Learning Generative Models of 3D Structures

Siddhartha Chaudhuri[1], Daniel Ritchie[2], Kai Xu[3], Hao (Richard) Zhang[4]

[1] Adobe Research and IIT Bombay, India
[2] Brown University, USA
[3] National University of Defense Technology, China
[4] Simon Fraser University, Canada

Figure 1: *Some results of data-driven generative modeling of 3D structures. From top-left to bottom-right: samples from a probabilistic model learned for part assembly [KCKK12], a probabilistic model of 3D object arrangements [FRS* 12], a deep model for room layouts [WSCR18]; a structural interpolation in a continuous latent space of 3D shape structures learned from a generative autoencoder network [LXC* 17].*

Abstract

Many important applications demand 3D content, yet 3D modeling is a notoriously difficult and inaccessible activity. This tutorial provides a crash course in one of the most promising approaches for democratizing 3D modeling: learning generative models of 3D structures. Such generative models typically describe a statistical distribution over a space of possible 3D shapes or 3D scenes, as well as a procedure for sampling new shapes or scenes from the distribution. To be useful by non-experts for design purposes, a generative model must represent 3D content at a high level of abstraction in which the user can express their goals—that is, it must be structure-aware. In this tutorial, we will take a deep dive into the most exciting methods for building generative models of both individual shapes as well as composite scenes, highlighting how standard data-driven methods need to be adapted, or new methods developed, to create models that are both generative and structure-aware. The tutorial assumes knowledge of the fundamentals of computer graphics, linear algebra, and probability, though a quick refresher of important algorithmic ideas from geometric analysis and machine learning is included. Attendees should come away from this tutorial with a broad understanding of the historical and current work in generative 3D modeling, as well as familiarity with the mathematical tools needed to start their own research or product development in this area.

CCS Concepts

• ***Computing methodologies*** → *Probabilistic reasoning; Neural networks; Shape analysis;*

DOI: 10.2312/egt.20191038

1. Introduction

Creating 3D content is a critical task in many important domains, including including computer-aided design (CAD), virtual and augmented reality, film and animation, computer games, architecture, simulation and training, advertising, smart homes, and digital storytelling. Yet despite the great demand for 3D content, creating it has remained a notoriously inaccessible skill: one typically needs extensive training in the use of esoteric software to perform the necessary detailed manipulations of 3D geometry.

To help alleviate this burden of expertise and to democratize the practice of 3D modeling, modern research has focused on incorporating domain knowledge and automatic low-level geometric synthesis into 3D modeling tools. Such tools allow users to focus on high-level, goal-driven specifications (which are natural for people to provide), while allowing the machine to take care of low-level geometric manipulations. The knowledge to enable this modeling paradigm is typically encapsulated in *generative models* of 3D structure: statistical distributions over spaces of possible 3D shapes or 3D scenes. To be useful in design applications, these distributions need to be both *structure-aware* (i.e., they directly represent the shape at a high level of abstraction suitable for layering design semantics) as well as *generative* (i.e., allowing for sampling, either unconditionally or conditionally, new shapes or scenes).

This tutorial will give the audience a crash course on generative 3D modeling, introducing them to the most important techniques and directions for building generative models of 3D structures. The first half of the tutorial will be introductory, providing both a broad overview of the field as well as a quick refresher of important algorithmic ideas from geometric analysis and machine learning. The second half will consist of a deep dive into the most exciting methods for building generative models of both individual shapes as well as composite scenes. We will highlight how standard data-driven methods need to be adapted, or new methods developed, in order to create models that are both generative and structure-aware. At the end, attendees should come away with a historical context, a high-level understanding of all relevant work in the area, and familiarity with the mathematical tools to explore further.

Presentation slides and other course notes will be archived at https://3dstructgen.github.io/ for public reference.

Tutorial Outline

Section Title	Presenter	Length
Introduction	Hao (Richard) Zhang	40 mins
Geometric and Generative Modeling Basics	Siddhartha Chaudhuri	40 mins
Break		10 mins
Generative Modeling of 3D Shapes	Kai (Kevin) Xu	40 mins
Generative Modeling of 3D Scenes	Daniel Ritchie	40 mins

2. Necessary Background and Target Audience

Background This tutorial is designed to be as accessible as possible. It will require a basic background in the fundamentals of geometric modeling, e.g., shape representations, transformations, etc. Some topics covered will assume familiarity with slightly more advanced concepts from linear algebra, e.g. singular value decomposition. As the tutorial focuses on machine learning models, comfort with the basics of probability and statistics is highly desirable.

Target Audience This tutorial is aimed at two audiences:

1. New graduate students in computer graphics, or more experienced researchers in other fields who are looking for an entry point into this exciting field.
2. Graphics software engineers and developers looking to understand these technologies and how they might fit into their own pipelines, products, or services.

At the end of the day, our goal is for people in group 1 to be well equipped with the necessary background and resources to start independent research of their own in this area, and for people in group 2 to feel confident that they know where to look (and with whom to consult) if they want to integrate research results into their work.

3. Tutorial Outline

Introduction (40 mins)
Presenter: Hao (Richard) Zhang
Richard will first introduce the presenters, the topics to be covered by each, and the learning goal for the tutorial: at the end of the day, attendees should understand the main motivations for studying generative models of 3D structures, feel equipped to start their own research projects in generative modeling of 3D shape and scene structures, or to integrate existing research results into their own systems. The technical content of his talk will start by explaining the importance of 3D modeling and content creation in computer graphics and other relevant fields, outlining the key challenges, and motivating why structural analysis and modeling would play a central role in addressing these challenges. He will then define some fundamental terms and go over the relevant history of structure-aware shape analysis, which is foundational for modern-day work on structure-aware shape synthesis and generative modeling. Finally, he will briefly highlight the latest and state-of-the-art results in this field, using deep neural networks and deep generative models, to set the stage for talks by subsequent presenters.

Geometric and Generative Modeling Basics (40 mins)
Presenter: Siddhartha Chaudhuri
Sid will present the relevant technical basics that audience members must understand to do work in generative geometric modeling. The first part of this section focuses on different representations of geometry — meshes, volumes, point clouds, part assemblies, parametric models, etc — and their relative strengths and weaknesses from the perspective of machine learning and generative modeling. The second part of this section focuses on the types of machine learning models used for generative 3D design, and how those models may vary from their application in other (non-3D) domains. Specifically, Sid will discuss how models must handle "non-regular" structures which are the foundations of 3D representations: graphs, sets and trees.

Sid will begin the ML section by introducing classical generative statistical models. The discussion will focus on various probabilistic graphical models: Bayesian networks, random fields, factor graphs, and stochastic grammars. Next, Sid will discuss more recent models based on deep neural networks. The discussion will focus on two themes: (a) generic approaches for building generative nets, e.g. variational autoencoders and generative adversarial networks, and (b) specific network models developed for structural representations, e.g. recursive neural networks and graph convolutional networks. Finally, Sid will briefly touch upon more ambitious approaches for learning complex generative priors, such as probabilistic program induction.

BREAK (10 mins)

Generative Modeling of 3D Shapes (40 mins)
Presenter: Kai (Kevin) Xu
In this section, Kevin will present and explain different data-driven methods for synthesizing structured shapes. He will briefly recap early approaches to the problem, including largely handcrafted models, move on to relatively low-dimensional data-driven models (parametrized templates, graphical models, grammars, etc), and then focus on modern approaches based on deep neural networks. He will explain why structure-aware generative models present particular difficulties for standard statistical architectures (fixed-dimensional graphical models or neural networks), since they need to process graphs with varying topologies and complexities. He will highlight various important approaches to this problem, including deep hierarchical models.

One of the key insights Kevin will share is that 3D structures can be effectively modeled by a hierarchical organization of parts encompassing part relationships such as adjacency and symmetry. In particular, he will introduce GRASS [LXC*17], a recursive neural network architecture for the hierarchical encoding and synthesis of 3D shape structures. The network can be tuned in an adversarial setup to yield a generative model of plausible structures, from which novel 3D structures can be sampled. Following GRASS, he will talk about several works on utilizing hierarchical structure representation and recursive encoding/decoding architecture, in achieving structure-aware 3D shape composition, shape-structure translation, and image-to-structure reconstruction. Finally, he will also outline directions for the future, such as shape generation conditioned on functional and semantic objectives, for goal-driven design.

Generative Modeling of 3D Scenes (40 mins)
Presenter: Daniel Ritchie
This final section will transition from generative models of individual object shapes to generative models of scenes made of multiple objects. Individual objects are well-characterized by physical attachment and regular structures such as symmetries. While some classes of scene exhibit similar strong regularities (e.g. dense urban scene layouts), many other types of scene do not, instead being characterized by looser arrangement patterns (e.g. indoor scenes). This fundamentally different character of scenes necessitates different approaches to generative modeling.

This section focuses in particular on indoor scene modeling, which is becoming increasingly important not only for traditional graphics applications in gaming, animation, and simulation, but also for creating the large quantities of training data required to power today's state-of-the-art machine learning systems in computer vision and robotics [Lan18]. It begins with a look back at the history of the sub-field of *indoor scene synthesis*, starting with early rule-based systems [XSF02, Ger09, MSL*11] and continuing into the era of partially and fully data-driven methods [YYT*11, FRS*12, KLTZ16]. We will discuss the strengths and limitations of these approaches, including tasks for which they are still well-suited today. Most of this section will be devoted to the state-of-the-art methods in indoor scene synthesis, which are primarily based on deep neural networks [WSCR18, RWaL19, ZYM*18, LPX*18a]. Here, we will compare and contrast the different scene and model representations used by these methods.

Finally, the section will conclude with a look toward the future, examining the open problems around scene generative modeling. The current state-of-the-art models have made large strides in their ability to generate *plausible* scenes—i.e. getting the right objects in the right place—but other important aspects of scenes remain understudied and difficult to generate. Such aspects include stylistic compatibility, functionality, and how these can be connected to natural language descriptions of scenes.

4. Presenters

Siddhartha Chaudhuri Siddhartha Chaudhuri is Senior Research Scientist in the Creative Intelligence Lab at Adobe Research, and Assistant Professor of Computer Science and Engineering at IIT Bombay. He obtained his Ph.D. from Stanford University, and his undergraduate degree from IIT Kanpur. He subsequently did postdoctoral research at Stanford and Princeton, and taught for a year at Cornell. Siddhartha's work combines geometric analysis, machine learning, and UI innovation to make sophisticated 3D geometric modeling accessible even to non-expert users. He also studies foundational problems in geometry processing (retrieval, segmentation, correspondences) that arise from this pursuit. His research themes include probabilistic assembly-based modeling [CKGK11, KCKK12, KLM*13, SSK*17], semantic attributes for design [CKGF13, YCHK15], and generative neural networks for 3D structures [LXC*17, ZXC*18, LPX*18b], and other applications of deep learning to 3D geometry processing [KAMC17, HKC*18, MKC18, LAK*18]. He is the original author of the commercial 3D modeling tool Adobe Fuse, and has taught tutorials on data-driven 3D design (SIGGRAPH Asia 2014) and shape "semantics" (ICVGIP 2016).

Kai (Kevin) Xu Kai Xu is an Associate Professor at the School of Computer Science, National University of Defense Technology, where he received his Ph.D. in 2011. He conducted visiting research at Simon Fraser University (2008-2010) and Princeton University (2017-2018). His research interests include geometry processing and geometric modeling, especially on data-driven approaches to the problems in those directions, as well as 3D vision and its robotic applications. He has published over 60 research papers, including 21 SIGGRAPH/TOG papers. He organized a SIGGRAPH Asia course [XKHK17] and a Eurographics STAR tutorial [XKH*16], both on data-driven shape analysis and processing. He is currently serving on the editorial board of Computer Graphics Forum, Computers & Graphics, and The Visual Computer. He also served as paper co-chair of CAD/Graphics 2017 and ICVRV 2017, as well as PC member for several prestigious conferences including

SIGGRAPH, SIGGRAPH Asia, SGP, PG, GMP, etc. Kai has made several major contributions to structure-aware 3D shape analysis and modeling with data-driven approach [XLZ*10, XZZ*11, XZCOC12, VKXZ*13, ZYL*17a], and recently with deep learning methods [LXC*17, LPX*18b, ZXC*18, NLX18].

Daniel Ritchie Daniel Ritchie is an Assistant Professor of Computer Science at Brown University. He received his PhD from Stanford University, advised by Pat Hanrahan and Noah Goodman. His research sits at the intersection of computer graphics and artificial intelligence, where he is particularly interested in data-driven methods for designing, synthesizing, and manipulating visual content. In the area of generative models for structured 3D content, he co-authored the first data-driven method for synthesizing 3D scenes [FRS*12], as well as the first method applying deep learning to scene synthesis [WSCR18]. He has also worked extensively on applying techniques from probabilistic programming to procedural modeling problems [RLGH15, RMGH15, RTHG16], including to learning procedural modeling programs from examples [RJT18]. In related work, he has developed systems for inferring generative graphics programs from unstructured visual inputs such as hand-drawn sketches [ERSLT18].

Hao (Richard) Zhang Hao (Richard) Zhang is a professor in the School of Computing Science at Simon Fraser University, Canada. He obtained his Ph.D. from the Dynamic Graphics Project (DGP), University of Toronto, and M.Math. and B.Math degrees from the University of Waterloo, all in computer science. Richard's research is in computer graphics with special interests in geometric modeling, analysis and synthesis of 3D contents (e.g., shapes and indoor scenes), machine learning (e.g., generative models for 3D shapes), as well as computational design, fabrication, and creativity. He has published more than 120 papers on these topics. Most relevant to the proposed tutorial topic, Richard was one of the co-authors of the first Eurographics STAR on structure-aware shape processing [MWZ*13] and taught SIGGRAPH courses on the topic. With his collaborators, he has made original and impactful contributions to structural analysis and synthesis of 3D shapes and environments including co-analysis [XLZ*10, SvKK*11, VKXZ*13], hierarchical modeling [WXL*11, VKXZ*13, LXC*17], semi-supervised learning [WAvK*12, YZXZ18], topology-varying shape correspondence and modeling [AXZ*15, ALX*14, ZYL*17b], and deep generative models [LXC*17, LPX*18b, ZXC*18].

References

[ALX*14] Alhashim I., Li H., Xu K., Cao J., Ma R., Zhang H.: Topology-varying 3d shape creation via structural blending. *ACM Transactions on Graphics (Special Issue of SIGGRAPH) 33*, 4 (2014), Article 158. 4

[AXZ*15] Alhashim I., Xu K., Zhuang Y., Cao J., Simari P., Zhang H.: Deformation-driven topology-varying 3d shape correspondence. *ACM Transactions on Graphics (Special Issue of SIGGRAPH Asia) 34*, 6 (2015), Article 236. 4

[CKGF13] Chaudhuri S., Kalogerakis E., Giguere S., Funkhouser T. A.: AttribIt: Content creation with semantic attributes. In *Proc. UIST* (2013). 3

[CKGK11] Chaudhuri S., Kalogerakis E., Guibas L., Koltun V.: Probabilistic reasoning for assembly-based 3D modeling. *ACM Transactions on Graphics (Proc. SIGGRAPH) 30*, 4 (2011). 3

[ERSLT18] Ellis K., Ritchie D., Solar-Lezama A., Tenenbaum J. B.: Learning to Infer Graphics Programs from Hand-Drawn Images. In *NIPS 2018* (2018). 4

[FRS*12] Fisher M., Ritchie D., Savva M., Funkhouser T., Hanrahan P.: Example-based Synthesis of 3D Object Arrangements. In *SIGGRAPH Asia 2012* (2012). 1, 3, 4

[Ger09] Germer T. and Schwarz M.: Procedural Arrangement of Furniture for RealâĂŘTime Walkthroughs. *Computer Graphics Forum 28*, 8 (2009). 3

[HKC*18] Huang H., Kalogerakis E., Chaudhuri S., Ceylan D., Kim V. G., Yumer M. E.: Learning local shape descriptors with view-based convolutional neural networks. *ACM Transactions on Graphics 37*, 1 (2018). 3

[KAMC17] Kalogerakis E., Averkiou M., Maji S., Chaudhuri S.: 3D shape segmentation with projective convolutional networks. In *Proc. CVPR (oral)* (2017). 3

[KCKK12] Kalogerakis E., Chaudhuri S., Koller D., Koltun V.: A probabilistic model for component-based shape synthesis. *ACM Transactions on Graphics (Proc. SIGGRAPH) 31*, 4 (2012). 1, 3

[KLM*13] Kim V. G., Li W., Mitra N. J., Chaudhuri S., DiVerdi S., Funkhouser T. A.: Learning part-based templates from large collections of 3D shapes. *ACM Transactions on Graphics (Proc. SIGGRAPH) 32*, 4 (2013). 3

[KLTZ16] Kermani Z. S., Liao Z., Tan P., Zhang H.: Learning 3D Scene Synthesis from Annotated RGB-D Images. In *Eurographics Symposium on Geometry Processing* (2016). 3

[LAK*18] Lin H., Averkiou M., Kalogerakis E., Kovacs B., Ranade S., Kim V. G., Chaudhuri S., Bala K.: Learning material-aware local descriptors for 3D shapes. *Proc. 3DV* (2018). 3

[Lan18] Lange D.: Unity and DeepMind partner to advance AI research. https://blogs.unity3d.com/2018/09/26/unity-and-deepmind-partner-to-advance-ai-research, 2018. Accessed: 2018-10-13. 3

[LPX*18a] Li M., Patil A. G., Xu K., Chaudhuri S., Khan O., Shamir A., Tu C., Chen B., Cohen-Or D., Zhang H.: Grains: Generative recursive autoencoders for indoor scenes. *CoRR arXiv:1807.09193* (2018). 3

[LPX*18b] Li M., Patil A. G., Xu K., Chaudhuri S., Khan O., Shamir A., Tu C., Chen B., Cohen-Or D., Zhang H.: Grains: Generative recursive autoencoders for indoor scenes. *arXiv preprint arXiv:1807.09193* (2018). 3, 4

[LXC*17] Li J., Xu K., Chaudhuri S., Yumer E., Zhang H., Guibas L.: GRASS: Generative recursive autoencoders for shape structures. In *SIGGRAPH* (2017). 1, 3, 4

[MKC18] Muralikrishnan S., Kim V. G., Chaudhuri S.: Tags2Parts: Discovering semantic regions from shape tags. *Proc. CVPR* (2018). 3

[MSL*11] Merrell P., Schkufza E., Li Z., Agrawala M., Koltun V.: Interactive Furniture Layout Using Interior Design Guidelines. In *SIGGRAPH 2011* (2011). 3

[MWZ*13] Mitra N., Wand M., Zhang H., Cohen-Or D., Bokeloh M.: Structure-aware shape processing. In *Eurographics State-of-the-art Report (STAR)* (2013), pp. ???–??? 4

[NLX18] Niu C., Li J., Xu K.: Im2struct: Recovering 3d shape structure from a single rgb image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 4521–4529. 4

[RJT18] Ritchie D., Jobalia S., Thomas A.: Example-based Authoring of Procedural Modeling Programs with Structural and Continuous Variability. In *Eurographics 2018* (2018). 4

[RLGH15] Ritchie D., Lin S., Goodman N. D., Hanrahan P.: Generating design suggestions under tight constraints with gradient-based probabilistic programming. In *Eurographics 2015* (2015). 4

[RMGH15] RITCHIE D., MILDENHALL B., GOODMAN N. D., HANRAHAN P.: Controlling Procedural Modeling Programs with Stochastically-Ordered Sequential Monte Carlo. In *SIGGRAPH 2015* (2015). 4

[RTHG16] RITCHIE D., THOMAS A., HANRAHAN P., GOODMAN N. D.: Neurally-Guided Procedural Models: Amortized Inference for Procedural Graphics Programs using Neural Networks. In *NIPS 2016* (2016). 4

[RWaL19] RITCHIE D., WANG K., AN LIN Y.: Fast and Flexible Indoor Scene Synthesis via Deep Convolutional Generative Models. In *CVPR 2019* (2019). 3

[SSK*17] SUNG M., SU H., KIM V. G., CHAUDHURI S., GUIBAS L.: ComplementMe: Weakly-supervised component suggestion for 3D modeling. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 36*, 6 (2017). 3

[SvKK*11] SIDI O., VAN KAICK O., KLEIMAN Y., ZHANG H., COHEN-OR D.: Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. *SIGGRAPH Asia 30*, 6 (2011), 126:1–126:9. 4

[VKXZ*13] VAN KAICK O., XU K., ZHANG H., WANG Y., SUN S., SHAMIR A., COHEN-OR D.: Co-hierarchical analysis of shape structures. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 69. 4

[WAvK*12] WANG Y., ASAFI S., VAN KAICK O., ZHANG H., COHEN-OR D., CHEN B.: Active co-analysis of a set of shapes. *ACM Transactions on Graphics (Special Issue of SIGGRAPH Asia) 31*, 6 (2012), Article 165. 4

[WSCR18] WANG K., SAVVA M., CHANG A. X., RITCHIE D.: Deep Convolutional Priors for Indoor Scene Synthesis. In *SIGGRAPH 2018* (2018). 1, 3, 4

[WXL*11] WANG Y., XU K., LI J., ZHANG H., SHAMIR A., LIU L., CHENG Z., XIONG Y.: Symmetry hierarchy of man-made objects. *Computer Graphics Forum 30*, 2 (2011), 287–296. 4

[XKH*16] XU K., KIM V. G., HUANG Q., MITRA N., KALOGERAKIS E.: Data-driven shape analysis and processing. In *SIGGRAPH ASIA 2016 Courses* (2016), ACM, p. 4. 3

[XKHK17] XU K., KIM V. G., HUANG Q., KALOGERAKIS E.: Data-driven shape analysis and processing. *Computer Graphics Forum 36*, 1 (2017), 101–132. 3

[XLZ*10] XU K., LI H., ZHANG H., COHEN-OR D., XIONG Y., CHENG Z.-Q.: Style-content separation by anisotropic part scales. *ACM Transactions on Graphics (TOG) 29*, 6 (2010), 184. 4

[XSF02] XU K., STEWART J., FIUME E.: Constraint-Based Automatic Placement for Scene Composition. In *Graphics Interface 2002* (2002). 3

[XZCOC12] XU K., ZHANG H., COHEN-OR D., CHEN B.: Fit and diverse: set evolution for inspiring 3d shape galleries. *ACM Transactions on Graphics (TOG) 31*, 4 (2012), 57. 4

[XZZ*11] XU K., ZHENG H., ZHANG H., COHEN-OR D., LIU L., XIONG Y.: Photo-inspired model-driven 3d object modeling. *ACM Transactions on Graphics (TOG) 30*, 4 (2011), 80. 4

[YCHK15] YUMER M. E., CHAUDHURI S., HODGINS J. K., KARA L. B.: Semantic shape editing using deformation handles. *ACM Transactions on Graphics (Proc. SIGGRAPH) 34*, 4 (2015). 3

[YYT*11] YU L.-F., YEUNG S.-K., TANG C.-K., TERZOPOULOS D., CHAN T. F., OSHER S. J.: Make It Home: Automatic Optimization of Furniture Arrangement. In *SIGGRAPH 2011* (2011). 3

[YZXZ18] YU F., ZHANG Y., XU K., ZHANG H.: Semi-supervised co-analysis of 3d shape styles from projected lines. *ACM Transactions on Graphics 37*, 2 (2018), Article 21. 4

[ZXC*18] ZHU C., XU K., CHAUDHURI S., YI R., ZHANG H.: SCORES: Shape composition with recursive substructure priors. In *SIGGRAPH Asia* (2018). 3, 4

[ZYL*17a] ZHU C., YI R., LIRA W., ALHASHIM I., XU K., ZHANG H.: Deformation-driven shape correspondence via shape recognition. *ACM Transactions on Graphics (TOG) 36*, 4 (2017), 51. 4

[ZYL*17b] ZHU C., YI R., LIRA W., ALHASHIM I., XU K., ZHANG H.: Deformation-driven shape correspondence via shape recognition. *ACM Transactions on Graphics 36*, 4 (2017), Article 51. 4

[ZYM*18] ZHANG Z., YANG Z., MA C., LUO L., HUTH A., VOUGA E., HUANG Q.: Deep generative modeling for scene synthesis via hybrid representations. *CoRR arXiv:1808.02084* (2018). 3