# SYNTHETIC DATA GENERATION PIPELINE FOR GEOMETRIC DEEP LEARNING IN ARCHITECTURE

S. Fedorova[1*], A.Tono[2†], M. S. Nigam[3†], J. Zhang[4], A. Ahmadnia[1], C. Bolognesi[1], D. Michels[4]

[1]Politecnico di Milano, Milan, Italy - (stanislava.fedorova, amirhossein.ahmadnia)@mail.polimi.it, cecilia.bolognesi@polimi.it
[2]CDI, Computational Design Institute, San Francisco - alberto.tono@cd.institute
[3]IIIT Hyderabad, India - meher.shashwat@students.iiit.ac.in
[4]KAUST, Saudi Arabia - jiayao.zhang@kaust.edu.sa

**KEY WORDS:** Synthetic 3D Dataset, Procedural Generation, Dataset Generation Pipeline, Architecture, Geometric Deep Learning, 3D Reconstruction.

**ABSTRACT:**

With the growing interest in deep learning algorithms and computational design in the architectural field, the need for large, accessible and diverse architectural datasets increases. Due to the complexity of such 3D datasets, the most widespread techniques of 3D scanning and manual building modeling are very time-consuming, which does not allow to have a sufficiently large open-source dataset. We decided to tackle this problem by constructing a field-specific synthetic data generation pipeline that generates an arbitrary amount of 3D data along with the associated 2D and 3D annotations. The variety of annotations, the flexibility to customize the generated building and dataset parameters make this framework suitable for multiple deep learning tasks, including geometric deep learning that requires direct 3D supervision. Creating our building data generation pipeline we leveraged the experts' architectural knowledge in order to construct a framework that would be modular, extendable and would provide a sufficient amount of class-balanced data samples. Moreover, we purposefully involve the researcher in the dataset customization allowing the introduction of additional building components, material textures, building classes, number and type of annotations as well as the number of views per 3D model sample. In this way, the framework would satisfy different research requirements and would be adaptable to a large variety of tasks. All code and data is made publicly available: cdinstitute.github.io/Building-Dataset-Generator.

## 1. INTRODUCTION

The recent advances in CAD software showed the advantages of using 3D models to the specialists working in various fields. The advantages include the possibility of a 3D visualization of a project, inclusion of all the relevant metadata related to it, a relatively quick transformation to the desired representation, and the modifications that are reflected on all the stages and views. It is especially relevant to the field of architecture and urban planning that involves a lot of different data required to take into consideration as well as the spatial representation of a project. At the same time, the time and resources required for a full 3D model creation for a building project are one of the obstacles that prevent some of the specialists of the field from using 3D modeling in their pipelines.

A potential solution to this issue could be leveraging the Geometric Deep Learning techniques for the generation of 3D models from a given input. This input could potentially be represented in various ways, for instance, as a set of constraints determining the project or an image representation. In the case of an architectural project, it would be quite time-consuming to determine a set of rules due to the complexity of the subject. The use of the images would take advantage of the ability of Deep Neural Networks to learn visual features in a self-supervised manner.

One of the challenges the researchers face when dealing with this question is availability of annotated 3D datasets. The complexity of the 3D model creation process, a variety of software and approaches to building modeling, differences in the modeling quality, and the lack of access for the completed and finished 3D models make it difficult for the neural networks to learn successfully. At this point, to our knowledge, there are no existing architectural datasets at the building scale with a sufficient amount of samples to perform Geometric Deep Learning. Due to these reasons we provide the architectural community with a data generation pipeline that automatically creates a synthetic building dataset suitable for various Deep Learning tasks. Due to the specificity of our contribution and the lack of the datasets on this architectural scale, there are no benchmarks or downstream metrics existing so far.

The establishment of a 3D dataset of building envelopes with relevant information and development of a 3D reconstruction framework could benefit a manifold of different industries such as robot navigation, drone delivery, retails, urban planning, AR/VR gaming experiences.

## 2. RELATED WORKS

**3D Datasets** The complexity of the urban planning and architecture fields create a variety of Deep Learning applications applicable to them. The drawback of this aspect is that different tasks require different input and ground truth data not only in terms of format but also in terms of internal requirements within one format. For instance, generative design of a building facade would require a semantic segmentation annotation, while 3D reconstruction would require a 2D representation of a building as well as a 3D representation, the format of which could vary based on the researcher's approach (mesh, voxels, point cloud). The intricacy of the subject of interest makes the

---

dataset creation time-consuming, so while there are quite a few 3D datasets (Sun et al., 2018, Chang et al., 2015, Wu et al., 2015, Kim et al., 2020, Xiang et al., 2014, Xiang et al., 2016, Mo et al., 2019, Koch et al., 2019, Lim et al., 2013) (Table 1), not many of them are related to architecture. The format issue makes this scarce number even smaller when applied to a specific problem, such as facade reconstruction, 3D classification or semantic segmentation of the building parts.

Urban 3D datasets are the most common ones in the field. This is due to the availability of data at the city level (shapefiles), open-source geospatial software and the relative ease of production, as most of these datasets contain Level of Detail 1 (LoD1) models (Gao et al., 2021). Level of Detail(LoD) is an important characteristic of the architectural datasets that allows to specify the amount of detail and generalization present in the 3D model (Biljecki et al., 2016b). LoD1 refers to a building envelope without any additional details while LoD4 indicates a detailed building model with an internal structure.

The work that that aims to solve the same problem is 3DCityDB (Yao et al., 2018), which contains the 3D representations of the real cities. The advantage their system has, is the use of the real-world data and texturing that gives more precise and realistic representations of buildings. The drawbacks of their framework include the limited amount of samples (as for the time of writing the database mentions the cities of Berlin and New York), the lack of meaningful information about the structures, absence of the semantic segmentation related to the parts of the buildings and the simplified models of the buildings (extruded polygons) that do not include finer details. The models in 3DCityDB are given in the urban context which could be seen as an advantage or disadvantage based on the task; moreover, there are no image annotations or separate 3D files the researchers could use directly for Deep Learning purposes.

Another approach that is similar to ours is Random3DCity (Biljecki et al., 2016a), that exploits procedural generation in order to create a variety of building forms. However, the authors aim to get a simplified building representation without the use of the real-world texturing which makes it inapplicable to single image to 3D reconstruction problem as we approach it. Moreover, the tool is intended to be used in CityGML format (Gröger and Plümer, 2012) that imposes certain limitations on the researchers. The main advantage of this dataset is the possibility to decide the LoD level up to LoD4 which includes the internal building structure and modularity that allows a relatively easy scaling of the dataset and its variations.

The dataset that is closest to ours is Structured3D (Zheng et al., 2020) that involves synthetic 3D models and image annotations for the architectural interior space. The annotations of this dataset include segmentation, depth and rendered images as well as the objects' structure and several interior configurations. There are a few more synthetic interior 3D scene datasets such as SceneNet (Handa et al., 2015) or InteriorNet (Li et al., 2018) (does not include 3D models). Unfortunately, all of them tackle the problem at a different architectural scale concentrating on one interior space while our task requires building exterior models with the information related to the structure.

Another disadvantage of the mentioned architectural datasets is related to their construction process, as the building objects they contain are not parametric, even the ones containing synthetic data. This aspect puts some limitations on the researchers using these datasets. We address this issue in our solution by

providing a parametric building generation using programmatic procedural approach.

## 3. DATASET GENERATION PIPELINE

### 3.1 Existing Datasets

| Dataset | Classes | Samples | Images | Std-dev |
|---|---|---|---|---|
| IKEA | 7 | 219 | 759 | 16.4 |
| ShapeNetCore | 55 | 51,300 | 307,800 | 1589.7 |
| ShapeNetSem | 970 | 9,033 | 72,000 | 343.3 |
| PartNet | 24 | 26,600 | NA | 322.4 |
| ObjectNet3D | 100 | 44,147 | 90,127 | 42.0 |
| Pascal3D+ | 12 | 79 | 30,899 | 1.9 |
| ModelNet10 | 10 | 4,899 | NA | 243.9 |
| ModelNet40 | 40 | 12,311 | NA | 215.7 |
| MCB | 68 | 58,700 | +300,000 | 1040.1 |
| ABC | NA | 1,000,000 | NA | NA |
| Pix3D | 9 | 395 | 10,069 | 64.5 |

Table 1. Datasets with the statistics on the number of classes and samples in each. Standard deviation is calculated for the number of 3D models per main class, sub-classes were not considered.

The datasets that are used by the majority of the researchers studying geometric deep learning usually consist of a heterogeneous set of items divided into several categories. While being a good fit for research frameworks, these datasets could hardly be used in industry due to the lack of specificity. One major issue that arises while dealing with 3D datasets is class imbalance which introduces a bias in the learning process if not managed properly, as in Table 1.

| Dataset | Scale | LoD | Class | 3D Models | Images |
|---|---|---|---|---|---|
| Random3DCity | 0 | 1-4 | 31 | NA | NA |
| 3DCityDB (Berlin) | 0 | 2 | NA | 550.000 | NA |
| Structured3D | 2 | 4 | NA | NA | 196K |
| InteriorNet | 2 | 4 | 158 | 1M models | 20M |
| SceneNet RGB-D | 2 | 4 | 37 | NA | 5M |
| Matterport3D | 2 | 4 | 20 | 90 scenes | 194K |
| Replica | 2 | 4 | 88 | 18 scenes | 20M |
| Gibson | 1,2 | 4 | NA | 572 scenes | NA |
| HyperSim | 2 | 4 | NA | 461 | 77.4K |

Table 2. Comparison of architectural datasets acoording to scale, data available, format, number of classes and LoD. Scale is presented as a set of categories {0, 1, 2}, where:
0 - city scale, 1 - building scale, 2 - interior scale

| Dataset | Render | Depth | Seg | Nrm | Format |
|---|---|---|---|---|---|
| Random3DCity | ✓ | | | | CityGML |
| 3DCityDB | ✓ | | | | CityGML |
| Structured3D | ✓ | ✓ | ✓ | | Mesh |
| InteriorNet | ✓ | ✓ | ✓ | ✓ | Mesh |
| SceneNet | ✓ | | | | RGB-D |
| Matterport3D | ✓ | ✓ | ✓ | | RGB-D |
| Replica | ✓ | | ✓ | | Mesh |
| Gibson | ✓ | ✓ | ✓ | ✓ | Mesh |
| HyperSim | ✓ | ✓ | ✓ | | Mesh |

Table 3. Comparison of annotations provided by popular architectural Datasets: Render(RGB), Depth maps, Segmentation masks, Normal maps and 3D model formats.

As can be observed from the Table 2 the datasets that could be used for the application of Deep Learning in the architectural field are skewed to urban scale and interior scale (Chang et al., 2017, Straub et al., 2019, Roberts and Paczan, 2020, Li et al., 2018, Handa et al., 2015, Zheng et al., 2020) leaving a building scale gap. To our knowledge there is only one open-source dataset (Xia et al., 2018) that provides 3D models of

the building structures. This dataset is rather rich in annotations as it provides depth, normals and segmentation annotations, rendered images as well as full 3D mesh models. Unfortunately, this dataset, too, is focused primarily on the interior space while leaving the building exterior without semantic segmentation and texturing. Our dataset concentrates on this limitation by providing a dataset similar in structure but applied to the external appearance of the buildings rather than the internal one. Another drawback of Gibson dataset is its size, as it does not provide a sufficient number of building samples for geometric deep learning frameworks. Our solution for this problem is generation of an arbitrary amount of building samples depending on the need of the research project.

At the same time, the unavailability of the datasets or data generation pipelines in the field of architecture results in unavailability of the benchmarks related to the data generation pipeline evaluation.

### 3.2 Requirements

From the analysis of the existing datasets we have inferred the requirements for the new framework:

- Sufficient dataset size
- Class balance to avoid bias
- Modularity of data
- Variation and diversity in data
- Generating required ground truth signals
- Associated metadata generation
- Extensibility, to accommodate more classes/models

Since this dataset is mostly intended to be used in Deep Learning tasks, it is necessary for it to have a sufficient amount of samples for the artificial neural networks to learn successfully. The dataset should also be rich in details and alterations of materials, modules, shapes, dimensions. At the same time all the classes and variations should not only be ample in the number of samples but also balanced in order not to introduce an additional bias to the learning process.

The modularity requirement of the dataset that refers to the parametric nature of a generated building allows for the substitution of different parameters and architectural components, randomization, and modification of the object's structure.

Architecture is a wide field with multiple problems that can be solved via deep learning. To accustom to as many problems from the spectrum as possible it fundamental to provide different ground truth signals and metadata that would help in the quantitative evaluation of the models' performance. Moreover, the wide range of possible applications requires the dataset to be modular and extendable to be adjusted to different problems based on the researchers' needs.

### 3.3 Proposed framework

The synthetic data generation pipeline developed by us addresses the prerequisites mentioned in the Subsection 3.2. It is intended primarily for the 3D reconstruction based on parts assembly task (Li et al., 2020), as we have identified that the image to 3D translation would have a major impact in many applications in the architectural field. Although, it should be noted that the annotations and modularity of the framework make it possible to use the generated data for the other deep learning tasks as well.
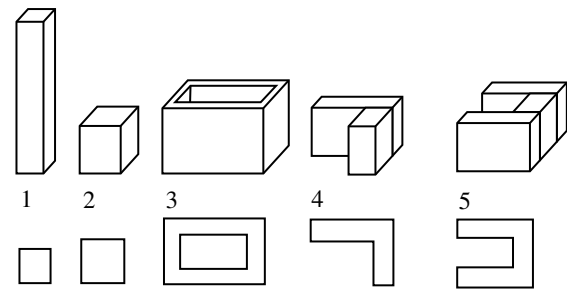


Figure 1. Initial types of the building envelopes. 1 - skyscraper; 2 - isolated building; 3 - patio; 4 - L-shaped building; 5 - C-shaped building.

In order to adjust for all the possible applications of this dataset it was decided to partially engage the user in the dataset creation and give the freedom of setting a number of parameters that define the content of the dataset and its variations.
It has been decided to define five classes at the initial stage of the framework development, as can be seen in Figure 1.

### 3.4 Implementation details

In contrast to the existing generic 3D datasets, and alike many architectural datasets, ours contains synthetic data only. This difference allows generating the data relative to a particular research problem in a customized manner. This characteristic automatically lifts the problem of image-shape alignment present in the other datasets (Sun et al., 2018) as the images, their segmentation masks, and depth annotations are generated from the mesh directly. For each sample we provide the 3D model in .obj format, the point cloud consisting of 2048 points, the rendered image, the ground truth annotations (segmentation mask, depth image in .exr format, surface normals), and the metadata, which will be expanded in the future editions.

### 3.5 Generation

The dataset size does not have an upper bound as it is created in a generative manner, the user can define the number of samples needed to be created and the classes the dataset should contain, as well as the set of component modules (e.g. windows and balconies) and textures to include. The user can add his own modules as .obj files to be used in the building dataset generation.

The input parameters include level of detail (1 or 2), use of materials, textures used (if any), modules used (if any), building types, generated image size, number of points per point cloud, some of the output formats and the general characteristics of the buildings defined by common knowledge and city regulation base, for instance, minimum and maximum building height, width and length. It is also possible to indicate whether it is necessary to generate several views for one 3D building model and how many. All the parameters need to be specified in the configuration file.

The output consists of the following set of data, also illustrated in Figure 2:

- Rendered image of a building, [.png format]
- Semantic segmentation mask of a building (each component has its own color) [.png format]
- Depth map of a building [.png and .exr format; .png contains normalized values and .exr contains the actual values]
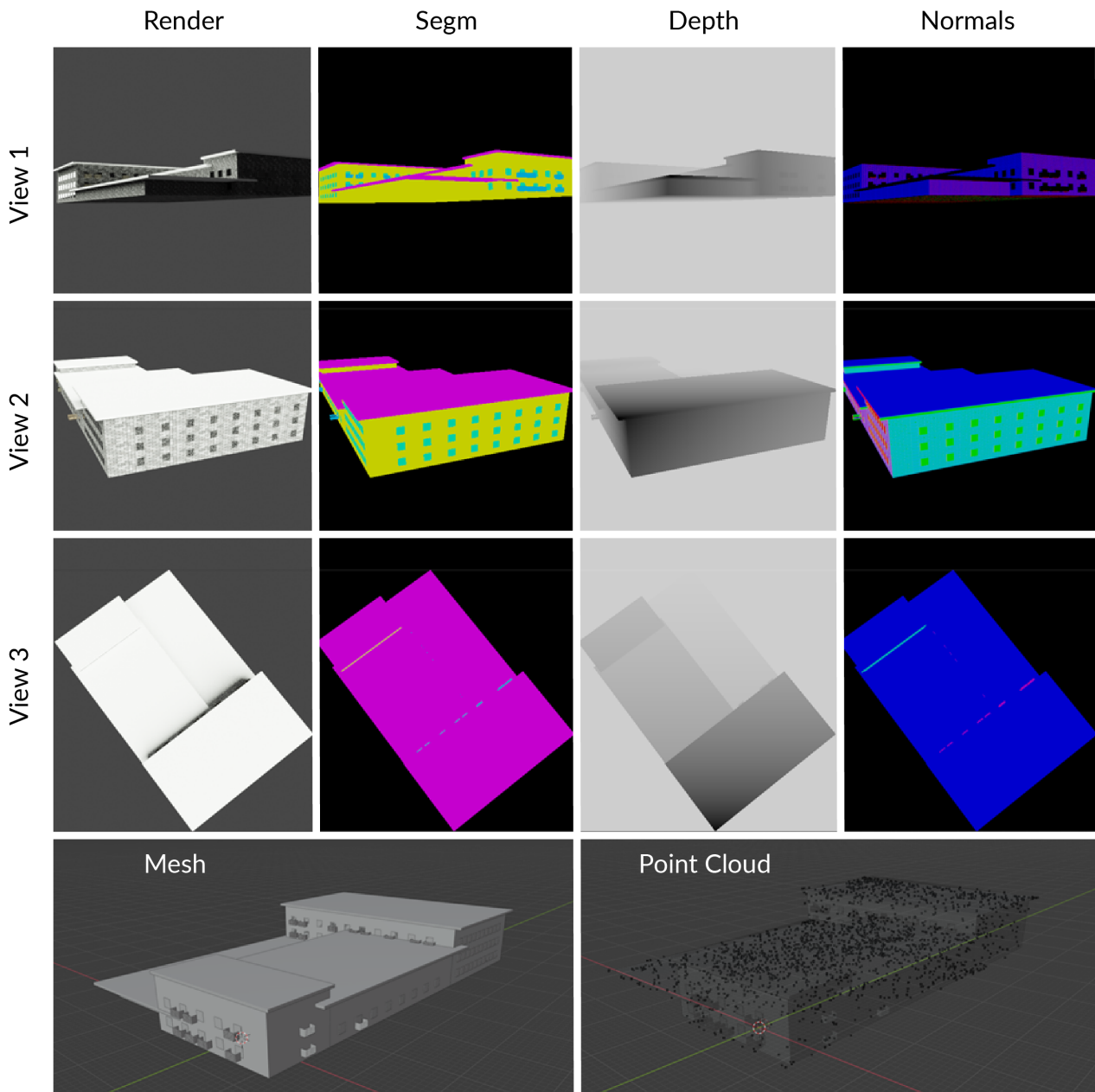
Figure 2. Sample of one 3D model generated by the pipeline with its annotations. 3 rows are 3 views of the model, fourth row shows the mesh model and the point cloud generated for the building model. The annotation columns are: 1 - rendered image, 2 - segmentation mask, 3 - depth image, 4 - image of surface normals.

- Map of surface normals showing the angle of each surface with respect to the camera [.png format]
- 3D model of a building (mesh) [.obj format]
- Point cloud of a building, by default point cloud consists of 2048 points which can be changed by the user [.ply format]

Dataset Generation code, and sample dataset renders has been made publicly available. [1]

### 3.6 Domain Randomization

In order to adjust for the domain adaptation of the real-world building data, it was necessary to introduce domain randomization, which is considered one of the most successful techniques

to handle the transfer from synthetic to real data(Tremblay et al., 2018). The pipeline uses majorly the parameters and the textures that are relatively close to the real-world ones in order to be able to adjust for the domain adaptation task. Our dataset generation framework has a parameter that allows to create randomized sample images from the same 3D model sample.

The randomization parameters are:

- Texture, reflectance of various building components
- Light color, strength and position
- Camera position and angle

Randomization of the camera angle is illustrated in Figure 3. The camera assumes a random position on the XY plane, while its rotation along the local Z axis is limited by the range (40,
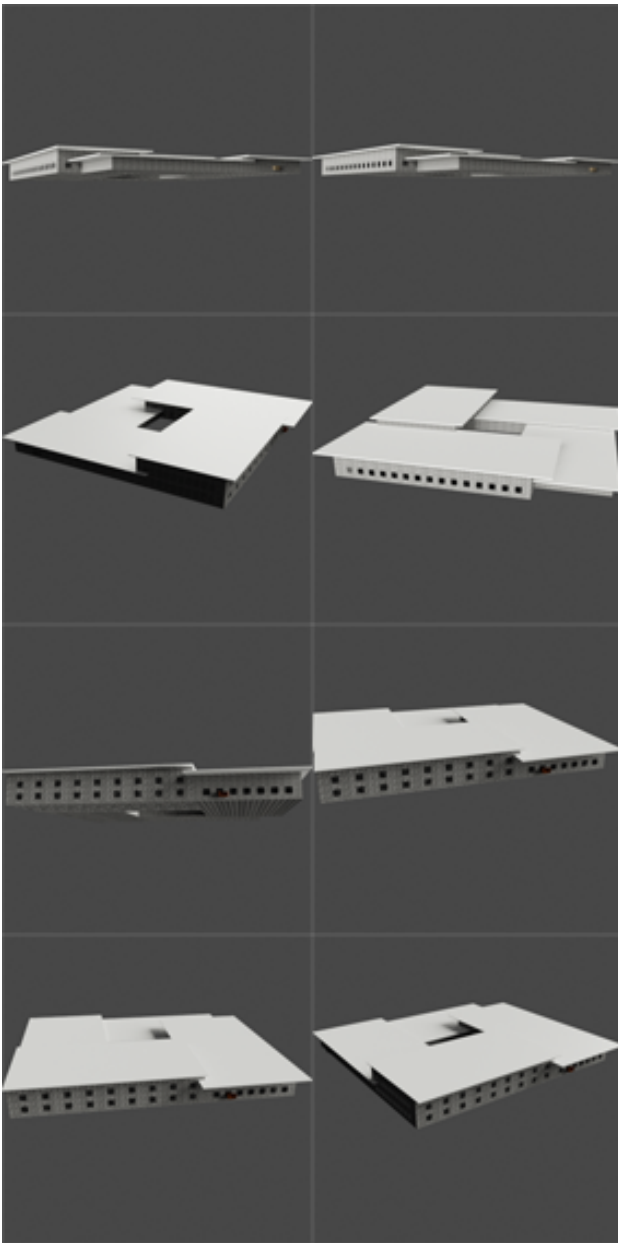
---

[1]

Figure 3. One 3D model sample with multiple views illustrating camera view randomization. The number of possible views per model is defined by the user.



Figure 4. One 3D model sample with multiple views illustrating light randomization. To better illustrate the changes in light the Figure features one view from one model, while in the dataset generation the light changes with the other randomization parameters.

100) in order to take the most significant views of the building model, close to the ones that could be taken in the real world. Texture is randomized per building volume, as in Figure 5, with a user-defined probability that all the volumes will be textured with the same material. The textures are selected randomly from the user-added textures. Light randomization is illustrated in Figure 4.

### 3.7 Performance

We ran the dataset generation algorithm with different input parameters on Windows 10 OS on CPU and GPU using AMD Ryzen 7 3800-X 8-Core Processor and GeForce GTX 1080 and report the performance in the Table 4. The images were rendered with 500x500 dimensions. The dataset generation time heavily depends on the number of components in the models, as generation of more components requires more time; thus, larger or

taller buildings require more time to be generated with respect to the models with smaller dimensions.

As previously mentioned, due to the lack of accessible datasets on the architectural scale and the specificity of the contribution, there are no benchmarks and downstream metrics available in the field.

### 4. LIMITATIONS

Building design can be considered a product design task of a very high complexity as it involves multiple parameters densely interconnected between them. This aspect makes the generation of lifelike synthetic data rather complicated. One of the limitations of the proposed dataset generation framework is the lack
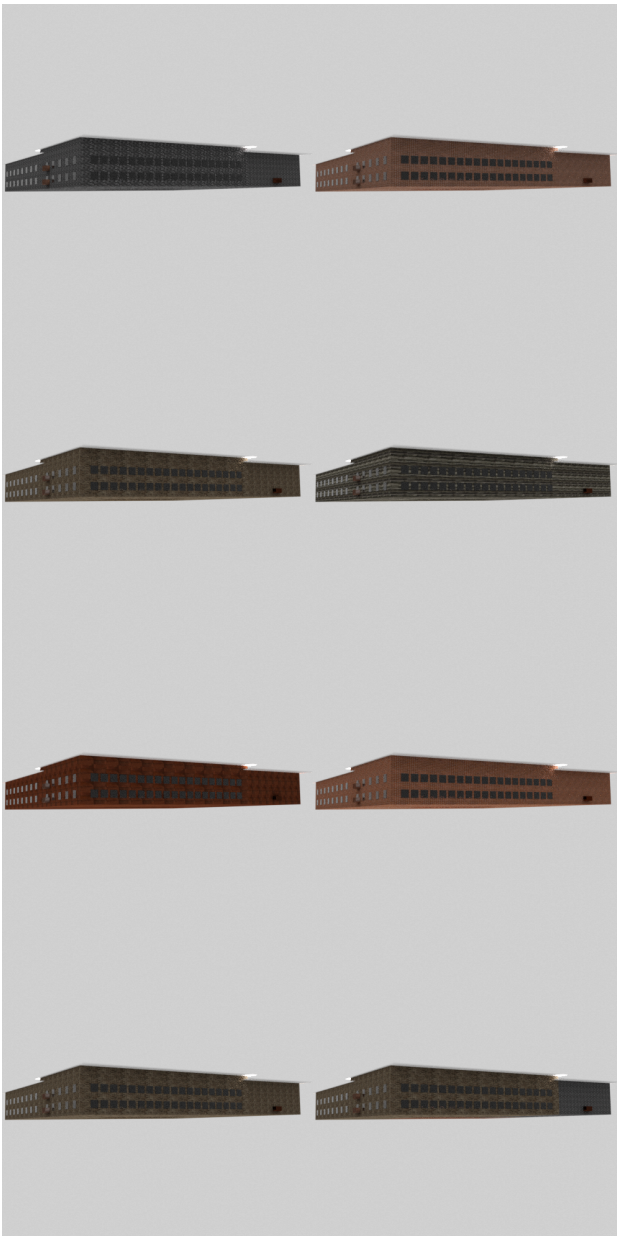
of authenticity, as it does not provide the real-world data and many visual features present in the photos made in an urban environment lack in the rendered images.

Another limitation is the lack of the urban surroundings around the building. The object is rendered in an empty scene which does not occur in the real world and thus makes domain transfer task more complicated. Moreover, the buildings generated with the framework account for the most typical design patterns present in the major cities, while the more creative and outstanding operas of architecture are not included in the generation pipeline due to the high complexity of the problem. This limitation is the most severe one as it narrows the range of possible building solutions and the entire variety of the architectural forms is not presented in this work. Not only does it affect the generalization capability of the neural networks, but it also introduces a bias in the training process.

We are planning to address the mentioned limitations in our future research work.

## 5. PERSPECTIVES

The proposed dataset is not free from limitations, which we intend to tackle in our future works. This work is the first step towards the use of Geometric Deep Learning in the field of Architecture, consequently, it opens multiple possible ways for future development. One of them is to improve the quality and variation of the building models by adding various building types, components and textures and by making the renders look more realistic. Another possible development direction could be the extension of the level of detail and the addition of the building parameters into the dataset as well as the internal building structure. Moreover, it is important to develop the framework to account for the limitation of not having the urban surroundings. Expanding the framework in order to have the nearby buildings, urban furniture, trees and other elements of urban environment is fundamental for 3D reconstruction in architecture.

The exploration of the Geometric Deep Learning Frameworks intended for single image to 3D reconstruction, their performance on the present dataset is equally important to the development of the dataset generation framework, as it is the task the dataset was intended for. This direction of research could bring multiple practical implications related to the fields of architecture, gaming, AR/VR, simulation and others. Moreover, with the variety of annotation this dataset provides, it would be possible to use it in the other Deep Learning tasks as well.

Finally, this dataset could serve as a base for the advances in the field of generative architectural design via the exploitation of Generative Adversarial Networks and their ability to learn from the visual features. Moreover, the possibility to generate synthetic data in various ways could facilitate domain adaptation from synthetic to real data.

## 6. CONTRIBUTION

The contribution of this work consists of two principal parts:

- A detailed overview of the existing datasets related to the architectural and urban fields as well as the main geometric deep learning frameworks aimed at solving single image to 3D reconstruction task for buildings;



Figure 5. One 3D model sample with multiple views illustrating texture randomization. To better illustrate the changes in texture the Figure features one view from one model, while in the dataset generation volume textures change with the other randomization parameters.

| GPU | EXR | Multiview | Time (hours) |
|:---:|:---:|:---:|:---:|
| | | | 1.7 |
| | | ✓ | 2.7 |
| | ✓ | | 1.17 |
| | ✓ | ✓ | 2.5 |
| ✓ | | | 0.34 |
| ✓ | | ✓ | 0.8 |
| ✓ | ✓ | | 0.92 |
| ✓ | ✓ | ✓ | 0.88 |

Table 4. Comparison of dataset generation performance across different configuration settings(on a Windows 10 OS). The generated dataset consisted of 100 model each, with 3 captured views in multi-view setting.

- Providing the architectural and deep learning community with a field-specific dataset generation pipeline targeted at various tasks.

## 7. CONCLUSION

The overview of the frameworks related to the Single Image to 3D Reconstruction task demonstrate the necessity of a field-specific dataset due to their inability to generalize to unseen data. The exploration of the existing architectures has suggested the necessary requirements and highlighted the weak points of the existing architectural datasets, which later became the foundation for our tool.

Presented dataset generation framework provides a field-specific tool for the generation of 2D and 3D data in various formats. This instrument gives the researchers the possibility to apply the Deep Learning and Geometric Deep Learning techniques to the architecture-related tasks. The framework proved to be able to generate large amounts of various data in a short time compared to the traditional methods previously exploited in the building 3D modelling and the consequent image rendering. Moreover, the way the tool was built allows for extendability and customization of the dataset to the specific tasks within the architectural field using task-specific building components, textures and building typologies. It also addresses the limitations of the existing urban datasets we have given an extensive overview of.

However, this dataset is not without drawbacks, the main of which is the use of the synthetic data instead of the real-world data. This is the trade-off that was made due to the high complexity and time and effort consumption of the 3D modelling and 3D scanning processes. Our decision in favor for the synthetic data opens multiple directions for the future work as on the dataset framework as on the algorithms using this data to learn. Moreover, the provided synthetic data facilitates the research in 3D reconstruction related to the architecture field that was not available in the open access previously.

## ACKNOWLEDGEMENTS

## REFERENCES

Biljecki, F., Ledoux, H., Stoter, J., 2016a. Generation of multi-lod 3d city models in citygml with the procedural modelling engine random3dcity. IV-4/W1, 51–59.

Biljecki, F., Ledoux, H., Stoter, J., 2016b. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25-37.

Chang, A., Dai, A., Funkhouser, T., Halber, M., Niessner, M., Savva, M., Song, S., Zeng, A., Zhang, Y., 2017. Matterport3D: Learning from RGB-D Data in Indoor Environments. *International Conference on 3D Vision (3DV)*.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F., 2015. Shapenet: An information-rich 3d model repository. cite arxiv:1512.03012.

Gao, W., Nan, L., Boom, B., Ledoux, H., 2021. Sum: A benchmark dataset of semantic urban meshes.

Gröger, G., Plümer, L., 2012. CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12-33.

Handa, A., Patraucean, V., Badrinarayanan, V., Stent, S., Cipolla, R., 2015. SceneNet: Understanding Real World Indoor Scenes With Synthetic Data.

Kim, S., Chi, H.-g., Hu, X., Huang, Q., Ramani, K., 2020. A large-scale annotated mechanical components benchmark for classification and retrieval tasks with deep neural networks. *Proceedings of 16th European Conference on Computer Vision (ECCV)*.

Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., Panozzo, D., 2019. Abc: A big cad model dataset for geometric deep learning. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, W., Saeedi, S., McCormac, J., Clark, R., Tzoumanikas, D., Ye, Q., Huang, Y., Tang, R., Leutenegger, S., 2018. Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. *British Machine Vision Conference (BMVC)*.

Li, Y., Mo, K., Shao, L., Sung, M., Guibas, L., 2020. Learning 3d part assembly from a single image.

Lim, J. J., Pirsiavash, H., Torralba, A., 2013. Parsing IKEA Objects: Fine Pose Estimation. *ICCV*.

Mo, K., Zhu, S., Chang, A. X., Yi, L., Tripathi, S., Guibas, L. J., Su, H., 2019. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Roberts, M., Paczan, N., 2020. Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding.

Straub, J., Whelan, T., Ma, L., Chen, Y., Wijmans, E., Green, S., Engel, J., Mur-Artal, R., Ren, C., Verma, S., Clarkson, A., Yan, M., Budge, B., Yan, Y., Pan, X., Yon, J., Zou, Y., Leon, K., Carter, N., Briales, J., Gillingham, T., Mueggler, E., Pesqueira, L., Savva, M., Batra, D., Strasdat, H. M., Nardi, R. D., Goesele, M., Lovegrove, S., Newcombe, R. A., 2019. The Replica Dataset: A Digital Replica of Indoor Spaces. *ArXiv*, abs/1906.05797.

Sun, X., Wu, J., Zhang, X., Zhang, Z., Zhang, C., Xue, T., Tenenbaum, J. B., Freeman, W. T., 2018. Pix3d: Dataset and methods for single-image 3d shape modeling. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

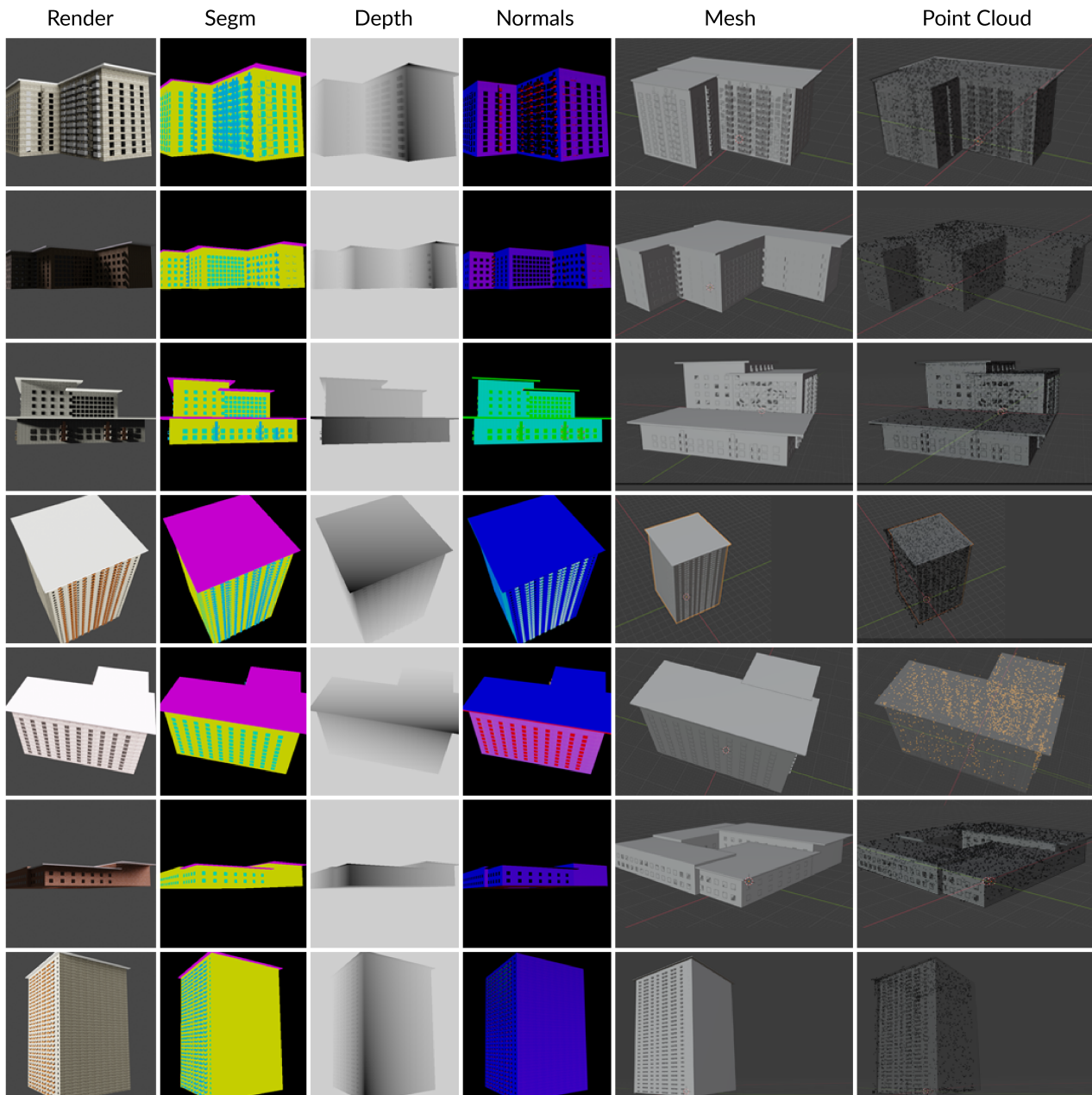| Render | Segm | Depth | Normals | Mesh | Point Cloud |
|--------|------|-------|---------|------|-------------|



Figure 6. Samples of the images generated by the synthetic pipeline. Columns from left to right: rendered images with material textures, semantic segmentation mask, depth map, map of surface normals, image of a mesh object, Point Cloud. Note: point cloud is represented on top of a semitransparent mesh object for the sake of clarity. The point cloud file does not contain any mesh objects.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S., 2018. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1082–10828.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3d shapenets: A deep representation for volumetric shapes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xia, F., R. Zamir, A., He, Z.-Y., Sax, A., Malik, J., Savarese, S., 2018. Gibson env: real-world perception for embodied agents. *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*, IEEE.

Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., Mottaghi, R., Guibas, L., Savarese, S., 2016. Objectnet3d: A large scale database for 3d object recognition. *European Conference Computer Vision (ECCV)*.

Xiang, Y., Mottaghi, R., Savarese, S., 2014. Beyond pascal: A benchmark for 3d object detection in the wild. *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

Yao, Z., Nagel, C., Kunde, F., 2018. 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML.

Zheng, J., Junfei, Z., Li, J., Tang, R., Gao, S., Zhou, Z., 2020. Structured3d: A large photo-realistic dataset for structured 3d modeling. *in proc. of European Conference on Computer Vision 2020 (ECCV)*.