Structure-Aware Learning for 3D Data

MINHYUK SUNG, Stanford University

3D data arising from modeling by designers or scanning with depth sensors has a unique characteristic – it matches the actual physical form of an object as it presents in the real world. Hence, unlike 2D images containing a projected view, it directly enables understanding of how the objects are composed and structured in the physical space. In this paper, I present our novel methodologies on learning the structure of 3D data from a collection and various downstream applications based on them. I first illustrate how a part-based representation of 3D objects, fusing a discrete and combinatorial global structure with continuous local geometry space, can facilitate creating and editing shapes by efficiently exploring in the 3D object space. Such a structural representation is achieved from a co-analysis of 3D shapes, but the co-analysis generally requires to have correspondence information, which is expensive and difficult to obtain. I also introduce how different shapes and information associated with the shapes can be bridged through neural network training without the supervision of correspondences. Lastly, I discuss some model fitting problems discovering structure in 3D shapes and propose a way to combine supervised learning with the best features of classical optimization for more robust estimation.

CCS Concepts: • **Computing methodologies** → **Shape modeling**; **Shape analysis**; **Machine learning approaches**.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Minhyuk Sung. 2019. Structure-Aware Learning for 3D Data. 1, 1 (May 2019), 4 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

My research is concerned with the world of 3D shapes, and especially 3D geometric data processing and analysis. Such data arise both in content creation by designers and artists as well as through sensing the real world by devices such as depth sensors, akin to images. 3D data is, however, distinguished from other data modalities by its unique characteristics: First, 3D data is commonly represented as the *surface* of a 3D object – and not the volume inside. This is innately a sparse and irregular representation, unlike images. In fact, over many decades, a diverse set of representations for 3D data has been developed: e.g., volume-based octrees, BSPs, and CSGs, and surface-based splines, meshes, and point clouds. Second, 3D data is the closest digital representation we have of real objects, aiming to match their actual physical form. Hence, unlike 2D images containing a projected view, 3D data enables to understand the entities as they are composed and structured in the real world. Third, 3D geometries (surfaces) are not only data themselves but also domains wherein other information is defined: e.g., physical attributes such as texture, material and reflectance, and semantic annotations including keypoints and parts associated with labels. Since the information is defined on independent domains, co-analyzing it is not possible without bridging one domain to the other — and this correspondence information is expensive and difficult to obtain. **Due to such unique characteristics, conventional tools for processing or analyzing typical regularly sampled signal data cannot be directly applied to 3D data. The goal of my research is to develop novel methodologies specialized in 3D based upon a profound understanding of its nature.**

Over the course of my Ph.D. studies, I have pursued research aiming to balance between (1) developing fundamental tools, for instance, discovering structures in 3D and relating 3D shapes in a collection, and (2) solving more concrete, tangible problems, such as shape synthesis [2, 4], partial scan completion [2, 3], segmentation [5, 6], keypoint correspondences [5], and geometric primitives fitting [1]. The underlying ideas in my research are encapsulated in the following fundamental themes:

(1) **Part-based Structural 3D Representation**
3D objects have structure and naturally decompose the objects into parts that are related by adjacencies, or symmetries and regularities. The space of 3D objects can thus be efficiently expressed by fusing a discrete, combinatorial global structure with continuous local geometry spaces. I introduce the ways to create new shapes from scratch or from a given partial shape by exploring the joint space with the part-based representations.

(2) **Deep 3D Shape Co-analysis Without Relational Supervision**
Co-analyzing a set of 3D shapes and their associated information generally requires relational information mapping shapes to each other. Such relational information is however coarse, inconsistent, and incomplete in most 3D databases. I propose novel neural network frameworks that can discover relationships among shapes and their associated information without the supervision of them.

(3) **Supervised Learning for Model Estimation**
Several problems on discovering structure in 3D shapes are formulated as optimization problems with objective functions that are non-convex and at times ill-posed. Hence existing solutions suffer from the problem of fine-tuning algorithm parameters for every input shape. I develop a data-driven framework learning hyper-parameters from supervision, combining the best features of classical optimization with learning.

In what follows, I illustrate more details of my work and also elaborate my future goals in each of the above themes.

Author's address: Minhyuk Sung, mhsung@cs.stanford.edu, Stanford University, 318 Campus Drive, Stanford, California, 94305.

XXXX-XXXX/2019/5-ART $15.00
https://doi.org/10.1145/1122445.1122456

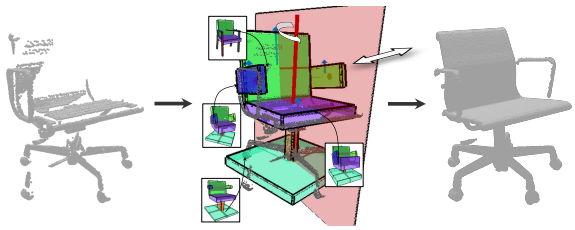

Fig. 1. A part structure and symmetries predicted from the input partial scan data are used to exploit geometry from both symmetry and database sources and complete the missing area [3].

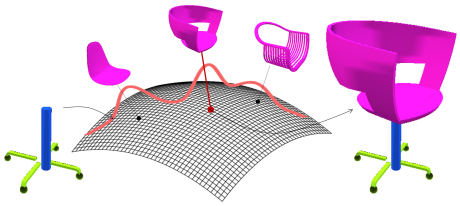

Fig. 2. Multiple candidates of complementary parts are suggested in an iterative process of new shape assembly [4].

2 PART-BASED STRUCTURAL 3D REPRESENTATION

Learning a space of data variation is being extensively studied in diverse domains including image, audio, and language processing, building on the advances of deep learning techniques. Based on the understanding of the data space, one can develop an automated system to synthesize and edit data. Such a system is essential for 3D shapes since creating a new shape requires high expertise. Generative models learning the data space in other data domains typically represent the data distribution with the most fine-grained unit of the data: pixels in images and waveform samples in audio. Such an approach is however inappropriate in 3D shapes, for two reasons. First, most 3D shapes are highly structured and regularized with symmetries. Second, the variation is often disentangled into local parts, e.g., wings and fuselage of airplanes. For these reasons, **it is more effective in 3D shapes to represent the variation with part-based structures associated with global regularities.**

I leverage the part-level structure in the problem of partial scan completion. In the completion, geometries filling the missing areas can come from either the input data itself (based on symmetries) or 3D models in the database. In my research [3], I introduce a system that jointly utilizes both sources by predicting a part structure represented with bounding boxes and global symmetries relating them to each other (Figure 1). This part-level analysis enables the system to discover the symmetry patterns even with the presence of severe missing information and also to borrow geometries from existing models at the part-level. The other application I also worked on is part-based shape synthesis. Note that artists mostly create 3D CAD models as assemblies of components. Hence it is natural to assist the modeling process by suggesting new parts and their locations in each iteration of assemblies. My framework [4] takes a partial shape as input and provides multiple suggestions of complementary parts (Figure 2). This can be either incorporated in an interactive modeling system to assist the users or adapted to a fully automatic system synthesizing new shapes. In the literature of the assembly-based shape synthesis, this is the first work that learns the complementary relationships without manually labeling the parts and thus enables to utilize a large-scale database of raw CAD models with minimal preprocessing.

The potential of such part-based shape analysis is in producing high-quality 3D models suitable for graphics applications by reusing local fine geometries in exemplars while varying global combinatorial structures. The research in this direction can be further expanded by associating each part in the structure with deformation operations such as scaling and rotation. Also, data in various modalities including images, sketches, and languages can be used to supervise the part assemblies. Ultimately, I wish to solve a reverse-engineering problem of converting a raw 3D geometry to a CAD form containing both the global structural information and per-part parameterized shapes.

3 DEEP 3D SHAPE CO-ANALYSIS WITHOUT RELATIONAL SUPERVISION

One of the main factors in the success of deep learning is the availability of large data collections. In 3D, there are also many large repositories of CAD models, indoor/outdoor scenes, human/animal character shapes, and medical images, decorated with metadata. In many problems of 3D analysis, however, it is not sufficient to have a large collection of data but what is needed is to see how the data are *related* to each other, particularly in lower levels such as for parts or points of objects. Such relational information is often annotated in 3D shapes with labels in the databases, but typically many issues arise when exploiting them. For example, **labels can be inconsistent across models** — in terms of both syntax and semantics; e.g., in a bike, 'seat' or 'saddle' label can be interchangeably used for the same part, and 'seat' part may or may not include 'seat post'. **Also, annotations can be too coarse for some applications;** e.g., a 'rim' and 'spokes' in a 'wheel' may need to be distinguished. **There can even be missing annotations.** Such difficulties motivate us to develop neural network frameworks that analyze a set of 3D shapes without relational supervision.

In the assembly-based shape synthesis introduced in the previous section, the challenge is in describing which parts can coexist and contact each other in an object. This part compatibility is determined by both functionality and style of parts, which are not clearly classified and thus cannot be adequately expressed with labels. My neural-network-based framework [4] learning the compatibility does not rely on any label on parts but leverages geometry and contact information. It produces latent codes of parts encoding the compatibility, which also allows comparing their functional and style-wise roles in the objects and relating them across different 3D models. In my following work [2], the idea of learning the part relationships is further expanded with a more principled approach. Here, I introduce a network encoding the sets of compatible parts with latent codes equipped with algebraic set operations (Figure 3). Then, when defining interchangeable parts as ones sharing the same set of compatible parts, they are easily retrieved with the simple set operations in the latent space.

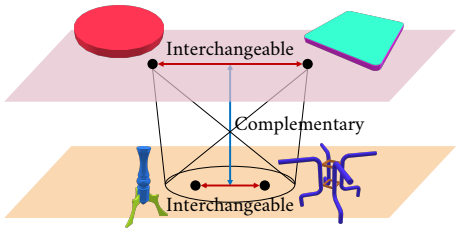

Fig. 3. The embedding of parts enables comparing sets of complementary parts with an algebraic operation and retrieving interchangeable parts based on the set comparison [2].

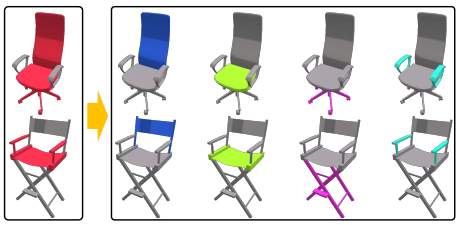

Fig. 4. My framework [5] takes a set of *unrelated* functions on shapes as input (e.g. subsets of parts and keypoints) and predicts *synchronized* dictionaries of atomic functions (e.g. atomic parts).

These works also led me to consider a more general problem of relating any type of per-point functions defined on different 3D shapes. **The fundamental difficulty of analyzing the 3D shape information in a collection is that they are not defined on a common parameterizing domain like a regular 2D lattice for images.** Building a canonical space mapping to all 3D shapes is only available with dense correspondences and is also sometimes infeasible due to heterogeneous groups with only a partial common structure. In my research [5], I instead leverage canonicalization that appears in the training of a neural network learning shape-dependent basis functions. The output dictionary of atomic functions for each shape composes the input functions as linear combinations of the dictionary elements. What is surprising is that the network orders the atoms consistently across the different shapes, thus effectively learning shape correspondences, even though it was given no supervision towards this end (Figure 4). This framework has been successfully adapted to various applications such as part/keypoint matching, instance segmentation in 3D scenes, and functional basis synchronization with arbitrary smooth input functions.

This line of my work opens up several future research problems. For example, in a case when each object has a *hierarchy* of parts or per-point functions, it becomes a *vertical* network of parts in an object while the relationships among parts across objects build a *horizontal* network. Then one can ask how to jointly analyze these two types of networks, which is non-trivial. There can also be multiple horizontal networks layered for each level of vertical networks. Moreover, my previous work considers the case when the relationships are fully unsupervised, but a partially supervised case can also be considered in future research.

4 SUPERVISED LEARNING FOR MODEL ESTIMATION

In geometry processing, there are many classic problems generally formulated as optimization problems, such as alignment, deformation, parameterization, and symmetry detection. The large body of literature in these problems can make one believe that these are fully solved. **However, many existing algorithms suffer from the difficulty of fine-tuning algorithm parameters for each of the input shapes.** This is mostly due to the non-convexity of the problem — even sometimes it needs to find multiple local minima corresponding to multiple solutions — and ill-posedness meaning that the error is sensitive to the change in the model space. This difficulty prevents processing a large volume of shapes without significant user controls. In practice, such a case of iterating the same process mostly happens for a specific kind of input data, such as scans of objects in a particular category and with a known noise pattern. Thus, in my research, I develop a data-aware framework leveraging deep learning techniques to learn the hyper-parameters from supervision.

A problem I have worked on is fitting geometric primitives to point clouds [1] (Figure 5); primitives include plane, sphere, cylinder, and cone. This is an essential process for converting a raw 3D geometric data to a parametric shape. The problem, however, becomes complex in optimization when the number of primitives and their corresponding areas on the shape are unknown. A threshold of fitting errors also needs to be very carefully tuned when noises are present. In my work, a neural network is proposed to understand the data through supervision. **The key in the neural network design is to find a proper space of outputs.** Directly regressing in primitive parameter spaces is not appropriate since a subtle difference in this space can lead to a significant fitting error. Instead, my framework predicts point cloud segments corresponding to each primitive and their types, which derive the final primitive parameters with a closed-form expression. I believe that such supervised learning ideas can be further applied to many of the other 3D geometry processing problems, where a deep network is used to learn structures and noise characteristics of a particular type of input data and in turn set parameters and guide a more classical optimization algorithm to complete the task.

5 FUTURE VISION

The core objective of my research is to facilitate processing and synthesis of 3D data via discovering the underlying structure from a data collection. My commitment to this direction has enabled new capabilities of automating or simplifying various geometry processing tasks, which were only available with high expertise or a vast effort of the users. My research is also dedicated to overcoming the hurdle of big data curation and preparation in the structure learning procedures and enabling them without direct supervision. I believe that such efforts can significantly change the landscape of possibilities in both future research and industrial usages.

In the following, in addition to the future work of each theme illustrated in the previous sections, I describe more future directions that I would like to explore:

(1) **3D Scene Understanding with Object Relation Priors**
Structures present not only in a single object but also in a

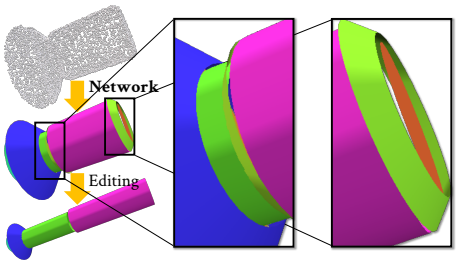

Fig. 5. A set of geometric primitives predicted to be fitted to the input point cloud enables downstream applications such as shape editing [1].

scene, including a set of objects. Typical scenes contain repeating objects of the same model (e.g., chairs/tables in an office), and the spatial arrangement of the objects is determined mainly by their functional interactions. Such relationships among objects have been extensively utilized in 3D scene synthesis. However, the other way around, parsing raw 3D scenes with the object relation priors, remains almost unexplored. I plan to extend my works on 3D scene analysis [5, 6] by learning and leveraging the structural scene priors.

(2) **Functionality Learning from Geometric and Spatial Structures**
In 3D data, the geometry of objects and their spatial arrangement in scenes often reflect their functionalities, as expressed in a maxim "form follows function". The functionality of objects typically refers to the way of interacting with humans or other objects, which is valuable knowledge in robotics and graphics applications. Directly acquiring such information, however, requires to conduct expensive simulations of interactions in real or virtual worlds. I am interested in facilitating a data-driven analysis of the functionalities by discovering correlations between them and geometric/spatial structures.

(3) **Discovering Intrinsic Structures from Deformable Shapes**
Like human-made objects and scenes, most deformable shapes such as human and animal bodies also contain structures of part adjacencies and symmetries, but these structures are not represented in an extrinsic way in the Euclidean space. The variation of such shapes are also constrained in multiple ways: e.g., the skin stretch of human/animal bodies and the angle of joint movement are limited in specific ranges. I aim to expand my research scope to the deformable shapes and discover the intrinsic structures.

(4) **Deep Learning 3D Data Variations with Interpretable Transformations**
Unlike other data modalities, 3D data often relate to each other via explicit and intuitive transformations: e.g., rigid transformations of parts in articulated objects and pose space deformations of human/animal shapes. For applications guiding users to explore and manipulate 3D data, it is required to represent plausible variations with such transformations understanding shape structures. General generative models, however, are not capable of specifying the type of transformations in representing the data distribution. I am interested in developing deep 3D generative models that enable user controls in 3D data exploration and manipulation with desired transformations.

REFERENCES

[1] Lingxiao Li*, **Minhyuk Sung***, Anastasia Dubrovina, Li Yi, and Leonidas Guibas. 2019. Supervised Fitting of Geometric Primitives to 3D Point Clouds. In *CVPR*. (* equal contribution).
[2] **Minhyuk Sung**, Anastasia Dubrovina, Vladimir G. Kim, and Leonidas Guibas. 2018. Learning Fuzzy Set Representations of Partial Shapes on Dual Embedding Spaces. In *Symposium on Geometry Processing (SGP)*.
[3] **Minhyuk Sung**, Vladimir G. Kim, Roland Angst, and Leonidas Guibas. 2015. Data-driven Structural Priors for Shape Completion. In *SIGGRAPH Asia*.
[4] **Minhyuk Sung**, Hao Su, Vladimir G. Kim, Siddhartha Chaudhuri, and Leonidas Guibas. 2017. ComplementMe: Weakly-Supervised Component Suggestions for 3D Modeling. In *SIGGRAPH Asia*.
[5] **Minhyuk Sung**, Hao Su, Ronald Yu, and Leonidas Guibas. 2018. Deep Functional Dictionaries: Learning Consistent Semantic Structures on 3D Models from Functions. In *NeurIPS*.
[6] Li Yi, Wang Zhao, He Wang, **Minhyuk Sung**, and Leonidas Guibas. 2019. GSPN: Generative Shape Proposal Network for 3D Instance Segmentation in Point Cloud. In *CVPR*.