Neurosymbolic Generation of 3D Animal Shapes through Semantic Controls

Vivian Liu[a], Lydia Chilton[a]

[a] *Columbia University, New York, USA*

Abstract

While there have been many advancements in generative models for 3D design, there has been a limited amount of user interface work in this co-creation domain. The user interface controls and interaction paradigms emerging in this field tend to be unintuitive and hard to standardize, as they are often based upon complicated work related to latent space disentanglement, dimensionality reduction, and other bespoke computational techniques.

We demo a user interface that provides intuitive controls for the generation of basic 3D animal shapes. These controls, a set of *semantic sliders*, map to simple and universal operations such as scale and rotation. By adjusting these parameters over animal limbs, users can semantically guide generative models towards their goals, optimizing the mapping between AI action and user intention.

Our user interface operates over a generative model that implements Wei et. al.'s semi-supervised architecture for learning semantically meaningful embeddings [1]. To train it, we collected artist data and generated synthetic data by authoring a parametric animal shape generator. This generator produces low-fidelity, abstracted animal shapes we refer to as *metashapes*.

Our system is an instance of a *neurosymbolic* generative system, which is when the generative system learns dually from data as well as from symbolic, algorithmic constraints. We conclude with an analysis of the benefits and drawbacks of neurosymbolic generation for 3D animal shapes and the utility of metashapes for user control over AI.

Keywords

neurosymbolic, generative models, semantic controls, human-AI interaction, 3D user interfaces

1. Introduction

1.1. Related Work

Prior work has shown that generative networks such as variational autoencoders and GANs can learn latent spaces and generatively produce 3D shapes [2]. However, the dimensions of these latent spaces are often highly entangled and too hyperdimensional to be human interpretable. Reverse engineering meaningful signals from latent spaces has become an active area of research. Techniques such as beta-variational autoencoders, InfoGAN [3], and latent space factorization have been developed for the purposes of disentangling the latent space. However, all these methods work to varying degrees of success and tend to be contingent upon the curation and training of large datasets.

Recently, Wei et. al. (2020) proposed an architecture that challenges latent spaces with a semi-supervised model that learns a semantic space. A semantic space can generate instances of 3D object classes after being trained jointly on synthetic and real data; within this space, users can carry out 3D shape editing operations [1].

The demo we present implements the ar-

Joint Proceedings of the ACM IUI 2021 Workshops, April 13–17, 2021, College Station, USA
vivian@cs.columbia.edu (V. Liu); chilton@cs.columbia.edu (L. Chilton)

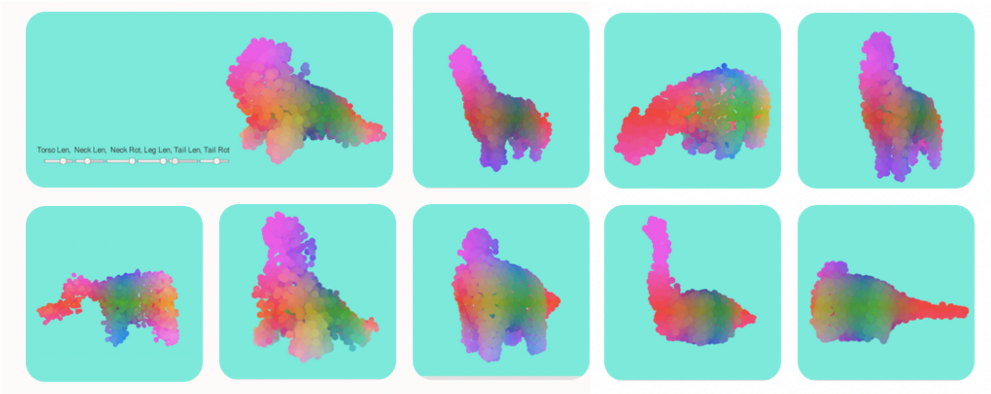

Figure 1: Our user interface generates animal metashapes, which are generic low-fidelity animal shapes. Above are nine animal metashapes arrived at from our user interface. In the top left rectangle, we picture six of the "semantic sliders" used to generate these shapes. These sliders give users control over the following semantically meaningful parameters such as *torso length, neck length, neck rotation, tail length, tail rotation, and leg length.* These parameters operate on the shape outputs using intuitive mental operations like scale and rotation.

chitecture and methods proposed by Wei et. al on the domain of quadruped animals. Wei et. al. demonstrate the success of their method using state of the art academic datasets on well-established generative task domains such as chairs, airplanes, and human bodies. We choose to focus on the domain of animals, because animals are one of the most common classes of 3D assets created. They display a high variance in their shapes, which make them far harder to statistically parameterize than easier shapes like human silhouettes. In spite of this variance, animals share structural similarities that humans can intuitively characterize. For example, we can generalize quadrupeds to be four-legged animals with a head, neck, and a tail, and teach this abstraction to our system. We refer to this abstraction as a *metashape*: a generic, low-fidelity shape that can abstractly characterize a class of 3D objects. We utilize these metashapes and the aforementioned architecture for our neurosymbolic generative system.

2. System

Our system utilizes point clouds as our 3D shape representations, as neural networks have shown success on a number of 3D tasks involving point clouds from instance segmentation [4] to shape editing to interpolation [2].

To create a dataset of animals, we first web-scraped 228 3D mesh assets made public on Sketchfab by artists [5] and sampled point clouds from these assets. Using open3d, a Python package for 3D graphics, we rotated, normalized, and scaled our data to fit within a unit sphere, center around the 3D origin, and face the same direction.

2.1. Metashape Generator

To create a synthetic dataset of 20,000 animal metashapes in accordance with Wei et. al's architecture, we utilized Blender's module for Python scripting to spawn metaballs that coagulate into basic animal shapes. Metaballs are 3D primitives common to computer graphics software that can additively or subtractively react to one another to form organic-looking shapes. Our metashapes were symbolically parameterized by vector directions, limb lengths, and limb rotations. Our inspiration for this approach comes from an idea long theorized by cognitive science that 3D shapes can be decomposed into more basic primitives known as geons [6].

We created two versions of the generative model for this demo that work with one consistent user interface. The first is parameterized by six semantic axes, the second by twenty one semantic axes. These parameters corresponded to length, width, height, rotation, radius, position, spacing and other labels that characterized the primitives corresponding to parts of the animal metashape. The exact labels can be found in the appendix.

These labels supervise the learning of the semantic space and teach the model 3D operations such as scale and rotation over specific parts. While generating a synthetic dataset from a template is a source of inductive bias, we attempted to mitigate this by informing our template with results from Superquadrics, a recent part-segmentation model that was applied successfully to animal meshes [7].

3. Results

We present a user interface designed in the Unity3D game engine, which abstracts over the generative model and allows users to interact with it in real-time using semantic sliders. These sliders map to the original axes of the semantic supervision and offer explainability for the model's actions. We additionally demo preliminary features for direct manipulation, camera view movement, user history, generative AI history, and transparency. The real-time nature of the interactions, the explainability of the model through the semantic sliders, and the concept of memory are all in accordance with best practices proposed by Llano et. al. for explainable computational creativity systems [8].

Author interactions with the system efficiently exposed the generative model's abilities and shortcomings. It was moderately successful at learning from its semantic supervision and able to produce transformations in scale and rotation over *torso length, neck length, and tail length.* Exploration through the user interface produced the varied results as pictured in Figure 1. However, the model, in its first iteration with six semantic parameters, failed to completely disentangle the semantic space. The model showed an on and off ability to control parameters such as *neck rotation, tail rotation, leg length, and tail length.* By on and off, we mean that while in certain clusters of parameters the 3D transformations over rotation and scale were accurate, in other clusters the sliders produced suboptimal behavior. For example, one specific problem was that the model sometimes mixed up the posterior extrusion of the neck with the anterior extrusion of the tail. Another problem was that the model seemed unable to capture the extreme ends of our synthetic training data (i.e. long legs). More examples of malformed edits are illustrated and captioned in Figure 2.

4. Discussion

In this section, we discuss the following learnings from this demo.

4.0.1. Efficient design space exploration

Though many generative models can now be interacted with in realtime [9, 10], it is often still intractable to completely visualize the design space these models sample over. However, our design space is low-dimensional; six-dimensional in one version, 21-dimensional

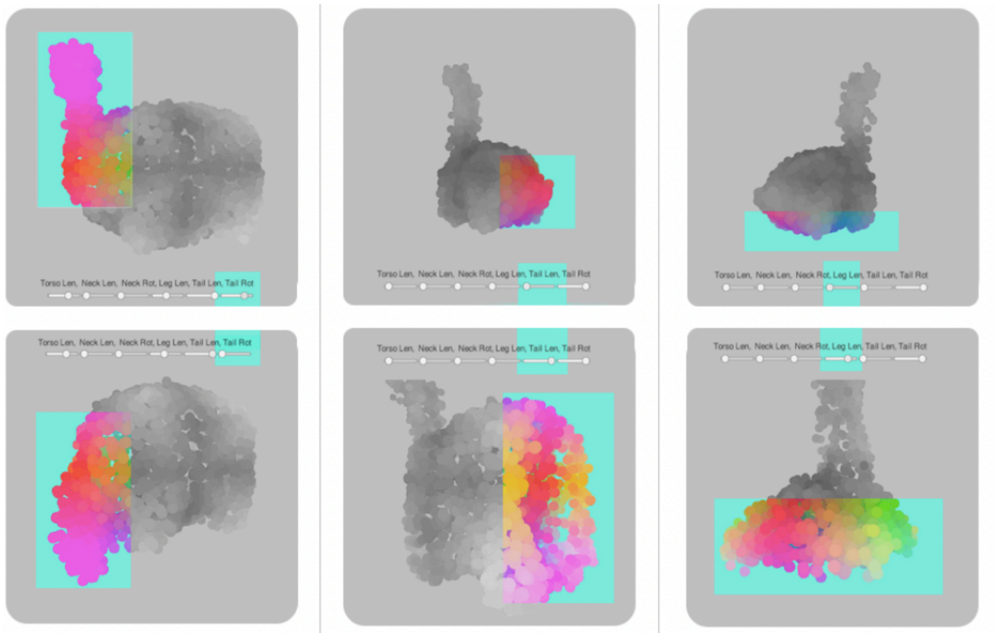

Figure 2: Cases of bad output animal metashapes for three semantic axes. Left: edits on *tail rotation* result in changes of *neck rotation*. The model mixes up posterior extrusion of neck with the anterior extrusion of the tail. Center: Editing *tail length* leads to a "negative" *tail length*, which appears as a posterior indent in the animal shape. Right: Maximizing the *leg length* parameter leads to an outward, noisy extension of legs. The affected areas in the images and parameters are saturated and highlighted respectively.

in another. Users have access to the entire design space and can traverse through it within minutes. They can find best and worst case outputs within seconds. The efficient exploration that we allow implements the following principle established by Gero et. al.: interactions between humans and AI are improved when humans can efficiently explore and understand the global knowledge distributions underlying generative models [11].

4.0.2. Using metashapes as abstractions between users and AI

We argue that metashapes are an ideal abstraction between users and AI. In our system, metashapes gave users ways to operate over the design space with universal concepts like scale and rotation, mental operations that are intuitive and shared by everyone. The semantic meaning attached to each slider optimized the translation of user intention into AI output. While metashapes are 3D concepts, they translated well to a user interface that would be intuitive even to non-technical end users. The minimal interface is a good counterexample to the many heavier generative model user interfaces which encourage users to interact with lower-level complexities like data distributions [12] and hyperparameters [10].

4.0.3. Challenges for neurosymbolic generation

We acknowledge that there are limitations to neurosymbolic generation with metashapes. One of the most significant challenges is finding the right metashape abstraction to encapsulate a class of 3D shapes. While methods to find these abstractions do exist [13, 14, 7], it is hard to evaluate the correctness of their abstractions. Furthermore, these methods do not often lend to intuitive user interface controls and metaphors.

Additionally, by incorporating symbolic constraints from a template, we put a concrete number on the degrees of freedom users may access. While we set this number to generate a tractable low-dimensional design space, a number that is too small limits users and a number that is too big can overwhelm them. Even if we gave users an exorbitant number of degrees of freedom, users could still very well want to go beyond what is offered and define their own axes.

The neurosymbolic generation in this system could also benefit from understanding the interplay between real artist and synthetic data. For example, the synthetic dataset behind our first iteration of the generative model created straight legs that pointed downwards. However, the generative model altered the parameter of leg length by extending them not only downwards but also *outwards*, reproducing patterns present in the shapes of amphibians and reptiles. We optimistically believe that the model was able to generalize some natural variance and establish some correlation between real and synthetic data. However, there were certainly corner cases in which the output was distinctly "real" or "synthetic". More work could be done to figure out how to investigate and mitigate overfitting, as the generative model was built on a highly asymmetric composition of datasets.

5. Conclusion

We present a demo of a neurosymbolic generative system that allows users to create 3D animal shapes with semantically meaningful controls. Additionally, we illustrate how symbolically generated metashapes can be a useful abstraction going forward for human-AI interaction.

Acknowledgments

Vivian Liu is supported by a National Science Foundation Graduate Research Fellowship. The authors thank Panos Achlioptas (@optas) for open-sourcing the loss function used for model training, Felix Herbst (@herbst) for contributing to the coloring function on the point clouds, and David O'Reilly for open sourcing a large pack of 3D animal assets.

References

[1] F. Wei, E. Sizikova, A. Sud, S. Rusinkiewicz, T. Funkhouser, Learning to infer semantic parameters for 3d shape editing, 2020. arXiv:2011.04755.

[2] P. Achlioptas, O. Diamanti, I. Mitliagkas, L. Guibas, Learning representations and generative models for 3d point clouds, 2018. arXiv:1707.02392.

[3] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, 2016. arXiv:1606.03657.

[4] C. R. Qi, H. Su, K. Mo, L. J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, 2017. arXiv:1612.00593.

[5] D. O'Reilly, Animals, 2020. URL: http://www.davidoreilly.com/library.
[6] I. Biederman, Recognition-by-components: a theory of human image understanding., Psychological review 94 2 (1987) 115–147.
[7] D. Paschalidou, A. O. Ulusoy, A. Geiger, Superquadrics revisited: Learning 3d shape parsing beyond cuboids, in: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.
[8] M. T. Llano, M. d'Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease, S. Colton, Explainable computational creativity, in: ICCC, 2020.
[9] A. Ghosh, R. Zhang, P. K. Dokania, O. Wang, A. A. Efros, P. H. S. Torr, E. Shechtman, Interactive sketch fill: Multiclass sketch-to-image translation, 2019. arXiv:1909.11081.
[10] Runway AI, Inc., Runwayml, ???? URL: https://runwayml.com.
[11] K. I. Gero, Z. Ashktorab, C. Dugan, Q. Pan, J. Johnson, W. Geyer, M. Ruiz, S. Miller, D. R. Millen, M. Campbell, S. Kumaravel, W. Zhang, Mental Models of AI Agents in a Cooperative Game Setting, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–12. URL: https://doi.org/10.1145/3313831.3376316.
[12] J. Matejka, M. Glueck, E. Bradner, A. Hashemi, T. Grossman, G. Fitzmaurice, Dream Lens : Exploration and Visualization of Large-Scale Generative Design Datasets (2018) 1–12.
[13] K. Genova, F. Cole, A. Sud, A. Sarna, T. Funkhouser, Local deep implicit functions for 3d shape, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4857–4866.
[14] R. Huang, P. Achlioptas, L. Guibas, M. Ovsjanikov, Limit shapes – a tool for understanding shape differences and variability in 3d model collections, Computer Graphics Forum 38 (2019) 187–202. doi:10.1111/cgf.13799.

A. Appendix

A.0.1. Generative model: Iteration 1

Iteration 1 with 6 semantic parameters consisted of the following parameters: torso length, neck length, neck rotation, tail rotation, leg length, tail length.

A.0.2. Generative model: Iteration 2

Iteration 2 with 21 semantic parameters from the following set of parameters: torso length, (front) torso width , (front) torso height, (back) torso width, (back) torso height, a choice between head type 1 (which emphasizes ear variation) and head type 2 (which emphasizes jaw variation), head size, head feature (ear / jaw) prominence, mouth angle, neck length, neck rotation, neck size, leg length, position of front legs, position of back legs, leg gap, leg angle, tail length, tail rotation, tail radius, tail variance, a choice between a tail that increases in width or decreases, and leg radius.