# Comparison of Machine Learning and Physics-Based Approaches for Thermal Infrared Simulation for Urban Digital Twins

Benedikt Kottler[1], Raphael Zipperer[1], Eva Strauss[1], Dimitri Bulatov[1]

[1] Fraunhofer IOSB, 76725 Ettlingen, Germany - (name.surname@iosb.fraunhofer.de)

**Keywords:** Thermal Infrared Simulation, RandLA-Net, InfraGAN, Multispectral, Remote Sensing, Validation

## Abstract

Thermal simulation in urban digital twins enables effective monitoring of surface urban heat islands and supports climate adaptation planning. This paper evaluates machine learning and physics based approaches for this task through a unified validation framework based on 3D point clouds applied to an urban region in Berlin. The framework enables comparison of RandLA Net for 3D point cloud processing, InfraGAN for 2D texture synthesis, and physics based simulation on triangulated mesh geometries. RandLA Net architecture is adapted for thermal prediction and tested with two feature sets: RGB only and RGB with physics derived material parameters. Deep learning methods demonstrate severe spatial overfitting: training errors are minimal (MAE less than 1 K), but test performance degrades significantly on unseen regions with MAE increasing by factors of 1.9 to 2.5. Unexpectedly, augmenting with material parameters worsens generalization, indicating inadequate feature integration. Physics based simulation maintains consistent predictions (MAE approximately 8 K) with systematic bias addressable through calibration. These results motivate hybrid approaches embedding physical constraints into neural architectures for robust urban thermal modeling.

## 1. Introduction

Digital twins have rapidly emerged as one of the most relevant topics in modern, application-oriented research. In both civil and military contexts, digital twin technology enables realistic simulation and consequent training for operational scenarios, image analysis and interpretation, efficient data analysis even with limited personnel, terrain assessment, decision support, and many other applications (see e.g. Zhong et al. (2023); Giberna et al. (2025)). Especially in airborne remote sensing scenarios, temporal modeling of temperatures represents a sensible and valuable tool at macro-, meso-, and micro-scales (see Li et al. (2023); Gonzalez-Caceres et al. (2025); Kong and Hucks (2023))

One important application within urban environments is the detection and monitoring of Surface Urban Heat Islands (SUHI), a phenomenon where surface temperatures within urban environments permanently exceed those of their local surroundings. The reasons for SUHI formation are manifold: The abundance of sealed, impervious materials, the scarcity of vegetation, the release of anthropogenic heat, and atmospheric pollution (Masson, 2006). Additionally, the three-dimensional structure of urban environments plays a critical role, as hot air becomes enclosed in street canyons through processes known as heat trapping and multiple radiations (Hidalgo et al., 2008). Because of the clearly three-dimensional character of urban scenery and the spatial propagation of heat phenomena, it would be highly efficient to possess a three-dimensional model of the area in question. Such a model would enable temporal simulation of heat development under varying material compositions, weather conditions, and other environmental factors. This is a typical scenario for the application of Digital Twins, as discussed in **?**Bulatov et al. (2020); Cárdenas-León et al. (2024) or Liu et al. (2025). In these publications, digital twins were computed from aerial data, which is particularly attractive since such data are relatively up-to-date, available at large scales, and cost-efficient compared to ground-based surveys. Physics-based modeling is considered the gold standard for measuring thermal properties of a scene, because physical laws can be applied, physical properties of all materials measured, and weather conditions retrieved for any point in time. In practice, however, we have to deal with incomplete, fragmentary information. The relevant scene parameters need to be estimated; additional, often less accurate, sources of information, such as building cadasters, must be taken into account; and simplified workflows are being applied to save computational time.

The generation of thermal and infrared images can increasingly be achieved using machine learning (ML) and, especially, deep-learning-based methods, which have become widely popular for style transfer tasks. While tremendous progress has been made in both 2D and 3D applications, the availability of up-to-date data for spectral ranges beyond the visual spectrum represents a particular challenge and constraint for scene modeling. For the visual spectrum (RGB), deep learning is well-established in image analysis thanks to the existence of numerous large-scale datasets that enable training of robust models. In contrast, the limited quantity and sometimes questionable quality of adequate thermal imagery remain the main counterargument against the use of deep learning for thermal modeling in digital twin contexts (Strauß et al., 2023; Pettirsch and Garcia-Hernandez, 2025). Only recently, thanks to sophisticated data augmentation, (cross)-domain adaptation techniques, and highly-specialized datasets (Mirabet-Herranz and Dugelay, 2024), few richly detailed entities, such as humans or vehicles, are increasingly being modeled using AI tools (Fischer et al., 2025).

ML-based methods promise both computational efficiency and spatially detailed results, but they depend heavily on the quality and quantity of training data. Critically, generalization challenges remain largely unexplored in thermal simulation contexts, particularly when models are applied to different urban environments or weather conditions not present in the training set. A systematic comparison between these different paradigms under controlled validation conditions is currently lacking.

This paper presents findings from a comparison of physics-based thermal simulation of Bulatov et al. (2024) and ML-based approaches within a unified validation framework. We evaluate InfraGAN (Özkanoğlu and Ozer, 2022) as a representative image-to-image translation method for 2D thermal synthesis, and a RandLA-Net-based approach (Hu et al., 2020) as a state-of-the-art method for 3D point cloud semantic segmentation. Style transfer is formulated as a classification problem with appropriately discretized temperature categories. We evaluate both the baseline feature set (XYZ coordinates and their RGB values) and an extended feature set that includes material classes retrieved from land cover classification and material databases. Thus, the input is given by the minimum requirements to keep the workflow as general as possible, allowing for replacing the aforementioned ML-based methods by more advanced (e.g., transformer-based) ones in the future. The Berlin-Moabit dataset, acquired by the DLR Institute of Optical Sensor Systems, includes synchronized RGB (summer 2017) and thermal (winter 2019) imagery with geometrically consistent, watertight 3D building models and corresponding thermal and RGB textures, that enable the validation on a common geometric domain.

We present a comparison of physics-based and ML-based approaches for thermal simulation through a unified validation framework based on 3D point clouds. Our first main contribution, described in Section 4, is the data preparation of the Moabit dataset into a format proper for such an evaluation: From textured meshes to a point cloud with multi-sensor color values. Section 5 is then dedicated to the calculation of temperatures using ML-based methods. Especially, an adapted RandLA-Net architecture is directly applied to point clouds for temperature prediction in Section 5.1 for the first time to the best of our knowledge, contribution our second contribution. But before we do that, most relevant sources of previous research and preliminary work describing the necessary input are provided in Sections 2 and 3, respectively. The findings, provided in Section 6, reveal both promising results and challenges, providing insights for future hybrid approaches that combine the interpretability of physics-based methods with the efficiency of machine learning. Finally, Section 7 summarizes the article content and outlines directions of future research.

## 2. Related Works

### 2.1 Physics-Based Thermal Simulation

Physics-based approaches compute surface temperatures from first principles using thermodynamic models, e.g. radiative transfer theory, and coupled heat transfer equations. The Thermal 4D Digital Twin framework of Bulatov et al. (2024) simulates urban thermal behavior by solving heat transfer equations with meteorological boundary conditions derived from weather data and surface properties. Kottler et al. (2019) demonstrated physically-based radiative transfer modeling for complex urban scenes. Recent studies have further demonstrated purely physics-based urban thermal simulations: CFD models that explicitly couple radiative, convective, and conductive heat fluxes for microclimate analysis (Nascimento et al., 2025), one-dimensional urban surface heat balance models applied to metropolitan areas (Hirano et al., 2024), and investigations highlighting the importance of radiative transfer processes in urban canopy models (Salim et al., 2022). Urban heat island identification and monitoring using physics-based simulation has been successfully demonstrated (Vogel et al., 2025; Joshi et al., 2025; Bulatov et al., 2024). These approaches provide interpretable results and possess a key advantage: They can compute temperatures for different weather conditions without requiring any training data. However, their accuracy is fundamentally limited by the availability of detailed material property information, which is often incomplete or unavailable at the building-element level. The primary limitations are computational cost and the uncertainty in material parameter acquisition, which can lead to spatially over-generalized results.

### 2.2 Deep Learning for Point Cloud Processing

Recent advances in deep learning for point cloud processing have demonstrated the potential of 3D neural architectures for physical parameter estimation from unstructured spatial data. RandLA-Net (Hu et al., 2020) established an efficient approach for large-scale point cloud segmentation through random sampling and local feature aggregation, forming the foundation for numerous subsequent applications. Building on this architecture, Oehmcke et al. (2024) achieved superior biomass estimation using 3D-Minkowski CNNs on unstructured LiDAR data, demonstrating the feasibility of 3D neural networks for environmental parameter prediction. In parallel, Liu et al. (2024) combined multiple foundation models for urban tree biomass prediction, showing the potential of ensemble approaches. More recently, Li et al. (2024) introduced hybrid NeFF-BioNet models combining 3D-CNNs with Transformer architectures, achieving 6–7% improvements over state-of-the-art methods on biomass estimation tasks. These works collectively demonstrate the feasibility of point cloud-based regression for environmental parameter estimation. However, thermal simulation represents a significantly more challenging task due to temporal variability in surface properties and the complex nonlinear nature of radiative transfer processes, which existing point cloud methods have not yet been applied to systematically.

### 2.3 Style Transfer for Thermal Image Generation

ML-based approaches for thermal image synthesis primarily employ generative adversarial networks (GANs) and image-to-image translation methods. InfraGAN (Özkanoğlu and Ozer, 2022) performs paired image-to-image translation from RGB to thermal imagery, effectively learning style transfer requiring perfectly aligned training pairs. Various extensions have incorporated improved loss functions for enhanced image fidelity (Han et al., 2024; Fischer et al., 2025) Recent advances in cross-spectral translation such as DGC_GAN demonstrate improved unpaired visible-to-thermal image synthesis for airborne and drone imagery (Yao et al., 2026), and GAN-based nighttime thermal image translation methods further illustrate the potential of deep generative models for thermal image generation (Yang et al., 2024).

Andriambololonaharisoamalala et al. (2025) demonstrated a high-resolution thermal image sharpening framework supported by generative techniques in an urban case study in Perth, Australia, further highlighting the potential of GAN-based methods for improving thermal imagery resolution and quality. These methods achieve significant computational efficiency compared to physics-based simulation, but they require extensive paired or unpaired training data and face generalization issues when applied to urban environments with different characteristics than those in the training set. The extension of style transfer methods to 3D point cloud representations remains largely unexplored in the thermal imaging context.

# 3. Preliminaries

## 3.1 Source Data

The Berlin-Moabit dataset originates from a summer and winter nighttime thermal infrared acquisition conducted by DLR Institute of Optical Sensor Systems. For this study, only the winter data from March 5 and 6, 2019 is used. Two Infratec VarioCam HD sensors captured 16-bit grayscale imagery encoding absolute surface temperatures at effective ground sampling distances of 35 cm to 2 m per pixel in nadir and forward-looking oblique (30° off-nadir) configurations. Following radiometric preprocessing and GPS/IMU bundle adjustment, Semi-Global Matching (SGM) processed 1669 TIR images at 80% overlap into dense 3D point clouds with linear intensity stretching applied prior to stereo matching for contrast enhancement.

Automated building reconstruction (Frommholz et al., 2016, 2017) generated watertight OBJ models with Constrained Delaunay Triangulation (CDT) tessellation of planar facade elements. The output comprises:

- **Geometry:** Shared 3D vertices $\mathbf{v} \in \mathbb{R}^3$ and face indices $\mathbf{f}_v$ of scene traingles across RGB (summer 2017) and TIR modalities
- **Textures:** RGB UV coordinates $\mathbf{v}_t^{\text{RGB}} \in [0, 1]^2$, texture faces $\mathbf{f}_{vt}^{\text{RGB}}$, and four 3-channel 16-bit PNG atlases. Furthermore, TIR UV coordinates $\mathbf{v}_t^{\text{IR}} \in [0, 1]^2$, texture faces $\mathbf{f}_{vt}^{\text{IR}}$, and four 1-channel 16-bit PNG atlases
- **Orthophotos/DSM:** RGB/TIR true-ortho mosaics at 10 cm and 1 m GSD respectively, DSM at 1 m resolution

The consistent spatial framework with identical vertices and topology across modalities enables geometry-preserving transformations via per-triangle affine UV mappings, uniform point sampling using barycentric coordinates, and bilinear texture interpolation. This dataset is the base for the subsequent multimodal point cloud generation pipeline described in section 4.2.

## 3.2 Thermal Simulation

The physics-based simulation in this dataset is described in Bulatov et al. (2024), where a thermodynamic model computing surface temperatures through the one-dimensional heat transfer equation solved with the Finite Volume Method (Kottler et al., 2019). The approach accounts for conductive, convective, and radiative fluxes for each triangle of the scene mesh, integrating historical weather data (air temperature, wind speed, cloud coverage), collected from Section 3.2, and material-specific physical parameters. Material assignment combines land cover classification via two-branch DeepLab v3+ (Chen et al., 2018; Qiu et al., 2022), roof material classification using Random Forest (Breiman, 2001) on interactively selected samples, and simplified per-building façade material assignment. The vector containing the material parameters (density, thermal conductivity, specific heat capacity, emissivity, albedo, core temperature, and surface element depth) is denoted by $\boldsymbol{\theta}$ and the typical ranges of the relevant parameters are contained in Table 1. The material parameter vector $\boldsymbol{\theta}$ is defined as constant per material, which leads to an overgeneralization of the intrinsic variability within a material class. Original validation against thermal infrared imagery yielded MAE 9.35 K, attributed to registration errors, measurement artifacts, simplified vegetation/convection models, and coarse material classification (one material per roof). The systematic bias and climate-specific calibration requirements motivate the investigation of ML-based alternatives that might learn data-driven corrections.

# 4. Dataset Preparation

## 4.1 Unified Texture Mapping

To enable direct comparison between RGB and TIR modalities despite differing UV parameterizations, TIR textures were transformed into the RGB coordinate system through per-triangle affine mappings. For each triangle face, the image-to-image transformation (homography) $\mathbf{T}$ was given by:

$$\mathbf{T} = [\mathbf{A}_{\text{RGB}}]^{-1}\mathbf{B}_{\text{IR}}, \tag{1}$$

where $\mathbf{A}_{\text{RGB}} \in \mathbb{R}^{3\times3}$ contains homogeneous RGB-UV coordinates $[\mathbf{v}_t^{\text{RGB}}, \mathbf{1}]$ of the triangle vertices, and $\mathbf{B}_{\text{IR}} \in \mathbb{R}^{3\times2}$ contains corresponding TIR-UV coordinates $\mathbf{v}_t^{\text{IR}}$. For each pixel within the RGB-UV triangle bounding box, barycentric coordinates determined interior membership, followed by an affine transformation to source TIR-UV space and bilinear interpolation of 16-bit temperature values. Regions outside TIR coverage were interpolated using harmonic inpainting, given, for example, by the `regionfill` funcion in MATLAB. This procedure unified the four texture atlases per modality into aligned texture maps at RGB resolution, enabling spatially consistent multi-modal point sampling.

## 4.2 Point Cloud Generation

Building upon the unified RGB-TIR texture mapping established in the previous step, point clouds were generated by transforming 2D texture coordinates back to 3D space and combining geometric, radiometric, and thermal attributes.

**4.2.1 Building Surfaces** For each triangle face, pixels within the triangle boundary in RGB-UV texture space are mapped back to 3D coordinates. The triangle vertices define a plane in 3D space, and a plane equation is fitted via Direct Linear Transformation (DLT) to obtain the plane normal and position.

A projection matrix $\mathbf{P}_0$ is computed that orthogonally projects 3D points onto this plane and represents them in a local 2D coordinate system. The triangle vertices are now available in two representations: as texture coordinates $(s, t)$ from the RGB-UV parameterization, and as 2D plane coordinates. A homography transformation $\mathbf{H}$ is computed using the three triangle vertices as correspondence pairs, enabling transformation of any texture pixel to plane coordinates.

Table 1. Ranges for material parameters $\boldsymbol{\theta}$.

| Parameter | Symbol | Typical Range |
|---|---|---|
| Density | $\rho$ | 500-2500 kg/m³ |
| Heat Capacity | $c_v$ | 800-2000 J/(kg·K) |
| Albedo | $\alpha$ | 0.1-0.9 |
| Thermal Conductivity | $\lambda$ | 0.1-2.0 W/(m·K) |
| Emissivity | $\varepsilon$ | 0.85-0.95 |
| Core Temperature | $T_{\text{core}}$ | 273-300 K |
| Depth | $d$ | 0.05-0.5 m |


https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-395-2026 | © Author(s) 2026. CC BY 4.0 License.

For each pixel $(s, t)$, the corresponding 3D point is recovered by solving an overdetermined system: the point must both project to the correct plane coordinates and lie within the fitted plane. This system is solved via Moore-Penrose pseudoinverse. RGB triplets and TIR temperatures are extracted via bilinear interpolation from the unified texture atlases at pixel coordinates $(s, t)$.

**4.2.2 Ground Surfaces** A binary building footprint mask was generated by rasterizing all building triangles onto the orthophoto grid using filled polygon rendering. Morphological operations (dilation with 4-pixel disk structuring element, closure with 200-pixel disk, hole filling) extracted contiguous ground regions after subtracting the building mask. Vegetation such as trees was implicitly included in the DSM rather than modeled as separate geometric entities. Points were sampled from the DSM at ground-classified pixel locations. RGB values were extracted from the visible orthophoto and TIR temperatures from the thermal orthomosaic at corresponding pixel positions. Coordinate transformation from image space to world coordinates (needed to perform thermal simulation) is computed using the geo-referenciation of the dataset. Ground points received a semantic label (class 5) to distinguish them from building surfaces. Each point retained its source image and UV coordinates for InfraGAN correspondence mapping and texture atlas identifier for material tracking.

### 4.3 Preparation of labels

**4.3.1 Outlier Removal and Normalization of Temperature Labels** Temperature outliers caused by sensor artifacts or transient heat sources (vehicle exhausts, street lights) were removed via quantile filtering at 1st and 99th percentiles. Gaussian noise ($\sigma = 0.2$ K) was added to ground temperatures to simulate intra-class variability and prevent exact memorization during ML training.

**4.3.2 Spatial Partitioning** Geographic train-test splitting was performed by interactively drawing a polygon boundary on the 2D point projection (Figure 1). The northern region (24%, ∼8.3M points) served as the test set with zero spatial overlap to the southern training region (76%, ∼ 33.9M points). This strict geographic separation prevents data leakage and enables assessment of model generalization to unseen urban structures, where thermal predictions must transfer to new building configurations and material compositions not present in training data.

## 5. Machine learning methods

### 5.1 RandLA-Net (3D)

**5.1.1 Input features** RandLA-Net (Hu et al., 2020) addresses semantic segmentation through multi-class classification. Two feature sets were trained to evaluate the contribution of physics-based material information. On the one hand, we consider the baseline feature set (XYZ coordinates and their RGB values). This variant relies solely on geometric coordinates and photogrammetric color information, representing a purely data-driven approach without explicit physical priors. On the other hand, we extend this feature set by the material parameter vector $\boldsymbol{\theta}$ from Section 3.2. The values of the $\boldsymbol{\theta}$ of individual points were mapped from the Physics Simulation Source in Bulatov et al. (2024) using KD-tree nearest-neighbor search. Parameters were augmented with 0.1% Gaussian noise to simulate realistic intra-class variations and prevent exact

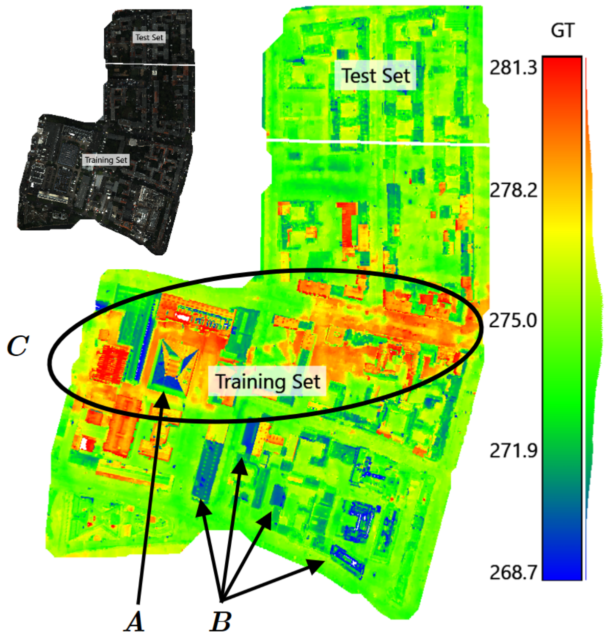


Figure 1. Test dataset (upper part) and training dataset (bottom part) with corresponding RGB view (top left). Marked challenges include (A) triangulation errors, (B) metal roofs/special materials, and (C) a high-temperature stripe, as well as issues in window detection, data gaps, and incomplete regions. The test dataset is more homogeneous than the training dataset.

memorization. This variant aims at examining the additional value provided by land cover classification. ???

Following RandLA-Net’s augmentation strategy, training applies geometric transformations to enhance robustness; Unlike the original implementation’s anisotropic scaling and symmetry operations, thermal scenes require rotation-invariant features while preserving vertical structure—buildings exhibit distinct thermal signatures at different heights and orientation. Additive Gaussian noise $\mathcal{N}(0, 0.0001)$ and color jittering (factor 0.8) were added to the XYZ coordinates and the RGB features, respectively.

**5.1.2 Network Architecture and Training** The baseline RandLA-Net architecture employs four encoding layers with progressive downsampling ratio 4 and feature dimensions $[16, 64, 128, 256]$. For thermal regression, encoder depth was extended to five layers with feature dimensions $d_{\text{out}} = [16, 64, 128, 256, 512]$ and variable decimation ratios $[4, 4, 4, 4, 2]$. This adaptation increases representational capacity with only moderate parameter growth ( 1.5M vs. 1.24M in the original).

Initial experiments with standard batch normalization ($\epsilon = 10^{-6}$, momentum 0.01) revealed gradient flow issues in regression tasks. Systematic ablation studies showed that batch normalization in encoder blocks (dilated residual blocks) provides stability benefits, while decoder and output layers benefit from identity mappings. This hybrid configuration balances normalization effects with unrestricted gradient flow.

Continuous temperature values are discretized into 40 bins using quantile-based quantization computed from the training distribution, enabling balanced multi-class formulation while preserving continuous ground truth values for evaluation.

Instead of the 8-way semantic classifier, four fully-connected layers with residual connections were implemented: $128 \rightarrow 128 \rightarrow 128 \rightarrow 128 \rightarrow 40$, with dropout (rate 0.2) between layers. Log-softmax activation produces class probabilities over 40 temperature bins.

Continuous temperature values are discretized into $K = 40$ bins for classification using quantile-based binning. Quantile boundaries are computed from the training temperature distribution $T_{\text{train}}$ such that each bin contains approximately equal numbers of training samples

$$q_0 = -\infty, \quad q_i = Q_{\frac{i}{K}} \ \ (i = 1, \ldots, K-1), \quad q_K = \infty$$

where $Q_p$ denotes the p-quantile. Each continuous temperature $T$ is assigned to bin index

$$c(T) = i \quad \text{if } T \in (q_i, q_{i+1}) \tag{2}$$

This ensures balanced class distributions across temperature ranges. Continuous ground truth values are retained for evaluation metrics; only training labels are discretized for weighted cross-entropy loss computation.

### 5.2 InfraGAN (2D)

InfraGAN Özkanoğlu and Ozer (2022) implements thermal infrared simulation through adversarial image-to-image translation. Unlike point cloud-based methods, InfraGAN operates directly on texture atlases generated through the unified RGB-TIR mapping framework established in Section 4.2. This 2D approach leverages the spatial consistency and high-resolution photogrammetric RGB textures to generate corresponding thermal representations, which are subsequently remapped to 3D geometry for validation.

**5.2.1 InfraGAN Data Preparation** InfraGAN operates on 2D texture images rather than 3D point clouds. Points from the unified point cloud are rasterized into 2D texture atlases using their UV coordinates. For each texture atlas index, points are projected to 2D pixel coordinates $(i, j)$ and their RGB values and thermal ground truth are written to the corresponding pixel positions.

RGB atlases are stored as 8-bit PNG images (intensity range [0, 255]), while thermal atlases are stored as floating-point TIFF with lossless compression to preserve absolute temperature values. For processing with InfraGAN, textures are decomposed into $256 \times 256$ patches with stride 128 (50% overlap), excluding masked-out regions based on the geographic train-test partition. At inference, RGB test patches are processed through the trained generator to produce synthetic thermal textures. Overlapping regions are reconstructed via weighted averaging at patch boundaries. Full-resolution thermal atlases are then reprojected onto point clouds for validation.

**5.2.2 Architecture and Training of InfraGAN** The generator architecture follows the U-Net design of Ronneberger et al. (2015) with encoder-decoder structure and skip connections at corresponding resolution levels to preserve fine-grained spatial features during thermal translation. The discriminator implements PatchGAN Isola et al. (2017) with $70 \times 70$ receptive field, enforcing local texture realism rather than global image consistency. The combined loss function

$$\mathcal{L} = \mathcal{L}_{\text{adv}}(G, D) + \lambda \mathcal{L}_{\text{L1}}(G) \tag{3}$$

balances adversarial objectives ($\mathcal{L}_{\text{adv}}$) with pixel-wise L1 reconstruction loss ($\mathcal{L}_{\text{L1}}$). The weighting factor $\lambda = 100$ prioritizes accurate temperature reproduction over adversarial realism, reflecting the regression nature of thermal prediction. Training proceeds over 200 epochs with batch size 16 patches using Adam optimizer (learning rate $2 \times 10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$).

**Inference and 3D Remapping.** Test texture patches are processed through the trained generator to produce synthetic thermal predictions. Overlapping regions are reconstructed via weighted averaging based on distance to patch boundaries, reducing boundary artifacts at seams. The resulting full-resolution thermal texture atlases are transferred back to the 3D point cloud representation through the inverse UV mapping established in Section 4.2: For each point, the corresponding atlas identifier and integer pixel coordinates stored during point generation enable direct temperature lookup via nearest-neighbor sampling. This bidirectional 2D-3D transformation framework preserves spatial consistency across modalities and enables direct comparison with physics-based simulation and RandLA-Net predictions through the unified validation pipeline.

## 6. Results

### 6.1 Quantitative Evaluation

Table 2. Thermal prediction performance on training and test datasets.

| Method | Training/Test MAE (K) | RMSE (K) | MI (norm.) | Pearson $r$ |
|---|---|---|---|---|
| RA-Net +Mat | 0.87 / 1.76 | 1.50 / 2.43 | 0.29 / 0.09 | 0.77 / 0.36 |
| RA-Net | 0.84 / 1.57 | 1.48 / 2.17 | 0.31 / 0.10 | 0.78 / 0.45 |
| Phys Sim | 7.77 / 9.05 | 8.40 / 9.44 | 0.02 / 0.03 | $-0.03/-0.04$ |
| IR InfraGAN | 0.72 / 1.79 | 0.85 / 2.31 | 0.57 / 0.01 | 0.98 / 0.16 |

Table 2 presents thermal prediction performance using four metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) measure absolute accuracy with different outlier sensitivity, Pearson correlation ($r$) quantifies linear correlation strength, and Mutual Information (MI) captures nonlinear dependencies.

**Training Performance:** ML methods demonstrate strong fitting capabilities on training data. InfraGAN achieves lowest absolute errors (MAE 0.72 K, RMSE 0 K) with near-perfect linear correlation ($r = 0.98$) and the highest normalized mutual information (0.57). RandLA-Net variants show comparable performance (MAE 0.84 to 0.87 K, $r = 0.77$ to 0.78) with moderate information capture (normalized MI 0.29 to 0.31). Physics-based simulation exhibits systematic bias with large errors (MAE 7.77 K) and negligible correlation ($r = -0.03$).

**Test Performance and Generalization:** Spatial generalization to unseen urban regions reveals severe overfitting across all ML methods. InfraGAN suffers a dramatic decay in performance: MAE increases by a factor of 2.5, while Pearson correlation and normalized MI drop from 0.98 to 0.16 and from 0.57 to 0.01, respectively, indicating near-complete loss of predictive information. RandLA-Net variants show comparable degradation with MAE increasing by a factor of 1.9 to 2.0 from 0.84 and 0.87 K to 1.57 and 1.76 K. Normalized MI dropping by a factor of 3.1 to 3.2 from 0.29 and 0.31 to 0.09 and 0.10. Physics-based simulation maintains consistent performance (MAE 7.77 K to 9.05 K), exhibiting systematic bias without spatial overfitting

**Material Parameter Integration:** Augmentation with physics-derived material parameters provides no performance benefit. RGB-only configuration outperforms RGB+Materials on test data across all metrics (MAE 1.57 K vs. 1.76 K, Pearson $r$ 0.45 vs. 0.36, normalized MI 0.10 vs. 0.09), suggesting inadequate feature integration strategies rather than insufficient physical information.

## 6.2 Qualitative Evaluation

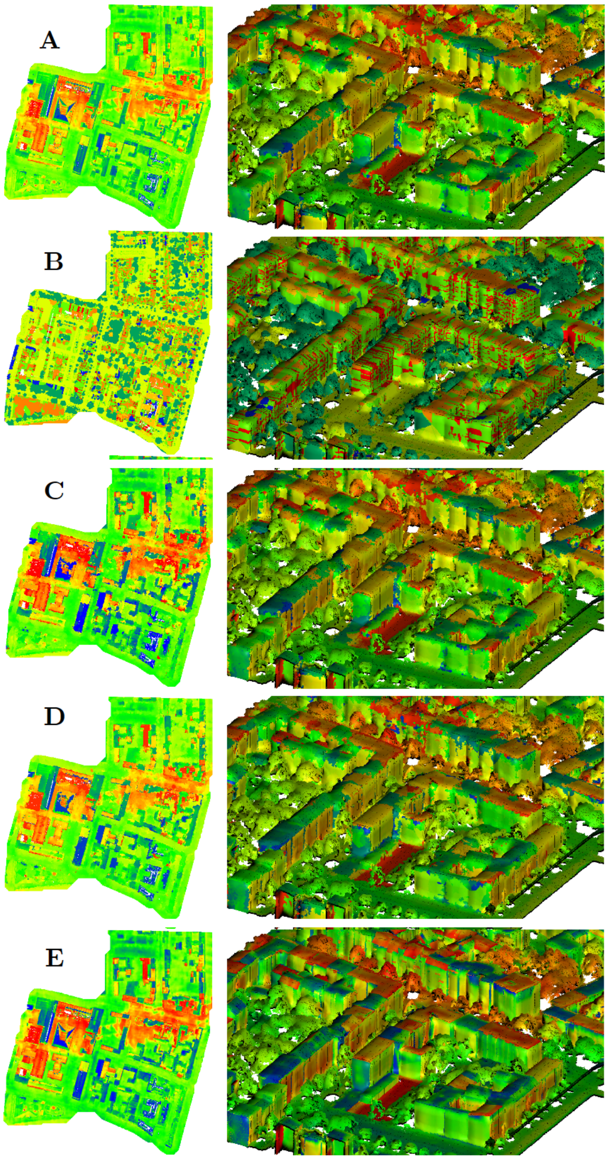

Figure 2. Thermal predictions for training region (Moabit district). (A) Ground truth TIR. (B) Physics-based simulation. (C) RandLA-Net RGB-only. (D) RandLA-Net RGB+Materials. (E) InfraGAN. Color scale as in Figure 1.

**6.2.1 Training Results** Figure 2 reveals qualitative differences between prediction methods on the training dataset. Ground truth data (A) exhibits a horizontal stripe with elevated temperatures traversing the scene from left to right, probably attributed to sensor calibration drift or atmospheric effects during acquisition. This artifact distorts spatial temperature distribution and introduces non-physical features. Metal roofs appear as blue signatures due to low emissivity ($\varepsilon \approx 0.2$–$0.4$), causing reduced apparent temperatures in TIR measurements.

ML-based methods reproduce the artifact with varying intensity. InfraGAN (E) and RandLA-Net RGB+Materials (D) overestimate temperatures in the stripe region, demonstrating memorization of training data anomalies without physical plausibility constraints. RandLA-Net RGB-only (C) produces moderate temperature estimates in this area, showing reduced artifact amplification compared to the augmented variant. Physics-based simulation (B) eliminates the stripe entirely through thermal equilibrium enforcement, the artifact has no representation in the geometric-material model and thus cannot manifest in predictions. Note that the train-test partition is methodologically irrelevant for physics-based simulation. Deterministic equations apply identically across all locations by design, explaining the invariant performance.

This behavior confirms quantitative findings from Table 2: ML-based methods achieve high training fidelity, including spurious measurement features, while physics simulation maintains consistency independent of input data artifacts at the cost of systematic bias.

**6.2.2 Test Results** Figure 3 visualizes predictions on the spatially disjoint test region, revealing generalization characteristics absent in training data. Ground truth (A) exhibits homogeneous temperature distribution without the artifact stripe present in training data (Figure 2). The oblique 3D view (right column) shows characteristic thermal patterns: West-facing facades appear warmer than north-facing surfaces, windows exhibit lower temperatures than surrounding walls, and metal roofs (blue) show reduced apparent temperatures compared to tile roofs (green) due to emissivity differences.

Physics-based simulation (B) systematically underestimates vegetation temperatures while overestimating window temperatures—modeled fenestration appears uniformly warm across all orientations, contradicting ground truth where windows consistently appear cooler than adjacent wall surfaces. Material classification correctly identifies metal versus tile roofing, producing appropriate temperature differentiation.

RandLA-Net variants (C,D) exhibit spatial artifacts inherited from training data despite geographic separation. RGB-only configuration (C) shows scattered high-temperature spots where local feature distributions resemble the training artifact stripe, indicating inappropriate feature-to-temperature associations. RGB+Materials (D) amplifies this effect with more extensive warm regions, demonstrating that material augmentation increases overfitting rather than improving physical consistency. Both variants successfully capture facade orientation effects (west warmer than north) and produce plausible vegetation temperatures, but fail to resolve window boundaries.

InfraGAN (E) produces visually coherent predictions despite low quantitative metrics (Table 2), exhibiting regression-to-mean behavior where extreme temperatures compress toward dataset average—hot surfaces appear cooler, cold surfaces warmer. The method completely fails to distinguish facade orientations: North-facing walls appear warmer than west-facing surfaces in the 3D view (right), contradicting physical expectations and ground truth. This failure stems from InfraGAN's 2D texture-based processing lacking explicit 3D geometric context—surface normal information required for solar irradiance modeling is absent from input features.

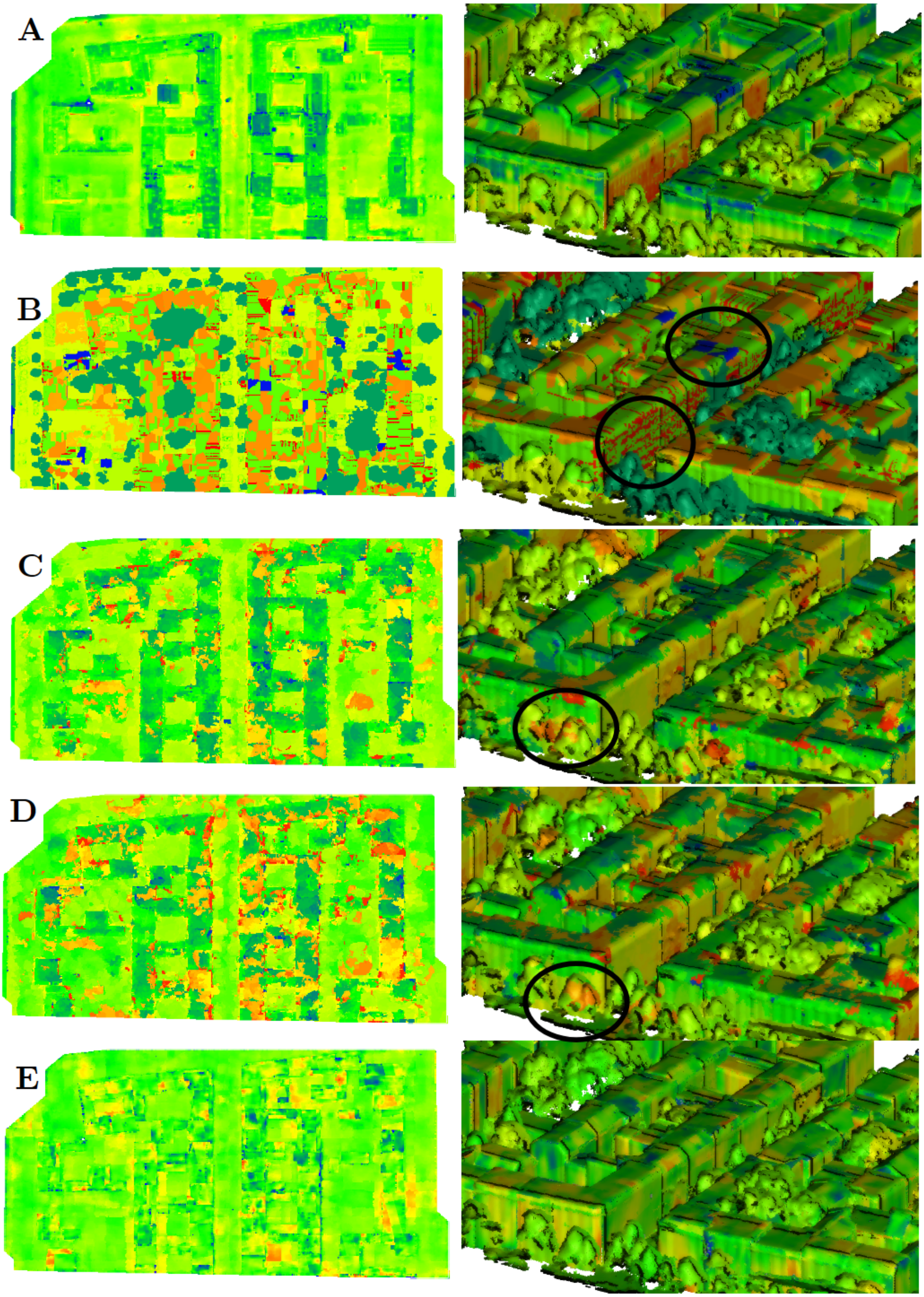


Figure 3. Thermal predictions for test region (northern Moabit). Rows: (A) Ground truth TIR, (B) Physics-based simulation, (C) RandLA-Net RGB-only, (D) RandLA-Net RGB+Materials, (E) InfraGAN. Left column: Nadir view. Right column: Oblique 3D visualization (northeast to southwest perspective). Color scale as in Figure 1.

## 7. Conclusions and Future Work

This paper establishes a unified validation framework comparing physics-based and ML-based thermal simulation for urban digital twins. The evaluation on the Berlin-Moabit dataset with strict geographic train-test splitting revealed complementary characteristics of both paradigms.

Methods based on machine learning (RandLA-Net, InfraGAN) achieve excellent training accuracy but exhibit severe spatial generalization degradation on test data. Material parameter augmentation degrades rather than improves performance—RGB+Materials underperforms RGB-only across all test metrics. This result seems counterintuitive because, apparently, more information flows in. In reality, for surface from above, we have certain estimates on land cover and material classes, but we do not know them for sure, and for the material parameters, we take standard values from literature sources (Bulatov et al., 2020), which may partly be valid for an Australian dataset, but need an additional tuning for the Moabit dataset. For non visible surface elements, such as walls, land cover classes, material classes, and parameters can only be guessed.

Physics-based simulation maintains consistent performance across spatial domains with systematic bias from climate parameterization mismatches (Australian models applied to Berlin). Validation against in-situ measurements Bulatov et al. (2024) demonstrated 2 to 3 K accuracy after region-specific calibration, confirming model adequacy with appropriate configuration. However, in this paper, no fine-tuning for physical simulation has been performed at all, which manifests in significantly worse results even compared with the test data for ML-based results in all metrics. In part, suboptimal results may be explained by multi-day acquisitions of reference data, which introduce temporal inconsistencies that destroy global correlations for training and simulations.

In view of these shortcomings, the following ideas for future research can be proposed.

• **Temporally Synchronized Acquisition:** Collecting complete scenes within single flight missions under consistent atmospheric conditions.

• **Feature Enhancement:** For 3D ML-based approaches, extending the feature set for point clouds with local and global geometric features (surface normals, height-above-ground). Furthermore, application of Markov and Conditional Random Field as a post-processing step helps to enforce spatial smoothness and thermodynamic consistency in predictions.

• **Physics-Informed Architectures:** Implementing soft constraints enforcing Stefan-Boltzmann law and heat diffusion through loss terms may robustify the performance. The RandLA-Net method, chosen as our default 3D method, failed at computing regression result, which made it necessary to discretize the temperatures. Thus, necessity for better architectures that allow regression and, ideally, offer cross-attention mechanisms between visual and material embeddings instead of simple feature concatenation becomes evident.

• **Multi-Temporal Modeling:** We expect that in the future, static predictions may be extended to time-series forecasting with recurrent architectures (LSTM, Transformer) or neural ODEs, capturing diurnal cycles and transient weather events while maintaining 3D spatial context.

• **Hybrid Physics-ML Frameworks:** Finally, especiall tempting direction of future research would to train ML models to predict residuals from physics baselines rather than absolute temperatures, combining thermodynamic consistency with data-driven bias correction.

The established validation infrastructure with unified point cloud representation, geographic splitting and multi-metric evaluation provides a foundation for systematic advancement. Identified failure modes define concrete targets for achieving operational thermal digital twins in urban monitoring applications.

### Aknowledgements

We particularly would like to thank Dirk Frommholz, Dennis Dahlke, Magdalena Linkiewicz, Sebastian Pless, Karsten Stebner and Matthias Geßner from the Institute of Space Research of the German Aerospace Center (DLR) in Berlin/Germany for acquiring, pre-processing, and providing us the Moabit dataset.

### References

Andriambololonaharisoamalala, R., Kottler, B., Bulatov, D., Fischer, S., Zipperer, R., Helmholz, P., 2025. High-resolution thermal sharpener supported by generative techniques: Case study in Perth, Australia. *Earth Resources and Environmental Remote Sensing/GIS Applications XVI*, 13671, SPIE, 244–256.

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.

Bulatov, D., Burkard, E., Ilehag, R., Kottler, B., Helmholz, P., 2020. From multi-sensor aerial data to thermal and infrared simulation of semantic 3D models: Towards identification of urban heat islands. *Infrared Physics & Technology*, 105, 103233.

Bulatov, D., Frommholz, D., Kottler, B., Qiu, K., Strauss, E., 2024. Using passive multi-modal sensor data for thermal simulation of urban surfaces. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-2-2024, 17-24.

Cárdenas-León, I., Morales-Ortega, R., Koeva, M., Atún, F., Pfeffer, K., 2024. Digital twin-based framework for heat stress calculation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 67–74.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Fischer, S., Kottler, B., Strauß, E., Bulatov, D., 2025. Exploration and Validation of Specialized Loss Functions for Generative Visual-Thermal Image Domain Transfer. 527–534.

Frommholz, D., Linkiewicz, M., Meißner, H., Dahlke, D., 2017. Reconstructing buildings with discontinuities and roof overhangs from oblique aerial imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 42(W1), 465–471.

Frommholz, D., Linkiewicz, M., Poznanska, A.-M., 2016. Inlining 3D reconstruction, multi-source texture mapping and semantic analysis using oblique aerial imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 41(B3), 605–612.


https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-395-2026  

Giberna, M., Voos, H., Tavares, P., Nunes, J., Sorg, T., Masini, A., Sanchez-Lopez, J. L., 2025. On digital twins in defence: Overview and applications. *arXiv preprint arXiv:2508.05717*.

Gonzalez-Caceres, A., Hunger, F., Forssén, J., Somanath, S., Mark, A., Naserentin, V., Bohlin, J., Logg, A., Wästberg, B., Komisarczyk, D., Edelvik, F., Hollberg, A., 2025. Towards digital twinning for multi-domain simulation workflows in urban design: A case study in Gothenburg. *Journal of Building Performance Simulation*, 18(3), 311–332.

Han, S., Mo, B., Xu, J., Sun, S., Zhao, J., 2024. TransImg: A Translation Algorithm of Visible-to-Infrared Image Based on Generative Adversarial Network. *International Journal of Computational Intelligence Systems*, 17(1), 264.

Hidalgo, J., Masson, V., Baklanov, A., Pigeon, G., Gimeno, L., 2008. Advances in urban climate modeling. *Annals of the New York Academy of Sciences*, 1146(1), 354–374.

Hirano, Y., Ichinose, T., Ohashi, Y., Shiraki, Y., Onishi, A., Yoshida, Y., 2024. Simulation of urban surface temperature and surface heat balance in the Tokyo metropolitan area. *Sustainable Cities and Society*, 112, 105596.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11108–11117.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134.

Joshi, P., Lin, T.-S., He, C., Lamer, K., 2025. Urban weather modeling using WRF: linking physical assumptions, code implementation, and observational needs. *Geoscientific Model Development*, 18(20), 7869–7890.

Kong, X., Hucks, R. G., 2023. Preserving our heritage: A photogrammetry-based digital twin framework for monitoring deteriorations of historic structures. *Automation in Construction*, 152, 104928.

Kottler, B., Burkard, E., Bulatov, D., Haraké, L., 2019. Physically-based thermal simulation of large scenes for infrared imaging. *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 1, 53 – 64.

Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S. et al., 2023. Big Data in Earth system science and progress towards a digital twin. *Nature Reviews Earth & Environment*, 4(5), 319–332.

Li, X., Hayder, Z., Zia, A., Cassidy, C., Liu, S., Stiller, W., Stone, E., Conaty, W., Petersson, L., Rolland, V., 2024. NeFF-BioNet: Crop biomass prediction from point cloud to drone imagery. *arXiv preprint arXiv:2410.23901*.

Liu, H., Mou, C., Yuan, J., Chen, Z., Zhong, L., Cui, X., 2024. Estimating urban forests biomass with LiDAR by using deep learning foundation models. *Remote Sensing*, 16(9), 1643.

Liu, Y., Chen, X., Yan, S., Cui, Z., Xiao, H., Liu, Y., Zhang, M., 2025. Thermalgs: Dynamic 3D thermal reconstruction with gaussian splatting. *Remote Sensing*, 17(2), 335.

Masson, V., 2006. Urban surface modeling and the meso-scale impact of cities. *Theoretical and applied climatology*, 84(1), 35–45.

Mirabet-Herranz, N., Dugelay, J.-L., 2024. Beyond the visible: Thermal data for facial soft biometric estimation. *EURASIP Journal on Image and Video Processing*, 2024(1), 27.

Nascimento, D., Ascenso, A., Lopes, D., Lopes, M., Rodrigues, V., 2025. CFD modelling of an urban microclimate during a heat wave: Coupling solar radiation and heat transfer. *Urban Climate*, 62, 102486.

Oehmcke, S., Li, L., Trepekli, K., Revenga, J. C., Nord-Larsen, T., Gieseke, F., Igel, C., 2024. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. *Remote Sensing of Environment*, 302, 113968.

Özkanoğlu, M. A., Ozer, S., 2022. InfraGAN: A GAN architecture to transfer visible images to infrared domain. *Pattern Recognition Letters*, 155, 69-76.

Pettirsch, A., Garcia-Hernandez, A., 2025. Overcoming data scarcity in roadside thermal imagery: A new dataset and weakly supervised incremental learning framework. *Sensors*, 25(7), 2340.

Qiu, K., Budde, L. E., Bulatov, D., Iwaszczuk, D., 2022. Exploring fusion techniques in u-net and deeplab v3 architectures for multi-modal land cover classification. *Earth Resources and Environmental Remote Sensing/GIS Applications XIII*, 12268, SPIE, 190–200.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 234–241.

Salim, M. H., Schubert, S., Resler, J., Krč, P., Maronga, B., Kanani-Sühring, F., Sühring, M., Schneider, C., 2022. Importance of radiative transfer processes in urban climate models: A study based on the PALM 6.0 model system. *Geoscientific Model Development*, 15(1), 145–171.

Strauß, E., Bulatov, D., Stütz, P., 2023. Synthetic thermal imagery for UAV-based reconnaissance by change detection. *Artificial Intelligence for Security and Defence Applications*, 12742, 197–209.

Vogel, J., Stadler, S., Chockalingam, G., Afshari, A., Henning, J., Winkler, M., 2025. SanDyPALM v1. 0: static and dynamic drivers for the PALM model to facilitate urban microclimate simulations. *Geoscientific Model Development*, 18(18), 6063–6094.

Yang, S., Sun, M., Lou, X., Yang, H., Liu, D., 2024. Nighttime thermal infrared image translation integrating visible images. *Remote Sensing*, 16(4), 666.

Yao, S., Sun, X., Song, B., Wei, Y., Wang, Y., Li, Y., Liu, X., 2026. DGC_GAN: An Unpaired Method for Cross-Spectral Image Translation from Visible to Thermal Infrared. *Remote Sensing*, 18(4), 569.

Zhong, C., Cheng, S., Kasoar, M., Arcucci, R., 2023. Reduced-order digital twin and latent data assimilation for global wildfire prediction. *Natural Hazards and Earth System Sciences*, 23(5), 1755–1768.