Multi-Modal 3D Neural Representations for Scene Modeling: Towards Building Health Management and Energy Evaluation

Chenghao Xu[1], Malcolm Mielle[2], and Olga Fink[3]

[1,3] *EPFL – Intelligent Maintenance and Operations Systems (IMOS)*
Lausanne, CH-1015, Switzerland
chenghao.xu@epfl.ch   olga.fink@epfl.ch

[2] *Schindler EPFL Lab, Lausanne, Switzerland, CH-1015, Switzerland*
malcolm.mielle@ik.me

ABSTRACT

Prognostics and Health Management (PHM) for buildings requires accurate digital representations that support energy-efficiency assessment and predictive analysis throughout the asset lifecycle. This is particularly important for building envelopes, whose condition directly affects thermal performance, energy efficiency, occupant comfort, and long-term structural health. However, most existing buildings lack simulation-ready digital models, while current inspection and modeling workflows remain labor-intensive, geometrically limited, and inadequate for integrating thermal and physical information. This thesis proposes a cost-efficient and automated framework for reconstructing simulation-ready, multi-modal three-dimensional building representations from visual inputs alone. The research is organized around three main contributions. First, it develops semantic scene reconstruction methods that combine neural implicit surface modeling with semantic prediction to recover detailed 3D building envelopes and estimate key characteristics such as window-to-wall ratio and footprint. Second, it investigates learning-based estimation of thermophysical parameters through implicit thermal field reconstruction and differentiable heat-transfer simulation. Third, it addresses practical sensing constraints by enabling multi-modal reconstruction from limited and unsynchronized thermal observations.

1. MOTIVATION AND PROBLEM STATEMENT

Prognostics and Health Management (PHM) aims to monitor asset conditions, identify degradation, diagnose faults, and support predictive decision-making throughout the asset lifecycle (Fink et al., 2026). In the built environment, PHM is increasingly important because buildings are long-lived, energy-intensive assets whose envelope conditions strongly affect thermal performance, energy efficiency, occupant comfort, and long-term structural health (Nardi, Lucchi, Rubeis, & Ambrosini, 2018). This need is especially urgent in Europe, where building operations account for 40% of total energy consumption and 35% of greenhouse gas emissions, as reported by the European Commission (EC, 2024). These figures underscore the importance of both energy-efficient design and effective retrofitting for achieving climate neutrality.

However, effective PHM for buildings remains limited by the lack of accurate, scalable, and updatable digital representations of existing assets. Most buildings lack detailed Building Information Modeling (BIM) or Building Energy Modeling (BEM) models, while generating such models remains labor-intensive, expert-driven, and costly (C. Xu et al., 2025). This limitation is particularly critical for building envelopes, where degradation mechanisms such as insufficient insulation, thermal bridges, material aging, facade damage, and window-related heat losses often remain undetected without detailed geometric and thermal characterization.

Moreover, a functional digital building model for energy analysis and renovation requires additional inputs, including key building characteristics, such as orientation, window-to-wall ratio, and building footprint, as well as physical parameters, such as thermal resistance and material properties. These variables must be accurately measured or estimated to faithfully represent real-world building systems and enable continuous monitoring and predictive forecasting. However, existing inspection and modeling workflows still rely heavily on manual surveys, sparse measurements, or simplified assumptions (Marino, Nucara, & Pietrafesa, 2017; Chi et al., 2020). Together, these challenges underscore the urgent need for **a cost-efficient and automated solution that can rapidly generate energy-oriented digital building models from scratch, particularly for existing structures**

---

without prior digital documentation.

Recent advances in visual 3D representation techniques, such as Neural Radiance Fields (Fan et al., 2022) and Gaussian Splatting (Kerbl, Kopanas, Leimkühler, & Drettakis, 2023), have demonstrated remarkable capability in capturing detailed representations of complex environments and advancing 3D scene understanding. These neural representations effectively overcome the limitations of traditional point-cloud-based methods in handling sparse image inputs and complex geometries. However, their applications have largely been confined to domains such as robotics (Wang et al., 2024), human avatars (Sun et al., 2022), and autonomous driving (Tonderski et al., 2024), leaving a significant research gap in their application to building digitalization. Addressing this gap is critical, as these techniques hold substantial potential for creating highly accurate and detailed 3D building models that can be directly integrated into simulation workflows and support energy-efficiency assessment.

Moreover, current neural representations primarily focus on geometry and appearance, which are insufficient for thermal simulation and energy-oriented analysis. In contrast, PHM requires the integration of multi-modal information, as degradation is rarely observable from a single modality alone (Fink et al., 2026). In building applications, combining RGB, semantic, and thermal data can provide complementary information: RGB imagery captures appearance and geometric cues, semantic information identifies building components, and thermal imagery reveals heat-transfer patterns and anomalies associated with envelope defects or material deterioration. Beyond multi-modal scene understanding, effective PHM also depends on estimating latent physical parameters, such as thermal conductivity, diffusivity, and boundary conditions, which govern heat-transfer behavior. However, vision-based estimation of such parameters has received limited attention in the context of buildings. Without these physically grounded parameters, PHM systems remain largely confined to descriptive inspection, rather than enabling reliable diagnosis and prognosis.

In summary, this doctoral thesis aims to develop a cost-efficient and automated framework for generating digital building models for energy-efficiency evaluation from image-based observations. The proposed research will address scene reconstruction for complex geometries, accurate estimation of building characteristics and thermophysical parameters, and integration with heat-transfer simulation. The significance of this research lies in its potential to make advanced building energy analysis more scalable and accessible. By combining recent advances in 3D reconstruction, multi-modal learning, and physical modeling, the proposed framework is expected to support decision-making, renovation planning, and performance optimization for both existing and future buildings.

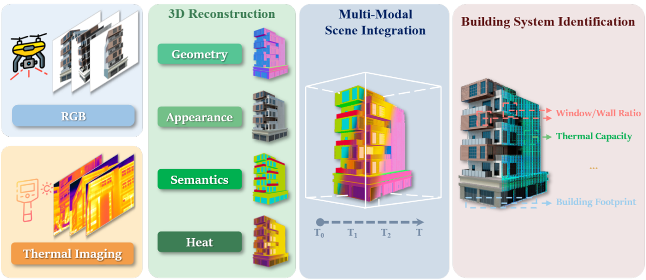

Figure 1. Overview of the proposed framework for generating multi-modal 3D building representations and identifying building system parameters from RGB and thermal imagery.

2. EXPECTED CONTRIBUTIONS

The primary objective of this thesis is to **integrate multi-modal visual information into three-dimensional scene representations for building modeling and analysis**. Beyond conventional approaches that focus mainly on geometry or appearance reconstruction, this thesis investigates how high-fidelity 3D representations can support downstream building applications. Specifically, the research addresses two central questions: how multi-modal visual inputs can be efficiently integrated into implicit scene representations, and how the resulting representations can be exploited for building-specific analysis tasks. The main pipeline of the proposed framework is illustrated in Fig. 1.

2.1. Semantic Scene Reconstruction for Building Characteristics Estimation

Building characteristics, such as orientation, dimensions, and window-to-wall ratio (WWR), are key inputs for energy-consumption analysis (Choi, Cho, & Kim, 2012), sustainable building design (Bank, McCarthy, Thompson, & Menassa, 2010; Al-Saggaf, Taha, Hegazy, & Ahmed, 2020), and retrofit planning (Pichugin, 2018). However, these applications typically assume that such characteristics are known a priori. Existing methods remain largely limited to 2D planar analysis or simplified 3D approximations, making them insufficient for capturing the geometric complexity and component-level details of real-world buildings.

The first contribution of this thesis addresses this gap by *developing a framework to estimate building characteristics from a fully 3D spatial perspective*. By integrating neural implicit surface reconstruction with semantic modeling, the proposed approach recovers detailed geometry and component-level information, even for buildings with complex geometries or partial occlusions. Based on the reconstructed surfaces, a mesh-based analysis method will be introduced to estimate key building characteristics, including the window-to-wall ratio and building footprint. To support systematic evaluation, the thesis will also provide a dedicated synthetic building dataset with ground-truth characteristics and a corresponding benchmarking pipeline.

2.2. Learning-based Physical Parameter Estimation for Thermal Simulation

Accurate identification of material-specific physical parameters is essential for heat-transfer simulation in buildings. Parameters such as thermal conductivity, thermal diffusivity, and boundary conditions govern how heat propagates through building components and play a critical role in replicating real-world thermal behavior. However, obtaining these parameters through direct measurement is often impractical, particularly for existing buildings, where observations are typically limited to sparse surface-level thermal data.

The second contribution focuses on estimating such thermophysical parameters from thermal image sequences. Since these latent properties cannot be directly inferred from surface temperatures alone, the proposed framework reconstructs continuous implicit 3D thermal fields from sparse observations and couples them with differentiable heat-transfer simulation. By combining implicit neural representations with physics-based constraints, the method aims to infer internal thermal states and material-related parameters relevant to building performance. In addition, the integration of semantic scene information will support parameter estimation across different building components and materials, thereby improving the realism and applicability of thermal simulation for building analysis.

2.3. Multi-Modal Neural Thermal Scene Reconstruction

Existing thermal scene reconstruction methods (Hassan, Forest, Fink, & Mielle, 2024; J. Xu, Liao, Prabhakar, & Patel, 2024; Lin, Pan, Fridovich-Keil, & Wetzstein, 2024; Ye et al., 2024) primarily focus on thermal novel view synthesis, aiming to generate high-fidelity thermal images from unseen viewpoints rather than explicit 3D representations. As a result, these methods do not directly provide surface-based or volumetric models, such as meshes or structured thermal fields, that are suitable for simulation-driven energy analysis. Without such structured geometric representations, NeRF-based thermal reconstructions lack the spatial fidelity and physical interpretability required for building envelope characterization and heat-transfer simulation.

The final contribution of this thesis addresses the practical challenge of reconstructing multi-modal scenes when thermal inputs are sparse, unsynchronized, or not tightly calibrated with RGB imagery. To this end, this thesis will develop a thermal-specific alignment pipeline that combines temperature-derived structural features with geometric priors from monocular 3D reconstruction to estimate thermal camera poses under challenging sensing conditions. In addition, uncertainty-aware strategies will be investigated to identify and filter unreliable thermal observations, thereby improving reconstruction consistency and robustness. Finally, the work will explore methods for converting implicit thermal scene representations into explicit 3D thermal models, enabling their integration into downstream energy analysis and simulation workflows.

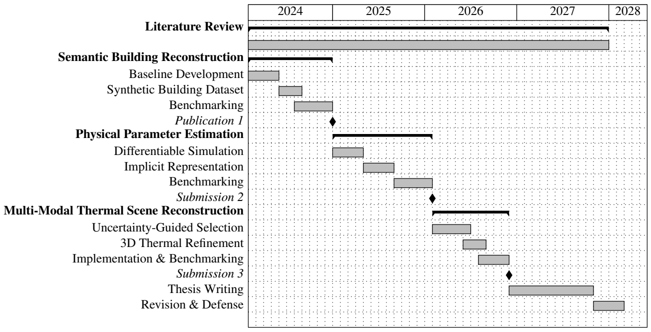

Figure 2. Planned research timeline and major milestones.

3. PROPOSED RESEARCH PLAN

3.1. Time Plan

This PhD research began in February 2024. As of March 2026, the first contribution has been completed and published in *Building and Environment*, while the second contribution is currently being prepared for submission. The remaining work will focus on multi-modal thermal scene reconstruction and thesis writing. The detailed schedule is shown in Fig. 2.

3.2. Current Progress

To date, the first contribution has been successfully completed and published. This work focused on semantic scene reconstruction for estimating building envelope characteristics from sparse visual observations. Specifically, it introduced a two-stage framework that reconstructs semantic building envelopes in 3D and estimates key envelope characteristics from the reconstructed scene. The framework combines an SDF-based neural surface representation with an additional semantic network, enabling the joint recovery of building geometry, appearance, and semantic attributes within a unified 3D representation. On top of this representation, a mesh-based analysis pipeline was developed to estimate characteristics such as the window-to-wall ratio (WWR) and building footprint from a fully 3D spatial perspective. In addition, a synthetic benchmark dataset was created for evaluation, comprising 28 diverse 3D building models, each rendered with 200 multi-modal images and annotated with ground-truth building characteristics. Experimental results demonstrated that the proposed BuildNet3D framework generalizes well across different building types and achieves accurate WWR estimation, thereby establishing a strong foundation for the subsequent stages of the thesis.

The second work package is currently in draft form and focuses on learning-based estimation of thermophysical parameters from imagery. It proposes a physically grounded framework that treats neural thermal scene representations not merely as rendering tools, but as intermediate physical models for inferring material and boundary parameters from temperature observations. Specifically, the method jointly reconstructs geometry and time-resolved temperature fields, and couples them with a heat-transfer model to estimate thermal properties directly from the reconstructed scene. The framework consists of three main components: SDF-based geometric reconstruction from RGB images, differentiable forward heat-transport simulation, and inverse estimation of thermal diffusivity through this differentiable formulation. The draft also includes a preliminary experimental design using both synthetic and real data, together with comparisons against existing inverse estimation baselines. Overall, this second contribution extends the thesis from geometric and semantic scene understanding toward physically grounded thermal modeling, laying the methodological basis for predictive simulation and PHM-oriented analysis.

REFERENCES

Al-Saggaf, A., Taha, M., Hegazy, T., & Ahmed, H. (2020). Towards sustainable building design: the impact of architectural design features on cooling energy consumption and cost in saudi arabia. *Procedia Manufacturing*, *44*, 140–147.

Bank, L. C., McCarthy, M., Thompson, B. P., & Menassa, C. C. (2010). Integrating bim with system dynamics as a decision-making framework for sustainable building design and operation. In *Proceedings of the first international conference on sustainable urbanization (icsu)* (pp. 15–17).

Chi, F., Wang, Y., Wang, R., Li, G., Peng, C., & A. (2020). An investigation of optimal window-to-wall ratio based on changes in building orientations for traditional dwellings. *Solar Energy*, *195*, 64-81. doi: https://doi.org/10.1016/j.solener.2019.11.033

Choi, I. Y., Cho, S. H., & Kim, J. T. (2012). Energy consumption characteristics of high-rise apartment buildings according to building shape and mixed-use development. *Energy and Buildings*, *46*, 123–131.

EC. (2024). *Energy performance of buildings directive* (Tech. Rep.).

Fan, Z., Wang, P., Gong, X., Jiang, Y., Xu, D., & Wang, Z. (2022). Nerf-sos: Any-view self-supervised object segmentation from complex real-world scenes. *arXiv e-prints*, arXiv–2209.

Fink, O., Sharma, V., Nejjar, I., Von Krannichfeldt, L., Garmaev, S., Zhang, Z., … Steiner, K. (2026). From physics to machine learning and back: Part i - learning with inductive biases in prognostics and health management (phm). *Reliability Engineering System Safety*, *271*, 112213. doi: https://doi.org/10.1016/j.ress.2026.112213

Hassan, M., Forest, F., Fink, O., & Mielle, M. (2024). *Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis.*

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023, July). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, *42*(4).

Lin, Y. Y., Pan, X.-Y., Fridovich-Keil, S., & Wetzstein, G. (2024). ThermalNeRF: Thermal radiance fields. In *Ieee international conference on computational photography (iccp).*

Marino, C., Nucara, A., & Pietrafesa, M. (2017). Does window-to-wall ratio have a significant effect on the energy consumption of buildings? a parametric analysis in italian climate conditions. *Journal of Building Engineering*, *13*, 169-183. doi: https://doi.org/10.1016/j.jobe.2017.08.001

Nardi, I., Lucchi, E., Rubeis, T., & Ambrosini, D. (2018). Quantification of heat energy losses through the building envelope: A state-of-the-art analysis with critical and comprehensive review on infrared thermography. *Building and Environment*, *146*, 190–205.

Pichugin, S. (2018). Reliability estimation of industrial building structures. *Magazine of Civil Engineering*(7 (83)), 24–37.

Sun, M., Yang, D., Kou, D., Jiang, Y., Shan, W., Yan, Z., & Zhang, L. (2022). Human 3d avatar modeling with implicit neural representation: A brief survey. In *2022 14th international conference on signal processing systems (icsps)* (pp. 818–827).

Tonderski, A., Lindström, C., Hess, G., Ljungbergh, W., Svensson, L., & Petersson, C. (2024). Neurad: Neural rendering for autonomous driving. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 14895–14904).

Wang, G., Pan, L., Peng, S., Liu, S., Xu, C., Miao, Y., … Wang, H. (2024). Nerf in robotics: A survey. *arXiv preprint arXiv:2405.01333.*

Xu, C., Mielle, M., Laborde, A., Waseem, A., Forest, F., & Fink, O. (2025). Exploiting semantic scene reconstruction for estimating building envelope characteristics. *Building and Environment*, *275*, 112731. doi: https://doi.org/10.1016/j.buildenv.2025.112731

Xu, J., Liao, M., Prabhakar, K. R., & Patel, V. M. (2024). *Leveraging thermal modality to enhance reconstruction in low-light conditions.* Retrieved from `https://arxiv.org/abs/2403.14053`

Ye, T., Wu, Q., Deng, J., Liu, G., Liu, L., Xia, S., … Pei, L. (2024). Thermal-nerf: Neural radiance fields from an infrared camera. In *2024 ieee/rsj international conference on intelligent robots and systems (iros)* (p. 1046-1053). doi: 10.1109/IROS58592.2024.10802480