



Hybrid Immediacy: Designing with Artificial Neural Networks Through Physical Concept Modelling

Mathias Bank¹(✉), Viktoria Sandor¹, Robby Kraft¹, Stephan Antholzer²,
Martin Berger², Tilman Fabini¹, Balint Kovacs³, Tobias Hell², Stefan Rutzinger¹,
and Kristina Schinegger¹

¹ Department of Design, i.s.d | Structure and Design, University of Innsbruck, Innsbruck, Austria
{mathias.bank-stigsen, viktoriasandor}@uibk.ac.at

² Department of Mathematics, University of Innsbruck, Innsbruck, Austria

³ Center for Geometry and Computational Design, Vienna University of Technology,
Vienna, Austria

Abstract. In digital design practice, the connection and feedback between physical and digital modelling is receiving increasing attention and is seen as a source of creativity and design innovation. The authors present a workflow that supports real-time design collaboration between human and machine intelligence through physical model building. The proposed framework is investigated through a case study, where we test the direct connectivity of physical and digital modelling environments with the integration of artificial neural networks. By combining 3D capturing tools and machine learning algorithms, the research creates an instant feedback loop between human and machine, introducing a hybrid immediacy that puts physical model building back at the centre of the digitally focused design process. By fusing physical models and digital workflows, the research aims to create interactivity between data, material and designer already at the early stage of the design.

Keywords: Computational immediacy · Early design phase · ANN · Physical model building · Hybrid design tool · 3D capturing · Machine intelligence

1 Introduction

Modelling is the core technique in any architectural design process. A model materialises design intentions and “objectifies” them by embedding design knowledge in the object (Oxman 2008). Herbert Stachowiak defined the general nature of models through key characteristics: They are illustrations of content that they are abstracting or “abbreviating” in order to record only the aspects that are relevant to their user. They are designed for a special purpose or task and are therefore evaluated according to their “usefulness” for the model-maker (Stachowiak 1973).

The original version of this chapter was previously published non-open access. The correction to this chapter can be found at https://doi.org/10.1007/978-3-031-13249-0_48

© The Author(s) 2023, corrected publication 2024

C. Gengnagel et al. (Eds.): DMS 2022, Towards Radical Regeneration, pp. 13–23, 2023.
https://doi.org/10.1007/978-3-031-13249-0_2

The “concept model” has a very unique role in the design process, it does not illustrate a design outcome but the initial and essential design intention. It is a tool in itself for materialising and communicating design ideas that keeps a purposeful ambiguity in order to leave space for imagination and further development¹. According to Vera Bühlmann “models maintain a relation with ideas, and seek to sustain and communicate their power”; they do not determine the concept but rather enrich it through a “surplus capacity” (Bühlmann 2013). In the research, the concept model is an instrument for communication and exploration and in that sense a medium and a tool at the same time.

The feedback between physical and digital modelling has received heightened attention as a source of design innovation (Stavrić et al. 2013). The integration of the physical model into digital workflows opens the door for interactivity between data, material and designer (Thomsen and Tamke 2012). A common interest in many physical-digital experimentations is to increase creative capacity and immediate/intuitive control of the process without the need to explicitly define underlying geometrical rules of design objects as Mario Carpo suggested designers might directly „use“ chunks of scanned objects (Carpo 2017). Combined with artificial neural networks and growing archives of 3D point clouds of objects this vision could reach a further dimension - designers will not simply sample chunks, but they will be able to learn from spatial objects, deduce features and apply them to architectural designs without remodelling.

1.1 Related Work

Due to the continued introduction of new 3D capturing tools and scanning devices, there are promising opportunities to re-integrate physical design processes into digital design workflows. In the field of Reverse Engineering (RE), scanning technologies are used to build digital models from physical objects (Hsieh 2015). Although scanning technologies have been around for decades, their performance has improved significantly in recent years. After Microsoft developed Kinect in 2010 a much more affordable scanning device than LIDAR could offer, there has been a growing interest in its use in architectural modelling. While some research projects were focusing on the materialisation of the sensed point clouds in digital environments (Hsieh 2015), others utilised Kinect devices as navigation tools in VR environments allowing immediate ways of connecting the physical to the digital (Souza et al. 2011).

While connecting the physical with the digital through scanning devices is crucial for hybrid concept modelling, the project aims additionally to introduce artificial neural networks as interpreters in the design process. Some of the earliest attempts to implement AI as a participatory system in the early stage of design are the projects by John and Jane Frazer presented in the book *An Evolutionary Architecture* (Frazer 1995). In their work, the role of AI is to recognize patterns and react to the design suggestions, generating a responsive design loop. A more contemporary take on this approach is developed by SPACEMAKER where immediacy and real-time feedback become a key feature of human-machine collaborative design (Jeffrey et al. 2020). Although this project situates the process entirely in the digital space, the input of ANNs can be channelled in various

¹ A famous example of such a concept model is the digital model for the Moebius house by UN Studio. See image in (Hirschberg et al., 2020).

captured formats of the physical space. Just like the project, Deep Perception by Fernando Salcedo (Leach 2021) showcases through real-time cameras how captured folded textiles can be directly connected to a trained network.

1.2 Objectives

The authors present a workflow that supports real-time design collaboration between human and machine intelligence through physical model building. In the tested use case the immediate connectivity of the physical and digital modelling environments is challenged through artificial neural networks. Kinects are used as 3D capturing devices and a machine learning network capable of processing 3D point cloud data directly from the Kinects is established.

2 Conceptual Overview

The following workflow (Fig. 1) outlines a real-time design collaboration between human and machine intelligence through physical model building. The proposed workflow is centred around a physical installation consisting of a large black table, two Kinects, ring lights, a screen, an Arduino control panel and a computer (Fig. 2) - and is organised into four interlinked modules:

- Build
- Capture
- Machine Learning
- Post-processing

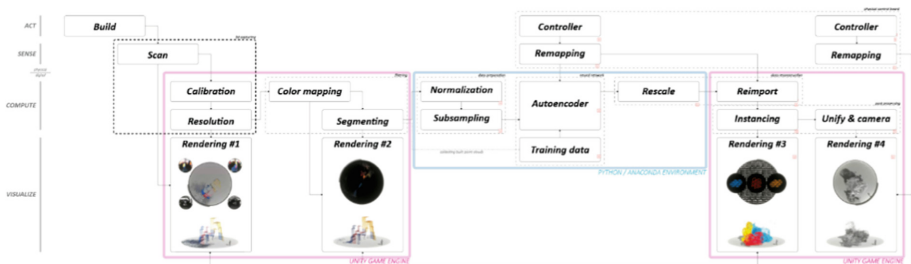


Fig. 1. Diagram of the implemented workflow. From left to right, the build, capture, machine learning and post-processing modules.

The game engine Unity is used as the platform for connecting the components in real-time while allowing users to switch between four different visualisation modes to see the immediate outcome of each step.

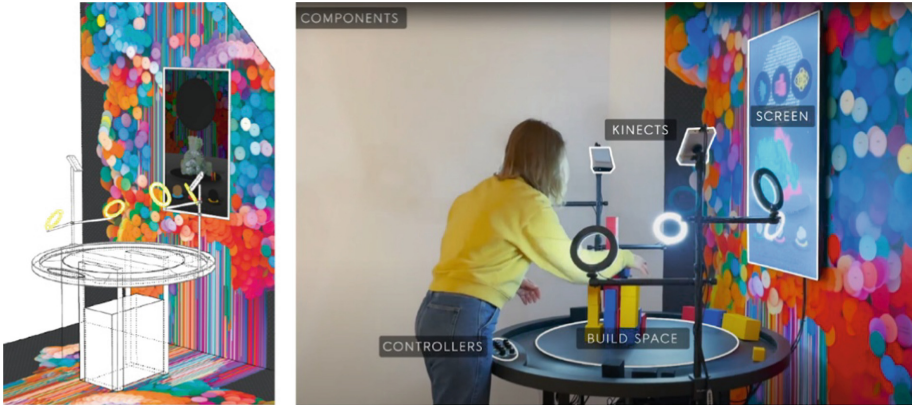


Fig. 2. Illustration and image of the physical installation depicting the main components: A large black table, two Kinects, ring lights, a screen, an Arduino control panel and a computer.

2.1 Build/Physical Setup

The build module is centred around a custom table (Fig. 2) where users can assemble architectural compositions from physical blocks in different colours. To test the workflow, three different colours were selected and applied to a limited number of different building blocks. Each colour represented a unique geometric group, which, through 3D capturing, could be manipulated by an artificial neural network on screen. A control panel embedded in the custom table was used to control the interpolation factor and target, while an additional three potentiometers controlled the scale, rotation and selection of the four visualisation modes (Fig. 3).

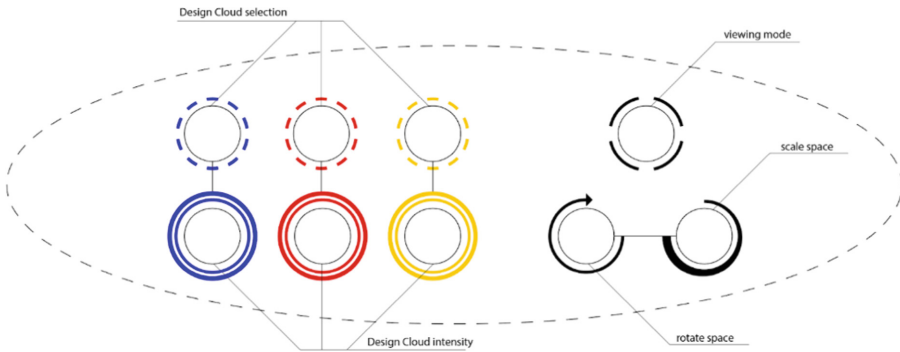


Fig. 3. The graphic depicts the control panel embedded within the custom table.

2.2 Capture

The system of capturing and processing the points required accommodating the Microsoft Azure Kinect and its SDK. A custom application was built using Unity because the Azure Kinect SDK's C# bindings interfaced seamlessly with Unity's C# front end-user code, and Unity supports custom shaders, including compute shaders, which allowed us to make heavy optimizations during the point cloud processing phase.

Homography

We ran a simple homography to virtually align the Kinect cameras in a shared 3D space. Our homography was a custom algorithm that required one scalene triangle to be placed in the centre of the table. The triangle was detected using Sobol edge detection (Kanopoulos et al. 1988), and the Kinect's depth camera was used to bring these edges into 3D. The triangles were then aligned, resulting in one 4×4 position and orientation matrix for each camera.

Point Clouds

Each Kinect device comes with one depth camera and one colour camera. A compute shader converts these raw Kinect images into a point cloud. First, each depth camera pixel is treated as a magnitude that scales its corresponding 3D vector (precomputed in a lookup table) and then this resulting point is transformed by the camera's homography matrix so that each camera's point clouds appear in the same 3D space. This is the first visualisation mode, the "raw Kinect point cloud", where each point in the cloud contains the colour as seen by the colour camera.

Colour Segmentation

The next step uses another compute shader to filter the points, both according to colour and position. The position filtering is simple: we crop points that lie outside a vertically-aligned cylinder with an infinite height and a specified radius that matches the circular installation table in physical space. The colour filtering works by analysing the colours according to the HSV values. First, the user provides a set of reference colours and \pm ranges on all three HSV dimensions which define a bounding box with the reference colour at its centre. The shader rejects points based on their relations to the bounding boxes. The colours that match are categorised and coloured according to their reference colour. This is the second visualisation mode, the "segmented colour clouds". Additionally, these segmented colour point cloud arrays contain the data which is sent to the machine learning application.

2.3 Machine Learning

An artificial neural network is used to transform the Kinect scans of the physically assembled architectural compositions through a user-selected design point cloud. In the workflow, machine learning performs the following steps. We get a preprocessed point cloud from a Kinect scan and a selection of a design point cloud that was chosen by the user together with an interpolation factor. First, we cluster the point clouds into equal-sized clusters with ($n = 2048$) points and then apply the encoder to each of them.

The next step is to interpolate between each cluster from the Kinect point cloud and the design point cloud and then apply the decoder to get new (clustered) point clouds. Finally, the individual clusters are reassembled into one big point cloud. This process can also be applied to multiple point clouds simultaneously. In our case, the setup processes three separate instances at a time which encode the three colours of the used building blocks.

Training Data

A collection of more than 10000 3D point clouds were assembled to provide the training data for the machine learning network. Each point cloud was part of a category of geometric structures (Fig. 4) from which the design point clouds were selected. The intent with the used training data, was to apply the geometric structures of the selected design point clouds within the captured Kinect point clouds - enhancing the appearance of physically assembled compositions on the table.



Fig. 4. Training data and a subset of data from three categories From left to right: Basic volumes, orthographic planes and lattices.

Autoencoder

Our approach is based on the FoldingNet introduced by Yang et al. (2018). However, we make some adjustments that yield better results for our applications. First, the approach uses a 3D grid for the fixed folding input instead of a 2D grid and secondly an extended loss function was implemented, resulting in the biggest improvement. Namely, we use the Chamfer distance (as in the original) together with the Wasserstein (or earth-mover) distance which we calculate due to Feydy et al. (2019) (Fig. 5).

Thus we get a neural network that yields good abstract representations of point clouds. Since these are just vectors we can process them however we want and then use the decoder part to reconstruct a new point cloud. To be more precise we linearly interpolate between two points i.e. two outputs of the encoder and then apply the decoder yielding a new point cloud, resulting in a mix of the two original point clouds.

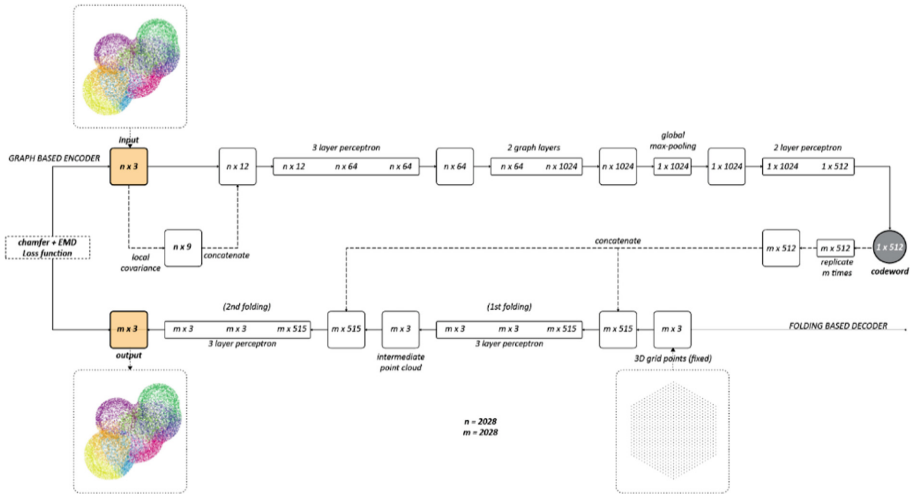


Fig. 5. FoldingNet autoencoder structure based on Yang et al. (2018).

2.4 Post-processing

After the Kinect data has gone through the artificial neural network, the three transformed points clouds are streamed back into Unity for post-processing. Here the representation of the points can be enhanced to better visualise the potential within the user-assembled concept model. This allows for the implementation of a variety of different strategies and algorithms to be applied according to the design use case.

3 Performance

The case study of the research is a hybrid instrument, a concept modelling station, tested in two different environments. The first, by being an interactive exhibition space, gives access to a wider audience for hybrid collaborations, while the second implements the workflow of the modelling station in a professional design context.

3.1 Exhibited Station

By placing the concept modelling station within an exhibition space (Fig. 6), we aimed to test its durability and user-friendliness with a non-professional audience. The station was exhibited in a gallery for three months, where it operated six days a week, ten hours a day. The age range of the visitors was very diverse, with guided tours from pre-school children to university staff.

The interactive aspect of the installation was very popular in this setup. The building blocks were in constant use, while the circular shape of the table allowed up to 4–5 users to collaborate on one model. Although over the three months the Kinect devices had to be recalibrated due to the movement of the drafting table to which they were attached, the real-time connection between the Kinect and Autoencoder was never broken. The

network processed and generated three point clouds simultaneously every three seconds. While the number of navigation options via the control panel was quite limited compared to any digital 3D environment (turn-table and scale function), visitors often found it difficult to grasp the buttons' functionalities. In terms of "immediacy", the hybrid station met our expectations, but when we applied a higher interpolation factor to the scanned clouds, the resulting distortions made the real-time representation less obvious.



Fig. 6. Three images from the exhibition, where the concept modelling station was tested by a non-professional audience.

3.2 Tower Design

To test the workflow and the modelling station in an architectural design context, a brief for a tower design was defined. This more professional format of the study aims to explore the benefits of a collaborative, yet hybrid design approach at the early stages of the design. The design process starts with composing coloured wooden blocks on the table. In the example (Fig. 7), the colours are used to define different geometric characteristics in the compositions. As the blocks are placed on the table, the displayed point clouds give an alternative view of the composition. Although the building blocks and the table size limit the scale of the model, the scalability of the digital model allows for spatial interpretations regardless of the physical size of the blocks. The designers can select and assign specific 3D patterns, porosities, etc. to each of the assembled colour compositions on the table. The autoencoder (ML), controlled by the user modifies the compositions through the interpolation of selected features. The resulting Machine-interpreted point clouds are displayed on the screen in real-time.

During the collaborative tower sketching process a variety of the different compositions - both the original and the machine interpreted - were frequently saved in .xyz point cloud format. This opened up the integration into existing design software for further processing of promising concepts.



Fig. 7. Top: Variation of scanned concepts for a tower design. Middle: Transformation toward selected design point cloud via autoencoder. Bottom left: Outcome of the transformation. Bottom right: Post-processed point cloud in Unity. The depicted process is running in real-time and is responsive.

4 Conclusion

In this paper, we present a workflow that facilitates a real-time interaction with an artificial neural network through physical modelling. The user-assembled physical objects allows for an almost instant design dialogue with the trained data of the neural network through the support of capturing devices and point cloud notation. This establishes an immediate feedback loop between human and machine intelligence, introducing a hybrid

immediacy that reintegrates physical model building at the centre of a digital-focused design process. The work shows the potential of using Kinects to capture the physical boundary of a model, while simultaneously using machine learning to apply selected geometric structures, and 3D patterns to it. The immediacy of the setup creates an intuitive way to physically search for conceptual design ideas without the need to remodel design features digitally.

However, during this research, we have encountered a number of limitations that need to be acknowledged. To precisely capture a physical composition from all angles, two Kinects provide a limited field of view, resulting in several unscanned areas when the physical models get more complex. This could be remedied by more Kinects but would result in an increased performance cost. Concerning the developed artificial neural network, the challenges are equally conceptual and technical. Though the integration of the network with the scanned boundary compositions works very well, the conceptual idea of blending a selected geometric structure into it was difficult to execute technically. Partially, this is a result of the inherent dilemmas within computing on large datasets, where the network gets biased towards the most common data.

Nevertheless, the work is still in its early phases, and further developments to the demonstrated workflow could minimise the listed limitations and significantly expand on its useability within a design scenario. Additionally, we plan to expand the interactions with the produced concept models with the use of a mixed reality (MR) 3D sketching application. The sketching app combines a familiar stylus-on-tablet input and interaction paradigm with MR capabilities to allow sketching directly in 3D space, thus providing additional tools for amending and refining the concept models in real-time.

Acknowledgements. This work was funded by the Austrian Science Fund (FWF) project F77 (SFB “Advanced Computational Design”).

References

- Bühlmann, V.: Die Empörung des Modells // models, outraged“ – Zhdk conference “David and Goliath – Models between art and architecture. <https://monasandnomos.org/2013/11/19/die-emporung-des-modells-models-outraged-abstract-for-my-paper-at-the-zhdk-conference-david-and-goliath-models-between-art-and-architecture/> (2013)
- Carpó, M.: *The Second Digital Turn: Design Beyond Intelligence*. The MIT Press, Cambridge (2017)
- Frazer, J.: *An Evolutionary Architecture*. Architectural Association (1995)
- Hirschberg, U., Hovestadt, L., Fritz, O. (eds.): *Atlas of Digital Architecture: Terminology, Concepts, Methods, Tools, Examples, Phenomena*. Birkhauser (2020)
- Hsieh, C.T.: A new kinect-based scanning system and its application. *Appl. Mech. Mater.* **764–765**, 1375–1379 (2015). <https://doi.org/10.4028/www.scientific.net/AMM.764-765.1375>
- Jeffrey, L., Hakon, D., Hakon, F., Stanislas, C.: (2020)
- Kanopoulos, N., Vasanthavada, N., Baker, R.L.: Design of an image edge detection filter using the Sobel operator. *IEEE J. Solid-State Circuits* **23**(2), 358–367 (1988). <https://doi.org/10.1109/4.996>

- Leach, N.: *Architecture in the Age of Artificial Intelligence: An introduction to AI for architects*. Bloomsbury Publishing Plc (2021). <https://doi.org/10.5040/9781350165557>
- Oxman, R.: Digital architecture as a challenge for design pedagogy: theory, knowledge, models and medium. *Des. Stud.* **29**, 99–120 (2008). <https://doi.org/10.1016/j.destud.2007.12.003>
- Souza, L., Pathirana, I., Mcmeel, D., Amor, R.: *Kinect to Architecture* (2011)
- Stachowiak, H.: *Allgemeine Modelltheorie*. Springer, Vienna (1973). <https://doi.org/10.1007/978-3-7091-8327-4>
- Stavrić, M., Sidanin, P., Tepavčević, B.: *Architectural Scale Models in the Digital Age: Design, Representation and Manufacturing*. Springer (2013)
- Thomsen, M.R., Tamke, M.: The active model: a calibration of material intent. In: *Persistent Modelling: Extending the Role of Architectural Representation*, 1st edn. Routledge (2012)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

