11-685 Project Proposal - Team 52

David Chen
School of Architecture
Carnegie Mellon University
Pittsburgh, PA 15213
davidch2@andrew.cmu.edu

Chia Hui Yen
School of Architecture
Carnegie Mellon University
Pittsburgh, PA 15213
huiyenc@andrew.cmu.edu

Graham Felton
School of Architecture
Carnegie Mellon University
Pittsburgh, PA 15213
gtf@andrew.cmu.edu

Karthick Raja
School of Architecture
Carnegie Mellon University
Pittsburgh, PA 15213
kbangaru@andrew.cmu.edu

Abstract

This project will explore the generation of novel architectural styles through a 2D-to-3D deep learning models. The model will be trained on an architectural dataset, that encapsulates architectural forms. By manipulating this model, we can generate novel designs, interpolate between existing styles, and synthesize new architectural forms. The generated forms can be used for design inspiration and early-stage ideation as well as serve as a method for generating design iterations.

1 Overview and Context

1.1 Motivation

Iterative design techniques allow architects to explore possible design options without having to commit to a singular prototype, however current design technologies are limited in their ability to provide novel iterations based on prototypes.

1.2 Objective

We propose a new deep learning architecture based on traditional variational autoencoder models which seeks to allow for rapid prototyping by exploring a latent space of possible design iterations given a singular input. By encoding information from a dataset trained on specific architectural typologies, our model allows designers to explore a range of possible design iterations.

1.3 Literature Review

The papers below are primary sources while a more exhaustive list of our secondary and tertiary sources can be found in **Section 4.1**.

1 Rhee, J. Deeprise: A Novel Method to Model and Design a High-rise Building Form through Deep Neural Network.

2 Huang, H., Li, Z., He, R., Sun, Z., and Tan, T. 2018. IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis. `http://arxiv.org/abs/1807.06358`.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

2 Methodology

2.1 Model Description

The model is a variational autoencoder (VAE) that learns a mapping between 2D floor plan images and corresponding 3D geometry. The model enables exploration of design styles through the latent space.

2.2 Dataset

Eight possible datasets were identified from our literature review, listed below in the table column labelled **"References"**.

Table 1: Possible Architectural Datasets

Datasets	**Trade-offs**	**Uses**	**Data Types**	**Suitable Models**	**References**
ShapeNet Core Dataset	Large-scale, but may lack real-world variety; voxelized data can be memory intensive.	3D object generation, shape classification, auto-encoding, implicit decoders, generative models.	Voxel, Meshes	3D CNNs, VAE, GANs, Implicit Decoders	PointNet, Learning Implicit Fields for Generative Shape Modeling
Redwood Dataset	Real-world objects, but dominated by single views, small baselines.	3D reconstruction, can supplement other datasets.	Images	CNNs, GANs	NeRS
Stanford Products Dataset	Real-world objects, but dominated by single views, small baselines.	3D reconstruction, can supplement other datasets.	Images	CNNs, GANs	NeRS
Multi-view Marketplace Cars (MVMC)	Focused on cars, but contains diverse illumination and viewpoints.	Multi-view reconstruction, view synthesis, inverse rendering, 3D object generation.	Images	CNNs, NeRS, GANs	NeRS
OSM 3D High-Rise Buildings	May be noisy or incomplete due to crowd-sourced nature; requires conversion from 2D data.	Generative modeling of buildings, urban morphology studies, style-based generation.	2D images, 3D models	CNNs, VAE, GANs	Archi-Learning
Text Datasets for Style Descriptions	May require custom creation for specific styles; difficult to align with visual features.	Providing text prompts for conditional 3D object generation.	Text	RNNs, Transformers, Text Encoders	N/A
Facade and Roof Datasets	Specific to building parts, but useful for style transfer; requires element labels.	GAN training for generating building facades, style transfer on 3D models.	Images	GANs, CNNs	FrankenGAN
ImageNet (ILSVRC)	Not specific to 3D but useful for pre-training; 2D dataset for feature transfer.	Pre-training models for generic image features, 3D object style transfer.	Images	CNNs, Transfer Learning Models	Three Decades of ML in CAAD, Urban Morphology Meets ML
"Quick, Draw" Dataset	Focus on sketches, useful for abstract and biologically-inspired designs; needs 3D object pairing.	Generates a variety of object styles, especially abstract and biologically-inspired.	Sketches	GANs, VAEs	N/A
Paired Text and 3D Model Data	Does not currently exist; requires custom creation.	Training models for 3D object generation from text.	Text, Meshes, Point Clouds, Voxel	GANs, VAEs, Conditional GANs	N/A
Paired 2D Image and 3D Model Data	Can be difficult to align data accurately; does not currently exist; requires custom creation.	Style transfer between 2D images and 3D models.	Images, Meshes, Point Clouds, Voxel	GANs, VAEs, Conditional GANs	N/A
LIFULL HOME's Database	Contains floorplans but lacks bubble diagrams; limited to residential architecture; may lack detailed 3D models.	House layout generation from bubble diagrams; useful for graph-constrained generative models.	Images, Graphs	GANs, Graph Neural Networks (GNNs)	House-GAN

2.3 Evaluation Metrics

The rows in the table below list possible evaluation metrics, loss function, baselines, and experiments while the columns define from which reference we have gathered the information.

Category	Hasev et al.	Deeprise (Rhee et al.)	Kelly et al.
Evaluation Metric	• Reconstruction accuracy (visual + clustering) • Clustering validity in latent space • Geographic mapping validation	• Accuracy of generated high-rise building forms • Latent space structure analysis, t-SNE & DBSCAN clustering for high-rise types	• Perceptual user study • Style consistency across generated samples • Comparison against ground truth
Loss Function	• Reconstruction Loss (L1 loss) • KL Divergence (Latent Space Regularization)	• IntroVAE: Combines GAN-style adversarial loss with VAE-style latent space regularization • L1 loss for pixel similarity • KL Divergence for structured latent space	• Adversarial Loss (GAN) • L1 Loss (Pixel Similarity) • KL-Divergence (Style Vector Control) • Latent Reconstruction Loss (Prevents Mode Collapse)
Baseline	• Manual classification from existing architectural studies • Alternative clustering methods (DBSCAN, K-Means) • Ground truth comparison using historical architectural typologies	• Comparison between color-based encoding and monochrome representation for high-rise buildings • Evaluating latent space clustering against expert classification	• Comparison with Pix2Pix & BicycleGAN • Real vs. Generated evaluation through perceptual study
Implemented Extensions/Experiments	• Latent space interpolation • Multi-scale clustering (primary, secondary, tertiary) • Mapping to geographical context	• Experimented with RGB height encoding vs. monochrome encoding • Tested model across different floor sampling strategies • Evaluated latent space clustering on 4,596 buildings	• Style-conditioned synthesis • Multi-GAN architecture for improved detail • GAN-based super-resolution for fine-grained feature generation

2.4 Baseline

Our possible evaluation method includes both quantitative and qualitative assessments of the following models:

- **Baseline Model for Comparisons:**
  - **Pix2Pix** - Evaluates architectural style feature preservation.
  - **BicycleGAN** - Tests style diversity and latent space regularization.
  - **Clustering Methods:**
    - **DBSCAN** - Detects natural clusters without predefined categories.
    - **K-Means** - Evaluates separability between different styles.
- **Perceptual Study and Expert Classification:**
  - Compare clustering results against expert-annotated architectural typologies.

- Validate with historical architectural typologies as ground truth.
- **Latent Space Analysis:**
  - **t-SNE / UMAP** - Visualize latent space structure.
  - **Silhouette Score / Davies-Bouldin Index** - Measure cluster compactness and separation.

3 Anticipaated Challenges

Dataset curation and augmentation is expected to be an area of difficulty, along with evaluation of chosen network architecture. Evaluation of our model versus the models of our reference literature is additionally anticipated to be difficult with current computational resources.

4 Administrative Details

4.1 Bibliography

References

Memo Akten, Rebecca Fiebrink, and Mick Grierson. Learning to see: you are what you see. In *ACM SIGGRAPH 2019 Art Gallery*, pages 1–6, Los Angeles California, July 2019. ACM. ISBN 978-1-4503-6311-2. doi: 10.1145/3306211.3320143. URL `https://dl.acm.org/doi/10.1145/3306211.3320143`.

Imdat As, Siddharth Pal, and Prithwish Basu. Artificial intelligence in architecture: Generating conceptual design via deep learning. *International Journal of Architectural Computing*, 16(4): 306–327, December 2018. ISSN 1478-0771, 2048-3988. doi: 10.1177/1478077118800982. URL `https://journals.sagepub.com/doi/10.1177/1478077118800982`.

Shanaka Kristombu Baduge, Sadeep Thilakarathna, Jude Shalitha Perera, Mehrdad Arashpour, Pejman Sharafi, Bertrand Teodosio, Ankit Shringi, and Priyan Mendis. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141:104440, September 2022. ISSN 09265805. doi: 10.1016/j.autcon.2022.104440. URL `https://linkinghub.elsevier.com/retrieve/pii/S0926580522003132`.

Chenyi Cai, Zifeng Guo, Baizhou Zhang, Xiao Wang, Biao Li, and Peng Tang. Urban Morphological Feature Extraction and Multi-Dimensional Similarity Analysis Based on Deep Learning Approaches. *Sustainability*, 13(12):6859, June 2021. ISSN 2071-1050. doi: 10.3390/su13126859. URL `https://www.mdpi.com/2071-1050/13/12/6859`.

Zhiqin Chen and Hao Zhang. Learning Implicit Fields for Generative Shape Modeling, September 2019. URL `http://arxiv.org/abs/1812.02822`. arXiv:1812.02822 [cs].

Hang Chu, Daiqing Li, David Acuna, Amlan Kar, Maria Shugrina, Xinkai Wei, Ming-Yu Liu, Antonio Torralba, and Sanja Fidler. Neural Turtle Graphics for Modeling City Road Layouts, October 2019. URL `http://arxiv.org/abs/1910.02055`. arXiv:1910.02055 [cs].

Renaud Danhaive and Caitlin T. Mueller. Design subspace learning: Structural design space exploration using performance-conditioned generative modeling. *Automation in Construction*, 127:103664, July 2021. ISSN 09265805. doi: 10.1016/j.autcon.2021.103664. URL `https://linkinghub.elsevier.com/retrieve/pii/S0926580521001151`.

Huaibo Huang, Zhihang Li, Ran He, Zhenan Sun, and Tieniu Tan. IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis, October 2018. URL `http://arxiv.org/abs/1807.06358`. arXiv:1807.06358 [cs].

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks, November 2018. URL `http://arxiv.org/abs/1611.07004`. arXiv:1611.07004 [cs].

Tom Kelly, Paul Guerrero, Anthony Steed, Peter Wonka, and Niloy J. Mitra. FrankenGAN: Guided Detail Synthesis for Building Mass-Models Using Style-Synchonized GANs. *ACM Transactions on Graphics*, 37(6):1–14, December 2018. ISSN 0730-0301, 1557-7368. doi: 10.1145/3272127.3275065. URL `http://arxiv.org/abs/1806.07179`. arXiv:1806.07179 [cs].

Immanuel Koh. Architectural sampling: three possible preconditions for machine learning architectural forms. *Architectural Intelligence*, 2(1):7, March 2023. ISSN 2731-6726. doi: 10.1007/s44223-023-00024-1. URL `https://link.springer.com/10.1007/s44223-023-00024-1`.

Chongyang Ma, Nicholas Vining, Sylvain Lefebvre, and Alla Sheffer. Game level layout from design specification. *Computer Graphics Forum*, 33(2):95–104, May 2014. ISSN 0167-7055, 1467-8659. doi: 10.1111/cgf.12314. URL `https://onlinelibrary.wiley.com/doi/10.1111/cgf.12314`.

Markus Mathias, Anđelo Martinović, and Luc Van Gool. ATLAS: A Three-Layered Approach to Facade Parsing. *International Journal of Computer Vision*, 118(1):22–48, May 2016. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-015-0868-z. URL `http://link.springer.com/10.1007/s11263-015-0868-z`.

Paul Merrell and Dinesh Manocha. Continuous Model Synthesis.

Vahid Moosavi. Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet, September 2017. URL `http://arxiv.org/abs/1709.02939`. arXiv:1709.02939 [cs].

Pascal Muller, Peter Wonka, Simon Haegler, Andreas Ulmer, and Luc Van Gool. Procedural Modeling of Buildings.

Nelson Nauata, Kai-Hung Chang, Chin-Yi Cheng, Greg Mori, and Yasutaka Furukawa. House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation, March 2020. URL `http://arxiv.org/abs/2003.06988`. arXiv:2003.06988 [cs].

Gen Nishida, Ignacio Garcia-Dorado, Daniel G. Aliaga, Bedrich Benes, and Adrien Bousseau. Interactive sketching of urban procedural models. *ACM Transactions on Graphics*, 35(4):1–11, July 2016. ISSN 0730-0301, 1557-7368. doi: 10.1145/2897824.2925951. URL `https://dl.acm.org/doi/10.1145/2897824.2925951`.

Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, April 2017. URL `http://arxiv.org/abs/1612.00593`. arXiv:1612.00593 [cs].

Jinmo Rhee. Deeprise: A Novel Method to Model and Design a High-rise Building Form through Deep Neural Network.

Jinmo Rhee, Pedro Veloso, and Ramesh Krishnamurti. Three decades of machine learning with neural networks in computer-aided architectural design (1990–2021). *Design Science*, 9:e25, 2023. ISSN 2053-4701. doi: 10.1017/dsj.2023.21. URL `https://www.cambridge.org/core/product/identifier/S2053470123000215/type/journal_article`.

Daniel Ritchie, Ben Mildenhall, Noah D. Goodman, and Pat Hanrahan. Controlling procedural modeling programs with stochastically-ordered sequential Monte Carlo. *ACM Transactions on Graphics*, 34(4):1–11, July 2015. ISSN 0730-0301, 1557-7368. doi: 10.1145/2766895. URL `https://dl.acm.org/doi/10.1145/2766895`.

Julian Togelius, Georgios N. Yannakakis, Kenneth O. Stanley, and Cameron Browne. Search-Based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):172–186, September 2011. ISSN 1943-068X, 1943-0698. doi: 10.1109/TCIAIG.2011.2148116. URL `http://ieeexplore.ieee.org/document/5756645/`.

Carlos A. Vanegas, Tom Kelly, Basil Weber, Jan Halatsch, Daniel G. Aliaga, and Pascal Müller. Procedural Generation of Parcels in Urban Modeling. *Computer Graphics Forum*, 31(2pt3):681–690, May 2012. ISSN 0167-7055, 1467-8659. doi: 10.1111/j.1467-8659.2012.03047.x. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2012.03047.x`.

Abraham Noah Wu, Rudi Stouffs, and Filip Biljecki. Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales. *Building and Environment*, 223:109477, September 2022. ISSN 03601323. doi: 10.1016/j.buildenv.2022.109477. URL https://linkinghub.elsevier.com/retrieve/pii/S0360132322007089.

Yun Kyu Yi, Yahan Zhang, and Junyoung Myung. House style recognition using deep convolutional neural network. *Automation in Construction*, 118:103307, October 2020. ISSN 09265805. doi: 10.1016/j.autcon.2020.103307. URL https://linkinghub.elsevier.com/retrieve/pii/S0926580520300261.

Yuyang Yin, Dejia Xu, Zhangyang Wang, Yao Zhao, and Yunchao Wei. 4DGen: Grounded 4D Content Generation with Spatial-temporal Consistency, December 2024. URL http://arxiv.org/abs/2312.17225. arXiv:2312.17225 [cs].

Dawen Yu, Shunping Ji, Jin Liu, and Shiqing Wei. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:155–170, January 2021. ISSN 09242716. doi: 10.1016/j.isprsjprs.2020.11.011. URL https://linkinghub.elsevier.com/retrieve/pii/S092427162030318X.

Simiao Yu, Hao Dong, Pan Wang, Chao Wu, and Yike Guo. Generative Creativity: Adversarial Learning for Bionic Design. In Igor V. Tetko, Věra Kůrková, Pavel Karpov, and Fabian Theis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2019: Image Processing*, volume 11729, pages 525–536. Springer International Publishing, Cham, 2019. ISBN 978-3-030-30507-9 978-3-030-30508-6. doi: 10.1007/978-3-030-30508-6_42. URL http://link.springer.com/10.1007/978-3-030-30508-6_42. Series Title: Lecture Notes in Computer Science.

Jason Y Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild.

Pengcheng Zhang, Huagui He, Yun Wang, Yang Liu, Hong Lin, Liang Guo, and Weijun Yang. 3D Urban Buildings Extraction Based on Airborne LiDAR and Photogrammetric Point Cloud Fusion According to U-Net Deep Learning Model Segmentation. *IEEE Access*, 10:20889–20897, 2022a. ISSN 2169-3536. doi: 10.1109/ACCESS.2022.3152744. URL https://ieeexplore.ieee.org/document/9716925/.

Yuzhe Zhang, Chan Chi Ong, Jianmin Zheng, Seng-Tjhen Lie, and Zhendong Guo. Generative design of decorative architectural parts. *The Visual Computer*, 38(4):1209–1225, April 2022b. ISSN 0178-2789, 1432-2315. doi: 10.1007/s00371-021-02142-1. URL https://link.springer.com/10.1007/s00371-021-02142-1.

Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. Structured3D: A Large Photo-Realistic Dataset for Structured 3D Modeling. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, volume 12354, pages 519–535. Springer International Publishing, Cham, 2020. ISBN 978-3-030-58544-0 978-3-030-58545-7. doi: 10.1007/978-3-030-58545-7_30. URL https://link.springer.com/10.1007/978-3-030-58545-7_30. Series Title: Lecture Notes in Computer Science.

4.2 Github Repository

https://github.com/1gfelton/11-685-Project