscopy, super-resolution fluorescence and electron tomography
35 [19, 20]. For example, automation of modalities such as Serial Block-Face Scanning Electron Microscopy
36 is already enabling the production of data at rapid rates. The bottleneck lies in the fact that much of
37 the data generated from these imaging modalities need to be manually curated before it can be used for
38 physics-based simulations. This current *status quo* of manually processing and curating these datasets for
39 simulations is a major obstacle to our progress. In order to bridge the gap between abundance of cellular
40 ultrastructure data generated by 3D electron microscopy (EM) techniques and simulations in these realistic
41 geometries, innovations in machine learning (ML) methods will be necessary to reduce the time it takes to
42 go from structural datasets to initial models. There are already many similar efforts at the organ/tissue and
43 connectomics length scales [21, 22, 23]. In this work, we summarize the main steps necessary to construct
44 simulations with realistic cellular geometries (Fig. 1) and highlight where innovation in ML efforts are
45 needed and will have significant impacts. We further discuss some of the challenges and limitations in the
46 existing methods, setting the stage for new innovations for ML in physics-based cellular simulations.

2   SOURCES OF ERROR IN IMAGING MODALITIES

47 Images generated by the various microscopy modalities must undergo pre-processing to correct for errors
48 such as uneven illumination or background noise [25, 26]. The choice of suitable algorithms for error
49 correction depends on multiple factors, some of which are listed here – the length scale of the experiment
50 being conducted, scalability and reproducibility of the experiment, optical resolution of the microscope,
51 sensitivity of the detector, specificity of the staining procedure, imaging mode (2D, 3D, 3D time-series),
52 imaging modality (fluorescence, EM, ET etc.,) and other imaging artifacts like electronic noise, lens
53 astigmatism, mechanical tilting/vibration, sample temperature, and discontinuous staining [27, 25, 26, 28].
54 These sources of error are an important consideration for model implementation further downstream [6].

55 Electron tomography (ET) remains one of the most popular methods of cell imaging for modeling
56 purposes [29, 30, 31], as it retains the highest resolution of all the 3D cell imaging techniques [26] by
57 reconstructing a 3D object from a series of 2D images collected at different tilt angles [32]. However,
58 images from ET also have a low signal to noise ratio (SNR) and have anisotropic resolution (for example,
59 1 nm resolution in x, y and 10 nm resolution in z) [25]. This is partly because biological samples can
60 withstand only a limited dose of electron beam radiation (SNR is proportional to the square root of the
61 electron beam current) before the specimen is damaged [33]. Other sources of error such as misalignment of
62 projections and missing wedges from an incomplete tilt angular range can significantly affect the quality of

63 the reconstruction. To work with data such as these, image processing steps are required for high resolution
64 3D reconstruction [25, 34]. Commonly used software packages for image processing such as IMOD [35]
65 and TomoJ [36] use reconstruction algorithms such as Weighted Backprojection (WBP) and Simultaneous
66 Iterative Reconstruction Technique (SIRT). While these have been very effective at reconstruction, sources
67 of error can still accumulate, leading to further manual adjustment [37].

3 APPLICATIONS OF ML FOR THE SEGMENTATION AND RECONSTRUCTION OF CELLULAR STRUCTURES

68 Given a noisy 3D reconstruction, how can we segment cellular structures of interest? One approach is to
69 employ manual segmentation tools applied to 3D tomograms such as XVOXTRACE [32, 38], and more
70 generally, manual contouring, interpolation, and contour stacking (Fig. 1). The advantage of such methods
71 is that the human eye performs exceptionally well at detecting objects in an image [27, 39]. Consequently,
72 semi-manual and manual segmentation are widely adopted, favoring accuracy over efficiency. However,
73 such methods can be extremely tedious and not always reproducible. Alternatively, numerous semi-
74 automatic segmentation algorithms such as interpolation, watershed, thresholding, and clustering are
75 available as plugins in software packages like IMOD [35] and ImageJ [40] (Fig. 1, classical). However,
76 the accuracy of open platform algorithms is debatable [41] because of two main reasons – *(i)* Even with a
77 'perfect' ET reconstruction (no tilt misalignment, no missing wedge, no noise), the application of filtering
78 algorithms like Gaussian blur or non-linear anisotropic diffusion (NAD) [42] can cause artefacts that lead
79 to misclassifications, rendering the image unsuitable for downstream quantitative simulations and analysis.
80 *(ii)* Segmentation workflows are often designed for a specific structure and/or imaging modality, limiting
81 their generalizability and applicability.

82   Annual cell segmentation challenges are evidence of the demand for automatic segmentation [43, 44],
83 with many of its past winners responding with ML-based programs [45, 46]. Training labels for ML
84 techniques requires a relatively small percentage (as small as 10%) of manually segmented labels, allowing
85 for very large data sets to be processed significantly faster than previously implemented semi-automatic
86 segmentation methods. The most successful teams utilized ML techniques such as random forest classifiers,
87 support vector machines, or a combination of these to get segmentations comparable or often even better
88 than their human counterparts [45, 43, 46, 44] (Fig. 1, machine learning). These techniques function
89 by imputing several image features such as noise reduction, and texture and edge detection filters [47].
90 These filters are then used to train a classification algorithm in an interactive manner, achieving better
91 classification accuracy at the cost of increased training time compared to the direct application of a filter.
92 However, because the algorithm is interactive, it still requires manual input and both the training time and
93 accuracy can depend on the user.

94   More recently, deep learning-based ML algorithms (Fig. 1, deep learning), and more specifically,
95 convolutional neural networks (CNNs) have surged in popularity due to the success of AlexNet in the
96 ImageNet classification challenge [48]. CNNs are complex learnable non-linear functions that do not
97 require the imputation of data-specific features. Indeed, CNNs learn the feature mapping directly from
98 the image. The U-Net convolutional neural network architecture [46] further generalized deep learning,
99 winning the ISBI neuronal structure segmentation challenge in 2015 with a quicker speed and with fewer
100 training images. It functions by using the feature mapping imputed by a CNN to map the classification
101 vector back into a segmented image. Such is the achievement of the U-Net that its variants are now the
102 state-of-the-art in tasks like calling genetic variation from gene-sequencing data [49], brain tumor detection
103 [50] and segmentation of medical image datasets [51]. However, such deep learning based methods have

104 their own challenges. They require both quality and quantity of annotated training data, significant amount
105 of training time, graphics processing unit computing, and can generalize poorly to a different dataset.

106 Both the difficulty and cost of generating annotated training data increases exponentially when dealing
107 with Volumetric (3D) images compared with 2D, which are the desired inputs for biophysical simulations.
108 Since the U-Net is a 2D architecture [46], it cannot be applied directly to 3D images without modifications.
109 To this end, 3D U-net used sparsely annotated 2D slices to generate volumetric segmentations of brain
110 tumors [52]. Similarly, VoxRestNet [53] introduced residual learning using ResNet [54], a deep residual
111 network capable of training hundreds to thousands of layers without a performance drop, to a voxelwise
112 representation of 3D magnetic resonance (MR) images of the brain, paving the way for scalable 3D
113 segmentation.

114 Excitingly, such algorithms are being made openly accessible and easy-to-use. For example, iLastik
115 [45, 55] and Trainable Weka Segmentation [47] are both available as plugins in software packages like
116 ImageJ. These tools provide an interactive platform for segmentation, employing supervised classification
117 techniques like random forests as well as unsupervised clustering such as K-means [47]. Similarly, deep
118 learning tools such as DeepCell [56] and U-Net [57, 46] are also available in various bioimage software
119 packages. Other stand-alone tools like the Allen Cell Structure Segmenter provide a lookup table of
120 20 segmentation workflows that feed into an iterative deep learning model [58]. Cloud compute based
121 segmentation plugins like CDeep3M [59] leverage Amazon Web Services (AWS) images to provide an
122 efficient and compute-scalable tool for both 2D and 3D biomedical images.

123 Generating well-organized and annotated training data continues to be the major challenge for most ML
124 segmentation methods. Crowdsourced annotation tools like Amazon's Mechanical Turk can be useful in
125 this context, but are still limited by the difficulty of training naive users on tracing specific structural images.
126 Alternatively, many ML algorithms leverage transfer learning approaches using pre-trained networks
127 such as VGG-net [60, 61, 62], AlexNet [48] and GoogleNet [63]. In fact, popular semantic segmentation
128 and clustering networks like Fully Convolutional Networks (FCN) [64] and DECAF [65] are themselves
129 implemented using transfer learning approaches. Such transfer learning can also be used to generalize
130 models trained on biological data to a different cell type or experimental condition, significantly reducing
131 the time for training and accompanying computing resources required. More recently, label-free approaches
132 employing a U-net variant have been applied to predict cellular structure from unlabeled brightfield images
133 [66, 67]. These methods can serve as a platform for building low cost, scalable, and efficient segmentation
134 of 3D cellular structure.

4 APPLICATIONS OF ML FOR THE GENERATION OF SYNTHETIC CELLULAR STRUCTURES

135 There are two main aspects involved in the development of comprehensive biophysical models – (1) what is
136 the process being modeled? and (2) what is the geometry in which this process is being modeled? Answers
137 to the first question are based on experimental observations and specific biology. Answering the latter is
138 significantly more challenging because of the difficulties in – *(i)* obtaining accurate segmentations, *(ii)*
139 discovering new structure from experiments, and *(iii)* simultaneously visualizing multiple structures. The
140 use of synthetically generated geometries, which can probe different arrangements of organelles within
141 cells could be relevant for generating biologically relevant hypotheses.

142 A subset of ML models, called *generative* models, deal with the task of generating new synthetic but
143 realistic images that match the training set distribution. For our purposes, such methods are relevant in the

144 context of generating *(i)* noise-free images, *(ii)* images representative of a different cell type, structure, or
145 time-point, and *(iii)* unlikely images that represent the most unique shapes of the structure being imaged.
146 For example, by capturing the unlikely and likely shapes in our dataset, we could generate sequences
147 of synthetic images that transition from one shape to the next. These synthetic images can be used in
148 biophysical simulations to generate biologically relevant hypotheses.

149   In recent years, there has been rapid progress in applying deep generative models to natural images, text,
150 and even medical images. Popular classes of deep generative models like Variational Autoencoders [68],
151 Generative Adversarial Networks [69], and Autoregressive models such as PixelRNN [70] and PixelCNN
152 [71] have achieved state of the art performance on popular image datasets such as MNIST [72], CIFAR
153 [73] and ImageNet [74]. Each class of models has numerous modified implementations. For example,
154 GANs alone include models like deep convolutional GAN (DCGAN) [75], conditional GAN (cGAN)
155 [76], StackGAN [77], InfoGAN [78] and Wasserstein GAN [79] to name a few. Each model has its own
156 distinct set of advantages and disadvantages. GANs can produce photo-realistic images at the cost of
157 tricky training and no dimensionality reduction. VAEs allow for both generation and inference, but their
158 naive implementation results in less photo-realistic generative examples. Autoregressive models obtain
159 the best log-likelihoods at the cost of poor dimensionality reduction. Importantly, all of these models
160 are unsupervised, implying that they are not limited by manual annotation that is otherwise a common
161 challenge to supervised learning approaches.

162   In cell biology, much of the work in building generative models of cellular structures has been associated
163 with the open source CellOrganizer [80, 81, 82, 83, 84, 85, 86], which uses a Gaussian Mixture Model
164 given reference frames like the cell and nuclear shape in order to predict organelle shape distribution.
165 These models also have the option to be parametric (parameters such as number of objects), which reduces
166 the complexity of the learning task, the training time and GPU computing resources required, while also
167 allowing for exploration and analysis of the parameters and their effect on the spatial organization of
168 cells. Aside from CellOrganizer, other recent efforts have begun to leverage deep generative models in
169 cell biology. We now have models that can predict structure localization given cell and nuclear shape [87],
170 extract functional relationships between fluorescently tagged proteins structures in cell images [88], learn
171 cell features from cell morphological profiling experiments [89], and interpret gene expression levels from
172 single-cell RNA sequencing data [90, 91].

173   The challenge going forward will be how best to use generative modeling given the data in hand. This will
174 depend on the question we want to ask of the data. For example, if we are modeling processes associated
175 with cell and nuclear shape, spherical harmonics based generative models might be more appropriate than
176 deep learning based methods [92]. If we are interested in inpainting a missing wedge from a tomogram
177 using a generative model, then GANs might be more appropriate [93]. Generated images can also be used
178 as a source of training data for segmentation and classification tasks [94]. Taken together, these choices will
179 help develop efficient end-to-end pipelines for segmentation and shape generation, and provide a platform
180 for running biophysical simulations. Already, CellOrganizer can export spatial instances to cell simulation
181 engines such as MCell [95] and VirtualCell [96], allowing us to simulate chemical reactions in different
182 spatial compartments. Similar pipelines for deep generative models will need to be implemented in order
183 to fully realize their downstream interpretations.

5 APPLICATIONS OF ML FOR MESHING, SIMULATION, AND DATA ANALYSIS

184 ML is commonly applied to mesh segmentation and classification; examples include PointNet [98]
185 (segments and classifies a point cloud), and MeshCNN [99] (segments and classifies edges in a mesh).
186 However, although the term *machine learning* was not traditionally used to describe meshing techniques,
187 in fact algorithms for mesh generation (cf. [100]), mesh improvement (such as mesh smoothing [101]), and
188 mesh refinement [102, 103, 104, 105] all fundamentally involve local (cf. [106]) and/or global (cf. [107])
189 optimization of an *objective function* (see Fig. 2). Mesh point locations, and/or edge/face connectivity
190 decisions are viewed as parameters that are determined (or *learned*) as part of an iterative algorithm
191 that extremizes a local or global objective function (usually involving constraints as well) in an effort
192 to generate, improve, or refine a given mesh. In addition, *adaptive numerical methods* for simulation of
193 physical systems involving the solution of ordinary (ODE) and partial (PDE) differential equations are
194 again an early example of the application of ML techniques in computational science, long before the
195 terminology was widely used. A classic reference from the 1970's in the context of adaptive finite element
196 methods is [108, 109]; all modern approaches to adaptive numerical methods for ODE and PDE systems
197 continue to follow the same general framework outlined in that work: *(i)* Solve the ODE/PDE on the
198 current mesh; *(ii)* Estimate the error using *a posteriori* indicators; *(iii)* Refine the mesh using provably
199 non-degenerate local refinement with closure; *(iv)* Go back to step *(i)* and repeat the iteration until a target
200 quality measure is obtained (a standard approach is to approximately minimize a *global error function*,
201 through the use of local error estimates). These types of adaptive algorithms are effectively *machine*
202 *learning* the best possible choice (highest accuracy with least cost) of mesh and corresponding numerical
203 discretization for the target ODE/PDE system. Recent work in the area is now moving toward a more
204 explicit and sophisticated use of modern ML techniques (cf. [110, 111]).

205 Given a high quality and high resolution mesh representation of a structural geometry (see Fig. 2), we
206 can begin to probe structure-function relationships through mathematical modeling [11, 112]. However, a
207 single realistic geometry is not enough since it only captures a snapshot in time of the cellular geometry.
208 Furthermore, structural variability is a hallmark of cell biology [14, 24]. Dimensionality reduction
209 techniques like principal component analaysis (PCA) or Variational Autoencoders (VAEs) can help
210 determine both the average and extreme representations of a distribution of shapes, providing a starting
211 point for running simulations. Generative models (discussed section 4) can then be used to populate a
212 single cell with multiple learned distributions – for example, generate cell shapes from EM images overlaid
213 with protein structure shapes learned from fluorescence images [10, 87, 80].

214 To facilitate population studies, it is important that structural datasets be made publicly available, as they
215 commonly are in neuroscience [24, 113, 23]. This is following in the footsteps of -omics datasets such as
216 genomics, proteomics and transcriptomics [114], which have traditionally been made public in large scale
217 projects like the Cancer Genome Atlas [115], the Human Microbiome Project [116] and the ENCODE
218 project consortium [117]. ML can then be used to identify structure-phenotype associations, in much the
219 same way as genotype-phenotype relationships are predicted from -omics studies [118, 119].

220 Importantly, by running simulations on distributions of realistic shapes, we can generate experimentally
221 testable hypotheses. This is much harder in -omics datasets, where mechanistic insight is usually obtained
222 via constraint based modeling [119]. Further, we can also explore the implications of assuming an idealistic
223 geometry – a common assumption in hypothesis-driven modeling. For example, idealized geometries have
224 been the starting point of many signaling models that explore spatio-temporal dynamics using deterministic
225 reaction-diffusion formulations[120, 4, 17, 5, 18, 121, 122, 123] or using Brownian dynamics or other
226 formulations[124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134]. An excellent example of insights

227 gained using these idealized geometries is in exploring how changing the diffusion distances can affect the
228 dynamics of signaling molecules such as Ca^{2+} [123]. Other physical systems that are commonly modeled
229 and simulated include structural mechanics [7, 2], fluid mechanics [135, 112] and thermodynamics [136].

230   A major bottleneck in setting up accurate computational simulations of biophysical systems, idealistic or
231 otherwise, revolve around the choice of constitutive equations, estimation of the free parameters such as
232 reaction rate constants, diffusion coefficients, and material properties, and computational algorithms for
233 solving the resulting governing equations numerically on these domains. While there is a large history of
234 mathematical modeling in biology to set the stage for constitutive equations, estimation of free parameters
235 remains a major challenge. Another major challenge for physically realistic models of signaling is knowing
236 the location of the various molecules involved. Realistic geometries pose the additional challenge of
237 requiring us to first understand the distribution of shapes, followed by analyzing simulation results across
238 that distribution. Similar to how ML can be used in adaptive numerical methods to output a good mesh, ML
239 can also be used for adaptive nonlinear data fitting to determine biophysical parameters with uncertainty
240 estimates [137, 138]. Incorporating domain knowledge such as stress-strain relationships [139] or statistical
241 molecular dynamic states [140] into ML algorithms can also improve interpretability while closing the
242 loop between ML frameworks and biophysical modeling.

6 PERSPECTIVES AND FUTURE DIRECTIONS

243 In this perspective, we have discussed three key aspects of a pipeline for realistic cellular simulations:
244 *(i)* Reconstruction and segmentation of cellular structure; *(ii)* Generation of cellular structure; and *(iii)*
245 Mesh generation, refinement and simulation. While these were discussed separately, neural networks
246 like Pixel2Mesh demonstrate the feasibility of end-to-end pipelines from a single black box [141]. Of
247 course, black boxes are not interpretable, and recent ML frameworks using Visible Neural Networks [142]
248 demonstrate the potential of incorporating extensive prior knowledge to create a fully interpretable neural
249 network capable of highlighting functional changes to every neuron/subsystem upon perturbing the input.
250 Other ML frameworks like SAUCIE use regularizations to enforce mechanistic intepretability in the hidden
251 layers of an autoencoder neural network [143]. We anticipate that future endeavours will implement a fully
252 interpretable and end-to-end pipeline for biophysical simulations.

CONFLICT OF INTEREST STATEMENT

253 The authors declare that the research wasedonducted in the absence of any commercial or financial
254 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

255 All co-authors contributed to the writing of the manuscript, and also provided area-specific expertise.

256 *R. Vasan*: Provided expertise on generation and simulation.

257 *M. Rowan*: Provided expertise on imaging, segmentation, and reconstruction.

258 *C.T. Lee*: Provided expertise on reconstruction, meshing and simulation.

259 *G.R. Johnson*: Provided expertise on imaging, segmentation, reconstruction and generation.

260 *P. Rangamani*: Provided expertise on imaging, segmentation, reconstruction, modeling, and simulation.

M. Holst: Provided expertise on meshing, simulation, and analysis.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

[1] Lee CT, Amaro RE. Exascale Computing: A New Dawn for Computational Biology. *Computing in Science & Engineering* **20** (2018) 18–25. doi:10.1109/MCSE.2018.05329812.

[2] Alimohamadi H, Vasan R, Hassinger JE, Stachowiak JC, Rangamani P. The role of traction in membrane curvature generation. *Molecular biology of the cell* **29** (2018) 2024–2035.

[3] Vasan R, Akamatsu M, Schöneberg J, Rangamani P. Intracellular membrane trafficking: Modeling local movements in cells. *Cell Movement* (Springer) (2018), 259–301.

[4] Bell M, Bartol T, Sejnowski T, Rangamani P. Dendritic spine geometry and spine apparatus organization govern the spatiotemporal dynamics of calcium. *The Journal of general physiology* **151** (2019) 2221.

[5] Ohadi D, Rangamani P. Geometric control of frequency modulation of camp oscillations due to ca2+-bursts in dendritic spines. *Biophysical journal* (2019). doi:https://doi.org/10.1016/j.bpj.2019.10.004.

[6] Vasan R, Maleckar MM, Williams CD, Rangamani P. Dlite uses cell-cell interface movement to better infer cell-cell tensions. *Biophysical journal* (2019). doi:https://doi.org/10.1016/j.bpj.2019.09.034.

[7] Vasan R, Rudraraju S, Garikipati K, Akamatsu M, Rangamani P. A mechanical model reveals that non-axisymmetric buckling lowers the energy barrier associated with membrane neck constriction. *Soft Matter* (2019).

[8] Rudraraju S, Van der Ven A, Garikipati K. Mechanochemical spinodal decomposition: a phenomenological theory of phase transformations in multi-component, crystalline solids. *npj Computational Materials* **2** (2016) 16012.

[9] Mihai LA, Budday S, Holzapfel GA, Kuhl E, Goriely A. A family of hyperelastic models for human brain tissue. *Journal of the Mechanics and Physics of Solids* **106** (2017) 60–79.

[10] Murphy RF. Building cell models and simulations from microscope images. *Methods* **96** (2016) 33–39.

[11] Lee CT, Laughlin JG, Moody JB, Amaro RE, McCammon JA, Holst MJ, et al. An open source mesh generation platform for biophysical modeling using realistic cellular geometries. *arXiv preprint arXiv:1909.04781* (2019).

[12] Rangamani P, Lipshtat A, Azeloglu EU, Calizo RC, Hu M, Ghassemi S, et al. Decoding information in cell shape. *Cell* **154** (2013) 1356–1369.

[13] Deuling H, Helfrich W. Red blood cell shapes as explained on the basis of curvature elasticity. *Biophysical journal* **16** (1976) 861–868.

[14] Bartol Jr TM, Bromer C, Kinney J, Chirillo MA, Bourne JN, Harris KM, et al. Nanoconnectomic upper bound on the variability of synaptic plasticity. *Elife* **4** (2015) e10778.

[15] Ritz R, Sejnowski TJ. Synchronous oscillatory activity in sensory systems: new vistas on mechanisms. *Current opinion in neurobiology* **7** (1997) 536–546.

[16] Harris KM, Kater S. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annual review of neuroscience* **17** (1994) 341–371.

[17] Cugno A, Bartol TM, Sejnowski TJ, Iyengar R, Rangamani P. Geometric principles of second messenger dynamics in dendritic spines. *Scientific reports* **9** (2019) 1–18.

[18] Ohadi D, Schmitt DL, Calabrese B, Halpain S, Zhang J, Rangamani P. Computational modeling reveals frequency modulation of calcium-camp/pka pathway in dendritic spines. *Biophysical journal* (2019). doi:https://doi.org/10.1016/j.bpj.2019.10.003.

[19] Huang F, Sirinakis G, Allgeyer ES, Schroeder LK, Duim WC, Kromann EB, et al. Ultra-high resolution 3d imaging of whole cells. *Cell* **166** (2016) 1028–1040.

[20] Graf BW, Boppart SA. Imaging and analysis of three-dimensional cell culture models. *Live cell imaging* (Springer) (2010), 211–227.

[21] Lichtman JW, Pfister H, Shavit N. The big data challenges of connectomics. *Nature neuroscience* **17** (2014) 1448.

[22] Maher G, Wilson N, Marsden A. Accelerating cardiovascular model building with convolutional neural networks. *Medical & biological engineering & computing* **57** (2019) 2319–2335.

[23] Januszewski M, Kornfeld J, Li PH, Pope A, Blakely T, Lindsey L, et al. High-precision automated reconstruction of neurons with flood-filling networks. *Nature methods* **15** (2018) 605.

[24] Wu Y, Whiteus C, Xu CS, Hayworth KJ, Weinberg RJ, Hess HF, et al. Contacts Between the Endoplasmic Reticulum and Other Membranes in Neurons. *Proc. Natl. Acad. Sci. USA* **114** (2017) E4859–E4867. doi:10.1073/pnas.1701078114.

[25] van Aarle W, Palenstijn WJ, De Beenhouwer J, Altantzis T, Bals S, Batenburg KJ, et al. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **157** (2015) 35–47.

[26] Lidke DS, Lidke KA. Advances in high-resolution imaging–techniques for three-dimensional imaging of cellular structures. *J Cell Sci* **125** (2012) 2571–2580.

[27] Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. *Nature methods* (2019) 1.

[28] Koster A, Chen H, Sedat J, Agard D. Automated microscopy for electron tomography. *Ultramicroscopy* **46** (1992) 207–227.

[29] Mazel T, Raymond R, Raymond-Stintz M, Jett S, Wilson BS. Stochastic modeling of calcium in 3d geometry. *Biophysical journal* **96** (2009) 1691–1706.

[30] West M, Zurek N, Hoenger A, Voeltz GK. A 3d analysis of yeast er structure reveals how er domains are organized by membrane curvature. *The Journal of cell biology* **193** (2011) 333–346.

[31] Noske AB, Costin AJ, Morgan GP, Marsh BJ. Expedited approaches to whole cell electron tomography and organelle mark-up in situ in high-pressure frozen pancreatic islets. *Journal of structural biology* **161** (2008) 298–313.

[32] Perkins G, Renken C, Martone M, Young S, Ellisman M, Frey T. Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. *Journal of structural biology* **119** (1997) 260–272.

[33] Baker LA, Rubinstein JL. Radiation damage in electron cryomicroscopy. *Methods in enzymology* (Elsevier), vol. 481 (2010), 371–388.

[34] Phan S, Boassa D, Nguyen P, Wan X, Lanman J, Lawrence A, et al. 3d reconstruction of biological structures: automated procedures for alignment and reconstruction of multiple tilt series in electron tomography. *Advanced structural and chemical imaging* **2** (2017) 8.

[35] Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using imod. *Journal of structural biology* **116** (1996) 71–76.

[36] Messaoudi C, Boudier T, Sorzano COS, Marco S. Tomoj: tomography software for three-dimensional reconstruction in transmission electron microscopy. *BMC bioinformatics* **8** (2007) 288.

[37] Leary R, Saghi Z, Midgley PA, Holland DJ. Compressed sensing electron tomography. *Ultramicroscopy* **131** (2013) 70–91.

[38] Yin X, Kidd GJ, Ohno N, Perkins GA, Ellisman MH, Bastian C, et al. Proteolipid protein–deficient myelin promotes axonal mitochondrial dysfunction via altered metabolic coupling. *J Cell Biol* **215** (2016) 531–542.

[39] Le Borgne H, Guyader N, Guérin-Dugué A, Hérault J. Classification of images: Ica filters vs human perception. *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.* (IEEE) (2003), vol. 2, 251–254.

[40] Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with imagej. *Biophotonics international* **11** (2004) 36–42.

[41] Jerman T, Pernuš F, Likar B, Špiclin Ž. Enhancement of vascular structures in 3d and 2d angiographic images. *IEEE transactions on medical imaging* **35** (2016) 2107–2118.

[42] Frangakis AS, Hegerl R. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *Journal of structural biology* **135** (2001) 239–250.

[43] Arganda-Carreras I, Turaga SC, Berger DR, Cireşan D, Giusti A, Gambardella LM, et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in neuroanatomy* **9** (2015) 142.

[44] Maška M, Ulman V, Svoboda D, Matula P, Matula P, Ederra C, et al. A benchmark for comparison of cell tracking algorithms. *Bioinformatics* **30** (2014) 1609–1617.

[45] Sommer C, Straehle C, Koethe U, Hamprecht FA. Ilastik: Interactive learning and segmentation toolkit. *2011 IEEE international symposium on biomedical imaging: From nano to macro* (IEEE) (2011), 230–233.

[46] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention* (Springer) (2015), 234–241.

[47] Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, et al. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33** (2017) 2424–2426.

[48] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (2012), 1097–1105.

[49] Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal snp and small-indel variant caller using deep neural networks. *Nature biotechnology* **36** (2018) 983.

[50] Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. *annual conference on medical image understanding and analysis* (Springer) (2017), 506–517.

[51] Weng Y, Zhou T, Li Y, Qiu X. Nas-unet: Neural architecture search for medical image segmentation. *IEEE Access* **7** (2019) 44247–44257.

[52] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3d u-net: learning dense volumetric segmentation from sparse annotation. *International conference on medical image computing and computer-assisted intervention* (Springer) (2016), 424–432.

[53] Chen H, Dou Q, Yu L, Qin J, Heng PA. Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage* **170** (2018) 446–455.

[54] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *European conference on computer vision* (Springer) (2016), 630–645.

[55] Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, et al. ilastik: interactive machine learning for (bio) image analysis. *Nature Methods* (2019) 1–7.

[56] Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology* **12** (2016) e1005177.

[57] Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods* **16** (2019) 67.

[58] Chen J, Ding L, Viana MP, Hendershott MC, Yang R, Mueller IA, et al. The allen cell structure segmenter: a new open source toolkit for segmenting 3d intracellular structures in fluorescence microscopy images. *bioRxiv* (2018) 491035.

[59] Haberl MG, Churas C, Tindall L, Boassa D, Phan S, Bushong EA, et al. Cdeep3m—plug-and-play cloud-based deep learning for image segmentation. *Nature methods* **15** (2018) 677.

[60] Bruna J, Sprechmann P, LeCun Y. Super-resolution with deep convolutional sufficient statistics. *arXiv preprint arXiv:1511.05666* (2015).

[61] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. *European conference on computer vision* (Springer) (2016), 694–711.

[62] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[63] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), 1–9.

[64] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), 3431–3440.

[65] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, et al. Decaf: A deep convolutional activation feature for generic visual recognition. *International conference on machine learning* (2014), 647–655.

[66] Ounkomol C, Seshamani S, Maleckar MM, Collman F, Johnson GR. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature methods* **15** (2018) 917.

[67] Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G, Lipnick S, et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173** (2018) 792–803.

[68] Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[69] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Advances in neural information processing systems* (2014), 2672–2680.

[70] Oord Avd, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759* (2016).

[71] Van den Oord A, Kalchbrenner N, Espeholt L, Vinyals O, Graves A, et al. Conditional image generation with pixelcnn decoders. *Advances in neural information processing systems* (2016), 4790–4798.

[72] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine* **29** (2012) 141–142.

[73] Krizhevsky A, Hinton G, et al. Learning multiple layers of features from tiny images. Tech. rep., Citeseer (2009).

[74] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition* (Ieee) (2009), 248–255.

[75] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).

[76] Antipov G, Baccouche M, Dugelay JL. Face aging with conditional generative adversarial networks. *2017 IEEE International Conference on Image Processing (ICIP)* (IEEE) (2017), 2089–2093.

[77] Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision* (2017), 5907–5915.

[78] Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems* (2016), 2172–2180.

[79] Arjovsky M, Chintala S, Bottou L. Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017).

[80] Johnson GR, Buck TE, Sullivan DP, Rohde GK, Murphy RF. Joint modeling of cell and nuclear shape variation. *Molecular biology of the cell* **26** (2015) 4046–4056.

[81] Johnson GR, Li J, Shariff A, Rohde GK, Murphy RF. Automated learning of subcellular variation among punctate protein patterns and a generative model of their relation to microtubules. *PLoS computational biology* **11** (2015) e1004614.

[82] Shariff A, Kangas J, Coelho LP, Quinn S, Murphy RF. Automated image analysis for high-content screening and analysis. *Journal of biomolecular screening* **15** (2010) 726–734.

[83] Rohde GK, Ribeiro AJ, Dahl KN, Murphy RF. Deformation-based nuclear morphometry: Capturing nuclear shape variation in hela cells. *Cytometry Part A: The Journal of the International Society for Analytical Cytology* **73** (2008) 341–350.

[84] Shariff A, Murphy RF, Rohde GK. Automated estimation of microtubule model parameters from 3-d live cell microscopy images. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE) (2011), 1330–1333.

[85] Peng T, Murphy RF. Image-derived, three-dimensional generative models of cellular organization. *Cytometry Part A* **79** (2011) 383–391.

[86] Zhao T, Murphy RF. Automated learning of generative models for subcellular location: building blocks for systems biology. *Cytometry Part A* **71** (2007) 978–990.

[87] Johnson GR, Donovan-Maiye RM, Maleckar MM. Generative modeling with conditional autoencoders: Building an integrated cell. *arXiv preprint arXiv:1705.00092* (2017).

[88] Osokin A, Chessel A, Carazo Salas RE, Vaggi F. Gans for biological image synthesis. *Proceedings of the IEEE International Conference on Computer Vision* (2017), 2233–2242.

[89] Caicedo JC, McQuin C, Goodman A, Singh S, Carpenter AE. Weakly supervised learning of single-cell feature embeddings. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), 9309–9318.

[90] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nature methods* **15** (2018) 1053.

[91] Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nature communications* **9** (2018) 2002.

[92] Ruan X, Murphy RF. Evaluation of methods for generative modeling of cell and nuclear shape. *Bioinformatics* (2018).

[93] Ding G, Liu Y, Zhang R, Xin HL. A joint deep learning model to recover information and reduce artifacts in missing-wedge sinograms for electron tomography and beyond. *Scientific reports* **9** (2019) 1–13.

[94] Yang D, Deng J. Learning to generate synthetic 3d training data through hybrid gradient. *arXiv preprint arXiv:1907.00267* (2019).

[95] Stiles JR, Bartol TM, et al. Monte carlo methods for simulating realistic synaptic microphysiology using mcell. *Computational neuroscience: realistic modeling for experimentalists* (2001) 87–127.

[96] Loew LM, Schaff JC. The virtual cell: a software environment for computational cell biology. *TRENDS in Biotechnology* **19** (2001) 401–406.

[97] Lee C, Laughlin J, Angliviel de La Beaumelle N, Amaro R, McCammon J, Ramamoorthi R, et al. GAMer 2: A system for 3D mesh processing of cellular electron micrographs (2019). Submitted for publication. Available as arXiv:1901.11008 [q-bio.QM].

[98] Qi CR, Su H, Mo K, Guibas LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), 652–660.

[99] Hanocka R, Hertz A, Fish N, Giryes R, Fleishman S, Cohen-Or D. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)* **38** (2019) 90.

[100] Lee C, Moody J, Amaro R, McCammon J, Holst M. The implementation of the colored abstract simplicial complex and its application to mesh generation. *ACM Trans. Math. Software* **45** (2019) 28:1–28:20. Available as arXiv:1807.01417 [math.NA].

[101] Bank RE, Smith RK. Mesh smoothing using *a posteriori* error estimates. *SIAM J. Numer. Anal.* **34** (1997) 979–997.

[102] Liu A, Joe B. Quality local refinement of tetrahedral meshes based on bisection. *SIAM J. Sci. Statist. Comput.* **16** (1995) 1269–1291.

[103] Maubach J. Local bisection refinement for N-simplicial grids generated by relection. *SIAM J. Sci. Statist. Comput.* **16** (1995) 210–277.

[104] Bey J. Tetrahedral grid refinement. *Computing* **55** (1995) 355–378.

[105] Arnold D, Mukherjee A, Pouly L. Locally adapted tetrahedral meshes using bisection. *SIAM J. Sci. Statist. Comput.* **22** (1997) 431–448.

[106] Gao Z, Yu Z, Holst M. Feature-preserving surface mesh smoothing via suboptimal Delaunay triangulation. *Graphical Models* **75** (2013) 23–38.

[107] Chen L, Holst M. Efficient mesh optimization schemes based on optimal Delaunay triangulations. *Comp. Meth. in Appl. Mech. Engr.* **200** (2011) 967–984.

[108] Babuška I, Rheinboldt W. A posteriori error estimates for the finite element method. *Internat. J. Numer. Methods Engrg.* **12** (1978) 1597–1615.

[109] Babuška I, Rheinboldt W. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.* **15** (1978) 736–754.

[110] Fritzen F, Fernández M, Larsson F. On-the-fly adaptivity for nonlinear twoscale simulations using artificial neural networks and reduced order modeling. *Frontiers in Materials* **6** (2019) 75. doi:10.3389/fmats.2019.00075.

[111] Manevitz LM, Bitar A, Givoli D. Neural network time series forecasting of finite-element mesh adaptation. *Neurocomputing* **63** (2005) 447–463.

[112] Updegrove A, Wilson NM, Merkow J, Lan H, Marsden AL, Shadden SC. Simvascular: an open source pipeline for cardiovascular simulation. *Annals of biomedical engineering* **45** (2017) 525–541.

[113] Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, et al. Saturated reconstruction of a volume of neocortex. *Cell* **162** (2015) 648–661.

[114] Perez-Riverol Y, Zorin A, Dass G, Vu MT, Xu P, Glont M, et al. Quantifying the impact of public omics data. *Nature communications* **10** (2019) 1–10.

[115] Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics* **45** (2013) 1113.

[116] Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* **449** (2007) 804.

[117] Maher B. Encode: The human encyclopaedia. *Nature News* **489** (2012) 46.

[118] Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-generation machine learning for biological networks. *Cell* **173** (2018) 1581–1592.

[119] Zampieri G, Vijayakumar S, Yaneske E, Angione C. Machine and deep learning meet genome-scale metabolic modeling. *PLoS computational biology* **15** (2019).

[120] Resasco DC, Gao F, Morgan F, Novak IL, Schaff JC, Slepchenko BM. Virtual Cell: Computational tools for modeling in cell biology. *WIREs Systems Biology and Medicine* **4** (2012) 129–140. doi:10.1002/wsbm.165.

[121] Louch WE, Hake J, Jølle GF, Mørk HK, Sjaastad I, Lines GT, et al. Control of Ca^{2+} Release by Action Potential Configuration in Normal and Failing Murine Cardiomyocytes. *Biophysical Journal* **99** (2010) 1377–1386. doi:10.1016/j.bpj.2010.06.055.

[122] Yang PC, Boras BW, Jeng MT, Docken SS, Lewis TJ, McCulloch AD, et al. A Computational Modeling and Simulation Approach to Investigate Mechanisms of Subcellular cAMP Compartmentation. *PLOS Computational Biology* **12** (2016) e1005005. doi:10.1371/journal.pcbi.1005005.

[123] Cheng Y, Yu Z, Hoshijima M, Holst MJ, McCulloch AD, McCammon JA, et al. Numerical analysis of ca2+ signaling in rat ventricular myocytes with realistic transverse-axial tubular geometry and inhibited sarcoplasmic reticulum. *PLoS computational biology* **6** (2010) e1000972.

[124] Stiles JR, Van Helden D, Bartol TM, Salpeter EE, Salpeter MM. Miniature endplate current rise times less than 100 microseconds from improved dual recordings can be modeled with passive acetylcholine diffusion from a synaptic vesicle. *Proceedings of the National Academy of Sciences of the United States of America* **93** (1996) 5747–5752. doi:10.1073/pnas.93.12.5747.

[125] Kerr RA, Bartol TM, Kaminsky B, Dittrich M, Chang JCJ, Baden SB, et al. Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces. *SIAM Journal on Scientific Computing* **30** (2008) 3126–3149. doi:10.1137/070692017.

[126] Stiles JR, Bartol TM. Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. De Schutter E, editor, *Computational Neuroscience: Realistic Modeling for Experimentalists* (CRC Press) (2001), 87–127.

[127] Chen W, De Schutter E. Parallel STEPS: Large Scale Stochastic Spatial Reaction-Diffusion Simulation with High Performance Computers. *Frontiers in Neuroinformatics* **11** (2017). doi:10.3389/fninf.2017.00013.

[128] Hepburn I, Chen W, Wils S, De Schutter E. STEPS: Efficient simulation of stochastic reaction–diffusion models in realistic morphologies. *BMC Systems Biology* **6** (2012) 36. doi:10.1186/1752-0509-6-36.

[129] Hepburn I, Chen W, De Schutter E. Accurate reaction-diffusion operator splitting on tetrahedral meshes for parallel stochastic molecular simulations. *The Journal of Chemical Physics* **145** (2016) 054118. doi:10.1063/1.4960034.

[130] Drawert B, Engblom S, Hellander A. URDME: A modular framework for stochastic simulation of reaction-transport processes in complex geometries. *BMC systems biology* **6** (2012) 76. doi:10.1186/1752-0509-6-76.

[131] Roberts E, Stone JE, Luthey-Schulten Z. Lattice microbes: High-performance stochastic simulation method for the reaction-diffusion master equation. *Journal of Computational Chemistry* **34** (2013) 245–255. doi:10.1002/jcc.23130.

[132] Hattne J, Fange D, Elf J. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics* **21** (2005) 2923–2924. doi:10.1093/bioinformatics/bti431.

[133] Andrews SS, Bray D. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology* **1** (2004) 137–151. doi:10.1088/1478-3967/1/3/001.

[134] Oliveira RF, Terrin A, Benedetto GD, Cannon RC, Koh W, Kim M, et al. The Role of Type 4 Phosphodiesterases in Generating Microdomains of cAMP: Large Scale Stochastic Simulations. *PLOS ONE* **5** (2010) e11725. doi:10.1371/journal.pone.0011725.

[135] Wang W, Diacovo TG, Chen J, Freund JB, King MR. Simulation of platelet, thrombus and erythrocyte hydrodynamic interactions in a 3d arteriole with in vivo comparison. *PLoS One* **8** (2013) e76949.

[136] Peskin CS, Odell GM, Oster GF. Cellular motions and thermal fluctuations: the brownian ratchet. *Biophysical journal* **65** (1993) 316–324.

[137] Verrelst J, Muñoz J, Alonso L, Delegido J, Rivera JP, Camps-Valls G, et al. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3. *Remote Sensing of Environment* **118** (2012) 127–139.

[138] Huys QJ, Paninski L. Smoothing of, and parameter estimation from, noisy biophysical recordings. *PLoS computational biology* **5** (2009) e1000379.

[139] Mendizabal A, Márquez-Neila P, Cotin S. Simulation of hyperelastic materials in real-time using deep learning. *arXiv preprint arXiv:1904.06197* (2019).

[140] Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365** (2019) eaaw1147.

[141] Wang N, Zhang Y, Li Z, Fu Y, Liu W, Jiang YG. Pixel2mesh: Generating 3d mesh models from single rgb images. *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), 52–67.

[142] Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, et al. Using deep learning to model the hierarchical structure and function of a cell. *Nature methods* **15** (2018) 290.

[143] Amodio M, Van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, et al. Exploring single-cell data with deep multitasking neural networks. *Nat. Methods* **16** (2019) 1139–1145.
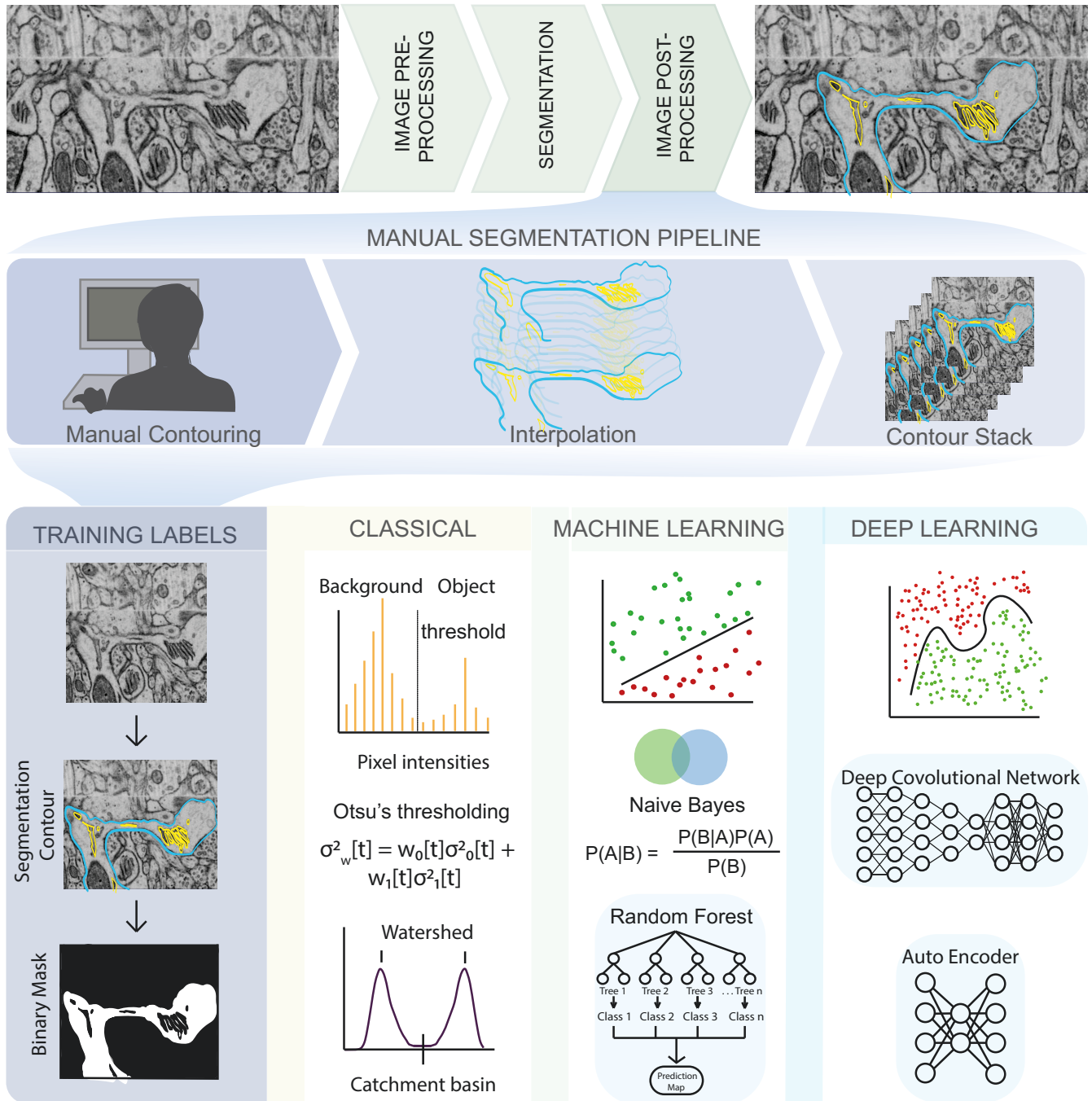
Figure 1. An illustration of the complex pipeline needed to go from imaging data to a segmented mesh, with various opportunities for emerging techniques in machine learning shown throughout the pipeline. Top row: EM images obtained from [24] of dendritic spines from mouse brain tissue. Middle row: Manual tracing or contouring, interpolation, and stacking of contours is extremely time consuming, prone to error, and relies of human judgement. Bottom row: On the other hand, development of training labels and different learning techniques can reduce both time and error, bridging the gap between biological data and simulations. Classical algorithms like Otsu's thresholding and watershed are widely used and convenient but prone to error. Traditional machine learning algorithms like Random Forest and Naive Bayes are less prone to error and easy to use but require manual painting/interaction. Deep learning algorithms are highly effective and require no manual interaction but are limited by large training sets and compute resources. The list of techniques described is representative only, and not exhaustive.
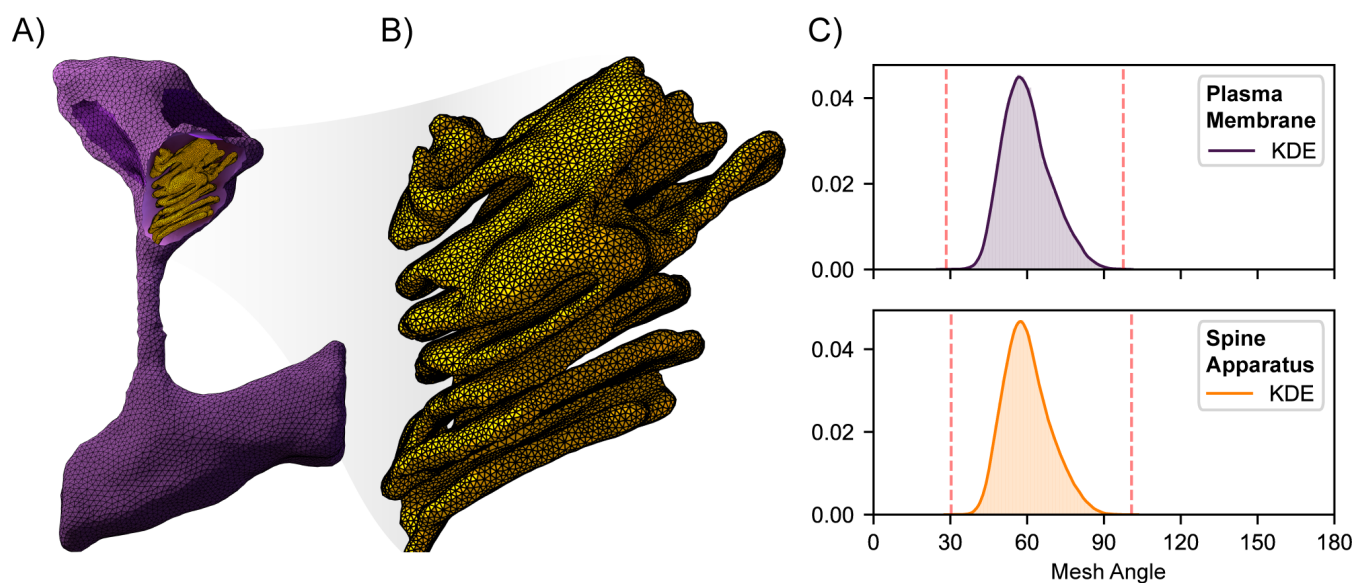
Figure 2. An illustration of complexity, size, quality, and local resolution of meshes typically needed for realistic simulation of biophysical systems. Meshes are generated using `GAMer 2` [97, 11]. A) Example surface mesh of a dendritic spine with geometry informed by electron micrographs from [24]. The plasma membrane is shown in purple with the post synaptic density rendered in dark purple. The spine apparatus, a specialized form of the endoplasmic reticulum is shown in yellow. B) A zoomed in view of the spine apparatus. Note that the mesh density is much higher in order to represent the fine structural details. C) Binned histogram distributions of mesh angles for both the plasma membrane and spine apparatus. The colored smooth lines are the result of a kernel density estimate. Dotted red lines correspond to the minimum and maximum angle values in each mesh. Both meshes are high quality with few high aspect ratio triangles (i.e., those deviating most from equilateral).